**The University of Sydney**

**Department of Infectious Diseases**

**Faculty of Medicine**

# Investigations of Sequence Variation in Human Papillomavirus Type 16 Cervical Cancer Isolates

**A thesis presented for the degree of**

**Doctor of Philosophy**

**2000**

**Author:**

**Kylie Jane Watts**

**B. Medical Science (Hons)**

**Supervisor: Dr. Barbara Rose**

**Associate Supervisor: Dr. Carol Thompson**

# DECLARATION

I hereby certify that this thesis incorporates original research which has not been previously submitted for a higher degree to any other University: and that to the best of my knowledge and belief, it does not contain any material previously published or written by other persons except where reference has been made in the text.

[redacted]

2.10.00

**Kylie Jane Watts** **Date 2/10/2000**

# PUBLICATIONS AND WORK PRESENTED FROM THE INVESTIGATIONS DESCRIBED IN THIS THESIS

## Journal Articles

1. "Upregulation of the HPV 16 E6/E7 promoter in cervical cancer isolates is mediated by specific sequence variations in the upstream regulatory region". Kylie J. Watts, Carol H. Thompson, Yvonne E. Cossart and Barbara R. Rose. Submitted July 2000.

2. "Sequence variation in the E6 and E2/E4 coding regions and physical state of human papillomavirus type 16 cervical cancer isolates from Australia and the South Pacific regions". Kylie J. Watts, Carol H. Thompson, Yvonne E. Cossart and Barbara R. Rose. Manuscript in preparation.

## Conference Abstracts

1. October 1998: From Cell to Society. College of Health Sciences and Medical Foundation Research Conference, Leura, NSW - Abstract title: "Evidence for the functional significance of mutations in the upstream regulatory region of HPV 16 cervical cancer isolates". Kylie J. Watts, Carol H. Thompson, Yvonne E. Cossart and Barbara R. Rose.

2. January 1999: 17$^{th}$ International Papillomavirus Conference, Charleston, South Carolina, USA - Abstract title: "Evidence for the functional significance of mutations in the upstream regulatory region of HPV 16 cervical cancer isolates". Kylie J. Watts, Carol H. Thompson, Yvonne E. Cossart and Barbara R. Rose.

3. August 1999: 11$^{th}$ International Congress of Virology, Sydney, NSW - Abstract title: "Investigations of mutations upregulating oncogene expression in episomal HPV 16 positive cervical cancers". Kylie J. Watts, Carol H. Thompson, Yvonne E. Cossart and Barbara R. Rose.

# ACKNOWLEDGMENTS

It is impossible to comprehend successfully completing a PhD without a number of supportive people who do deserve more than just a 'thank you' on this page.



To my supervisor, Barbara Rose, thank you for always going above and beyond what may generally be expected of a supervisor. You have set a precedent for me to follow in the future. A simple thank you is by no means sufficient. I have been extremely lucky to have you as my teacher.



To my associate supervisor, Carol Thompson, thank you for your supervision and assistance. Thank you also to Yvonne Cossart for many helpful writing suggestions.



To my family and extended family, in particular my father Michael Watts, thank you for being supportive as well as professional furniture removalists on short notice on more than one occasion!



To my fellow postgraduate students, in particular, Robert, Jim, Ji Hong and Aniko, thanks for always listening to me, for your support, and for always having happy, smiling faces through countless corridor crossings. Thank you also to Lilia for friendship and employment.



Thank you particularly to the University of Sydney for an APA scholarship that has supported me over three and a half years and has enabled me to undertake this PhD.

$

This PhD has been an amazing experience with a very large learning curve. It is hard to believe that I have come to the end and it is time to close this door and open another. I consider it a privilege to have been a student in the Department of Infectious Diseases over the last four years.

# SUMMARY

Human papillomavirus type 16 (HPV 16) is the major causative agent of cervical cancer worldwide, but the clinical aggression displayed by individual HPV 16 positive cancers is highly variable. This might reflect altered expression of the viral oncogenes resulting from sequence variations in the important regulatory region of the viral genome, the upstream regulatory region (URR). Initially, this was assessed by amplifying by polymerase chain reaction (PCR) and sequencing the entire URR of 34 cervical cancer isolates from Australia and New Caledonia. Only one isolate was identical to the HPV 16 prototype; one large-scale change (an 81bp duplication), 44 variant nucleotides, and one single base insertion were identified. Most of the variants were of European lineage (82%), whilst smaller proportions of Asian and Asian American lineage isolates were also identified (12% and 6% respectively).

The level of expression of the viral oncogenes, E6 and E7 is controlled by cellular transcription factors (TFs) binding to their response sites in the URR. A major objective of this study was to analyse the effects of URR sequence changes on oncogene promoter activity. Six of the variant URRs, selected mainly on the basis of potentially important sequence changes, were cloned into a luciferase vector. The constructs were transiently transfected into HeLa and HT3 cells and luciferase activities of variants, a surrogate for E6 and E7 expression, were compared with that of the prototype. Promoter activities controlled by three of the five URRs that contained only minor sequence variations were upregulated between three- and eleven-fold, while that mediated by the URR with the large-scale duplication was comparable to the prototype. Attempts were then made to identify specific sequence changes responsible for upregulated promoter activity by PCR-based site-directed mutagenesis. Changes in a yin and yang 1 (YY1) binding site and an overlapping octamer binding factor-1 (Oct-1)/papilloma enhancer binding factor-1 (PEF-1) site both mediated upregulation of the promoter in the isolate with the highest luciferase activity. Neither of these changes affected the binding of their corresponding TFs in electrophoretic mobility shift assays though. A nucleotide substitution in the papillomavirus silencing motif (PSM) partly explained promoter upregulation in the isolate with the second highest luciferase activity. Further analysis revealed that

sequence variation in the PSM affected the binding of CCAAT displacement protein (CDP). Sequence variation in other YY1 and transcriptional enhancer factor-1 (TEF-1) binding sites did not affect promoter activity in two of the upregulated isolates.

Deregulation of E6/E7 expression is central to HPV-associated malignant conversion. In many cases this is achieved by integration of viral DNA into the host genome, often disrupting the E1 and/or E2 regions. However, HPV 16-associated cancers may carry only episomal HPV. Sequence variation in the URR is believed to account for at least some instances of malignant conversion in the absence of integration. For this reason the physical state of HPV DNA in the 34 isolates was investigated by Southern hybridisation and PCR of the E2 gene. Overall, for the 25 isolates where results could be obtained, 44% were integrated into host chromosomes, 40% were episomal, and 16% contained both integrated and episomal DNA. The physical state of nine of the isolates could not be determined because of specimen inadequacy. The two isolates with the highest promoter activities and the isolate with the large-scale duplication contained only episomal HPV DNA. Twenty of 28 isolates (71%) contained the entire E2 gene. For the eight isolates lacking an entire E2 gene, deletion of the 3' region (four isolates), the 5' region (one isolate), the entire E2 gene (two isolates), or the presence of a discontinuous E2 gene (one isolate) was determined by PCRs spanning different lengths of the E2 region.

In addition to sequence variation in the URR, it has also been suggested that some amino acid substitutions in viral coding regions, such as in E2 and E6 could have biological significance. The potential implications of amino acid changes include changes to B or T cell epitopes and alterations to protein function. The E2 gene and the overlapping E4 gene from 13 of the isolates, and the E6 gene from 27 isolates were sequenced. Twenty one from 30, five from 14, and 12 from 20 nonsynonymous amino acid substitutions were identified in the E2, E4, and E6 genes respectively.

This study has provided one of the most comprehensive analyses of sequence variation within HPV 16 cancer isolates and the first HPV 16 sequence data from Australia and the South Pacific region. It has also provided evidence of new strategies by which HPV 16 may achieve malignant conversion without integration.

# TABLE OF CONTENTS


| | Page Number |
|---|---|
| Declaration | ii |
| Publications and Work Presented from the Investigations Described in This Thesis | iii |
| Acknowledgments | iv |
| Summary | v |
| Table of Contents | vii |
| List of Tables | xix |
| List of Figures | xxii |
| List of Abbreviations | xxv |

**CHAPTER 1**

**REVIEW OF THE LITERATURE**

| | |
|---|---|
| **1.1 GENERAL INTRODUCTION** | 2 |
| **1.2 CERVICAL CANCER** | 3 |
| 1.2.1 Incidence of Cervical Cancer | 3 |
| 1.2.2 Cervical Cancer as a Sexually Transmissible Disease (STD) | 4 |
| 1.2.3 Risk Factors for Cervical Cancer | 6 |
| **1.3 PAPILLOMAVIRUSES** | 9 |
| 1.3.1 Viral Structure | 9 |
| 1.3.2 Human Papillomavirus Genomic Organisation | 9 |
| 1.3.3 Classification | 11 |
| 1.3.4 Lesions Resulting from Infection with Different Human Papillomavirus | 12 |
| 1.3.5 High Risk Human Papillomavirus Types | 13 |
| 1.3.6 Epidemiology of Human Papillomavirus Infection | 16 |
| 1.3.7 Risk Factors for Cervical Human Papillomavirus Infection | 17 |
| 1.3.8 Immune Surveillance | 18 |
| 1.3.8.1 Cell-Mediated Responses | 18 |
| 1.3.8.2 Humoral Responses | 19 |
| **1.4 CELL BIOLOGY OF ANOGENITAL PAPILLOMAVIRUS INFECTION** | 20 |
| 1.4.1 Tissue Specificity | 20 |

1.4.2 Transmission 20
1.4.3 Life-Cycle and Replication Strategies 21
1.4.3.1 Latent Infection 21
1.4.3.2 Productive Infection 22
1.4.3.3 Transforming Infection 24
1.4.4 Immortalisation and Transformation of Cells by Human Papillomavirus 25
1.4.5 Physical State of Papillomavirus DNA in Cervical Cancers 27
**1.5 MOLECULAR VIROLOGY OF PAPILLOMAVIRUSES** 29
1.5.1 Papillomavirus Proteins Involved in Viral Replication and Gene Regulation 29
1.5.1.1 E1 29
1.5.1.2 E2 30
1.5.2 Papillomavirus Proteins with Functions that are not Well Defined 35
1.5.2.1 E3 35
1.5.2.2 E4 35
1.5.3 The Papillomavirus Oncoproteins 36
1.5.3.1 E5 36
1.5.3.2 E6 37
1.5.3.3 E7 41
1.5.3.4 E8 45
1.5.4 The Capsid Proteins 45
1.5.4.1 L1 45
1.5.4.2 L2 45
1.5.4.3 L3 and L4 46
1.5.5 The Upstream Regulatory Region (URR) 46
1.5.5.1 The 5' Region of the Upstream Regulatory Region 47
1.5.5.2 The Central Segment (Enhancer) 47
1.5.5.3 The 3' Region (Promoter) 49
1.5.6 Transcription and Regulatory Factors 51
1.5.6.1 Transcriptional Repressors and Activators 51
1.5.7 DNA Binding Domains 52
1.5.7.1 Zinc Finger Motifs 52
1.5.7.2 Steroid Receptors 52
1.5.7.3 Homeodomains – Helix-Turn-Helix Motifs 53
1.5.7.4 POU Domains 53
1.5.7.5 Leucine Zipper Motifs 53
1.5.7.6 Helix-Loop-Helix Motifs 54
1.5.8 Activation Domains 54

1.5.9 Transcription Factor Regulation 55
1.5.10 Factors that Affect Regulation by the Upstream Regulatory Region 55
- Focussing on Human Papillomavirus Type 16
1.5.10.1 Promoter Specific Factor (Sp1) 56
1.5.10.2 Yin Yang 1 (YY1) 56
1.5.10.3 Activator Protein 1 (AP1) 57
1.5.10.4 Nuclear Factor I (NFI)/CCAAT Transcription Factor (CTF) 58
1.5.10.5 Octamer Binding Protein-1 (Oct-1) 59
1.5.10.6 Transcriptional Enhancer Factor-1 (TEF-1) and TEF-2 59
1.5.10.7 Papilloma Enhancer Binding Factor-1 (PEF-1) 59
1.5.10.8 Papillomavirus Silencing Motif-Binding Protein (PSM-BP) 60
1.5.10.9 Glucocorticoids and Progesterone 60
1.5.10.10 Estrogen 61
1.5.10.11 Epidermal Growth Factor 61
1.5.10.12 Cytokines 62
1.5.10.13 Cellular Oncogenes 62
1.5.10.14 DNA Methylation 63
1.5.10.15 Viral Proteins (Other than Papillomavirus) 63
**1.6 MANIFESTATIONS OF HUMAN PAPILLOMAVIRUS INFECTION** 65
1.6.1 Benign Lesions 65
1.6.2 Cytological Classification of Cervical Lesions 66
1.6.2.1 Atypical Squamous Cells of Undetermined Significance (ASCUS) 67
1.6.2.2 Low Grade Cervical Intraepithelial Neoplasia (CIN 1) 67
1.6.2.3 High Grade Cervical Intraepithelial Neoplasia (CIN 2 and 3)/ Carcinoma *in situ* 68
1.6.2.4 Carcinoma 68
1.6.3 International Federation of Gynecologists and Obstetricians (FIGO) Staging System for Cervical Cancer 70
1.6.4 Classification of Invasive Cervical Tumours 72
1.6.5 Other Carcinomas Caused by Human Papillomavirus 72
**1.7 CERVICAL CANCER PREVENTION** 73
1.7.1 Detection of Cervical Lesions 74
1.7.1.1 Papanicolaou Smears and Cytology 74
1.7.1.2 The ThinPrep® Pap Test™ 75
1.7.1.3 Colposcopy 75
1.7.1.4 The Acetic Acid Test 76

1.7.2 Laboratory Identification of Human Papillomavirus 76
1.7.2.1 Routine Human Papillomavirus Testing 79
**1.8 MANAGEMENT OF CERVICAL PRECANCER AND CANCER** 80
1.8.1 Historically 80
1.8.2 Current Strategies 81
**1.9 DEVELOPING STRATEGIES FOR THE TREATMENT OF CERVICAL LESIONS** 82
1.9.1 Vaccines 82
1.9.1.1 Prophylactic Vaccines 82
1.9.1.2 Therapeutic Vaccines 83
1.9.2 Antivirals 83
1.9.3 Gene Therapy 84
**1.10 INTRODUCTION TO THE STUDY** 85

**CHAPTER 2**

**SEQUENCE VARIATION IN THE UPSTREAM REGULATORY REGION OF HUMAN PAPILLOMAVIRUS TYPE 16 CERVICAL CANCER ISOLATES**

**2.1 INTRODUCTION** 88

**2.2 METHODS** 90
2.2.1 Study Population 90
2.2.2 Preparation of Clinical Isolates 90
2.2.2.1 Phenol Extraction of Nucleic Acids from Biopsy Material 90
2.2.3 The Polymerase Chain Reaction 93
2.2.3.1 Principle of the Polymerase Chain Reaction 93
2.2.3.2 Precautions to Prevent Polymerase Chain Reaction Cross-Contamination 95
2.2.4 Polymerase Chain Reaction Components 96
2.2.4.1 Deoxyribonucleoside Triphosphates (dNTPs) 96
2.2.4.2 DNA Polymerase 96
2.2.4.3 Polymerase Chain Reaction Buffer 97
2.2.4.4 Magnesium Chloride ($MgCl_2$) 98
2.2.4.5 Water 98
2.2.4.6 Liquid Paraffin Oil 98

2.2.4.7 Thermal Cycler 98
2.2.4.8 Reaction Tubes 98
2.2.4.9 Primers 98
2.2.5 Polymerase Chain Reaction Preparation 99
2.2.5.1 Cocktail Preparation 99
2.2.5.2 Addition of Targets 100
2.2.5.3 Addition of Positive Control Nucleic Acids 100
2.2.5.4 Thermal Cycling 100
2.2.6 Optimisation of Polymerase Chain Reaction Protocols 100
2.2.7 Polymerase Chain Reaction Protocols for Amplification of the Upstream Regulatory Region 102
2.2.7.1 Hot Start Polymerase Chain Reaction 105
2.2.8 Preparation of Polymerase Chain Reaction Products for Electrophoresis 107
2.2.8.1 Preparation of Size Markers 107
2.2.8.2 Gel Electrophoresis 107
2.2.9 Polyethylene Glycol (PEG) Precipitation of Polymerase Chain Reaction Products for Sequencing 108
2.2.10 DNA Sequencing – Sanger Dideoxy Method 109
2.2.10.1 Computer Analysis 109

**2.3 RESULTS** 110
2.3.1 Polymerase Chain Reaction 110
2.3.1.1 Strategy 110
2.3.1.2 Polymerase Chain Reaction Results 110
2.3.2 Sequencing the Upstream Regulatory Region 111
2.3.2.1 Technical Considerations 112
2.3.2.2 Sequence Analysis 112

**2.4 DISCUSSION** 116
2.4.1 Sequence Variation in the Upstream Regulatory Region 116

## CHAPTER 3

## FUNCTION ANALYSIS OF SEQUENCE VARIATION WITHIN THE HUMAN PAPILLOMAVIRUS TYPE 16 UPSTREAM REGULATORY REGION


**3.1 INTRODUCTION** 120

**3.2 METHODS** 122

3.2.1 DNA Cloning 122

3.2.2 Selection of Human Papillomavirus Type 16 Upstream Regulatory Regions for Function Studies 122

3.2.3 Cloning Strategy 123

3.2.4 Cloning Upstream Regulatory Regions into pCR®-Blunt 124

3.2.4.1 Zero Blunt™ PCR Cloning 124

3.2.5 Preparation of Variant Upstream Regulatory Regions for Cloning into pCR®-Blunt 125

3.2.5.1 Production of Blunt-Ended Polymerase Chain Reaction Products 125

3.2.5.2 Purification of DNA by QIAquick Gel Extraction (QIAGEN®) 125

3.2.5.3 Ligation 126

3.2.6 Transformation of Bacterial Cells 126

3.2.6.1 Transformation Using One Shot™ Competent Cells (Invitrogen) 127

3.2.7 Screening Bacterial Colonies for pCR®-Blunt Transformants 128

3.2.7.1 Preparing Miniprep Plasmid DNA 128

3.2.7.2 Restriction Endonuclease Analysis of Miniprep Plasmid DNA 129

3.2.8 Reporter Genes 130

3.2.9 Subcloning Strategy 131

3.2.9.1 Restriction Endonuclease Cleavage of pALuc 132

3.2.9.2 Dephosphorylation of pALuc 132

3.2.10 Preparation of Upstream Regulatory Regions for Subcloning into pALuc 133

3.2.11 Subcloning 133

3.2.11.1 Ligation 133

3.2.12 Transformation 134

3.2.12.1 Preparation of XLI Blue Competent Cells for Transformation 134

3.2.12.2 Transformation of XLI Blue *E.coli* 134

3.2.13 Screening Bacterial Colonies for pALuc Transformants 135

3.2.13.1 *Hae* II Digestion 136
3.2.13.2 Large Scale Preparation of Bacterial Plasmids – QIAGEN Maxiprep protocol (QIAGEN®) 136
3.2.14 pCH110 β-Galactosidase Vector 137
3.2.14.1 Preparation of pCH110 138
3.2.15 Transient Transfection of Mammalian Cells 139
3.2.15.1 Strategy 139
3.2.15.2 Seeding Cells 140
3.2.15.3 Transient Transfection 140
3.2.15.4 Glycerol Shock 141
3.2.15.5 Harvesting Cellular Protein 141
3.2.15.6 Protein Estimation (Bio-Rad Method) 142
3.2.15.7 β-Galactosidase Assay 142
3.2.15.8 Luciferase Assay 143
3.2.16 Site-Directed Mutagenesis 144
3.2.17 Constructs Created by Site-Directed Mutagenesis 145
3.2.17.1 Isolate K2 145
3.2.17.2 Isolates K4 and H1 145
3.2.18 Polymerase Chain Reaction for Site-Directed Mutagenesis 147
3.2.18.1 Polymerase Chain Reaction Components 147
3.2.18.2 Phosphorylation of Primers 147
3.2.18.3 Optimised Polymerase Chain Reaction Protocol 148
3.2.18.4 Restriction Endonuclease Digestion of Products 149
3.2.19 Preparation of pCR®- Blunt for Site-Directed Mutagenesis 149
3.2.19.1 Plasmid Mini Purification Protocol (QIAGEN®) 150
3.2.19.2 Ligation 150
3.2.20 Screening Bacterial Colonies for pCR®- Blunt Transformants 151
3.2.21 Subcloning Upstream Regulatory Regions into pALuc 151
3.2.22 Polymerase Chain Reaction for Mycoplasma 151
3.2.23 Electrophoretic Mobility Shift Assays (EMSAs) 154
3.2.24 Strategy for Electrophoretic Mobility Shift Assays 156
3.2.24.1 Preparation of HeLa Nuclear Extract 156
3.2.24.2 Preparation of 5% Acylamide Gel in Low Ionic x0.5 TBE Buffer 157
3.2.24.3 Oligonucleotides 157
3.2.24.4 Labelling Oligonucleotides with dATP $\gamma$-$^{32}$P 159
3.2.24.5 Annealing the Oligonucleotides 159

3.2.24.6 Purification of Radioactive Probes Using NAP 5 Columns 159
3.2.24.7 Measuring Specific Activity of Probes 159
3.2.25 Oct-1/PEF-1 Electrophoretic Mobility Shift Assay Protocol 160
3.2.25.1 Preparation of "Mix" 160
3.2.25.2 Reaction Setup 160
3.2.26 Oct-1, YY1, PSM and Sp1 Electrophoretic Mobility Shift Assay Protocol 160
3.2.26.1 Preparation of "Mix" 160
3.2.26.2 Reaction Setup 161
3.2.26.3 Competition Assays 161
3.2.26.4 Supershifts 161
3.2.27 Visualising and Measuring Complexes 162
3.2.27.1 Electrophoresis 162
3.2.27.2 Fixing and Exposing the Gel to X-Ray Film 162
3.2.27.3 Measuring Intensity of Bands on Autoradiographs 162

**3.3 RESULTS** 163
3.3.1 Cloning Human Papillomavirus Type 16 Upstream Regulatory Regions 163
3.3.1.1 Technical Considerations 163
3.3.1.2 Luciferase Assays 165
3.3.2 Investigations to Identify Sequence Changes Mediating Increased Promoter Activity 168
3.3.2.1 Technical Considerations 168
3.3.2.2 Luciferase Assays Using Constructs Created by Site-Directed Mutagenesis 171
3.3.3 Electrophoretic Mobility Shift Assays (EMSAs) 174
3.3.3.1 Technical Considerations 174
3.3.3.2 Nucleotide 7792 at the Periphery of a YY1 Binding Site (nts 7790-7804) 175
3.3.3.3 Nucleotide 7676 in an Overlapping Oct-1/PEF-1 Binding Site (nts 7670-7678 and 7675-7685 Respectively) 178
3.3.3.4 Nucleotide 7894 in the PSM (nts 7893-7901 and 3-11) 180
3.3.3.5 Nucleotide 24 Adjacent to an Sp1 Binding Site (nts 28-33) 182

**3.4 DISCUSSION** 184
3.4.1 The Potential Significance of Sequence Variation in the Upstream Regulatory Region 184

## CHAPTER 4

# DETERMINATION OF THE PHYSICAL STATE OF HUMAN PAPILLOMAVIRUS TYPE 16 DNA IN CERVICAL CANCERS AND SEQUENCE ANALYSIS OF THE E2 AND E4 GENES


**4.1 INTRODUCTION** 193

**4.2 METHODS** 195

4.2.1 Determining the Physical State of Human Papillomavirus DNA 195

4.2.2 Southern Blot Hybridisation 195

4.2.2.1 Clinical Isolates 196

4.2.2.2 Restriction Endonuclease Digestion of DNA 196

4.2.2.3 Rapid Alkaline Transfer 196

4.2.2.4 Preparing Human Papillomavirus Type 16 Probe DNA 198

4.2.2.5 Labelling the Probe Using a Multiprime DNA Labelling Kit (Amersham) 198

4.2.2.6 Separation of Labelled Human Papillomavirus Type 16 DNA from Unincorporated dCTP $\alpha$-$^{32}$P 199

4.2.2.7 Measuring Specific Activity of the Probe 200

4.2.2.8 DNA Prehybridisation 200

4.2.2.9 DNA Hybridisation 200

4.2.2.10 Washing Membranes and Autoradiography 201

4.2.3 Polymerase Chain Reaction of the E2 Gene for Physical State Determination 202

4.2.3.1 Strategy 202

4.2.3.2 Clinical Isolates 202

4.2.3.3 Primers 202

4.2.4 Optimised E2 Polymerase Chain Reaction Protocols 203

4.2.5 Analysis of Full Length E2 Polymerase Chain Reaction Products by *Hinc* II Digestion 206

4.2.6 Sites of Disruption in E2 Genes 206

4.2.7 Polymerase Chain Reaction of the E2 and E4 Genes for Sequence Analysis 207

4.2.7.1 Strategy 207

4.2.7.2 DNA Sequencing 208

4.2.7.3 E2 and E4 Protein Analysis 208

**4.3 RESULTS** 209

4.3.1 The Physical State of Human Papillomavirus DNA by Southern Blot Hybridisation 209

4.3.1.1 Technical Considerations 209

4.3.1.2 Results by Southern Hybridisation 209

4.3.2 The Physical State of Human papillomavirus DNA by Polymerase Chain Reaction of the E2 Gene 212

4.3.2.1 Sensitivity Assay for the E2 Gene Polymerase Chain Reaction 212

4.3.2.2 Results for E2 Polymerase Chain Reaction 213

4.3.2.3 Restriction Endonuclease Analysis of Polymerase Chain Reaction Products 213

4.3.3 Overall Results for Physical State 213

4.3.4 Site of Disruption of E2 in Integrated Papillomavirus Genomes 214

4.3.5 Sequence Variation in the E2 and E4 Genes 215

4.3.5.1 Amino Acid Analysis of the E2 Protein 217

4.3.5.2 Amino Acid Analysis of the E4 Protein 218

4.3.5.3 Sequence Analysis of Noncancer Human Papillomavirus Type 16 Isolates 220

**4.4 DISCUSSION** 222

4.4.1 The Physical State of Human Papillomavirus Type 16 DNA in Cervical Cancers 222

4.4.2 Disruption of the E2 Gene 225

4.4.3 The Sequence of the E2 and E4 Genes 227

4.4.4 The Significance of E2 Sequence Variants 227

**CHAPTER 5**

**SEQUENCE ANALYSIS OF THE HUMAN PAPILLOMAVIRUS TYPE 16 E6 ONCOGENE**

**5.1 INTRODUCTION** 232

**5.2 METHODS** 234

5.2.1 Clinical Isolates 234

5.2.2 Polymerase Chain Reaction of the E6 Gene 234

5.2.3 DNA Sequencing 234

5.2.4 E6 Protein Analysis 236

**5.3 RESULTS** 237

5.3.1 E6 Polymerase Chain Reaction 237

5.3.1.1 General Polymerase Chain Reaction Results 237

5.3.2 Sequence Variation in the E6 Gene 237

5.3.2.1 Amino Acid Analysis 240

**5.4 DISCUSSION** 243

5.4.1 The Sequence of the E6 Gene 243

5.4.2 Potential Functional Significance of E6 Variation 244

**CHAPTER 6**

**GENERAL DISCUSSION AND FUTURE DIRECTIONS**

**6.1 GENERAL DISCUSSION** 250

**6.2 FUTURE DIRECTIONS** 254

**CHAPTER 7**

**APPENDICES**

**A7.1 SOLUTIONS** 257

**A7.2 TECHNIQUES** 264

A7.2.1 Nucleic Acid Quantitation 264

A7.2.2 Guidelines for the Design of Primers for Polymerase Chain Reaction 264

A7.2.3 Agarose Gel Electrophoresis 265

A7.2.3.1 Casting a Gel 265

A7.2.4 Long Term Storage of Plasmid Constructs 266

A7.2.5 Maintaining Cell Lines 266

A7.2.6 Removing Cells from Flasks 267

A7.2.7 Storing Cells in Liquid Nitrogen 267

A7.2.8 Counting Cells 268

**A7.3 SIZE MARKERS FOR GEL ELECTROPHORESIS** 269

**A7.4 SEQUENCING METHODS** 270

A7.4.1 Sequencing Using ABI PRISM® Big Dye™ Terminator Chemistry (PE Applied Biosystems) 270

A7.4.2 Cycle Sequencing 270

A7.4.3 Preparation of the Reaction Mix 271

A7.4.4 Cycling Conditions 271
A7.4.5 Purifying Extension Products to Remove Excess DyeDeoxy Terminators 272
A7.4.6 Preparing/Loading Samples and Analysis on the Applied Biosystems 377 DNA Sequencer 272
**A7.5 SELECTED SEQUENCE ELECTROPHEROGRAMS** 273
**A7.6 SEQUENCE ALIGNMENTS FOR THE HUMAN PAPILLOMAVIRUS TYPE 16 URR, E2/E4 AND E6 GENES** 282
A7.6.1 URR 282
A7.6.2 E2/E4 Genes and Proteins 295
A7.6.3 E6 Gene and Protein 308

**CHAPTER 8**

**REFERENCES**

**8.1 REFERENCES** 316

# LIST OF TABLES


Page Number

## CHAPTER 1

Table 1.1: Milestones in Papillomavirus Research 5

Table 1.2: Clinical Lesions Associated with Human Papillomavirus Types 13

Table 1.3: Physical State of High Risk Genital Human Papillomavirus DNA in Cervical Lesions 28

Table 1.4: Binding Sites in the Human Papillomavirus 16 Upstream Regulatory Region 64

Table 1.5: A Comparison of Viral Characteristics in Benign versus Malignant Cervical Disease 70

Table 1.6: General Guidelines for International Federation of Gynecologists and Obstetricians Staging 71

Table 1.7: Histological Classification of Malignant Cervical Cancers 72

Table 1.8: Overview of the Advantages and Disadvantages of Methods for Detecting Human Papillomavirus Infection 78

Table 1.9: Historical Overview of Management and Treatment of Cervical Precancer and Cancer 80

## CHAPTER 2

Table 2.1: Clinicopathological Variables 91

Table 2.2: Primers Used for Polymerase Chain Reaction of the Human Papillomavirus Type 16 Upstream Regulatory Region 99

Table 2.3: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-1 and URR-2, or Primers URR-1 and URR-4 102

Table 2.4: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-3 and URR-4 103

Table 2.5: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-3 and URR-2 104

Table 2.6: Hot Start Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-5 and URR-2 105

Table 2.7: Hot Start *UlTma*™ Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-1and URR-2 106

## CHAPTER 3


Table 3.1: Nucleotides Altered by Site-Directed Mutagenesis and Resulting Constructs 146

Table 3.2: Polymerase Chain Reaction Primers Used for Site-Directed Mutagenesis 147

Table 3.3: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Site-Directed Mutagenesis 148

Table 3.4: Mycoplasma Polymerase Chain Reaction Primers 152

Table 3.5: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Detection of Mycoplasma 153

Table 3.6: Oligonucleotides Used for Electrophoretic Mobility Shift Assays 158

Table 3.7: Raw Luciferase Values in HeLa Cells 166

Table 3.8: Raw Luciferase Values in HT3 Cells 166

Table 3.9: Final Values for Constructs in Both HeLa and HT3 166

Table 3.10: Luciferase Values in HeLa Cells for Constructs Created by Site-Directed Mutagenesis 172

Table 3.11: Final Values in HeLa Cells for Site-Directed Mutagenesis Constructs 172


## CHAPTER 4


Table 4.1: Primers Used for Polymerase Chain Reaction of the Human Papillomavirus Type 16 E2 Gene 203

Table 4.2: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers E2-1 and E2-2 (Full Length E2 Gene Product) 204

Table 4.3: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers E2-3 and E2-2 205

Table 4.4: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers E2-1 and E2-4 205

Table 4.5: The Physical State of Human Papillomavirus DNA in Each Isolate 211

Table 4.6: Amplification of Integrated Human Papillomavirus Type 16 DNA for Different Regions of the E2 Gene 215

Table 4.7: Amino Acid Changes in the E2 Protein 218

Table 4.8: Amino Acid Changes in the E4 Protein 219

Table 4.9: Sequence Variation in the Human Papillomavirus Type 16 E2 Gene from Noncancer Isolates 221

Table 4.10: Sequence Variation in the Human Papillomavirus Type 16 Upstream Regulatory Region from Noncancer Isolates 221

## CHAPTER 5

Table 5.1: Primers for Polymerase Chain Reaction of the E6 Gene 234

Table 5.2: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers E6-1 and E6-2 (E6 Gene Product) 235

Table 5.3: Amino Acid Changes in the E6 Protein 240

Table 5.4: Significance of Changes in Amino Acids in the E6 Protein in Terms of Hydrophilicity, Surface Probability, Flexibility, and Antigenic Index. 242

## CHAPTER 7

Table A7.3.1: Fragment Sizes for DNA Markers Used in Gel Electrophoresis 269

Table A7.4.1: Reagent Mix for Cycle Sequencing Reaction 271

Table A7.4.2: Cycling Conditions 271

Table A7.5.1: Primers and Corresponding Examples of Sequence Electropherograms 273

# LIST OF FIGURES


Page Number

## CHAPTER 1

Figure 1.1: Computerised diagram of bovine and human papillomavirus particles 9

Figure 1.2: Organisation of the human papillomavirus type 16 genome 10

Figure 1.3: The most common human papillomavirus types found in cervical cancer around the world 14

Figure 1.4: Illustration of the epidermal layer and stages of productive infection by human papillomavirus 23

Figure 1.5: Structure and function of the bovine and human papillomavirus E2 proteins 33

Figure 1.6: Cell cycle and disruption by the E6 and E7 proteins of carcinogenic human papillomavirus types 39

Figure 1.7: The human papillomavirus type 16 E6 protein 41

Figure 1.8: The human papillomavirus type 16 E7 protein 44

Figure 1.9: Expanded view of the human papillomavirus type 16 upstream regulatory region showing transcription factor binding sites 50

Figure 1.10: a) Wart on the finger, b) Plantar wart on the sole of the foot, c) Anal warts 66

## CHAPTER 2

Figure 2.1: Principles of the polymerase chain reaction 94

Figure 2.2: Different size polymerase chain reaction products from the upstream regulatory region 110

Figure 2.3: The human papillomavirus type 16 upstream regulatory region illustrating the extent of sequencing data obtained using each of the primers 111

Figure 2.4: Sequence variation in the human papillomavirus type 16 upstream regulatory region 114

Figure 2.5: A comparison of human papillomavirus type 16 variant lineages in cervical carcinomas from different regions of the world 115

## CHAPTER 3

Figure 3.1: The cloning vector pCR®-Blunt showing the location of genes and the multicloning site 124

Figure 3.2: Creation of bioluminescence by luciferase 131
Figure 3.3: The cloning vector pALuc showing relevant genes and the multicloning site 132
Figure 3.4: pCH110 β-galactosidase expression vector showing genes and restriction endonuclease sites 138
Figure 3.5: Introduction of a point mutation into the upstream regulatory region 144
Figure 3.6: Flow diagram protocol for site-directed mutagenesis 146
Figure 3.7: The electrophoretic mobility shift assay 155
Figure 3.8: pALuc constructs digested with *Bam*H I to determine the presence of a successfully cloned upstream regulatory region 163
Figure 3.9: pALuc constructs digested with *Hae* II to determine which contain upstream regulatory regions cloned in the correct orientation 164
Figure 3.10: pALuc constructs pALucS1, pALucH1, pALucK2, pALucK4, pALucO2, and pALucR1, digested with *Bam*H I to release the upstream regulatory region 165
Figure 3.11: Luciferase activity under the control of the selected (H1, K2, K4, O2, and R1) human papillomavirus type 16 upstream regulatory regions 167
Figure 3.12: Schematic representation of individual sequence mutations created in each of the K2, K4, and H1 constructs in the context of the transcription factor binding sites in which they were located 169
Figure 3.13: Example of polymerase chain reaction products used for site-directed mutagenesis cloning 170
Figure 3.14: Expression of luciferase under the control of human papillomavirus type 16 upstream regulatory regions in HeLa cells 173
Figure 3.15: Electrophoretic mobility shift assay autoradiograph showing the binding of HeLa nuclear extract to $^{32}$P-labelled prototype YY1 oligonucleotide (nts 7790-7804) 176
Figure 3.16: Percentage of bound YY1 to unlabelled prototype (16P7792) or mutant (16M7792) oligonucleotides derived from Molecular Analyst™ software on the phosphoimager 177
Figure 3.17: Electrophoretic mobility shift assay autoradiograph showing the binding of HeLa nuclear extract to $^{32}$P-labelled prototype Oct-1/PEF-1 oligonucleotide (nts 7667-7693) 179
Figure 3.18: Electrophoretic mobility shift assay autoradiograph showing the binding of HeLa nuclear extract to $^{32}$P-labelled prototype PSM oligonucleotide (nts 7893-15) 181
Figure 3.19: Electrophoretic mobility shift assay autoradiograph showing the binding of HeLa nuclear extract to $^{32}$P-labelled prototype Sp1 oligonucleotide (nts 20-38) 183

## CHAPTER 4

Figure 4.1: Diagrammatic representation of the Southern hybridisation apparatus 198
Figure 4.2: Location of primers for amplification of the E2 gene and corresponding products that can be amplified when segments of E2 are disrupted 207
Figure 4.3: An example of cellular nucleic acids digested with *Bgl* II or *Bam*H I, prior to alkaline transfer for Southern Hybridisation 210
Figure 4.4: Southern hybridisation demonstrating integrated and episomal forms of human papillomavirus DNA 210
Figure 4.5: Sensitivity assay for E2 polymerase chain reaction (using primers E2-1 and E2-1) 212
Figure 4.6: The physical state of human papillomavirus type 16 DNA in 34 cervical cancer isolates from Australia and the South Pacific 214
Figure 4.7: Sequence variation in the human papillomavirus 16 E2 gene from cervical cancer isolates 216
Figure 4.8: Distribution of amino acid changes in the E2 protein 217
Figure 4.9: Distribution of amino acid changes in the E4 protein 220

## CHAPTER 5

Figure 5.1: An example of E6 polymerase chain reaction products 237
Figure 5.2: Sequence variation in the human papillomavirus type 16 E6 gene 239
Figure 5.3: Distribution of amino acid changes in the E6 protein 241

## CHAPTER 7

Figure A7.2.1: The squares counted in a Neubauer chamber in order to determine the number of cells in a suspension 268
Figure A7.5.1: Representative DNA sequence electropherogram for primer URR-1 274
Figure A7.5.2: Representative DNA sequence electropherogram for primer URR-2 275
Figure A7.5.3: Representative DNA sequence electropherogram for primer URR-5 276
Figure A7.5.4: Representative DNA sequence electropherogram for primer URR-4 277
Figure A7.5.5: Representative DNA sequence electropherogram for primer E6-1 278
Figure A7.5.6: Representative DNA sequence electropherogram for primer E6-2 279
Figure A7.5.7: Representative DNA sequence electropherogram for primer E2-1 280
Figure A7.5.8: Representative DNA sequence electropherogram for primer E2-2 281

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AA | Asian American |
| AAV | Adeno-associated virus |
| Af | African |
| ALV | Avian leukosis virus |
| ANGIS | Australian National Genomic Information Service |
| AP | Alkaline transferase |
| AP1 | Activator protein 1 |
| As | Asian |
| ASCUS | Atypical squamous cells of undetermined significance |
| ATP | Adenosine triphosphate |
| ATPase | Adenosine triphosphatase |
| bp | Base pairs |
| BPV | Bovine papillomavirus |
| BRK | Baby rat kidney cells |
| BSA | Bovine serum albumin |
| C- | Carboxyl- |
| CAT | Chloramphenicol acetyltransferase |
| CBP | cAMP response element-binding protein-binding protein/ or CREB-binding protein |
| CDC | Centers for Disease Control and Prevention |
| CDK | Cyclin-dependent kinase |
| CDP | CCAAT displacement protein |
| c/EBP | CCAAT enhancer binding protein |
| CIN | Cervical intraepithelial neoplasia |
| CK | Cytokeratin |
| CKII | Casein kinase II |
| CMI | Cell-mediated immunity/ immune |
| CMV | Cytomegalovirus |
| COPV | Canine oral papillomavirus |
| CREB | cAMP response element-binding protein |
| CRPV | Cottontail rabbit papillomavirus |
| CTL | Cytotoxic T lymphocyte |
| CTF | CCAAT transcription factor |
| cys | Cysteine |
| Da | Dalton |

| | |
|---|---|
| dATP | 2'-Deoxyadenosine-5'-triphosphate, disodium |
| dCTP | 2' -Deoxycytidine-5'-triphosphate, triphosphate |
| ddNTP | Dideoxyribonucleoside triphosphate |
| dGTP | 2' -Deoxyguanosine-5'-triphosphate, trisodium |
| DMEM | Dulbecco's modified Eagle's medium |
| DNA | Deoxyribonucleic acid |
| DNase | Deoxyribonuclease |
| dNTP | Deoxyribonucleoside triphosphate |
| DPV | Deer papillomavirus |
| ds | Double stranded |
| DTT | Dithiothreitol |
| dTTP | 2' -deoxythymidine-5'-triphosphate, trisodium |
| E | Early/ or European |
| E2BS | E2 binding site |
| E2F | Elongation factor 2 |
| E2TA | E2 Transcriptional activator |
| E2TR | E2 Transcriptional repressor |
| E6-AP | E6-activator protein |
| E6BP | E6 binding protein |
| EBV | Epstein-Barr virus |
| *E.coli* | *Escherichia coli* |
| EDTA | Ethylenediaminetetraacetic acid |
| EEPV | European elk papillomavirus |
| EGF | Epidermal growth factor |
| EGFR | Epidermal growth factor receptor |
| EGFRE | EGF response element |
| EV | Epidermodysplasia verruciformis |
| FDA | Food and Drug Administration |
| FIGO | International Federation of Gynecologists and Obstetricians |
| GR | Glucocorticoid receptor |
| GRE | Glucocorticoid response element |
| HBSS | HEPES buffered saline solution |
| HBV | Hepatitis B virus |
| HC | Hybrid Capture |
| HCMV | Human cytomegalovirus |
| HG SIL | High grade squamous intraepithelial neoplasia |
| HIV | Human immunodeficiency virus |
| HLA | Human leucocyte antigen |
| HPV | Human papillomavirus |
| HSS | Hypersensitive site |

| | |
|---|---|
| HSV | Herpes simplex virus |
| IBSCC | International Biological Study of Cervical Cancer |
| IFN | Interferon |
| IL | Interleukin |
| K | Keratinising |
| k | Kilo |
| kbp | Kilobase Pairs |
| L | Late |
| LB | Luria-Bertani |
| LCR | Long control region |
| LG SIL | Low grade squamous intraepithelial lesion |
| Luc | Luciferase |
| MOPS | 3-(N-Morpholino) propane-sulphonic Acid |
| mRNA | Messenger ribonucleic acid |
| N- | Amino- |
| n | Nano |
| NA | North American |
| NCR | Noncoding region |
| NFA | NFI-associated |
| NFI | Nuclear factor I |
| NH&MRC | National Health and Medical Research Council |
| NIH | National Institute of Health (USA) |
| N/K | Non-keratinising |
| NSW | New South Wales |
| Oct | Octamer binding factor |
| OD | Optical Density |
| ONPG | o-nitrophenyl-β-D-galactopyranoside |
| ORF | Open reading frame |
| Ori | Origin of replication |
| p97 | HPV 16 early promoter |
| PAIN | Perianal intraepithelial neoplasia |
| Pap | Papanicolaou |
| PCR | Polymerase chain reaction |
| PEF-1 | Papilloma enhancer binding factor-1 |
| PEG | Polyethylene glycol |
| PIN | Penile intraepithelial neoplasia |
| PMSF | Phenylmethylsulfonyl fluoride |
| PR | Progesterone receptor |
| PSM | Papillomavirus silencing motif |
| PSM-BP | Papillomavirus silencing motif-binding protein |

| | |
|---|---|
| PV | Papillomavirus |
| Rb | Retinoblastoma protein |
| RNA | Ribonucleic acid |
| RNase | Ribonuclease |
| RPAH | Royal Prince Alfred Hospital |
| rpm | Revolutions per minute |
| RRP | Recurrent respiratory papillomatosis |
| RSV | Rous sarcoma virus |
| RT-PCR | Reverse transcriptase-PCR |
| SAP | Sheep alkaline phosphatase |
| SCC | Squamous cell carcinoma |
| SCJ | Squamo-columnar junction |
| SDS | Sodium dodecyl sulphate |
| Sp1 | Promoter specific factor 1 |
| SSC | Sodium citrate sodium chloride |
| STD | Sexually transmitted disease |
| SV40 | Simian virus 40 |
| TA | Transactivator |
| TAE | Tris acetate electrophoresis |
| TBE | Tris borate electrophoresis |
| TBP | TATA-binding protein |
| TBS | The Bethesda System |
| TE | Hydroxymethyl methylamine ethylene diamine tetra acetic acid |
| TEF-1 | Transcriptional enhancer factor-1 |
| TF | Transcription factor |
| TFIID | Transcription factor IID |
| TFBS | Transcription factor binding site |
| TGF | Transforming growth factor |
| Th | T-helper |
| $T_m$ | Melting temperature |
| TNF | Tumour necrosis factor |
| TPA | 12-o-tetradecanoyl phorbol-13-acetate |
| Tris | Hydroxymethyl methylamine |
| *Tth* | *Thermus* |
| U | Units |
| µ | Micro |
| URR | Upstream regulatory region |
| USA | United States of America |
| UV | Ultraviolet |
| V | Volts |

| | |
|---|---|
| v/v | Volume per volume |
| VAIN | Vagina intraepithelial neoplasia |
| VIN | Vulvar intraepithelial neoplasia |
| VLP | Virus-like particle |
| WHO | World Health Organisation |
| YY1 | Yin and yang 1 |

# CHAPTER 1

# REVIEW OF THE LITERATURE

# 1.1 GENERAL INTRODUCTION

Cervical cancer, the end-point in a continuum of cervical abnormalities known as cervical intraepithelial neoplasia (CIN), is a major health problem for women worldwide, second only to breast cancer in both incidence and mortality (National Institutes of Health [NIH] consensus statement, 1997). Even though it is potentially preventable, cervical cancer accounts for approximately 6% of all human cancers (zur Hausen and de Villiers, 1994), and is the most common cancer in the developing countries of the world.

Human papillomaviruses (HPVs) have long been recognised as the cause of benign cutaneous warts, a universal human affliction, so the discovery of the association of HPV with cervical cancer, and other anogenital cancers, was initially somewhat surprising. Cervical cancer is one of the few human cancers with well-defined virological associations. Both the cancer and its precursor lesions contain HPV DNA, and in 1996 the World Health Organisation (WHO) declared that HPV types 16 and 18 are carcinogens (Major, 1996). Other DNA viruses known to be aetiological agents of human cancer include hepatitis B virus (HBV) causing liver cancer, and Epstein-Barr virus (EBV) causing Burkitt's lymphoma and nasopharyngeal carcinoma. The process that leads to these cancers appears to result not only from the presence of an oncogenic virus but also from the interaction of the virus with environmental, immunologic and genetic factors. HPVs are found throughout the world and most people become infected at some time during their lives, even if the infection is not clinically apparent. HPV is believed to be the most common sexually transmitted virus in the world (Jenison *et al.*, 1990, Galloway, 1994, and Lowy *et al.*, 1994). The majority of clinical HPV infections are benign, and it is only on rare occasions that infection results in carcinoma. Sadly, even with adequate intervention, many patients still die from HPV-related cancer, especially cervical cancer. The past two decades have seen enormous progress in our understanding of papillomavirus (PV) biology, particularly due to the advent of molecular technology, and it is hoped that further research will continue to increase understanding, and new and more effective detection procedures and treatments will result.

# 1.2 CERVICAL CANCER

## 1.2.1 Incidence of Cervical Cancer

Worldwide, greater than 471,000 cases of cervical cancer are diagnosed each year (NIH consensus statement, 1997). The highest rates are found in the countries of Central and South America (up to 48 cases per 100,000 in some), Africa, and Asia (excluding Japan) (Armstrong *et al.,* 1992). Lower rates are found in Southern Europe, North America, Australia, New Zealand, and Israel, which are attributed in part to the effectiveness and availability of Papanicolaou screening programs introduced in the 1950's and 1960's.

Australia reported rates in the order of 10 cases of cervical cancer per 100,000 during the 1990's, although ethnic minority groups such as the Australian aborigines have considerably higher rates (Cancer Epidemiology Research Unit of the New South Wales [NSW] Central Cancer Registry, 1992, and the NSW Cancer Council, 1996). This amounts to approximately 1,000 cases of cervical cancer which are diagnosed each year and 350 deaths annually (Australasian Institute of Health and Welfare and Australasian Association of Cancer Registries, 1994, and The National Cervical Screening Program, 1998), making cervical cancer the eighth most common cancer in Australian women (Jelfs, 1995). In the state of NSW, cervical cancer accounts for approximately 2.8% of all cancers in women (NSW Cancer Council, 1996).

The incidence of cervical cancer increases from age 20 years to a peak at 40 to 60 years and thereafter slowly declines (Barber, 1982, and Singer *et al.,* 1995). Intraepithelial lesions, the precursors to invasive cervical disease, frequently occur in younger women (Cramer and Cutler, 1974), which is consistent with the notion that malignant disease of squamous epithelial cells requires a long latency period (Cannistra and Niloff, 1995). The lifetime risk of cervical cancer is approximately 0.3% to 0.5% in women who are screened frequently by Papanicoloau smear and approximately 3% to 4% in unscreened populations (Muir *et al.,* 1987, Franco, 1996, and Myers *et al.,* 2000).

### 1.2.2 Cervical Cancer as a Sexually Transmissible Disease (STD)

The first clues that cervical cancer might be the result of a sexually transmitted infection came from the epidemiological observations of Rigoni-Stern who reported that the incidence of cervical cancer in prostitutes and married women was substantially higher than that in nuns and virgins (Rigoni-Stern, 1842). For the greatest part of the 20th century different microbiological agents were suggested as causative agents of cervical cancer, although there was little evidence to implicate most of them. These included: *Treponema pallidum*, *Neisseria gonorrhoeae*, *Chlamydia trachomatis*, Herpes Simplex virus (HSV), Cytomegalovirus (CMV), and EBV. From the 1960's to the 1970's, HSV type 2 was thought to be the cause of cervical cancer with some evidence to support its role, particularly as the oncogenic ability of HSV was realised (zur Hausen, 1982). However, HSV DNA could not be demonstrated consistently in cervical tumours and a research group lead by zur Hausen (zur Hausen, 1986) hypothesised that the papilloma virus found in genital warts would be a more likely candidate (zur Hausen, 1975 and 1976). The identification of PVs associated with animal cancers, eg. Shope, and the pathological association of squamous cell carcinoma (SCC) arising in warts with the frequent co-occurrence of koilocytes (see section 1.6.2.2) and dysplastic cells, drew attention to the possibility that PVs might be involved in cervical cancer. The infectious nature of skin warts was known from observations made early in the 20th century (Ciuffo, 1907), but it was the advent of electron microscopy and DNA hybridisation that finally permitted demonstration of PV-like particles in nuclear basophilic inclusions in cells (Della Torre *et al.*, 1978, and Hills and Laverty, 1979). Early attempts to culture PVs *in vitro* though were unsuccessful, resulting in the dormancy of PV research until the advent of molecular biology in the 1970's. HPV types 16 and 18 were subsequently discovered in 1983 and 1984 (Durst *et al.*, 1983 and Boshart *et al.*, 1984). A summary of some of the milestones in PV research is given in Table 1.1.

Table 1.1: Milestones in Papillomavirus Research

(Adapted from Gross, 1990, and zur Hausen, 1994)

| YEAR | MILESTONE |
|---|---|
| 1842 | Incidence of cervical cancer depends on sexual contact and promiscuity (Rigoni-Stern) |
| 1907 | The viral nature of human skin warts was demonstrated by transmission studies (Ciuffo) |
| 1933 | The cottontail-rabbit PV (CRPV) was isolated and characterised (Shope and Hurst) |
| 1934 | The development of rabbit skin carcinomas as a synergistic interaction between CRPV and cocarcinogens was demonstrated (Rous and Beard) |
| 1949 | HPV identified as a viral particle by Strauss *et al.*, from skin papillomas |
| 1963-1965 | The structure and molecular weight of PV DNA from human warts was determined (Crawford, 1963 and 1965) |
| 1969 | Viral particles in human genital warts detected by electron microscopy (Almeida *et al.* and Dunn and Ogilvie) |
| 1973 | Hypothesis that genital cancers are caused by HPV infections (zur Hausen, 1975 and 1976) |
| 1974 | Initial hybridisation studies with cRNA probes derived from a plantar wart showed that different HPV types are responsible for common warts and for genital warts (zur Hausen *et al.*) |
| 1976 | Association made between PVs and cervical dysplasia by cytologic observation of koilocytotic cells (Meisels *et al.*) |
| 1977 | Existence of warts converting into SCCs and the suggestion of a role for specific viral genes in malignant transition (zur Hausen) |
| 1978-1979 | Electron microscopic demonstration of PV particles in cervical lesions (Della Torre *et al.*, 1978, and Hills and Laverty, 1979) |
| 1983 | Characterisation of HPV 16 DNA and identification in cervical carcinoma biopsies (Dürst *et al.*) |
| 1984 | Characterisation of HPV 18 DNA (Boshart *et al.*) |
| 1985 | Identification of specific viral transcripts in cervical cancer cells and of a specific mode of viral DNA integration (Schwarz *et al.*, and Yee *et al.*) |
| 1986 | Characterisation of HPV 31 and HPV 33 (Lörincz *et al.*, and Beaudenon *et al.*) |
| 1986 | Malignant transformation of NIH 3T3 cells by HPV 16 DNA (Yasumoto *et al.*) |
| 1987 | Immortalisation of human foreskin keratinocytes by HPV 16 (Dürst *et al.*, and Pirisi *et al.*) |
| 1988-1992 | Epidemiological studies reveal high risk HPV infections as major risk factors for cervical cancer development (Koutsky *et al.*, 1988, and Muñoz *et al.*, 1992a) |
| 1988-1992 | Malignant phenotype of human cervical carcinoma cells and malignant transformation of rodent cells depends on HPV E6/E7 gene expression (von Knebel Doeberitz *et al.*, 1988 and 1992) |
| 1989 | E7 gene product binds to the Rb protein (Dyson *et al.*, and Münger et al, 1989a) |
| 1990 | E6 gene product binds to the p53 protein (Werness *et al.*) |
| 1992 | E7 disrupts pRB interaction with cellular transcription factor E2F (Chellappan *et al.)* |
| 1997 | Alpha-6 integrin on cell membrane identified as a putative receptor molecule for HPV entry (Evander *et al.*) |
| 1999 | Almost 100% HPV DNA detectability in cervical cancers (Walboomers *et al.*) |

The association of HPV with cervical cancer was at first controversial. Extensive information has since accumulated from epidemiological and experimental studies, particularly as a result of the introduction of standardised methods such as PCR and sequence analysis that support HPV, in particular types 16 and 18, as the major cause of cervical cancer. In 1995 the International Biological Study of Cervical Cancer

(IBSCC) reported that HPV DNA is detectable in more than 93% of invasive cervical cancers worldwide (Bosch *et al.*, 1995). More recently, Walboomers and colleagues reassessed these results and claimed that the overall prevalence of HPV in cervical cancer is actually 99.7% (Walboomers *et al.*, 1999). Original estimates had been inaccurate due to a lack of sensitivity or specificity, from sampling inadequacy, or as a result of the detection system used.

More then 20 HPV types have now been associated with invasive tumours, with HPV 16 being the most prevalent. HPV negative carcinomas are extremely uncommon. Where they do seem to exist, a 'hit-and-run' mechanism is generally considered unlikely, as transcription of the viral oncogenes, E6 and E7, is necessary to maintain the transformed state (Crook *et al.*, 1989, and von Knebel-Doeberitz *et al.*, 1994). However, the possibility cannot be completely discounted and could potentially occur if the p53 and pRb pathways (see sections 1.5.3.2 and 1.5.3.3) suffered secondary genetic changes, removing the requirement for oncogene activity (Herrington, 1999). In addition, mutations of the cellular tumour suppressor protein p53, have been identified in HPV negative cervical carcinoma cell lines, for example HT3 (Crook *et al.*, 1991a).

### 1.2.3 Risk Factors for Cervical Cancer

Persistent infection with an oncogenic HPV type and high viral load have been established as the predominant risk factors for cervical cancer development (Munoz *et al.*, 1992, Schiffman *et al.*, 1993, Bosch *et al.*, 1995, and Ho *et al.*, 1995). Persistence of a single viral variant most commonly occurs (von Belkum *et al.*, 1995, and Yamada *et al.*, 1997) and the risk of cervical cancer in those women is over 100 times that of women negative for HPV (Lörincz *et al.*, 1992, and Munoz *et al.*, 1992b). It was also recently shown that the risk of carcinoma *in situ*, the immediate precursor to cervical cancer, is markedly increased with HPV 16 viral load (Josefsson *et al.*, 2000).

HPV stimulates cell proliferation and is responsible for the genetic instability of infected cells that leads to cancer (zur Hausen and de Villiers, 1994). Human cells

transfected with the DNA of HPV 16 or 18 may progress towards malignancy without further addition of exogenous carcinogens, merely as a function of time. However, chemical or physical carcinogens seem to accelerate the process (zur Hausen and de Villiers, 1994), consistent with the general hypothesis that carcinogenesis is a multistep process. Risk factors include:

- Cigarette smoking – a 2-fold increased risk of CIN and cancer is associated with smoking, which may be due to a reduction in the number of Langerhan's cells or constituents in the smoke that interact with HPV (Wigle *et al.*, 1980). In 1987, Schiffman and colleagues reported that nicotine can be found at high levels in the cervical mucosa (Schiffman *et al.*, 1987) and nicotine induces HPV transformed cells to proliferate (Waggoner and Wang, 1994).
- Dietary deficiencies – for example folic acid or vitamin C deficiency (Schneider and Shah, 1989).
- Oral contraceptive pill – a higher risk of cervical cancer for users of the oral contraceptive pill has been documented (Harris *et al.*, 1980, Brinton *et al.*, 1986, and Zandervan *et al.*, 1996). This could be due to the presence of progesterone receptor (PR) binding sites in the viral upstream regulatory region (see section 1.5.5 and in particular, section 1.5.10.9).
- Number of pregnancies/parity (Kjellberg *et al.*, 2000).
- Low socio-economic status – which usually means a woman is less likely to have regular Pap smears (Bosch *et al.*, 1992, and de Sanjose *et al.*, 1996).
- Immune status - a link between CIN and immunosupression is well established (Sillman *et al.*, 1984, and Carson *et al.*, 1986)
- HLA types - many examples of the association of particular HLA types with cervical cancer have been reported. For instance, the positive association of the DQB1*03 and DRB1*04 and DRB1*11 alleles with CIN and cervical cancer (Helland *et al.*, 1994, Nawa *et al.*, 1995, and Odunski *et al.*, 1996). However these findings have also been contradicted in different ethnic groups (Apple *et al.*, 1994, and Sastre-Garau *et al.*, 1996). HLA-DRB1*07 is significantly over-represented in cervical carcinoma patients whilst HLA-B15 is often under-represented (Brady *et al.*, 1999). A HPV type-specific association between HLA haplotypes and cervical lesions has also been demonstrated (Apple *et al.*, 1995).

- Telomere shortening – is normally a function of age and may contribute to senescence. Activation of telomerase restores telomere length and may contribute to carcinogenesis. Telomerase activation can be found in a high proportion of cervical cancers (Nagai *et al.*, 1999, and Zhang *et al.*, 1999).
- Number of extramarital sexual partners and number of extramarital prostitutes as sexual partners (Bosch *et al.*, 1996).

Infectious risks for cervical cancer, other than HPV, include:

- Human immunodeficiency virus (HIV)-induced immunosuppression - which exacerbates HPV-mediated cervical cytologic abnormalities. Women with HIV infection are more than 10 times as likely to develop CIN as those who are not infected (LaGuardia, 1993). Cervical neoplasia is more likely as HIV progresses, or as CD4 count falls. For example, in one study of women with CD4 counts less than 200/mm$^3$, 61% had CIN compared with a CIN prevalence of 38.4% in women with a CD4 count greater than 200/mm$^3$ (LaGuardia *et al.*, 1991). The risk associated with HIV could also be due, in part, to a direct or cytokine-mediated action (Dolei *et al.*, 1999). For instance the HIV tat gene can transactivate the HPV URR (Vernon *et al.*, 1993).
- Interaction of HPV with HSV – the presence of HSV 2 antibodies with HPV infection increases the risk of cervical cancer (Hildesheim *et al.*, 1991). Indeed, the introduction of part of the HSV genome into immortalised HPV 16 cells can induce transformation (DiPaolo *et al.*, 1990).

# 1.3 PAPILLOMAVIRUSES

## 1.3.1 Viral Structure

Papillomaviruses are small (52-55nm in diameter), non-enveloped, double-stranded DNA viruses (Figure 1.1). The structural proteins, L1 and L2, are arranged as 72 capsomeres with the overall ratio of L1:L2 being 4:1 (Baker *et al.*, 1991). The icosahedral protein capsid encloses a nucleohistone core incorporating the 8kb genome in a chromosome-like complex (Favre *et al.*, 1997a).

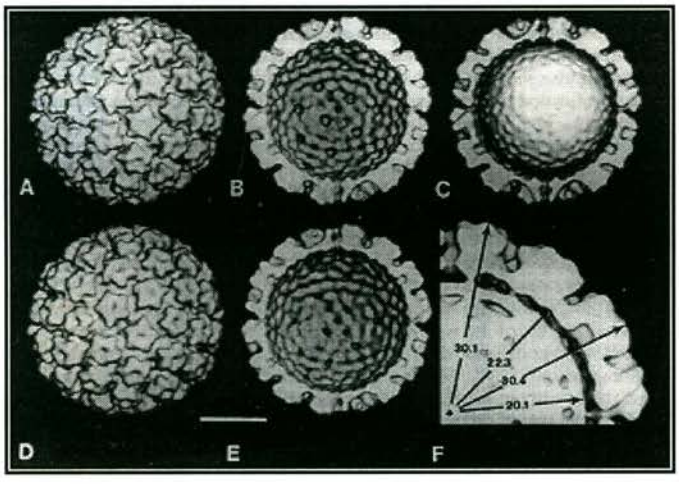



Figure 1.1: Computerised diagram of bovine and human papillomavirus particles.
(A, D) Surface shaded and (B, C, E, F) sectioned display of the bovine papillomavirus BPV 1 and human papillomavirus HPV 1 reconstructions viewed along a twofold axis of symmetry. (A-C and F) BPV 1 and (D and E) HPV 1. In B and E, the reconstructions were computationally sectioned at the equator and the nucleoprotein core was removed. In C the BPV 1 capsid shell was similarly sectioned but none of the nucleoprotein core was removed. F shows an octet of the BPV 1 reconstruction, which reveals a space between the core and capsid, and the narrow regions of density that connect the two. The bar = 20nm, Reproduced from Baker *et al.*, 1991, and used with permission).

## 1.3.2 Human Papillomavirus Genomic Organisation

PVs have a genome of approximately 8kbcomprisingeight to ten open reading frames (ORFs). These can be divided into early (E) and late (L) genes on the basis of when they are transcribed during an infection, and a noncoding region important in regulating gene expression, the upstream regulatory region (URR) (Figure 1.2).

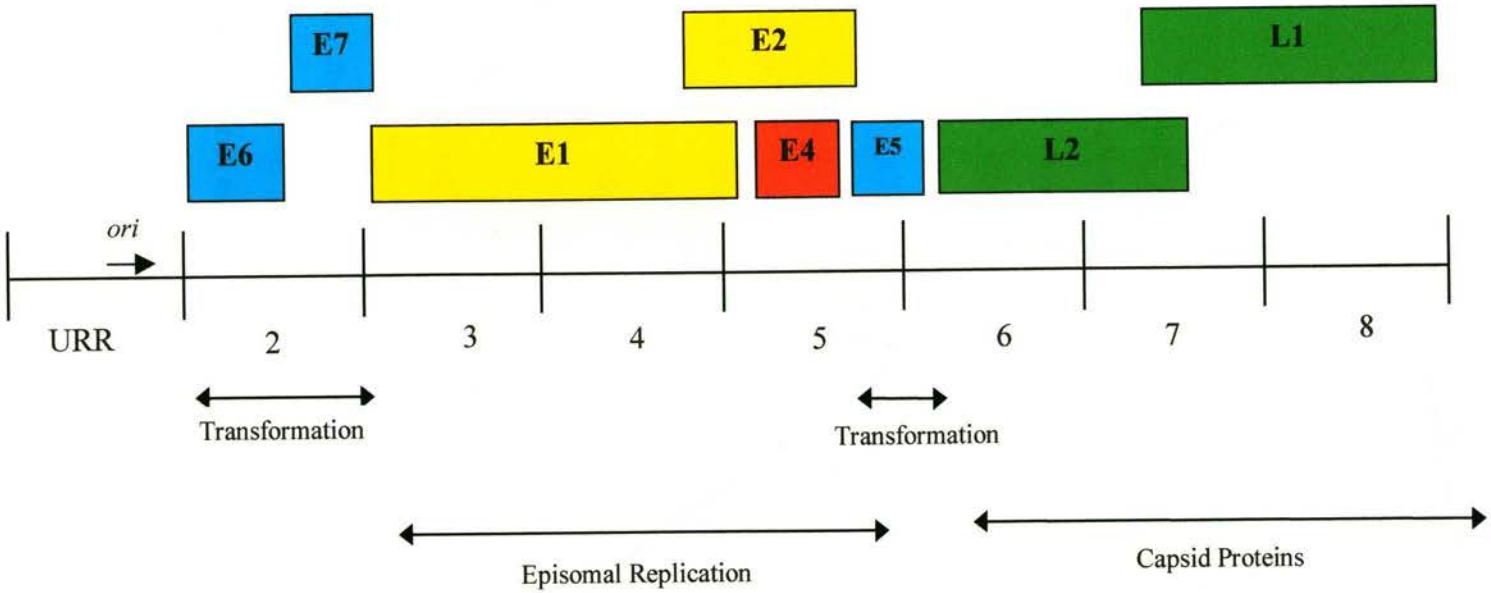



Figure 1.2: Organisation of the human papillomavirus type 16 genome
(Adapted from Pfister and Fuchs, 1994)

The genome is marked in kilobase pairs (kbp). The boxes represent open reading frames labelled 'E' or 'L' depending on their position in the 'early' or 'late' region of the genome. Basic gene functions are listed below the genome; '*ori*' represents the origin of replication (nt 61 in HPV 16).
Note: There is an early region promoter in the URR (p97 in HPV 16) and probably a late region promoter.
Box colours represent similarity of function ascribed to each gene product (as indicated by arrows beneath the figure).

### 1.3.3 Classification

On the basis of their capsid structure and nucleic acid composition, PVs constitute genus A in the Family *Papovoviridae*, Genus *Papillomavirus*. Each PV is allocated a species in the genus according to host range and is named after its natural host. For example, bovine papillomavirus (BPV), canine oral papillomavirus (COPV), cottontail rabbit papillomavirus (CRPV), and the human papillomaviruses (HPV). No interspecies transfer of HPV types has been detected.

Initial classification was based on the homology of PV DNA utilising liquid hybridisation or reassociation kinetics. The sequencing of viral genomes has now superseded this technique. HPV types are now defined by comparing DNA sequences of the L1 gene to those of all other known HPV types. If homology is less than 90%, the virus is classified as a new type; homology greater than 90% renders a viral isolate a subtype of another (de Villiers, 1994). Sequences varying from the reference strain by less than or equal to 2% are termed intratype variants (Bernard *et al.*, 1994). Genotypes are designated by number and subtypes by a letter following the chronological order of characterisation. HPV types, subtypes and variants are stable entities with no evidence for inter- or intra-typic recombinations (Favre *et al.*, 1997a).

Computer analysis of the L1 or E6 gene sequences has resulted in the distribution of the PVs into five phylogenetic supergroups that reflect tissue tropism and pathogenicity (Chan *et al.*, 1995, and Favre *et al.*, 1997a). There are 54 genital PV types, grouped into Supergroup A (Chan *et al.*, 1995). For genital HPVs, intratype variability does not exceed 2% over coding regions and 5% in the URR (Yamada *et al.*, 1995, and Stewart *et al.*, 1996).

### 1.3.4 Lesions Resulting from Infection with Different Human Papillomaviruses

At least 78 distinct HPV types have been identified and at least 40 more await full characterisation (Lörincz, 1999), the majority of which have a predilection for long-term epithelial infections. Approximately half of the identified types are mucosotropic HPVs while the rest are cutaneous or Epidermodysplasia verruciformis (EV) types (Table 1.2). Thirteen HPVs are commonly associated with cancer along with a number of rarer types, and these are collectively detectable in 95% to 100% of cervical cancer cases worldwide (Lörincz *et al.*, 1992, Lungu *et al.*, 1992, and Bosch *et al.*, 1995). However, it is difficult to classify many HPV types rigidly on the basis of their clinical manifestations, as they can be detected in diverse lesions at many body sites (Lörincz, 1999). HPV types affecting the anogenital region are classified as low, intermediate, or high risk according to their association with progression to cancer. Low risk HPV types are frequently found in low grade CIN and infrequently in cancers. For instance, HPV 6 and 11 are usually associated with exophytic warts on the external genitalia and the perianal region. Intermediate risk types are found in high grade CIN, but less frequently in cancers, while those types found frequently in high grade CIN and cancers are high risk types. HPV types 16, 18, 31, and 33 typically produce flat warts or dysplastic lesions on the cervix, subclinical infection of the penis or other anogenital sites. By far the majority of cervical cancers, and some anal, penile, vulval, and vaginal malignancies contain HPV 16 or 18.

Table 1.2: Clinical Lesions Associated with Human Papillomavirus Types (Adapted from Cossart *et al.*, 1995, Wieland and Pfister, 1997, and Vernon *et al.*, 1998)

| Lesion | Common HPV Type/s Involved |
|---|---|
| **CUTANEOUS SKIN WARTS** | |
| plantar warts | 1, 2, 4 |
| common or mosaic warts | 2, 4, 26, 27, 29, 57 |
| flat warts | 3, 10, 28, 49 |
| Butchers' warts | 7 |
| *EV benign | 2, 3, 10, 12, 15, 19, 36, 46, 47, 50 |
| EV benign and malignant | 5, 8, 9, 10, 14, 17, 20-25, 37 |
| **MUCOSAL/ANOGENITAL** | |
| Condylomata acuminata | 6, 11, 42-44, 54 |
| #CIN, VAIN, VIN, PAIN, and PIN | 6, 11, 16, 18, 30, 31, 33, 35, 39, 40, 42, 43, 51, 52, 55-59, 61, 62, 64 |
| carcinoma | 16, 18, 31, 33, 35, 39, 45, 51, 52, 54, 56, 66 |
| **OTHER** | |
| Focal epithelial hyperplasia | 13, 32 |
| Mouth/oral papillomas | 2, 6, 11, 16 |
| Laryngeal papilloma | 6, 11, 30 |
| Carcinoma (head/neck/lung) | 2, 6, 11, 16, 18, 30 |
| Conjunctival papilloma | 6, 11 |

*Epidermodysplasia verruciformis

#CIN = cervical intraepithelial neoplasia, VAIN = vaginal intraepithelial neoplasia, VIN = vulvar intraepithelial neoplasia, PAIN = perianal intraepithelial neoplasia, PIN = penile intraepithelial neoplasia.

### 1.3.5 High Risk Human Papillomavirus Types

HPV types 16, 18, 45, and 31 are the most prevalent HPV types detected in cervical cancers worldwide, although up to 18 HPV types can be found (Figure 1.3). Particular HPV types cluster in certain populations, in certain parts of the world and among different racial groups. The proportion of HPV 16 in cervical cancers, for example, varies from 31% to 78%, depending on the country of origin (Bosch *et al.*, 1995). HPV 18 predominates in South East Asia, whilst HPV 45 is predominant in West Africa. Typically HPV 16 is more commonly detected in SCC and HPV 18 and 45 in adenocarcinoma, or adenosquamous carcinoma (Chen *et al.*, 1994a, and Zhebe and Wilander, 1997). In addition, HPV 16 is more often found in well-differentiated cancers than HPV 18 (Walker *et al.*, 1989, and Bosch *et al.*, 1997).

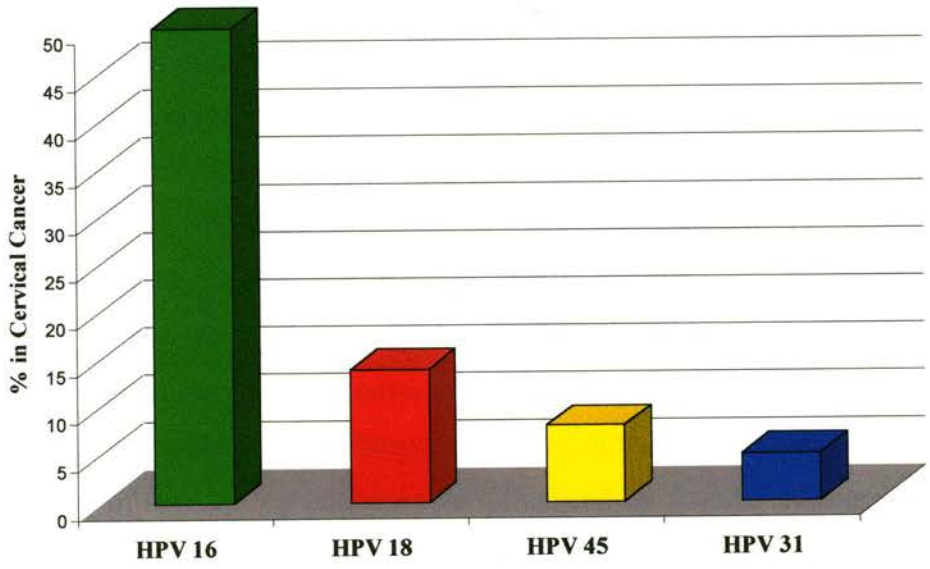



Figure 1.3: The most common human papillomavirus types found in cervical cancer around the world (adapted from Bosch *et al.*, 1995)

HPV 16 was originally cloned from a German tumour (Seedorf *et al.*, 1985). The virus has a circular DNA genome of 7906bp and is the most prevalent high risk HPV type among cytologically normal women as well as the most common cancer-associated HPV type (de Villiers, 1997). Extensive epidemiological investigations involving primarily the URR, but also the E6, L1 and L2 genes have revealed five major variant lineages, or phylogenetic clusters, for HPV 16 (Ho *et al.*, 1991 and 1993, Chan *et al.*, 1992a, Yamada *et al.*, 1997, and Wheeler *et al.*, 1997) including:

- European (E) – found in most European and North American isolates
- African (Af1 and Af2) – found in the majority of isolates from Africa
- Asian (As) – mainly found in South-East Asia
- Asian American (AA) – clustered in Central and South America, and Spain

Minor lineage variants such as North American (NA) variants from Central and South America have also been identified. There is an absence or rarity of genomic recombination between the HPV 16 lineages. It has been suggested that HPV 16 variants evolved consecutively while spreading throughout humankind (Ho *et al.*,

1991), and patterns of HPV phylogeny suggest coevolution of the virus with the three major human races, Africans, Caucasians, and East Asians (Ho *et al.*, 1993a). Minor patterns indicate smaller bottlenecks of viral evolution and spread, which may correlate with migration of ethnic groups. African isolates carry the largest number of sequence variations compared to the prototype, and the deepest phylogenetic branching, suggesting the evolution of HPV 16 in Africa (Ho *et al.*, 1993a). A similar evolution of HPV 18 has also been suggested (Ong *et al.*, 1993). Data on variants of the HPV 16 Af1, Af2 and AA branches indicate that some of these branches are closer to an unknown ancient HPV 16 precursor than to the variants of the E and As branches. It has been suggested that the geographic distribution of HPV types and of specific variants may be influenced by founder effects, coevolution of HPVs with human races, human migration patterns and viral fitness measures such as transmissibility (Yamada *et al.*, 1997). HPV 16 has been found in much higher copy number in lesions compared with HPV 18, 31, or 45 (Swan *et al.*, 1999). It has also been suggested that the median copy number of HPV 16 in cells increase with increasing epithelial abnormality, an effect not seen with other HPV types.

Of the high risk HPV types, HPV 18 is not as common in asymptomatic women and in high grade CIN as it is in invasive cervical disease (LaGuardia, 1993). HPV 18 positive tumours also recur more commonly after radical hysterectomy and have greater nodal involvement then tumours containing HPV 16 (Barnes *et al.*, 1988, Rose *et al.*, 1995, and Burger *et al.*, 1996). Similarly, the five year disease-free survival rate in one study was 100% for patients with intermediate-risk HPV-associated cervical tumours, 58% for patients with HPV 16 positive tumours, and 38% for patients with HPV 18 positive tumours (Lombard *et al.*, 1998). Collectively, these studies suggest that HPV 18 is superior at immortalising and transforming cells *in vivo* compared to HPV 16. There is certainly evidence that HPV 18 has more efficient immortalising properties than HPV 16 *in vitro* (Barbosa and Schlegel, 1989, and Berumen *et al.*, 1994). However the results of at least one study suggested that HPV type has no prognostic value (Ip *et al.*, 1992). This in part reflects the difficulties that are likely to arise when attempting to relate virological characteristics to biological outcomes, which may also be affected by variables such as the experience of the surgical team, facilities for patient care and treatment regimes.

### 1.3.6 Epidemiology of Human Papillomavirus Infection

HPV-induced skin warts are widespread and especially predominant in school children. Anogenital HPV infections on the other hand, are most common among sexually active adolescents and young women. Genital HPV infection in children although not common, can occur as a result of autoinoculation, transfer by a parent during normal contact (eg. bathing) or as a result of sexual abuse (Syrjanen and Puranen, 2000). Newborn babies can also be exposed to cervical HPV infection in the mother.

HPV prevalence is a measure of the percentage of individuals in a population that has new, persistent, or recurring HPV infection at a particular time, a measure of 'disease burden' (Wheeler, 1996). Cervical HPV prevalence is age-related, the peak prevalence of infection occurs at 15 to 25 years of age after a sharp rise in infections, and estimates of HPV infection range from 5% to 50% in cytologically normal women of reproductive age (Schiffman, 1994, Schiffman and Brinton, 1995, Franco *et al.,* 1997, and Torlolero-Luna, 1999). In one study of sexually active university students, HPV DNA was found in 43% of normal cervical tissue (Ho *et al.,* 1998). A similar study found one third of American female college students were HPV positive (Bauer *et al.,* 1991). HPV prevalence decreases in women over 30. This decrease may be due to immunologic clearance or suppression of existing infections combined with fewer new infections because of fewer new sexual partners. Lower prevalence rates are also found in monogamous women. It is likely that more than 50% of sexually active adults worldwide have been infected with at least one genital HPV type in their life (Syrjänen, 1990, and Koutsky, 1997). In the USA, genital HPV infection is the third most commonly diagnosed STD behind *Chlamydia trachomatis* infection and trichomoniasis. The Centers for Disease Control and Prevention in the USA (CDC) estimates that approximately 750,000 new cases of genital HPV infection are diagnosed annually in the USA, and the lifetime cumulative risk of genital HPV infection is approximately 10% (Tortolero-Luna, 1999). In postmenopausal women, HPV detection declines and remains low for the remainder of life (zur Hausen and de Villiers, 1994).

The most common outcome of cervical HPV infection is a "clinically inapparent, transient DNA detectability with no clinical sequelae" (Schiffman, 1994). Therefore most HPV infections are transient, although some are chronic, but only intermittently detectable. Mean infection durations have been estimated at 8.2 months and 13.5 months for nononcogenic and oncogenic HPV types respectively. After 12 months, 35% of individuals remained infected of those positive for HPV at enrolment in one study (Franco *et al.,* 1999). It has been estimated that approximately 3% of women infected with HPV develop cervical cancer in their lifetime, and the mean latency between symptomatic primary infection and cancer development is around 20 to 50 years (zur Hausen, 1989).

### 1.3.7 Risk Factors for Cervical Human Papillomavirus Infection

Risk factors for cervical HPV infection include:

- Age – probably the most important determinant. There is a decrease in prevalence after age 30, even independent of sexual activity (Meissels, 1992). This may be due to acquired immunity to HPV or it could represent a 'cohort effect'. That is, older women were less likely to be exposed to HPV in the past than now when HPV infection is more prevalent.
- Sexual behaviour – in particular the number of male sexual partners and male sexual partners who themselves have had multiple sexual partners. A young age of first sexual experience is also important. Men of female partners with CIN or cancer consistently have a higher prevalence of HPV infection then men of partners with no morphological signs of HPV infection (Pontén and Guo, 1998). Cervical or vulvar HPV can be detected in 17% to 21% of women with only one male sexual partner and in 69% to 83% of women with exposure to five or more sexual partners (Wieland and Pfister, 1997).
- Non use of barrier contraceptives (Kjaer *et al.,* 1996).
- Immunosuppression - for example as a result of HIV infection, renal-allograft transplant, or Hodgkin's disease.

### 1.3.8 Immune Surveillance

HPV has evolved mechanisms to evade (rather than subvert) the immune response by its tropism for superficial epithelial cells (Tindle, 1996). Strong immune responses to HPV infection therefore are often not generated. Latent infections do not induce an immune response either as no inflammation accompanies infection. Immune responses, when elicited, appear to be type-specific and are primarily cell-mediated with a small humoral component.

#### 1.3.8.1 Cell-Mediated Responses

Cell-mediated immunity (CMI) is the main line of defence against viral infections. The importance of CMI is demonstrated by individuals with impaired CMI, either genetic, iatrogenic (lymphoma or renal allograft recipients) or acquired (HIV), who have a higher rate of both subclinical and clinical HPV infection compared to those who are immunocompetent (Tornesello *et al.*, 1993). Successful CMI responses against HPV infections are believed to be elicited via the T-helper type 1 lymphocyte (Th1) pathway, mediated through the production of specific cytokines, such as interleukin-2 (IL-2), interferon-γ (IFN-γ) and tumour necrosis factor-α (TNF-α) (Coleman *et al.*, 1994, and Lörincz, 1999). The regression of warts is preceded by the infiltration of CD4 positive and CD8 positive T lymphocytes into the epidermis (Tagami *et al.*, 1985, Rogozinski *et al.*, 1988, and Edwards *et al.*, 1995).

The viral E6 and E7 proteins appear to be the target of the cytotoxic CD8-positive T lymphocytes (Chen *et al.*, 1992, Hilders *et al.*, 1993, and Gao *et al.*, 1994). Mice however, are immunologically ignorant of the HPV oncoproteins (Melero *et al.*, 1997), although they do mount a response to HPV L1 VLPs (Dupuy *et al.*, 1997). In humans the response to L1 decreases with increasing grade of cervical dysplasia (Luxton *et al.*, 1997), which is not surprising as there is usually loss of expression of L1 during increasing cellular abnormality. This provides some evidence for the importance of the immune response in the control of HPV infections. Interestingly, the E7 protein can also be found extracellularly, inducing the overexpression of immune suppressive IFNα cytokine from antigen presenting cells, which may effectively inhibit the T cell response (Le Buanec *et al.*, 1999). A CMI response against the E2 protein that is relatively conserved across HPV types has also been

detected (Kónya *et al.,* 1997). Many HPV-associated lesions regress spontaneously, probably as a consequence of CMI responses. Declining HPV prevalence with age is also probably due to the effects of CMI and possibly cross-protective responses. Infection with HPV 6/11 and HPV 16 for instance produces an antagonistic interaction which may be due to a cross-protective CMI response (Strang *et al.,* 1990, and Luostarinen *et al.,* 1999).

### 1.3.8.2 Humoral Responses

Humoral responses are mediated by B lymphocytes, which recognise antigenic determinants and produce antibodies. Conformational antigens on the surface proteins of HPV, L1 and L2, are likely to be important for inducing a neutralising antibody response to HPV and are the focus of prophylactic vaccine strategies (Kirnbauer *et al.,* 1994, Bontkes *et al.,* 1999a, and see section 1.9.1.1). Antibody responses to the E6 and E7 proteins of HPV have also been demonstrated (Müller *et al.,* 1992, and Dillner *et al.,* 1995). These responses are found among patients with invasive cervical cancer, but not among patients with CIN (Müller *et al.,* 1992, and Viscidi *et al.,* 1993) and show very little cross-reactivity (Roden *et al.,* 1996).

In individuals with humoral immunity disorders there is no increased susceptibility to HPV-induced lesions (Lutzner, 1985). Antibodies to HPV are also detected in only about half of women with invasive cervical cancer (Müller *et al.,* 1992, and Dillner *et al.,* 1995), which embodies the concept of 'immune surveillance escape', originally postulated by Burnet (Burnet, 1970). Antibodies directed against HPV proteins, conformational epitopes on intact viral particles, and the infected cells, therefore play only a small role in the immune response to HPV.

# 1.4 CELL BIOLOGY OF ANOGENITAL PAPILLOMAVIRUS INFECTION

## 1.4.1 Tissue Specificity

PVs are epitheliotropic. That is they infect squamous epithelial cells of the anogenital and/or oral mucous membranes, as well as skin, and replicate within cell nuclei. Some BPVs can also induce lesions in connective tissue and fibroblasts. The oncogenic genital HPVs infect the epithelia of the transformation zone of the cervix, which is a variable portion of ectocervix where the columnar (glandular) epithelium lining the endocervix joins the squamous epithelium of the ectocervix at the squamo-columnar junction (SCJ). The process of squamous metaplasia (changing of columnar to squamous epithelia) occurs during a woman's life. Most preneoplastic and neoplastic cervical processes originate within the transformation zone, mainly during the metaplastic process. The origin of infections at the SCJ may be mechanical (the SCJ is thinner), or biological (reserve cells may be particularly receptive to HPV infection) (Barrasso and Guillemotonia, 1997).

## 1.4.2 Transmission

Anogenital HPV infections are generally transmitted by sexual contacts (Ho *et al.*, 1993b), as the principal reservoirs of infection are the male and female genitalia. The presence of HPV DNA in a husband's penis confers a 9-fold increased risk of cervical cancer in wives (Bosch *et al.*, 1996) and identical HPV variants can be detected in sex partners (Xi *et al.*, 1993). Virgins, on the other hand, are usually free of detectable HPV infection. Although the majority of HPV infections remain clinically inapparent, they still contribute to the spread of infection. Direct skin contact is thought to be the major mode of transmission although it is believed that transmission by semen, non-venereal routes and fomite transmission may occur to a lesser extent. Genital HPV infection may also be passed from infected mothers to the oral cavities of their babies.

Infection with HPVs probably occurs though microtraumas in the epithelial architecture which exposes the basal cells of the epithelium to the virus. Oral-genital

contact may lead to oral site infection by anogenital HPV types. A putative receptor molecule for HPV entry has been identified on the cell membrane in alpha-6 integrin (Evander *et al.,* 1997). This integrin is expressed primarily during wound healing and is a logical candidate. HPV virions however also bind to heparin, a ubiquitous polysaccharide. It is not clear if the cellular proteins that mediate HPV entry also contribute to cell type specificity (Stubenrauch and Laimins, 1999).

### 1.4.3 Life-Cycle and Replication Strategies

The epidermis is a regenerative tissue composed of layers of keratinocytes with each layer representing an increased state of differentiation. The basal layer, the *stratum basale*, consists of proliferating, undifferentiated keratinocytes that migrate outward into the non-dividing suprabasal compartment and terminally differentiate. The second cell layer, the spinous layer, or *stratum spinosum*, has early markers of keratinocyte differentiation. The third layer is the *stratum granulosum* or granular layer. The *stratum lucidum* separates the living cells from the dead, cornified cells (corneocytes) that are the terminal products and occupy the *stratum corneum* (see Figure 1.4). The morphologically distinct epidermal layers are identified by the expression of genes that are only expressed at specific stages in keratinocyte differentiation, such as keratins. This profile is tightly regulated. In the cervix, HPV production occurs predominantly in flat and benign subclinical warts or CIN 1 lesions. The vast majority of these infections are self-limited, but the mechanisms leading to regression remain obscure. There are three states of PV replication – latent infection, productive infection, and defective, or transforming infection.

#### 1.4.3.1 Latent Infection

Immune impaired individuals have an increased incidence of genital warts and cervical cancer, probably resulting from the activation of existing latent infection (Sillman *et al.,* 1984, Blohme and Brynger, 1985, and Penn, 1986). The term latency is also used to describe the situation where HPV persists after a clinically apparent disease, or prolonged incubation periods. Latent HPV infection refers more specifically to the presence of HPV DNA in morphologically normal keratinocytes, while there is no differentiation-dependent synthesis of virions (Stubenrauch and

Laimins, 1999). Although not well understood, latent infection is believed to be the most common outcome of PV infection. The male urethra is often the site of latent infection.

#### 1.4.3.2 Productive Infection

Once the virus is endocytosed by epithelial cells, it is transported to the nucleus where it uncoats and the viral DNA is exposed. Following infection of basal cells an incubation period of between three and 18 months proceeds, and the viral genome becomes established in an episomal state of approximately 20 to 100 copies per cell. The length of the incubation period depends on the host rather than the virus and lesions may persist for weeks, months or years. The viral DNA replicates in S phase with the host cell in the basal layers (Chow and Broker, 1994) which is believed to take place by conservative bidirectional replication (Flores and Lambert, 1997). As infected cells divide, HPV DNA becomes distributed between both daughter cells. One migrates away from the basal layer and initiates a program of differentiation. The other cell continues to divide in the basal layer and provides a reservoir of viral DNA for further cell divisions (Stubenrauch and Laimins, 1999). PV genomes are transcribed into multiple, differentially spliced and partially overlapping RNA species. There are separate early gene and late gene promoters. The viral transcription patterns change considerably in different layers of the epithelium. In low grade lesions, all the PV ORFs are expressed and E4 and E5 are particularly abundant (Stoler *et al.,* 1992). E2 autoregulates its own expression, as high levels of E2 binding to the URR inhibit p97 activity, and the majority of transcripts encoding E1 and E2 initiate from here in the basal layers. As a result, viral copy number is maintained at a relatively constant level.

As cells differentiate and rise towards the surface, the virus switches to a productive rolling-circle method of replication that produces thousands of copies of genomes in the upper layers (Flores and Lambert, 1997). The key event in switching to productive infection is a change in promoter usage for E1/E2 expression to the late promoter. This promoter is not repressed by E2, and there is an increase in the levels of E1 and E2, which leads to an increase in the number of viral genomes. Thus HPV DNA is first detectable in the *stratum basale* and the signal gradually increases in intensity until it peaks in the *stratum spinosum* and *stratum granulosm.*

The second stage of productive infection is reached at a critical point near the surface of the epithelium where the keratinocytes are no longer responsive to growth stimulation. However, a large amount of viral DNA has accumulated so the production of virions can take place. Viral DNA replication, expression of structural proteins as well as PV particle assembly occur in the upper well-differentiated layers of the epithelia. The mechanism by which virion production is activated in the suprabasal layers is unknown. It may be the result of induction of the late promoter by differentiation-specific TFs. Replication of HPV in productive infection is dependent on the terminal differentiation of keratinocytes and is therefore very closely linked with the differentiation of this host cell; in particular, to the expression of a series of the sequentially higher molecular weight keratins which are produced as maturation proceeds. The relationship of the HPV life cycle to epithelial differentiation is illustrated in Figure 1.4.

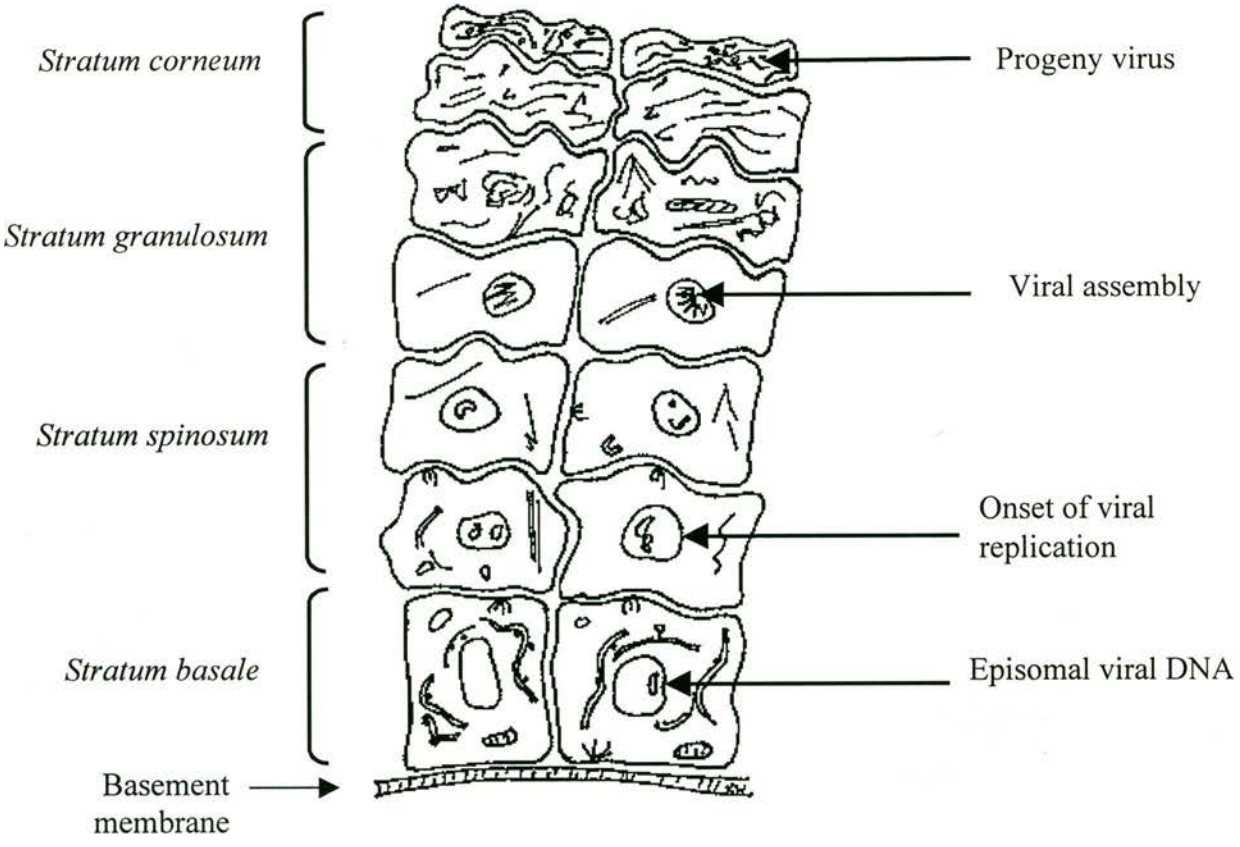



Figure 1.4: Illustration of the epidermal layer and stages of productive infection by human papillomavirus

The stages of epithelial differentiation are indicated on the left. Phases in the HPV life cycle during keratinocyte differentiation are indicated on the right.

In warts, the basal cells contain low levels of viral DNA and lack capsid proteins. The upper levels of the epithelium, by contrast, contain large amounts of viral proteins. The granular and cornified layers are packed with complete viral particles. There is a large difference in the large amount of virus found in papillomas affecting the highly keratinised areas of the body and the limited amount of virus in papillomas affecting mucosal surfaces. The number of virions produced is related to the amount of keratin expressed: the more keratins, the greater the viral load. In the lower epithelial layers the virus increases cell proliferation leading to papillomatosis and epithelial hyperplasia which is due to the E5, E6 and E7 proteins. Viral replication also interferes with the terminal differentiation of the host cell, leading to various cytopathic effects such as the accumulation of intracytoplasmic or nuclear inclusions observed in plantar warts, or perinuclear vacuolation that characterises koilocytotic cells of genital lesions. Nuclear remnants persist in the superficial layers (parakeratosis) due to the disordered keratin profile.

#### 1.4.3.3 Transforming Infection

Keratinocytes appear to have an inherent HPV promoter nonspecific mechanism of repression (Sun *et al.*, 1997) and cervical cancer is the result of failing host cell control over persisting viral genes. Viral transcription is subject to control by viral and cellular functions and is decisively determined by the differentiation state of the host keratinocyte (Fuchs and Pfister, 1994). The persistence of HPV E6 and E7 DNA in cancers suggests a role of HPV in maintenance of the transformed state. The E6 and E7 proteins of HPV 16 and 18 (and several other HPV types) disturb the physiological functions of cellular proteins, such as p53 and pRb. This leads to genetic instability which enhances the probability of mutations in the cellular proto-oncogenenes and tumour suppressor genes, contributing to tumour progression (Pfister, 1996). Cancer is a rare, unintended event in HPV infection, and is characterised by loss of production of viral particles, which is not in the virus's best interests.

In low grade cervical lesions there are few E6/E7 transcripts in basal layers, and an increase with epithelial differentiation in the upper layers. All ORFs are expressed in low grade lesions, with the E1/E4 and E5 transcripts being the most abundant. The L1 and L2 transcripts are present only in the terminally differentiated keratinocytes

(Stoler *et al.*, 1992, Böhm *et al.*, 1993, and Fuchs and Pfister, 1994). As the grade of neoplasia increases, cellular differentiation and overall viral transcription decreases with the L1 and L2 transcripts generally disappearing, and the E6 and E7 transcripts increasing (Stoler *et al.*, 1992). In high grade lesions/carcinomas, there are high levels of E6/E7 transcripts in all cell layers of the undifferentiated epithelium (zur Hausen, 1986, Durst *et al.*, 1992, and Stoler *et al.*, 1992). The viral transcription patterns are similar in precancerous lesions, or carcinoma *in situ* and invasive carcinomas, suggesting that additional alterations to host genes are necessary for progression (Stoler *et al.*, 1992, and Fuchs and Pfister, 1994).

The peak incidence of high grade lesions occurs 5 to 10 years after the peak occurrence of low grade lesions. Invasion occurs about a decade later (Schiffman, 1992). Severe dysplasia and invasive cancer often present as a monoclonal lesion. Genetic studies indicate that these pathologies arise by clonal selection from less advanced precursors and lateral expansion of that clone (Pontén and Guo, 1998). It is also possible that only a single basal cell becomes infected, and there is clonal expansion of that one cell, which will depend on the number of proliferating cells that are exposed to the virus and the dose of the infecting virus.

### 1.4.4 Immortalisation and Transformation of Cells by Human Papillomavirus

Human cells have a defined life span and growth and division are highly regulated. Immortalisation is a property of indefinite growth with changes in growth control. Transformation, on the other hand, is defined as a complete failure of the normal constraints of growth. It is acquired by stepwise accumulation of defects in genes regulating cell growth and homeostasis. There are two classes of genes in which mutations can cause transformation – oncogenes and tumour suppressor genes. Tumour suppressor genes impose a constraint on the cell cycle or cell growth; the release of that constraint is tumourigenesis. For lesions to progress from benign to cancer usually requires cofactors, eg. EV and ultraviolet (UV) light (Lutzner *et al.*, 1984), and X-irradiation in laryngeal papillomas (zur Hausen, 1977). Genetic aberrations usually occur in HPV infected cells at the transition from carcinoma *in*

*situ* to invasive cancer (Hesselmeyer *et al.*, 1996). However, even in mildly dysplastic lesions, tetraploidisation has often already occurred.

Only high risk HPVs immortalise human keratinocytes. However, the efficiency of immortalisation *in vitro* is generally low. Infections commonly enhance proliferation and result in a delay of cellular senescence and differentiation. Eventually most cells stop proliferating and differentiate. The binding of E6 and E7 to proteins that regulate the mitotic cycle suggests that continuous proliferation may result from viral protein interference with host cell proteins negatively regulating the mitotic cycle. Cells expressing E6 and E7 are also subject to a higher rate of mutation and chromosomal rearrangements than those that do not express E6 and E7 are. Cancers induced by oncogenic HPV types require the continued expression of E6 and E7. This is supported by experiments where antisense RNA inhibits translation of the E6 and E7 proteins. Consequently tumour progression is accompanied by an increase in E6 and E7 expression. Both E6 and E7 together from HPV 16 or 18, but not HPV 6, are sufficient to immortalise and transform primary human keratinocytes (Kaur and McDougall, 1988, Howley-Nelson *et al.*, 1989, Münger *et al.*, 1989b, Woodworth *et al.*, 1989, Hudson *et al.*, 1990, Barbosa *et al.*, 1991, and Halbert *et al.*, 1991). Each oncogene on its own (from high risk types 16, 18, 31, and 33, but not low risk types 6 and 11) plus *ras* can also cooperate to transform primary rodent cells (Matlashewski *et al.*, 1987, Phelps *et al.*, 1988, Storey *et al.*, 1988, Crook *et al.*, 1988 and 1989, Bedell *et al.*, 1989, Chesters and McCance, 1989, DiPaolo *et al.*, 1989, and Watanabe *et al.*, 1990). E6 and E7 of HPV 16 or 18 without additional oncogenes are also capable of inducing anchorage-independent growth and tumour formation in mice (Kanda *et al.*, 1988, Le *et al.*, 1988, Yutsudo *et al.*, 1988, and Bedell *et al.*, 1989). The expression of E6 also can prevent keratinocyte differentiation (Sherman and Schlegel, 1996). A blockage of differentiation results from rare immortalising events that depend on modifications of the host cell DNA. HPV 16 E6 also induces anchorage-independent growth in NIH 3T3 cells (Sedman *et al.*, 1991). Inhibition of E6/E7 gene expression in cancer cells leads to a significant reduction in cell growth (von Knebel Doeberitz *et al.*, 1988).

### 1.4.5 Physical State of Papillomavirus DNA in Cervical Cancers

"Epithelia expressing chromosomally integrated HPV oncogenes have a strong selective growth advantage and are more susceptible for clonal outgrowth as cancer cells" (Klaes *et al.,* 1999). Boshart *et al.* (1984) and Durst *et al.* (1985) published some of the first indications that malignant tumours containing HPV 16 or 18 are found in the integrated state as opposed to the episomal state in benign tumours. HPV DNA sequences integrated into the cellular genome have been identified in cervical cancers, cervical neoplasia-derived cell lines, and in immortalised human keratinocytes (Yee *et al.,* 1985, Lehn *et al.,* 1988, and Barbosa *et al.,* 1989). In contrast, integration of HPV DNA almost never occurs in EV-associated skin cancers, HPV 6 and 11-associated benign lesions, or warts. HPV 18, 31, and 35 can be found integrated in 100% of cervical carcinomas (Pirami *et al.,* 1997), however the proportion of HPV 16 found integrated is variable (Table 1.3).

When integration of HPV DNA occurs, it takes place before invasion of a cancer. Studies of established cell lines containing HPV 16 and HPV 31 show that episomal copies of HPV are lost or integrate into cellular DNA with time in tissue culture (Crook *et al.,* 1990, and Bedell *et al.,* 1991). It has been determined, at least for HPV 31, that extrachromosomal HPV DNA is necessary in rafts for production of viral particles, and that terminal differentiation alone is not sufficient (Frattini *et al.,* 1996). Integration of the HPV 16 genome into chromosomal DNA usually leads to disruption of the E1 and/or E2 ORFs, leading to the generation of mRNAs comprising E6 and E7 and cotranscribed cellular sequences (Klaes *et al.,* 1999). This results in increased stability of E6/E7 transcripts, which aids in transformation. Indeed, it is known that the growth rate of HPV-transformed cells is dependent on the expression levels of E6 and E7 (von Knebel Doeberitz *et al.,* 1988 and 1994). Disruption of the E2 ORF often occurs in the 5' region of the gene, resulting in the deletion of the E2 promoter and part of the transactivation domain, along with most of E2's replicative and E1 interaction functions.

No chromosomal integration locus is preferred for HPV 16 although the site of chromosomal integration may provide regulatory features. Integration adjacent to cellular promoters or mRNA processing signals can modify HPV expression (Smits

*et al.*, 1991, and von Knebel Doeberitz *et al.*, 1991). This may be due to dominant regulatory mechanisms that are able to override the response of the viral promoter to TFs.

Table 1.3: Physical State of High Risk Genital Human Papillomavirus DNA in Cervical Lesions

| Authors | HPV type/s | Normal Cervical Epithelia/CIN 1 | CIN 2/3 | Cervical Carcinoma |
|---|---|---|---|---|
| Pirami *et al.*, 1997 | 16, 18, 31 and 35 | - | 25% CIN 2/3 integrated (different HPV types) | 81% integrated (HPV 16, 18, 31, and 35). 100% HPV 18, 31, and 35 integrated, and 26% HPV 16 episomal. |
| Cullen *et al.*, 1991 | 16 and 18 | 100% episomal | 95% episomal<br>5% mixed* | HPV 16 – 53% integrated 27% episomal, and 20% mixed.<br>HPV 18 – 96% integrated, and 4% mixed. |
| Klaes *et al.*, 1999 | 16 and 18 | 100% episomal | CIN 2 – 95.5% episomal, 4.5% mixed. CIN 3 – 84.4% episomal, 11% mixed, and 4.6% integrated. | 12.5% episomal, 65% integrated and 22.5% mixed. |
| Daniel *et al.*, 1995 | 16 | - | CIN 3 – 100% integrated | 100% integrated |
| Vernon *et al.*, 1997 | 16 | - | - | 62% integrated, 16% episomal, and 22% mixed. |
| Park *et al.*, 1997 | 16 and 18 | - | - | HPV 16 – 70.6% integrated, 9.8% episomal, and 19.6% mixed.<br>HPV 18 – 100% integrated. |
| Das *et al.*, 1992a | 16 | 100% episomal | CIN 2 – 100% episomal. CIN 3 – 90% episomal, and 10% integrated. Carcinoma *in situ*# – 71.43% episomal, 23.8% integrated, and 4.76% mixed. | 69.05% integrated, 26.19% episomal, and 4.75% mixed. |
| Das *et al.*, 1992b | 16 | - | Carcinoma *in situ*# - 71% episomal, 23% integrated, and 5% mixed | 26% episomal, 69% integrated and 5% mixed. |

Physical state has been reported for HPV 16 and 18 together in some of the studies and a breakdown for HPV types cannot be given

*Mixed refers to both integrated and episomal HPV DNA forms in a lesion

# These reports have differentiated between CIN 3 and carcinoma *in situ*, whereas in this review they have been grouped together

# 1.5 MOLECULAR VIROLOGY OF PAPILLOMAVIRUSES

The PV ORFs are arranged on the positive strand of the viral DNA while the other DNA strand contains only short ORFs which are not conserved in terms of composition or localisation and are assumed to be non-coding. Transcription involves several different promoters and many of the RNAs are polycistronic. However there is much less information on HPV promoter elements than for BPV 1. A promoter in front of the E6 ORF is a common feature of all HPVs. There is also some indication of a late promoter in the 5' region of the E1 ORF in genital HPVs. Analysis of RNA transcripts reveals a complex arrangement of splicing and different transcripts generated by splicing within the same ORFs may have different biological roles. Each of the PV genes, their gene products, and the URR will now be discussed in more detail.

## 1.5.1 Papillomavirus Proteins Involved in Viral Replication and Gene Regulation

### 1.5.1.1 E1

E1 is the longest ORF in the PV genome. The full length 68kDa to 76kDa phosphorylated nuclear E1 protein is essential for PV replication. Some of the earliest indications that the PV E1 gene encoded a protein that could be necessary for viral replication came from the knowledge that the PV E1 protein shared regions of amino acid homology with the SV 40 large T-antigen, which is an important replication initiation protein (Clertant and Seif, 1984). It has since been determined that E1 is necessary to initiate HPV DNA replication and that it can bind ATP, DNA and the full length E2 protein, which strengthens the affinity of E1 for the origin of replication, *ori* (an A/T rich segment of DNA). The carboxyl terminus of E1 is highly conserved and is required for complex formation with the E2 protein (Yasugi *et al.*, 1997). In BPV 1, the E1 gene encodes several proteins, one of which is thought to possess DNA-dependent adenosine triphosphatase (ATPase) and ATPase-dependent helicase activity, which unwinds DNA at the origin to facilitate assembly of the replication machinery (Lambert, 1991, and Favre *et al.*, 1997b). The E1

protein also associates with DNA polymerase alpha, which is also a component of the cell replication machinery (Park *et al.*, 1994).

All genital HPVs contain an E1 binding site 25bp 5' of an Sp1 binding site, which is essential for transcriptional activation. Replication may be favoured after binding of E2 to its overlapping binding site and the concomitant displacement of Sp1. E1 only binds to the underphosphorylated form of E2 (the phosphorylated form of E2 in contrast is a transcriptional activator), and has a negative effect on transcriptional stimulation by E2. It is possible that a mechanism exists whereby viral function alternates between transcription of the oncogenes and replication of the genome (Tan *et al.*, 1994). It then follows that if the viral DNA becomes integrated into the cellular genome, and E1/E2 are disrupted, then replication ceases and the oncogene promoter can become constantly active. E1 certainly appears to suppress the transformation potential of HPV by negatively regulating the viral oncogenes. Disruption of the HPV 16 E1 protein also increases the immortalisation capacity of the E6 and E7 genes when transfected into keratinocytes (Romanczuk and Howley, 1992).

#### 1.5.1.2 E2

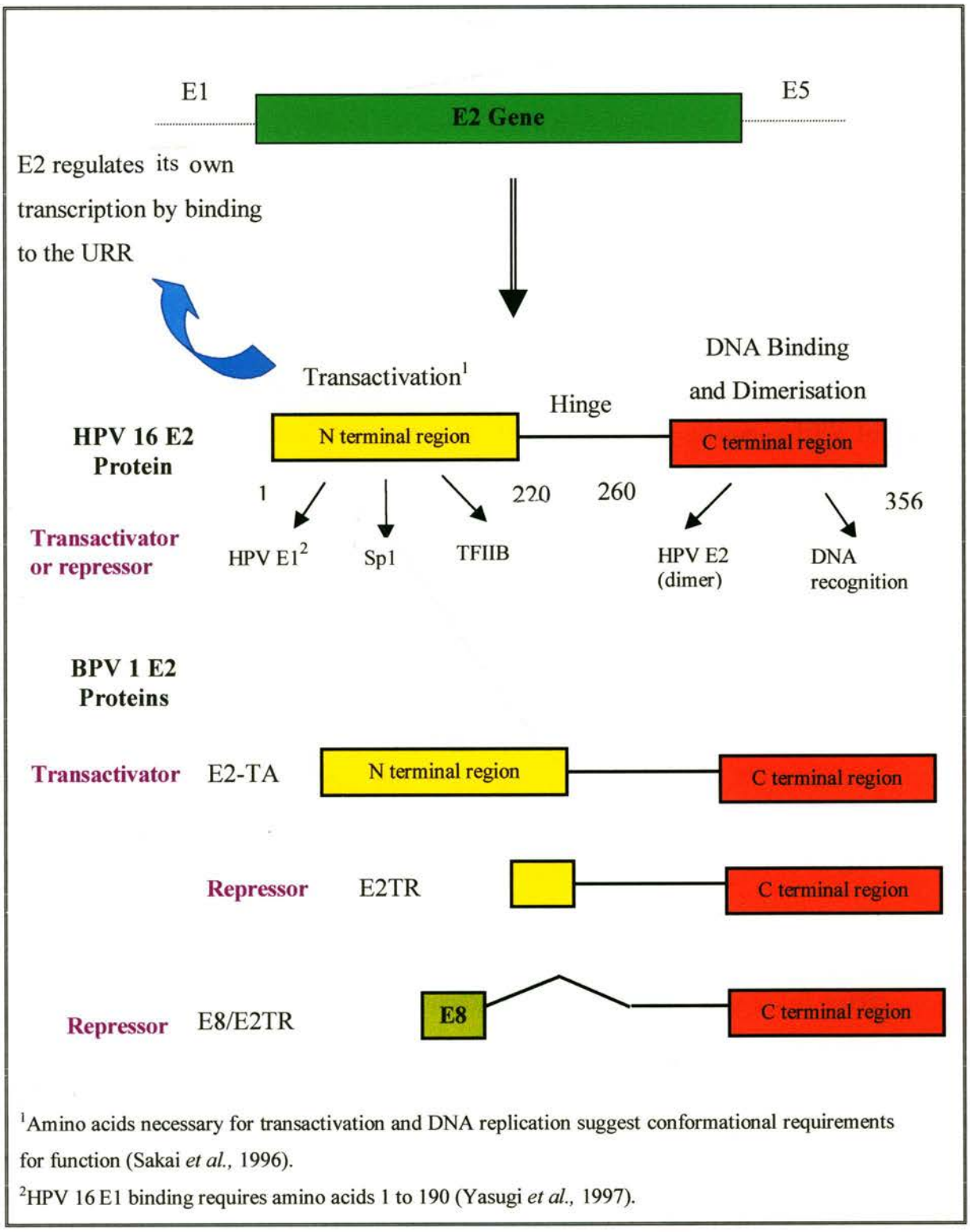
The E2 protein regulates viral transcription and replication, and therefore plays a central role in the virus life cycle. Many functions of E2 have been deciphered since its transactivating properties were discovered in 1985 (Spalholz *et al.*, 1985). Most of the current information on E2 has come from studies of the BPV 1 E2 protein, but the structure and function of E2 are comparable in all PV's so far examined. The full length E2 protein, a nuclear phosphoprotein of 45kDa to 48kDa, has two conserved domains, the amino and carboxyl terminal regions, separated by a non-conserved central region of variable length that serves as a flexible hinge. The amino terminal region (approximately 200 amino acids) is responsible for transactivation and the carboxyl terminal region (approximately 90 amino acids) mediates DNA binding and protein dimerisation. E2 binds DNA as a dimer to a palindromic sequence. The internal nucleotides and flanking sequences can greatly influence the affinity of binding.

In BPV 1, E2 encodes three nuclear proteins (shown in Figure 1.5). The full length 48kDa highly phosphorylated E2 protein (E2-TA) is a sequence-specific transcriptional activator and repressor of viral promoters through E2BSs in the URR (Bouvard *et al.*, 1994a, and Demeret *et al.*, 1994). There are also two amino terminally truncated E2 gene products (encoded by the 3' half of the ORF), the 30kDa E2 transcriptional repressor E2TR, and the 28kDa E8/E2 transcriptional repressor E8/E2TR (Lambert, 1991). These two proteins share the carboxyl terminal domain responsible for DNA binding and dimerisation. The shorter E2 protein is transcribed from a second E2 promoter within the E2 gene while the E2-E8 protein is produced by alternative splicing. Both shorter E2 proteins repress the activity of full length E2 through competitive DNA binding at the E2BSs or by forming heterodimers with the full length protein.

Although there is less known about HPV E2 proteins, there is no direct evidence that HPVs encode the two truncated E2 repressors (McBride *et al.*, 1991). The most obvious feature of all genital HPV URRs is the positioning of their four E2 binding sites (E2BSs), with highly conserved spacing of sites 2, 3 and 4, reflecting what is probably a common theme for regulation of the E6/E7 promoter by the E2 protein. In genital HPVs E2BS 2 is usually the most degenerate binding site for E2, whilst there is strict conservation of E2BS 1. The presence of two E2 sites close together in the enhancer of the PV early promoter constitutes a strong E2-dependent enhancer, and full length E2 products are able to activate a simple E2-dependent promoter (Ushikai *et al.*, 1994). One E2BS constitutes only a weak enhancer. Synergy may result from cooperative DNA binding (Ham *et al.*, 1991a). Cellular transcription factors (TFs) increase E2 activation by binding of E2 to weak E2BSs through protein-protein interactions. Once E2 has bound to DNA, TFs may also facilitate activation by E2 (Ham *et al.*, 1991a). It has been shown that a core promoter must have at least two elements to be able to respond to E2, such as the TATA box, or a binding site for a TF such as Sp1 (Ham *et al.*, 1991b, and Ushikai *et al.*, 1994). The E2 proteins from HPV 6 and 11 have a lower level of transcriptional activity compared to those from HPV 16 and 18 (Kovelman *et al.*, 1996 and 1999).

The full length HPV 16 or 18 E2 protein can act as a repressor of their respective E6/E7 promoters, an activity that is mediated through the proximal E2BSs (Romanczuk *et al.,* 1990). In HPV 16, both E2BSs 3 and 4 are required for p97 repression. Steger and Corbach (1997) suggested a dose-dependent regulation of the HPV 18 E6/E7 promoter by E2 due to variable occupancy of its binding sites, which have antagonising effects. In Steger and Corbach's model of HPV 18 early promoter activity and E2 involvement, HPV early gene expression is stimulated by the enhancer, leading to the production of E6 and E7 which pushes the cell into the cell cycle to allow for components required for DNA replication to be made. Low amounts of E2 further activate the early promoter via E2BS4, which generates more early proteins including E2, which will then occupy the weaker E2BSs. Binding of E2 to BSs 1, 2, and 3 leads to repression of transcription from the early promoter and activation of viral DNA replication, since E2BSs required for transcriptional silencing are required for DNA replication. The position of the HPV E2BSs suggests that E2 regulation principally involves a negative feedback loop. By repressing the promoter, E2 may also negatively regulate its own synthesis and is probably necessary for long term persistence (Bernard and Apt, 1994).

E2 can overrule the stimulatory effects of the viral enhancer, which seems to be the result of the unique topography of protein binding sites in the immediate vicinity of the basal promoter and depends on the distance between the TATA box and the proximal E2BSs (Garrido-Guerrero *et al.,* 1996). The two 3' E2BSs, 3 and 4, are flanked by an Sp1 binding site and the TATA box. Down regulation is most likely due to mutually exclusive binding of E2 and Sp1, impeding Sp1-mediated promoter activation and/or E2 and TATA-box-binding protein (TBP), preventing formation of the transcription preinitiation complex (Tan *et al.,* 1991 and 1994). However, Sp1 can also interact directly with E2 by its amino terminal domain and the E2 protein imposes a DNA conformational change that prevents factors such as TFIID from interacting (Garrido-Guerrero *et al.,* 1996). E2 is also able to directly target components of the general transcription machinery to repress promoter activity. Therefore E2 may function as an active repressor that directly inhibits HPV transcription at steps after TATA recognition by factors such as TFIID and TBP (Hou *et al.,* 2000). The general structure and functions of the E2 protein from HPV 16 and BPV 1 are presented in Figure 1.5.




[1]Amino acids necessary for transactivation and DNA replication suggest conformational requirements for function (Sakai *et al.*, 1996).

[2]HPV 16 E1 binding requires amino acids 1 to 190 (Yasugi *et al.*, 1997).

Figure 1.5: Structure and function of the bovine and human papillomavirus E2 proteins

E2 molecules can loop interspacing DNA sequences via their amino terminal region to interact with E2 bound at distant locations (Fuchs and Pfister, 1994, and McBride and Myers, 1996). It has been proposed that E2 dimer formation may serve to relocate distal DNA-bound TFs to the site of transcription initiation (Antson *et al.*, 2000). There has also been some suggestion that the hinge region of the E2 protein allows two E2 dimers located on opposite surfaces of the DNA helix, or separated by some distance, to interact efficiently.

The expression of E2 in HPV positive cervical cancer cell lines represses the expression of the viral oncoproteins, and causes cell cycle arrest and apoptosis (Dowhanick *et al.*, 1995, and Desaintes *et al.*, 1997). Indeed integration of HPV 16 DNA resulting in disruption of the E2 gene in cervical cancers has been associated with a decreased disease-free survival rate (Vernon *et al.*, 1998). HIV infection increases CIN rates (LaGuardia, 1993), and although the exact molecular mechanisms are unknown, it has been demonstrated that HIV *tat* protein can reverse E2-mediated repression of p97 (Vernon *et al.*, 1993). *Tat* may interact with E2 directly. Recently, it was also shown that the HPV E2 protein binds to the transcriptional co-activator cAMP response element-binding protein (CREB)-binding protein (CBP) *in vivo*, and that this interaction is necessary for E2-dependent transcription (Lee *et al.*, 2000). CBP has a histone acetyltransferase activity and has been implicated in chromatin remodelling. In normal cells, CBP inhibits cell cycle progression (Giles *et al.*, 1998) and it may also act as a gene expression activator by interacting with TFs. Therefore CBP may act as a tumour suppressor protein. E2 removes CBP from its functions by binding to it, an act that assists cell cycle progression. The BPV 1 E2 protein activation domain contains two likely amphipathic helices (Giri and Yaniv, 1988), which may be responsible for this interaction with CBP.

HPV 16 replication requires E2BS 3. Replication of viral DNA also requires E2-TA, E1 and *ori*. E1 and E2 form a complex at *ori* of which E2 probably plays an auxiliary role by enhancing the functions of E1 and helping to localise E1 onto the viral DNA. The E2 transactivation domain is required for these functions (Yasugi *et al.*, 1997), and the ability of E2 to complex with E1 appears to be essential for its function as an auxiliary replication factor (Sakai *et al.*, 1996). In addition, the viral DNA is

packaged more efficiently into virion particles in the presence of the E2 protein (McBride and Myers, 1996

## 1.5.2 Papillomavirus Proteins with Functions that are not Well Defined

### 1.5.2.1 E3

E3 is transcribed by only a few PVs, BPV 1, BPV 2, European Elk PV (EEPV) and BPV 4, and its functior is still a mystery. Except for BPV 4, this ORF overlaps E2 and E4. In BPV 4 E3 overlaps the E1 ORF. It is not known whether E3 is translated.

### 1.5.2.2 E4

The E4 gene is a region of maximum sequence diversity in PVs. It is completely contained within the central part of the E2 ORF and is translated in another reading frame. It many PV types, E4 lacks an initiation codon. The E4 protein is expressed from two mRNAs, E1^E4 and E1^E4 plus L1. All E4 proteins are proline-rich and may have a central hydrophilic region. The 17kDa E4 protein is non-structural and abundant in warts where it is predominantly cytoplasmic. In HPV 1-induced warts, E4 expression correlates with the onset of vegetative viral DNA replication and occurs after cells have left the basal layer, or higher layers for mucosal types (Doorbar and Myers, 1996). In HPV 16, E4 is expressed from the late promoter, and as such, it should be considered a late protein (Croissant *et al.,* 1985). During productive HPV infection the E4 protein aggregates in cytoplasmic and nuclear inclusion granules, which is most notable in cutaneous lesions where the E4 protein can constitute up to 30% of cell protein (Doorbar *et al.,* 1986, and Breitburd *et al.,* 1987). By contrast, the mucosal HPV types express much lower levels of E4 protein.

Interaction of the E4 protein with cytokeratins may induce collapse of the cell intermediate network and create cytoplasmic vacuolation (Doorbar *et al.,* 1991, and Roberts *et al.,* 1993 and 1997). The vacuolation associated with koilocytic cells is thought to be mediated by the E4 protein, and loss of E4 expression during malignant transformation may explain why koilocytes often disappear. HPV 16 E4 protein also associates with tonofilaments, resulting in the disruption of the cytokeratin filament

network (Roberts *et al.*, 1993, and Doorbar, 1996). There are two conserved motifs in the amino terminus of the E1^E4 fusion protein that are important for this interaction (Rogel-Gaillard *et al.*, 1992, and Roberts *et al.*, 1994) and in one study, no amino acid variations in this region were found in HPV 16 (Eriksson *et al.*, 1999). Involvement of E4 in virus maturation and/or vegetative viral DNA replication has also been suggested, as in warts, E4 expression correlates with L1 expression. The most probable role of E4 is interference with normal keratinocyte differentiation to favour production and release of viral particles (Rogel-Gaillard *et al.*, 1993, Doorbar, 1996, and Doorbar and Myers, 1996).

## 1.5.3 The Papillomavirus Oncoproteins

Three of the PV genes, E5, E6, and E7, are oncogenes (from the Greek *onkas* for tumour).

### 1.5.3.1 E5

The E5 ORF is the least conserved region of the PV genome. For instance, HPV 1 E5 has no similarity to E5 from genital HPVs, except that the resulting proteins are generally highly hydrophobic. The E5 ORF is not present in all HPVs and is truncated in some. When it is present it is located downstream of the E2 ORF. The corresponding E5 protein is a transmembrane protein, localised primarily to the endoplasmic reticulum, but it is also seen in the nuclear membrane and Golgi apparatus (Halpern and McCance, 1996).

E5 is the primary transforming protein of BPV 1, although even HPV 16 E5 can transform rodent fibroblasts (Leechanachai *et al.*, 1992, and Pim *et al.*, 1992). E5 can also stimulate anchorage-independent growth of cells, which is enhanced in the presence of EGF (Bouvard *et al.*, 1994b). The HPV 16, 6 and 11 E5 proteins have transforming abilities that include: increased DNA synthesis, anchorage-independent growth, focus formation leading to higher cell growth rates, and prolonging the lifespan of cells leading to immortalisation (Halpern and McCance, 1996).

As E5 is not usually expressed in human cancers, it is more likely that it plays a role in the early stages of transformation. Its major role appears to be the mitotic stimulation of cells through interaction with signal transduction pathways. For instance E5 stimulates the transforming activity of the epidermal growth factor receptor (EGFR) by enhancing signal transduction mediated by the receptor to the nucleus (Pim *et al.,* 1992). Signals that elicit cell growth, such as from EGF, arrive at the cell membrane and bind receptors, triggering a series of events that culminate in DNA synthesis. E5 may cause the cell to become permanently stimulated and low levels of EGF causes multiple rounds of cell division. This is an advantage to the virus because it increases the number of cells, and therefore increases DNA synthesis (Crawford and Tommasino, 1997). HPV 16 E5 has been shown to form a complex with the 16kDa subunit of the vacuolar $H^+$-ATPase (Conrad *et al.,* 1993). This leads to altered pH-dependent reactions, including those that affect internalisation and/or degradation of growth factor receptors. Cell-cell communication via gap junctions has also been found to be disrupted by the HPV 16 E5 protein, which may interfere with tumour-suppressor signalling (Oelze *et al.,* 1995).

**1.5.3.2 E6**

The 16kDa to 19kDa E6 protein is present in both the nucleus and cytoplasm of HPV-infected cells. HPV E6 shares functional similarities with the E1B protein of adenovirus type 5 and the SV 40 large T-antigen, which are both tumour-associated viruses. The E6 protein also contains four metal binding motifs, C-X-X-C, that constitutes two zinc fingers (Cole and Danos, 1987, Barbosa *et al.,* 1989, and Grossman and Laimins, 1989). Both zinc fingers are required for full E6 transactivation and transformation functions (Ullman and Emery, 1996). In HPV 5 and 8, E6 is the main transforming gene. However, in most HPV types, E6 alone is only weakly oncogenic.

When the other ORFs of HPV are either disrupted or deleted altogether upon integration of the viral DNA, the E6 and E7 genes are always maintained. The addition of antisense E6 and E7 inhibits the growth of CaSki and SiHa cancer cells (Tan and Ting, 1995), demonstrating the importance of both of these genes for maintaining the transformed state. HPV 16 E6-containing cells are polyploid and

also show evidence of translocations, amplifications and deletions (White *et al.*, 1994).

Inactivation of the cellular tumour suppressor protein p53 is a central component of malignant transformation in cervical cancer (Scheffner *et al.*, 1991). In normal cells, p53 acts by inducing the expression, or repressing transcription of certain genes, or interacting directly with certain proteins (Mantovani and Banks, 1999). p53 is involved in DNA damage repair by arresting cells in the G1 phase of the cell cycle. This allows for repair before replicative DNA synthesis in S phase and avoids setting mutations permanently into the genome. If the DNA damage is too great for repair, p53 induces apoptosis. Cells lacking wild-type (ie. mutated) p53 do not display appropriate G1 arrest after DNA damage and exhibit high rates of gene amplification and genomic instability. Approximately half of all human cancers do not have wild-type p53 (Mantovani and Banks, 1999).

HPVs have evolved mechanisms to counteract the apoptotic response of the cell by p53 to viral infection (Figure 1.6). E6 associates with a number of cellular proteins, including p53 (Werness *et al.*, 1990). This association is believed to be mediated through interaction of a cellular factor called E6-AP (E6-associated protein). This protein is a ubiquitin ligase, and together E6/E6-AP complex p53 and target it for degradation via the ubiquitin-mediated proteolysis pathway, thus inactivating its function, and resulting in increased cellular genetic instability (Scheffner *et al.*, 1990). It is the amino terminal sequences of the E6 protein that are necessary for directing p53 degradation, and these sequences are only conserved between the oncogenic HPV types (Crook *et al.*, 1991b). E6 also binds to E6 binding protein (E6BP), a calcium-binding protein found in the endoplasmic reticulum. The E6-E6-AP complex can bind *in vitro* to p53 or E6BP. At the present, the evidence for a physiological role of the E6-E6BP interaction is only circumstantial (Barbosa, 1996). E6BP appears to be a vitamin D receptor-associated factor (Imai *et al.*, 1997) and interaction with E6 may prevent mediation of vitamin D3's growth suppressive effects (Thomas *et al.*, 1999b). Alternative splicing of the E6 protein creates four truncated versions, designated E6*I-IV, commonly called the E6* proteins. One of their known functions is to modulate E6-mediated degradation of p53 during viral replication (Thomas *et al.*, 1999b).

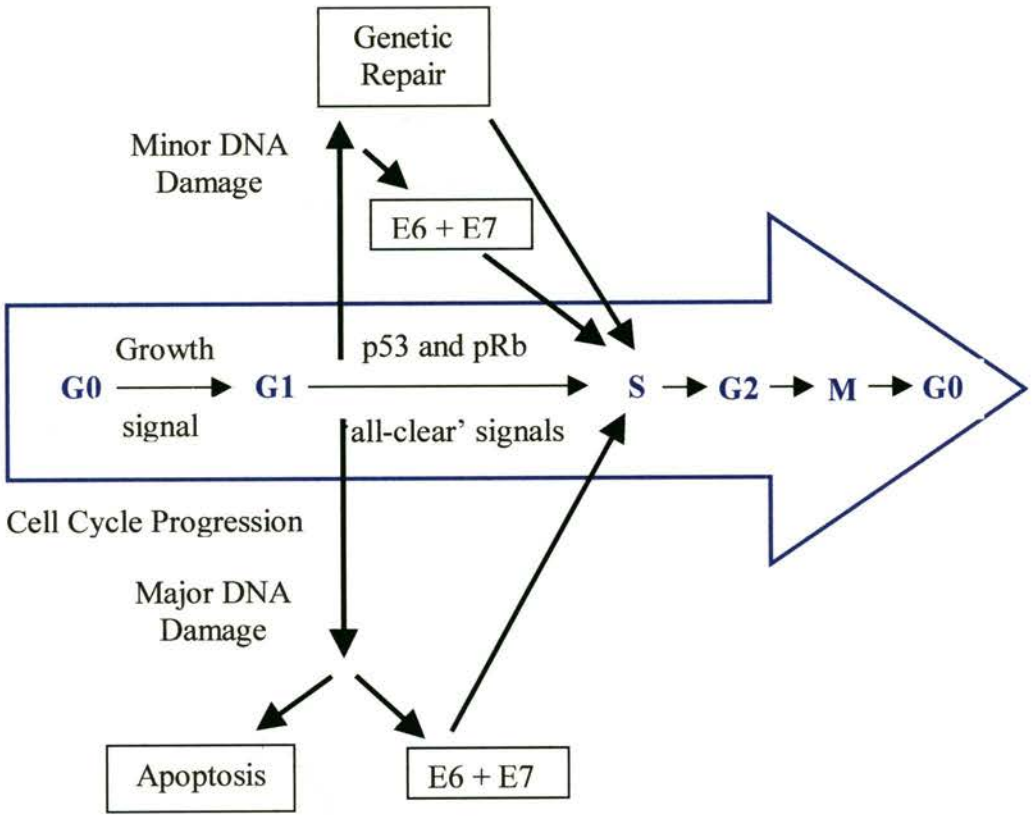



Figure 1.6: Cell cycle and disruption by the E6 and E7 proteins of carcinogenic human papillomavirus types (Adapted from Lörincz, 1999)

The cell cycle is the process by which the cell is activated from resting (G0) to grow and divide into two daughter cells (mitosis). An important checkpoint in the cell cycle is just prior to S phase (DNA synthesis) and is controlled by cellular proteins such as p53 and pRb. If an 'all-clear' signal is given, the cell cycle continues and the cell divides. If DNA damage is encountered, the cell attempts to repair the problem (if the damage is minor) or executes apoptosis (if DNA damage is major). The E6 and E7 proteins interfere with these pathways and drive the cells prematurely into the cell cycle, promoting the incorporation of mutations that speed the cell towards malignancy. G0, G1, and G2 = growth stages, and S = synthesis. M = mitosis.

The prevention of p53 DNA binding is also likely to play a role in abrogation of p53 transcriptional activity. The major function of E6 is to control p53 activities to assist the successful replication of the virus. Failure of the virus to balance p53 levels can lead to immortalisation and tumourigenesis. The E6 proteins of both oncogenic and benign HPV types associate with p53, however benign HPV types cannot target p53 for degradation and are much less potent inhibitors of p53 DNA binding (Crook *et*

*al.*, 1991b, and Lechner and Laimins, 1994). The carboxyl terminal region of the E6 protein is conserved among all HPV types and is important for p53 binding. However, the amino terminal sequences, necessary to direct p53 degradation, are conserved only between oncogenic HPV types which helps to explain the functional differences between the high and low risk genital HPVs (Crook *et al.*, 1991b, and Pim *et al.*, 1994). Similarly, E6BP does not associate with the low risk HPV E6 proteins (Chen *et al.*, 1995).

E6 can also overcome p53-independent apoptosis. This is partly the result of E6-induced ubiquitin-mediated degradation of the cellular *c-myc* and Bak proteins, both of which are associated with apoptosis (Gross-Mesilaty *et al.*, 1998, and Thomas and Banks, 1998 and Thomas *et al.*, 1999a). Recently it was recognised that E6 also targets the p53 coactivator, CBP/p300, an ability limited to high risk HPVs (Patel *et al.*, 1999, and Zimmermann *et al.*, 1999). CBP prevents G0/G1 transition during the cell cycle and facilitates p53-dependent gene expression. E6 binding to CBP represses the transcriptional activity of CBP and p53 transcriptional activity, and is another mechanism by which E6 abrogates p53 function; a function that is also shared by the other small DNA tumour viruses, SV 40 and adenoviruses. E6 also has transcriptional modulating activities that are independent of any effect on p53 (Crook *et al.*, 1991b and Desaintes *et al.*, 1992), and can stimulate telomerase activity (Klingelhutz *et al.*, 1996) which may lengthen the life span of the host cell and which may also contribute to transformation. A simplified figure of the HPV 16 E6 protein and the general location of its major functions are presented in Figure 1.7.

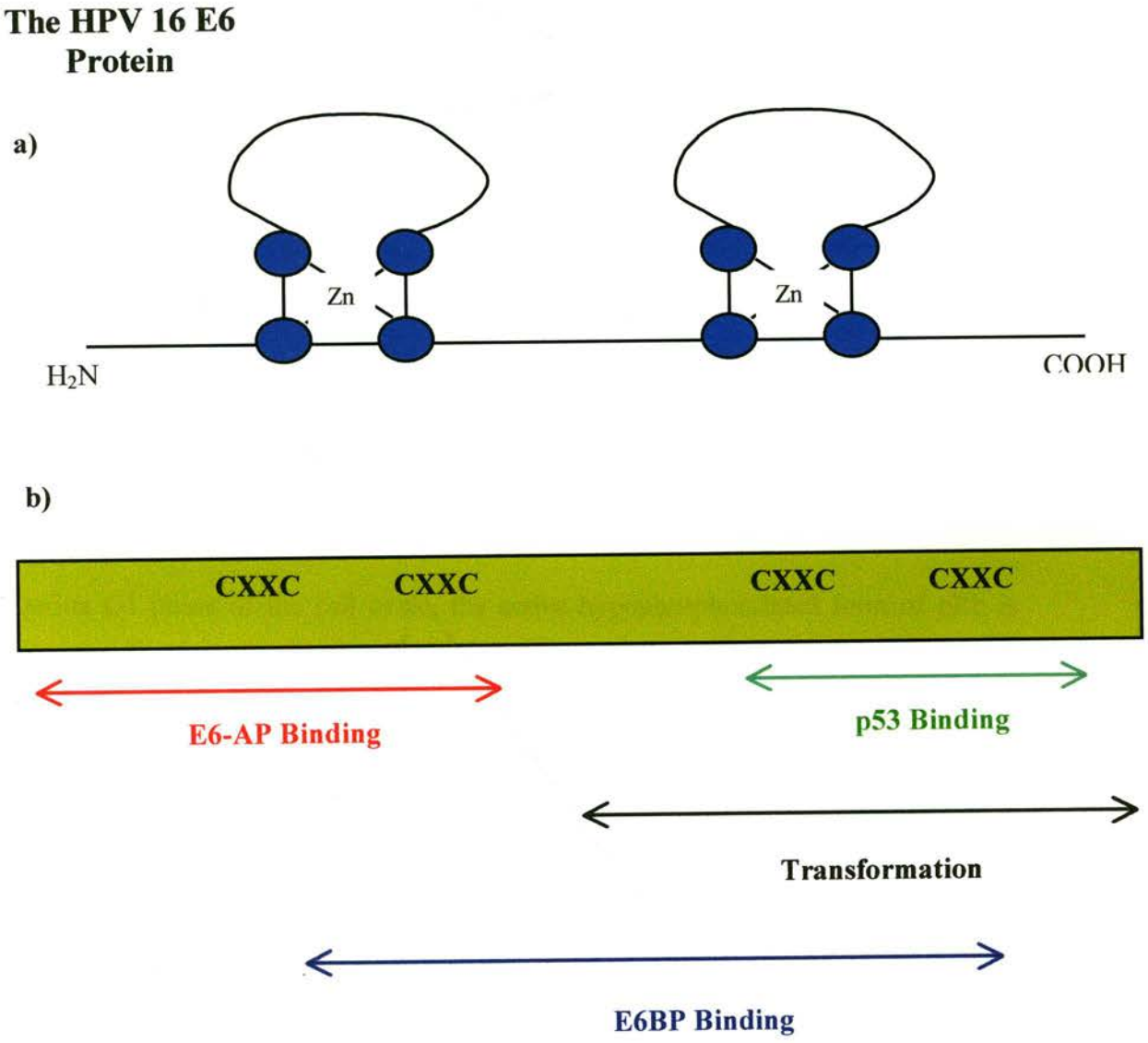



Figure 1.7: The human papillomavirus type 16 E6 protein

a) Diagrammatic representation of the two zinc fingers of the E6 protein, and b) The general location of the different regions of the E6 protein required for its major functions.

### 1.5.3.3 E7

The 10kDa to 14kDa E7 protein is a zinc-binding phosphoprotein that binds to the nuclear matrix. The amino terminus of the E7 protein has sequence and functional similarity to the adenovirus E1A protein and the SV 40 large T-antigen (Barbosa *et al.*, 1989, and Edmonds and Vousden, 1989). The major transforming capacity of the majority of HPVs resides in the E7 gene. This is mainly due to interaction with the

tumour suppressor protein retinoblastoma (pRb) (Dyson *et al.*, 1989, and Münger *et al.*, 1989a). The metal binding domain of E7, in particular a C-X-X-C motif, is important for this activity (Edmonds and Vousden, 1989, and McIntyre *et al.*, 1993). The E7 protein induces anchorage-independent growth of NIH 3T3 cells and cooperation with the oncoprotein *ras* transforms baby rat kidney (BRK) cells (Münger and Phelps, 1993, and Liu *et al.*, 1995). The transforming potential of E7 is directly correlated to its expression levels. E6 and E7 from high risk HPVs act cooperatively and are necessary and sufficient for immortalisation of cervical keratinocytes and for changing the differentiation pattern of keratinocytes (Howley-Nelson *et al.*, 1989, Münger *et al.*, 1989b, Hudson *et al.*, 1990, and Halbert *et al.*, 1991). The HPV 6 E6 and E7 proteins are unable to perform these functions.

During G1 phase of the cell cycle, the active hypophosphorylated form of pRb is complexed with E2F, resulting in inhibition of transcriptional activity. However in S phase, pRb is phosphorylated, the complex dissociates, and DNA replication occurs. E7 is also thought to disrupt this complex, releasing E2F to stimulate the transcription of cellular genes involved in cell cycle progression. Repression is normally relieved by cell cycle-regulated cyclin-dependent kinase (CDK) mediated phosphorylation of pRb. Therefore the E7 protein binding to pRB is similar to this process. It is possible that E7 targets pRb, which progresses the cell cycle through the G1/S checkpoint and the cell becomes unresponsive to signals that cause differentiation. Transformed cells expressing HPV 16 or 18 E6 lose the G1 checkpoint very early (Dulic *et al.*, 1994). Normally this leads to apoptosis, which is mediated primarily by p53. However E6 binds to p53 and the cell progresses to S phase instead, conferring a growth advantage to the cell. Senescence is overcome, resulting in immortalisation. Transformation may then result from additional abuse to the cellular genome. The G1/S interface "gatekeeper" is cyclin D, which regulates the CDK responsible for phosphorylating pRb, an event critical for transition into S phase. DNA damage leads p53 to induce expression of p21, a protein that inhibits cyclin D-CDK-mediated phosphorylation of pRb. This arrests cells in G1 phase. Therefore E7 inhibition of Rb leads to a loss of critical restriction points in the normal cell cycle.

The E7 protein is divided into three regions: CD1 (the amino terminal), which is required for immortalisation, CD2, which is required for pRb binding, and CD3 (the carboxyl terminal), which contains the zinc finger (Figure 1.8). The carboxyl terminus is also required for releasing E2F, is involved in dimerisation of E7, is necessary for transformation, protein stability and contributes to pRb binding. The carboxyl terminal sequences are virtually interchangeable between high and low risk types. The highly conserved amino terminus of E7 is the transcriptional transactivating domain and determines the affinity for binding to pRb and transformation. It is this region that is responsible for the biological differences between the high and low risk E7 proteins (Münger *et al.,* 1991). The E6 and E7 proteins of HPV 6 only have very weak immortalising activity (Halbert *et al.,* 1992). However sequence variation in the oncogenes of the low risk HPV types can confer oncogenicity (Heck *et al.,* 1992, and McGlennen *et al.,* 1992). The differences in transforming activity between HPV 6 and 16 E7 proteins is mainly determined by differences in the 30 amino terminal amino acid residues, particularly to the pRb binding region (Münger *et al.,* 1991, Heck *et al.,* 1992, Sang and Barbosa, 1992, and Takami *et al.,* 1992). E7s transactivation and transformation functions are interrelated, but separable, however pRb binding is necessary for both activities (Phelps *et al.,* 1992).

E7 also targets other cell cycle regulators such as $p21^{cip1}$, p107, p130, cyclin A, cyclin-associated kinase $p33^{CDK2}$, and TBP (Dyson *et al.,* 1992, Davies *et al.,* 1993, Tommasino *et al.,* 1993, and Ciccolini *et al.,* 1994). The exact roles of each of these interactions are not known. $p21^{cip1}$ is an important downstream inhibitor of cell cycle progression and HPV E7 binds to and abrogates its functions (Funk *et al.,* 1997 and Jones *et al.,* 1997). The association of E7 with pRB, p107 or p130 is believed to release E2F and allow transcription of genes regulated by it. A simplified figure of the HPV 16 E7 protein indicating the general location of its major functions is given in Figure 1.8.

**The HPV 16 E7 Protein**

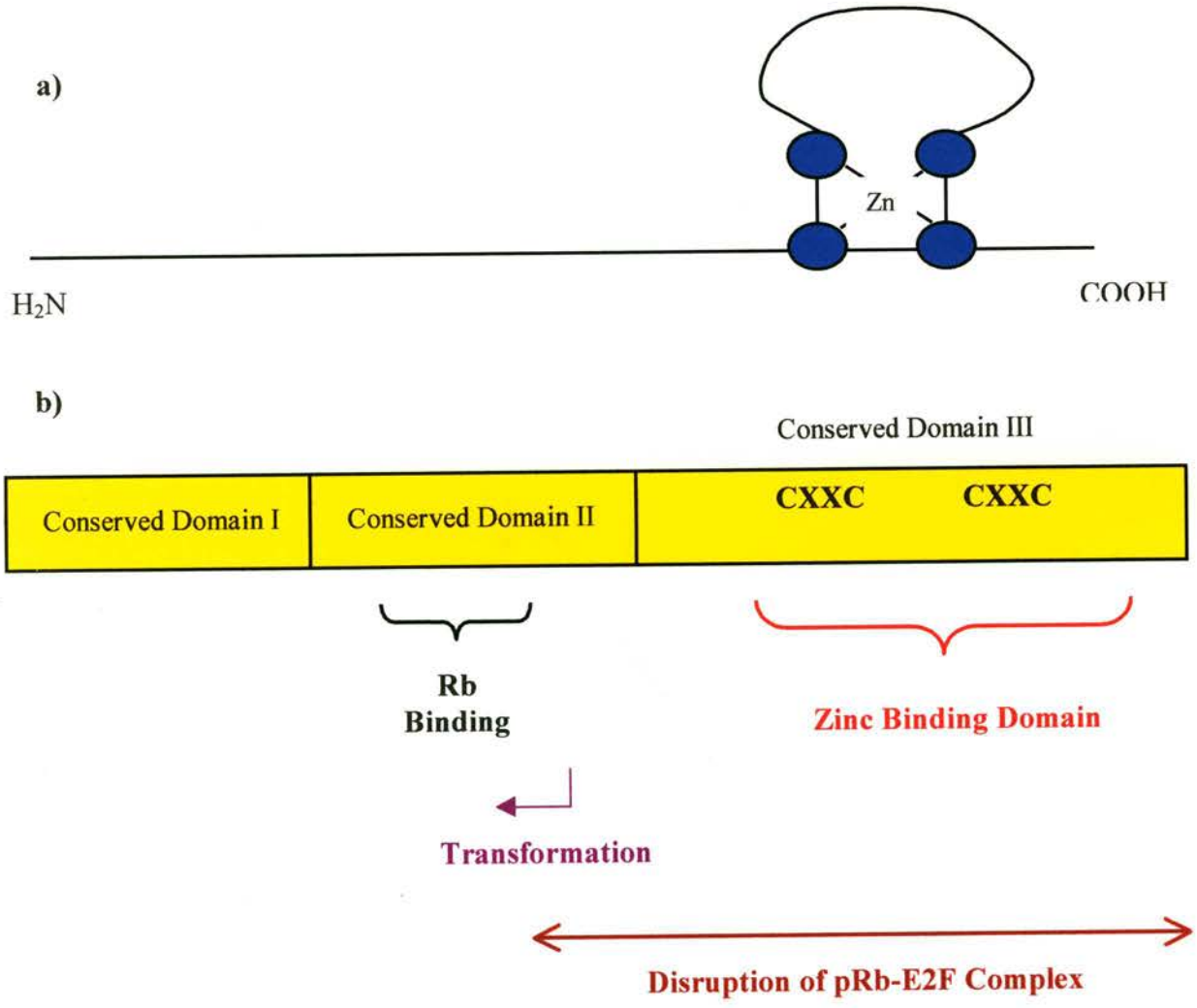



Figure 1.8: The human papillomavirus type 16 E7 protein

a) Diagram showing the zinc finger of the E7 protein, and b) The general location of the different segments of the E7 protein and different regions required for its major functions.

There is some experimental evidence that contradicts degradation of p53 by E6 and inactivation of Rb by E7 as mechanisms of HPV-induced transformation (Vernon *et al.*, 1998). Transfection of a mutant E7 that lacks Rb binding ability can still immortalise keratinocytes (Jewers *et al.*, 1992). Mice that lack p53 develop normally and therefore degradation by E6 may not be a direct cause of immortalisation

(Donehower *et al.*, 1992). Therefore E6 and E7 mediate host cell changes, eg. mutation and chromosomal abnormalities, by disrupting normal cell growth and these mutations may be able to persist because of E6 elimination of p53-mediated DNA repair (Vernon *et al.*, 1998). The major effects are likely to be in providing suitable conditions within the cell to enable viral DNA synthesis and replication to proceed, rather than transformation *per se* (Ullman and Emery, 1996).

#### 1.5.3.4 E8

The E8 ORF is present only among the BPVs and in HPV 6b. In BPV 1 a 28kDa E8/E2 fusion product is involved in transcriptional regulation by repressing E2 transactivation. The E8 ORFs of BPV 3, BPV 4 and BPV 6 are analogous to the E6 ORF, which is missing in these PVs.

### 1.5.4 The Capsid Proteins

#### 1.5.4.1 L1

Transcription of the PV late genes, L1 and L2 depends on poorly understood regulatory events associated with differentiation. L1 and L2 are only expressed in terminally differentiated keratinocytes within the upper layers of stratified epithelia. Promoters for L1 and L2 have been found in some PVs, but not in the genital HPVs. The L1 protein is the group antigen (common wart antigen) and is highly conserved among PVs from different species (Cowsert *et al.*, 1987, Pfister and Fuchs, 1987, and Yamada *et al.*, 1995). The L1 protein is 53kDa to 60kDa and makes up 80% of the viral capsid. It can be glycosylated *in vitro* and cross-links through disulphide bonds. L1 protein can self assemble into virus-like particles (VLPs) in expression systems to resemble virions (Kirnbauer *et al.*, 1993).

#### 1.5.4.2 L2

The 49kDa to 60kDa highly phosphorylated L2 protein is thought to be type-specific and is the minor structural component of PV capsids. The L2 ORF is the most variable coding segment in HPVs (Yamada *et al.*, 1995). Sequences at each end of the L2 gene, however, are highly conserved and may reflect the need to interact with the L1 protein. The amino terminal segments are also exposed on the viral surface,

are immunoreactive, and can generate neutralizing antibody responses (Yamada *et al.*, 1995). Amino acid changes in the L2 protein can alter its antigenicity (Yaegashi *et al.*, 1993). The central region of the protein is more species specific. L2 may be important for virus-cell interactions. A specific role for L2 in genome encapsidation has been suggested as L2, but not L1, binds HPV 16 DNA, and capsids self-assembled from L1 alone do not contain viral DNA (Zhou and Frazer, 1995 and 1996, and Roden *et al.*, 1996). The L2 protein does not self-assemble nor does it link to itself.

#### 1.5.4.3 L3 and L4

The L3 ORF is present only in BPV 4, deer papillomavirus (DPV), and HPV 5b. It is not known whether it is translated. The L4 ORF appears to be present only in BPV 4. Likewise, it is not known whether this ORF is translated.

### 1.5.5 The Upstream Regulatory Region (URR)

The URR is also referred to as the long control region (LCR) or noncoding region (NCR). The major transcription and replication controls of the viral life cycle are exerted from this region by complex interactions of viral DNA with hormones, cellular binding proteins and HPV gene products (Turek and Smith, 1996). The URR comprises 5% to 10% of the PV genome, varying from 400bp to approximately 1kb depending on the PV type. It encompasses the region from the termination of the L1 ORF to the first methionine of the E6 ORF. The URR contains between four and 12 E2BSs (four in HPV 16), one E1BS, and many binding sites for TFs (illustrated for HPV 16 in Figure 1.9). The URR also contains a nuclear matrix attachment region (Tan *et al.*, 1998), contains an epithelial-specific enhancer (Gloss *et al.*, 1987), the origin of replication, *ori*, and the early promoter. Regulation of this promoter may reflect the need to express the E6 and E7 proteins only when and where they are needed, particularly in terminally differentiating keratinocytes.

Different HPV types have homologous genes, but the URRs have diverged so much that they can't be aligned and nucleotide sequences lack long, similar segments. However, the URRs of more closely related HPV types, eg. genital HPVs, have

many short similar *cis*-responsive elements, suggesting conservation of function (O'Connor *et al.*, 1995). Therefore particular mechanisms must have been appropriate for the biology of these HPV types and could not undergo evolutionary change. The transcription of the major oncogenes, E6 and E7, is mostly regulated though these *cis*-responsive elements. Negative regulators such as E2 and the TF YY1 are particularly important as once their binding sites are removed, higher cellular levels of E6 and E7 are detected.

For transcription of the early genes to occur, RNA polymerase must recognise the promoter and form part of the transcription initiation complex. Transcription initiation is inefficient unless the complex is stimulated by transcriptional activators (Bernard and Apt, 1994). Many factors binding to the URR have been identified and cloned, although details of their molecular interactions with the transcription initiation complex remain obscure.

In genital HPVs, eg. HPV types 16, 18, 33, 31, 35, 45, 6, 11, and 34, there are four E2BSs that help divide the URR into three functionally distinct segments, the 5' region, the central segment, and the 3' region.

#### 1.5.5.1 The 5' Region of the Upstream Regulatory Region

The 5' region is approximately 300bp in length and is bounded by the termination codon of L1 and E2BS1. It contains the transcription termination and polyadenylation (nts 7321-7326) sites for late transcripts as well as a negative regulatory element acting at the level of late mRNA stability (Kennedy *et al.*, 1991, and O'Connor *et al.*, 1995).

#### 1.5.5.2 The Central Segment (Enhancer)

In general, clusters of transcription factor binding sites (TFBSs) form an enhancer that functions to 'enhance initiation' from a promoter element. The enhancer of genital HPVs is flanked by E2BSs 1 and 2 and is approximately 400bp in length. The enhancer contains many TFBSs of low or moderate activity acting in cooperation to generate a strong enhancer. Enhancers are found in many cellular, bacterial and viral genes. They can increase transcription by as much as several hundred-fold. In general, an enhancer can:

- Activate a promoter when placed up to several thousand bases from the promoter
- Activate a promoter when placed in either orientation relative to the promoter
- Activate a promoter when placed upstream or downstream of the transcribed region (Latchman, 1990).

Enhancers may activate gene expression either by changing the chromatin structure leading to nucleosome displacement or by direct interaction with the proteins of the transcriptional apparatus. In the case of chromatin structure changes, it is believed that such changes can be propagated over large distances, in both directions, causing the observed distance, position and orientation independence of the enhancer. Models involving the binding of protein factors to the enhancer followed by direct interaction with proteins of the transcriptional apparatus have been proposed to occur by several methods. For instance, by regulatory factors making contact with the promoter bound transcriptional apparatus by sliding along the DNA, or via a continuous scaffold of other proteins, or by looping out the intervening DNA (Latchman, 1990).

The HPV 16 enhancer has been mapped by DNase I protection studies and bandshifts to reveal TFBSs (Gloss *et al.*, 1989a, Cripe *et al.*, 1990, and Sibbet and Campo, 1990). TFBSs are necessary for modulating viral gene expression in response to the physiological and differentiation state of the host keratinocyte in different layers of the epithelium (O'Connor *et al.*, 1995). These changes are brought about by the differential binding of TFs to their sites in the enhancer and this in turn is dependent upon the quantitative and perhaps qualitative differences in TFs in a given cellular environment (O'Connor *et al.*, 1995). In a tissue containing an enhancer-binding protein, the enhancer will activate a promoter, whereas in other tissues where the protein is absent or inactive, the enhancer will have no effect. Some enhancers produce their action by relieving a negative effect. In these cases it appears that the enhancer binds a factor present in all tissues, which inhibits the promoter. Following induction, or in a particular tissue, this negative factor is inactivated or falls off the enhancer, allowing a positive factor to bind and activate gene expression. The HPV 16 and 18 enhancers display cell type specificity; for example, they are active in epithelial cells and inactive in fibroblast cells (Chong *et al.*, 1991).

More specifically, the cell type specificity of the HPV URR has been attributed to:

- The interplay between a number of different ubiquitous factors
- A combination of TF synergism or combinatorial regulation. Synergism is eliminated by mutations in alternative TFBSs, although the remaining sites in the enhancer are unaltered (Cripe *et al.,* 1990, and Chong *et al.,* 1991).
- Concentration differences of qualitatively identical factors
- Functional variation of factors through means other than their DNA recognition specificity (Apt *et al.,* 1993). For instance, TFs such as NFI, AP1, and Sp1, are derived from multigene families. Epithelial specificity originates from the fact that different members of these gene families differ in function and are expressed in a cell-type-specific manner.

In the HPV URR, degenerate TFBSs seem to be the rule rather than an exception, and could result in lower levels of HPV gene expression, which might be necessary for long-term viral persistence. Each enhancer element provides a quantitative contribution, rather than an all-or-none switch (Chong *et al.,* 1991). Degenerate binding sites would have an inherent sensitivity to changes in the transcriptional environment. Degenerate sites could also allow for a more subtle level of regulation, as different degeneracies may have different binding affinities, thus allowing for differential occupancy of sites at a given concentration of a TF (O'Connor *et al.,* 1995). Degeneracy of sites would also determine which particular sequences were bound by which factors where more than one site overlap (O'Connor *et al.,* 1995).

#### 1.5.5.3 The 3' Region (Promoter)

The 140bp 3' region is bounded by E2BS 2 and the ATG start codon of the E6 gene. The transcription start site is approximately 90bp downstream of the E1BS. Within this 90bp is an Sp1 site, two E2BSs, the TATA box and several other TFBSs. The precise spacing of these elements in the promoter is conserved in all genital HPVs.

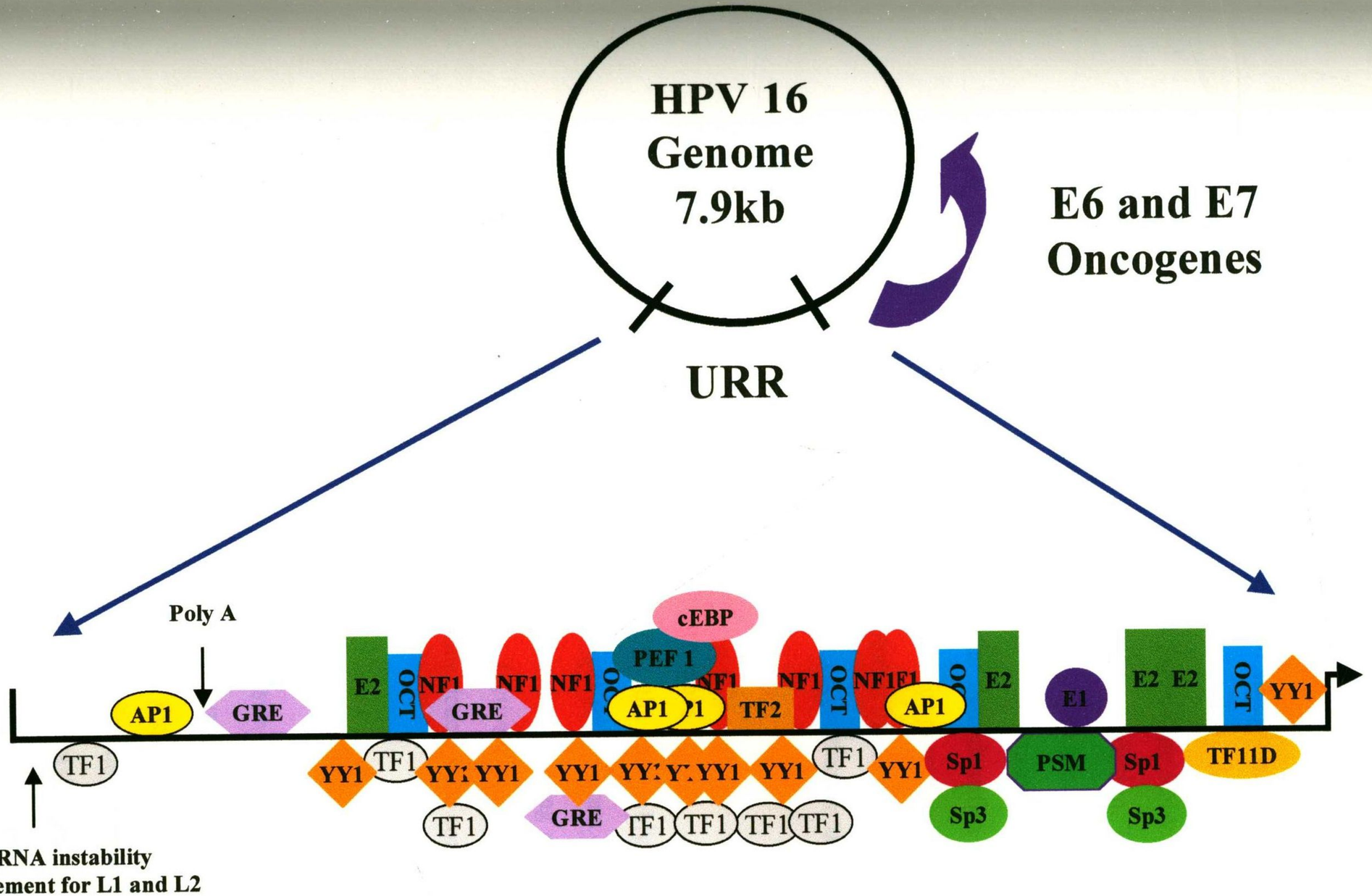



Figure 1.9: Expanded view of the human papillomavirus type 16 upstream regulatory region showing transcription factor binding sites

Sites represented: TF1=TEF-1, AP1, GRE, E2, YY1, OCT=Oct-1, NFI, PEF1, cEBP, TF2=TEF-2, Sp1, E1, E2, Sp3, PSM, and TFIID. Poly A = polyadenylation site for late transcripts. Diagram not to scale.

### 1.5.6 Transcription and Regulatory Factors

Control of transcription initiation rate is one means by which the expression of genes can be varied. Essentially this involves activating or repressing transcription in response to specific signals, by binding TFs to specific short DNA sequences in the promoter and/or enhancer of a gene. A TF is essentially defined as any protein that is needed for transcription initiation but is not part of the RNA polymerase complex. TFs can be repressors, activators or initiators of transcription. The accessory proteins required for RNA polymerase to initiate at a promoter are called basal TFs, involved in the mechanics of binding to DNA and initiating transcription. The nomenclature for basal factors is TFIIX, where 'X' is a letter that identifies the individual factor. Additional upstream factors are usually required for proper promoter function. These factors have casual names reflecting their histories of identification. TFs mediate the final steps in the relay of information from the cell surface to the level of expression of the gene (Eckert *et al.*, 1997).

#### 1.5.6.1 Transcriptional Repressors and Activators

Active or direct repression interferes with the assembly or action of the transcriptional preinitiation complex by physical interactions with its components. Activator-specific repression involves quenching by targeting an activator (passive), or quenching by targeting a general factor (active). Repressors may also alter chromatin structure by the recruitment of histone deacetylase complexes.

Passive repression involves:

- Direct competition for the same DNA binding site
- Interference with overlapping or neighbour activator-binding sites
- Modification of the DNA binding properties of the activator
- Titrating away limiting protein factors required for the activator function
- Masking and/or altering the function of the activation domain, or
- Blocking the DNA binding activity of the activator through protein-protein interactions (from Hou *et al.*, 2000).

A positive regulator (activator), after binding to DNA, exercises its functions by binding to other components of the transcription initiation apparatus. At a minimum, a TF contains a domain that binds to DNA and a domain that regulates transcription. They also frequently contain a domain that regulates their own activity, such as a phosphorylation site. These domains are usually functionally and physically separable. The DNA binding domain of a TF typically has an α-helical shape, within or adjacent to which resides clusters of positively charged amino acids (Papavassiliou, 1995). Recurrent structural motifs in DNA binding domains provide the basis for TF classification. The majority of TFs contain one of four motifs; helix-turn-helix (HTH) motif, zinc finger, leucine zipper, or helix-loop-helix (HLH) motif.

## 1.5.7 DNA Binding Domains

### 1.5.7.1 Zinc Finger Motifs

In a zinc finger motif a group of conserved amino acids bind a zinc ion. Zinc finger proteins generally have a series of zinc fingers, which comprise a DNA binding domain. The zinc is grasped by two cysteine (cys) and two histidine residues ( for example, Sp1 ) or four cysteines. The carboxyl terminal region of each finger forms an α-helix that binds DNA and the amino terminal region forms a β-sheet. Three α-helical stretches fit into one turn of the major groove of the DNA, and each α-helix, and thus each finger, makes two sequence specific contacts with DNA. Variations on the zinc finger include multi-cysteine zinc motifs found in steroid hormones.

### 1.5.7.2 Steroid Receptors

Binding of particular steroids activates steroid receptors. These zinc finger proteins usually contain non-repetitive cys/cys fingers with a central highly conserved domain for DNA binding. The carboxyl terminal region is involved with binding the hormone and the amino terminal region produces the maximal induction of transcription. Glucocorticoid and estrogen receptors have two such fingers. Two glucocorticoid receptors dimerise on binding to DNA and each finger controls one property of the receptor, ie. amino acids on the right side of the first finger control binding to the DNA and those on the left side of the second finger control the ability to form a dimer.

#### 1.5.7.3 Homeodomains - Helix-Turn-Helix Motifs

The homeodomain contains a helix-turn-helix (HTH) motif that consists of three closely linked α-helices. The third α-helix lies in the major groove of the DNA, while the helices 1 and 2 lie at an angle above the DNA.

#### 1.5.7.4 POU Domains

In many TFs of the HTH type, additional motifs flank the homeodomain and modulate its DNA binding affinity and specificity for particular DNA base sequences. For instance, the POU domain contains a homeobox-like sequence and a second conserved domain, the POU-specific domain. In the TF Oct-1, both domains are required for sequence-specific DNA binding, indicating that the POU homeodomain and the POU-specific domain can form two parts of a DNA binding element which are held together by a flexible linker sequence. The POU domain contains numerous α-helices (Latchman, 1990 and 1999). Other regions of the protein outside the POU domain then mediate changes in the transcription of genes. It has been suggested that POU factors play a critical role in determining the level of viral transcription following infection (Latchman, 1999). This could affect whether an initial viral infection produces a lytic infection, transformation, or an asymptomatic latent infection capable of reactivation.

#### 1.5.7.5 Leucine Zipper Motifs

Leucine zipper proteins consist of an α-helix in which leucines occur every two turns on the same side of the helix in every seventh position. A leucine zipper in one polypeptide interacts with a zipper in another polypeptide to form a dimer. The leucine residues facilitate dimerisation of two molecules of the TF by promoting interdigitation of two such helices. Leucine zippers do not directly bind DNA. By facilitating dimerisation of the protein they provide the correct protein structure for DNA binding by the adjacent region of the protein, which is rich in basic amino acids that can interact with the acidic DNA. Adjacent to each zipper is a stretch of positively charged resides involved in DNA binding, for example as found in c/EBP. In addition to the formation of a dimer by two identical factors, there can also be heterodimers between two different factors, which might have different properties in terms of sequence specific binding and gene activation compared to homodimers, eg.

JunB, JunD, and JunC all have leucine zippers that can interact to form homo- or heterodimers that bind AP1 sites. Another factor, Fos, has a leucine zipper that cannot form homodimers but can form heterodimers with Jun. The heterodimer binds to AP1 sites with greater than 30-fold affinity than the Jun homodimer (Latchman, 1990).

#### 1.5.7.6 Helix-Loop-Helix Motifs

Helix-loop-helix (HLH) proteins contain two helices linked by a loop, eg. proto-oncogenes like *c-myc*. The HLH TFs have some similarities to the leucine zipper family members. Both have a region of positively charged amino acids that contacts the DNA and an adjoining region that mediates dimer formation. Both families of TFs also have key roles in differentiation and development.

### 1.5.8 Activation Domains

Although binding to DNA is necessary for transcription, it is not in itself sufficient. Following DNA binding the TF must increase transcription, either by directly activating RNA polymerase or by facilitating the binding of other TFs and the assembly of a stable transcriptional complex. The 'activation domain' is usually distinct from the DNA binding domain. Activation domains in various TFs have no strong amino acid homology, but in many cases they have a very high proportion of acidic amino acids, resulting in a strong net negative charge, which is important for activation (Latchman, 1990). However Sp1 has two activation domains, neither of which are particularly acidic. Instead they are both rich in glutamine resides that are essential for transcriptional activation. The situation is similar in Oct-1 and Oct-2. In NFI the activation domain is rich in proline. There is a widespread interchangeability of activation domains from yeast, Drosophila and mammalian TFs, therefore one common mechanism may mediate activation in a wide range of organisms. Despite this, the nature of this universal mechanism is unclear. There are two suggested means – either activation by contact with RNA polymerase itself, or by contact with another TF, facilitating its binding into a stable transcription complex or by allowing it to interact with the RNA polymerase. There are regions on the RNA polymerase II large subunit that could interact with a negative domain or by hydrogen bonding via

glutamine-rich regions, however there is no evidence of this to date. In eukaryotes the TATA box is essential for transcription and binds TFIID. This could be the target for the activation domain, of which there is some evidence for (Latchman, 1990).

### 1.5.9 Transcription Factor Regulation

TF activity can be regulated in several ways. For instance by:

- Tissue specificity – ie. expressed in only certain tissues and therefore limiting what genes can be transcribed by it.
- Altered expression - eg. AP1 is overexpressed in cervical cancers (Nurnberg *et al.,* 1995).
- Modifications from active form by phosphorylation - eg. AP1 is converted to the active form by dephosphorylation of the Jun subunit.
- Binding of a ligand leading to activation or inactivation - eg. steroid receptors. This can influence localisation and ability to bind DNA.
- Alternate partners for a dimeric factor - one partner may cause the factor to be inactive, and synthesis of the active partner may displace the inactive partner.
- Controlling synthesis of the TF.

### 1.5.10 Factors that Affect Regulation by the Upstream Regulatory Region - Focussing on Human Papillomavirus Type 16

Approximately 20 different TFs have been described that bind to the URRs of PVs (Chong *et al.,* 1990 and 1991, and Sibbet and Campo, 1990). Cellular TFBSs in PV enhancers often do not bind as strongly as consensus motifs and it is the cooperation of these factors that produce enhancer activity (Chong *et al.,* 1990). The advantage to the virus is that it can be more sensitive to changes in the cellular environment and can modulate gene expression more readily. A list of factors known to bind the HPV 16 URR and the position of their binding sites are given (Table 1.4). Factors that affect the transcription of other PV types, but for which no role has been identified in the HPV 16 URR are not mentioned.

### 1.5.10.1 Promoter Specific Factor 1 (Sp1)

Sp1 is a ubiquitous zinc finger protein that activates RNA polymerase II dependent promoters of a large number of cellular and viral genes (Dynan and Tjian, 1983, and Hoppe-Seyler and Butz, 1992). Efficient transcription initiation depends on an Sp1 binding site that co-ordinates the signal from the enhancer. Sp1 facilitates the binding of components of the pre-initiation complex by competing with nucleosome assembly. The binding of E2 protein to the URR interferes with Sp1 binding in an overlapping motif (Gloss and Bernard, 1990). Sp3 is a transcriptional repressor that binds to the same motif. More recently, a second Sp1 site in the HPV 16 URR, upstream of the main Sp1 site was identified (Dong and Pfister, 1999). This site overlaps a YY1 site (YY1 9, Table 1.4) and competition between Sp1 and YY1 for DNA binding plays a role in YY1 repression. In Dong and Pfister's study, YY1 bound more efficiently than Sp1 at the same protein concentration. However, the Sp1 complex predominated in gel shifts using extracts from cervical carcinoma cell lines, suggesting a predominance of Sp1, which is in agreement with the high levels of Sp1 previously reported in cell lines (Apt *et al.,* 1996). The authors suggested that competition between Sp1 and YY1 could contribute to the course of epithelial differentiation and tumour progression.

### 1.5.10.2 Yin Yang 1 (YY1)

Yin and Yang 1 (YY1) is a multifunctional TF that regulates the expression of a number of cellular and viral genes. It can be an activator, repressor and initiator of transcription depending on context. It has separate activation and repression domains, which partly explains its multifunctionality. YY1 has multiple mechanisms for mediating repression, such as activator-specific repression (Galvin and Shi, 1997). Repression likely reflects the ability of YY1 to interfere with communication between activators and targets in the general transcriptional machinery. The YY1 protein has four carboxyl terminal zinc fingers, which constitutes the transcriptional repression domain. The minimal DNA binding domain of YY1 is composed of zinc fingers two and three. Of these, zinc finger two plays a central role in DNA binding and transcriptional repression. YY1 can also form heterodimers with other TFs. It may also act as a DNA-bending protein, which could regulate contacts between proteins (Natesan and Gilman, 1993).

YY1 was first implicated in the regulation of HPV 18 transcription by Bauknecht *et al.* (1992). It represses expression from the HPV 18 promoter by repressing AP1 activity, however the process was found to be complex and influenced by sequences upstream of the YY1 site called the switch region (Bauknecht *et al.,* 1995). In HeLa cells, the switch region causes YY1 to activate the promoter, however deletion of the switch region causes YY1 to act as a repressor of promoter activity. Bauknecht *et al.* (1995) suggested a cell-type specific and context-dependent activity of YY1 in HPV 18. That is, the ability of YY1 to repress the promoter is changed by its functional interaction with the switch region. More recently, a combination of YY1 and CCAAT enhancer binding protein (c/EBP) β was shown to bind to the switch region (Bauknecht *et al.,* 1996).c/EBP is a family of leucine zipper factors implicated in the regulation of many genes, including genes involved in cellular differentiation.

Dong *et al.* (1994) initially identified a transcriptional silencer upstream of the early promoter (p97) in HPV 16 that comprised several binding sites for YY1. YY1-mediated repression of p97 was more fully characterised by O'Connor *et al.* (1996) who identified 10 YY1 sites in the HPV 16 URR. Of these a cluster of five (designated YY1 4 to 9) located in the vicinity of an AP1 site was responsible for promoter repression, with all five sites being required for repression (O'Connor *et al.,* 1996). YY1 may repress AP1 activity as a result of YY1-CREB-binding (CBP) protein interaction. Interestingly, YY1-mediated repression is also observed in other viruses involved in persistent or latent infections, eg. adeno-associated virus (AAV), EBV, HSV 1, and CMV. In many of these cases, YY1 serves to keep viral gene expression at low levels until a specific trigger or phase in the viral life cycle when relief of YY1 repression occurs. Likewise, YY1-mediated repression may be important in order to maintain low levels of E6 and E7 in mitotically active basal epithelial cells, where the oncoproteins are not required (O'Connor *et al.,* 1996).

**1.5.10.3 Activator Protein 1 (AP1)**

AP1 binding sites are present in all HPV enhancers and there is some conservation in their positioning. There are seven AP1 sites in the HPV 16 URR (Gloss *et al.,* 1989a). AP1 binds as heterodimers made up of proteins derived from the *jun*, *fos*, and ATF gene families. Fos does not bind DNA on its own but can associate with any of the Jun proteins, with Jun:Fos heterodimers being more efficient at DNA

binding than Jun homodimers. AP1 sites in cellular genes usually serve as downstream targets of intracellular signalling mechanisms, eg. AP1 sites in HPV may modulate viral expression in response to growth factors such as epidermal growth factor (EGF). Indeed the HPV 16 early promoter, p97, is inducible by phorbol esters as a result of its AP1 sites (Chan *et al.,* 1990). Phorbol esters mimic diacylglycerol, which is a second messenger that induces protein kinase C to phosphorylate target proteins. This leads to stimulation of AP1 dependent promoters. Likewise, AP1 sites may translate increased mitogenic signals resulting from the actions of E5 into increased levels of viral gene transcription (as described in section 1.5.3.1). AP1 sites might also mediate the repressive effects of retinoids on HPV gene expression.

AP1 sites may also contribute to epithelial-specific activation since the genes that give rise to its subunits are differentially expressed (Welter and Eckert, 1995). Different AP1 factors are available to regulate gene expression at each differentiation stage in the epidermis. Likewise, AP1 sites with sequence variation from consensus are preferentially recognised by subsets of AP1 (Hoppe-Seyler and Butz, 1994), which could have important consequences because in cervical lesions, the pattern of HPV expression correlates with the pattern of AP1 expression (Kyo *et al.,* 1997).

#### 1.5.10.4 Nuclear Factor I (NFI)/CCAAT Transcription Factor (CTF)

All HPV enhancers contain low-affinity, half-palindromic NFI sites; HPV 16 has seven. The NFI family of genes consists of NFI-A, -B, -C, and -X. These different factors have a highly conserved amino terminal DNA binding and dimerisation domain. The carboxyl terminal transactivating domains are heterogeneous and are generated by differential RNA splicing of one gene product. NFI binds as a dimer and in epithelial cells, the major forms of the NFI dimers are made up of three splice products from the NFI-C gene, CTF-1, -2, and -3. In contrast, in fibroblasts, which don't support HPV transcription, spliced isoforms of the NFI-X gene predominate (Apt *et al.,* 1993 and 1994). Therefore the NF1 sites, like AP1 motifs, may determine epithelial-specificity of an enhancer.

#### 1.5.10.5 Octamer Binding Protein-1 (Oct-1)

Oct-1, a member of the POU factor family, regulates a large number of viral and cellular genes. Oct-1 recognises even quite degenerate motifs, however the three flanking base pairs play a role in determining binding affinity. At physiological levels, Oct-1 activates the promoter by binding to a conserved element at the 3' end of the enhancer, called NFA (NFI-associated). This octamer motif is separated by 2bp from an NFI site that is highly conserved among genital HPVs. A point mutation that prevents the binding of Oct or NFI causes a 10- to 12-fold decrease in enhancer activity (O'Connor and Bernard, 1995). A point mutation preventing the binding of both proteins does not cause any further decrease in activity, suggesting a cooperative interaction of both factors. Simultaneous binding of both factors to the composite element is required for efficient activation, suggesting that Oct-1 does not activate directly, but may act by tethering NFI. In contrast, Sibbet *et al.* (1995) reported another Oct-1 site in the URR that down regulates the promoter. Oct-1 may also modulate HPV expression in a cell-cycle dependent manner. In different layers of the epithelium, it may not be Oct-1, but other POU factors, which bind octamer motifs, eg. Oct-6 (O'Connor *et al.*, 1995).

#### 1.5.10.6 Transcriptional Enhancer Factor-1 (TEF-1) and TEF-2

TEF-1 is a ubiquitous TF, is not epithelial-specific and TEF-1 motifs are not present in all HPVs. However it requires a co-activator which is present in epithelial cells and absent in several other cells (Ishiji *et al.*, 1992, and Bernard and Apt, 1994). This constitutes functionally relevant variation as the co-activator is both limiting and cell specific. There are a number of potential TEF-2 sites in the majority of genital HPVs and a TEF-2 site in the HPV 16 URR has been identified (Chong *et al.*, 1991).

#### 1.5.10.7 Papilloma Enhancer Binding Factor-1 (PEF-1)

A motif with enhancer activity in the HPV 16 URR has been identified and called PEF-1 (Cuthill *et al.*, 1993). Its binding site is centred around a CCAAT box-like CCAAC element, and introduction of A to T transitions into the CCAAC element abolishes PEF-1 binding and enhancer function. More recently, the same group demonstrated that the PEF-1 site overlaps an Oct-1 site (Sibbet *et al.*, 1995). The URR is upregulated by PEF-1 and down regulated by Oct-1, with both proteins competing for DNA binding.

### 1.5.10.8 Papillomavirus Silencing Motif Binding Protein (PSM-BP)

O'Connor *et al.* (1998) identified a regulatory element in the HPV 16 URR that had silencing activity, and named it the PV silencing motif (PSM). It consists of two copies of a sequence that overlaps the origin of replication. Each sequence is 9bp in length separated by 7bp, and each copy of the sequence binds the same cellular factor, the PSM-binding protein (PSM-BP). Both copies of the binding sequence are required for transcriptional repression as they stabilise dimer formation, and only the dimer is capable of repression.

More recently O'Connor *et al.* (2000) identified PSMs in six other HPVs as well as identifying the PSM-BP as CCAAT displacement protein (CDP), or Cut protein. CDP is a homeoprotein expressed in basal epithelial cells, but not in differentiated primary keratinocytes. CDP/Cut is also known to repress cellular genes that are developmentally regulated. Tissue-specific activation of these types of promoters occurs upon terminal differentiation, when CDP/Cut availability decreases and promoter binding is lost. The PSM overlaps the origin of replication, *ori*, in the URR and the binding of E1 and CDP/Cut may be mutually exclusive. Repression by the PSM may couple HPV transcription to the stratification of epithelia. CDP functions by two distinct mechanisms. It can cause factor displacement, or it can have silencer activity without displacement of an activating TF. Association with histone deacetylase and potential structural alteration of nucleosomes may result in repression. CDP/Cut may be an important regulator of HPV transcription and replication during epithelial differentiation. Negative regulation of the HPV 6 early promoter by CDP/Cut has also been demonstrated (Pattison *et al.,* 1997, and Ai *et al.,* 1999).

### 1.5.10.9 Glucocorticoids and Progesterone

The importance of glucocorticoids and progesterone in HPV gene expression regulation and in transformation has been demonstrated (Pater *et al.,* 1988, Chan *et al.,* 1989, Dürst *et al.,* 1989, and Pater *et al.,* 1990). Binding sites for the activated glucocorticoid receptor (GR), which confers strong inducibility by glucocorticoids (dexamethasone) have been located in the HPV 16 URR (Gloss *et al.,* 1987). These binding sites, called glucocorticoid response elements (GRE), recognise activated

receptors whose half sites are arranged as palindromes. Both the GR and progestoerone receptor (PR) binds to the GREs (Chan *et al.*, 1989).

The activated GR, which consists of glucocorticoid bound to its receptor, has a central DNA binding domain. The amino terminal region is the least conserved part and the carboxyl terminal region binds the hormone. In the absence of the steroid, the steroid-binding domain prevents the receptor from recognising the GRE, a form of internal negative regulation. Induction of HPV gene expression by steroid hormones may account for the increased prevalence of HPV in pregnant women (Schneider *et al.*, 1987).

#### 1.5.10.10 Estrogen

Estrogen increases the transcription of the HPV 16 oncogenes in the cell line SiHa (Mitrani-Rosenbaum *et al.*, 1989) and promotes cell proliferation by increasing the response of the cell to growth factors and their receptors (Auborn *et al.*, 1991). This may help explain why PV lesions are exacerbated during pregnancy when estrogen levels are elevated (Schneider *et al.*, 1987).

#### 1.5.10.11 Epidermal Growth Factor

Epidermal growth factor (EGF) serves as a mitogen for epithelial cells. The EGFRE elicits positive regulation of the HPV 16 URR, an activity that has been mapped to the proximal 124bp of the URR (Yasumoto *et al.*, 1991). Increased expression of EGFR is associated with more aggressive behaviour and poorer prognosis in several types of cancers, including cervical cancer (Pfeifer *et al.*, 1989, Kim *et al.*, 1996, Kristensen *et al.*, 1996, and Kersemaekers *et al.*, 1999). It has been suggested that EGF induction of HPV 16 transcription is effected by AP1 as EGF is a potential natural activator of AP1 (Peto *et al.*, 1995). No EGF-dependent increase in p97 activity was found in immortalised keratinocytes in one study, suggesting that changes in growth factor signaling pathways may occur during tumour progression (Peto *et al.*, 1995).

### 1.5.10.12 Cytokines

The complicated network of positive and negative control mechanisms of the URR is also sensitive to extracellular signals. The cytokines transforming growth factor (TGF)-$\beta_1$ and $\beta_2$, leukoregulin, interferon (IFN)-$\gamma$, interleukin (IL)-1, and tumour necrosis factor (TNF)-$\alpha$ all suppress the expression of the E6 and E7 genes (Fuchs and Pfister, 1994, and Kyo *et al.*, 1994). NF-IL6/cEBP$\beta$ is another cytokine that represses the HPV 16 URR (Kyo *et al.*, 1993). The DNA binding domain of NF-IL6 is required for repression while its functional domain is not, suggesting that repression may be the result of competition with activators. This form of negative regulation probably contributes to the host defense against HPV infection. Normal cervical epithelial cells secrete many different cytokines, however, synthesis is significantly decreased in cell lines immortalised by HPV, and in HPV positive cervical cancer cell lines (Woodworth and Simpson, 1993).

### 1.5.10.13 Cellular Oncogenes

Proto-oncogenes encode growth factors that stimulate the growth of normal cells and other cellular proteins involved in transmitting the growth signal within the cell. For example, c-*myc* is normally involved in the regulation of DNA replication and cell division. Cancers exhibit extensive malregulation of gene expression, which results from the overexpression of proto-oncogenes due to errors in their regulation. Human cancers with no viral involvement are often due to the overexpression of individual proto-oncogenes or to their mutation. Amplifications of the c-*myc* and H-*ras* oncogenes have been found in cervical carcinomas as well (Pinion *et al.*, 1991, and Kinoshita *et al.*, 1993), and c-*myc* amplification is an early critical event in progression of cervical lesions (Aoyama *et al.*, 1998). *Ras* activating mutations have also been identified in HPV positive cancers (Anwar *et al.*, 1992, Enomoto *et al.*, 1990 and 1991, and Grendys *et al.*, 1997).

Several proto-oncogenes encode TFs that regulate the transcription of genes activated in growing cells. Increased levels of Jun and Fos are observed after treatment of cells with phorbol esters, which in turn causes increased transcription of other genes containing AP1 binding sites. Increased levels of Jun and Fos are also produced by treatment with serum or growth factors, which stimulates the growth of

quiescent cells. Therefore if viral infection causes increased and continual expression of these proteins, a cell can potentially be stimulated to grow continuously and will not respond to growth-regulating signals, which is sufficient to transform cells. The presence of AP1 sites in the URR brings the activation of E6 and E7 under the influence of cellular tumour genes and deregulation of nuclear proto-oncogene expression may contribute to overexpression of E6 and E7 (Nurnberg *et al.,* 1995).

#### 1.5.10.14 DNA Methylation

DNA methylation is involved in cellular senescence and has a role in viral latency in HIV infection. Integration of HPV DNA into the host genome subjects HPV to host-specific DNA methylation, resulting in down-regulation of transcriptional activity (Rösl *et al.,* 1993). This is not mediated by steric hindrance of methyl residues with TFs, but rather is due to an association with methyl-DNA binding proteins.

#### 1.5.10.15 Viral Proteins (Other than Papillomavirus)

Adeno-associated virus (AAV) is commonly found in the anogenital region and is negatively associated with cervical cancer. This is due to the fact that AAV can inhibit HPV-induced transformation (Hermonat, 1989 and 1994, and Hermonat *et al.,* 1997). This was recently shown to be due to the AAV regulatory protein Rep 78, which binds to the HPV 16 early promoter (nts 14 to 56), overlapping the Sp1 site necessary for transactivation (Zhan *et al.,* 1999). Rep 78 binding to p97 may also inhibit the E1/E2 complex binding to the origin of replication. Another example of a viral protein that can interact with the HPV URR is the HIV tat protein, which can transactivate the URR (Vernon *et al.,* 1993).

Table 1.4: Binding Sites in the Human Papillomavirus 16 Upstream Regulatory Region

| Factor | Consensus Sequence (5' to 3') | Effect on Promoter | Reference/s | Binding Sites in the HPV 16 URR |
|---|---|---|---|---|
| YY1 | TTGCGACATTTTGCG | Binding sites 4, 5, 7, 8 and 9 are repressors | O'Connor *et al.*, 1996 | 1. 7431-7446<br>2. 7589-7603<br>3. 7599-7608<br>4. 7772-7786<br>5. 7790-7804<br>6. 7812-7826<br>7. 7816-7830<br>8. 7823-7837<br>9. 7839-7853<br>10. 71-85 |
| NFI | TTGGC | Activator | O'Connor *et al.*, 1995 | 1. 7475-7479<br>2. 7556-7560<br>3. 7589-7593<br>4. 7677-7681<br>5. 7713-7717<br>6. 7744-7748<br>7. 7769-7773 |
| AP1 | TGANTCA | Activator | Chan *et al.*, 1990,<br>Sibbet and Campo, 1990<br>O'Connor *et al.*, 1995 | 1. 7305-7311<br>2. 7633-7639<br>3. 7650-7656<br>4. 7813-7819 |
| Sp1/Sp3 | GGGCGT | Sp1 = Activator<br>Sp3 = Repressor | O'Connor *et al.*, 1995,<br>Dong and Pfister, 1999 | 1. 28-33<br>2. 7845-7850 |
| GRE/PRE | TGTACANNNTGTCAT | Activator | Gloss *et al.*, 1987<br>Sibbet and Campo, 1990,<br>Mittal *et al.*, 1993 | 1. 7387-7402<br>2. 7477-7491<br>3. 7643-7657 |
| TEF-1 | TACATACTTC | Activator | O'Connor *et al.*, 1995, | 1. 7191-7200<br>2. 7463-7472<br>3. 7645-7654<br>4. 7687-7696<br>5. 7704-7713<br>6. 7737-7746<br>7. 7820-7829 |
| TEF-2/PVF | AGGCACATAT | | Chong *et al.*, 1991 | 1. 7701-7710 |
| Oct-1 | AATTGCAT | Activator/<br>Repressor | O'Connor and Bernard, 1995,<br>Sibbet *et al.*, 1995 | 1. 7468-7475<br>2. 7670-7678<br>3. 7734-7741 (NFA)<br>4. 7838-7845<br>5. 68-75 |
| PSM-BP/ CDP | 2x TAYAATAAT separated by 7bp | Repressor | O'Connor *et al.*, 1998 | 1. 7893-7901 and 3-11 |
| NFIL6/ cEBPβ | T(T/G)NNGNAA(T/G) | Repressor | Kyo *et al.*, 1993 | 1. 7256-7264<br>2. 7282-7320 (various)<br>3. 7356-7364<br>4. 7561-7553<br>5. 7594-7686<br>6. 7636-7628<br>7. 7649-7657<br>8. 7656-7648<br>9. 7745-7753 |
| PEF-1 | GCGCCAACGCC | Activator | Cuthill *et al.*, 1993 | 1. 7675-7685 |
| E2 | ACCG$(W)_4$CGGT | Repressor | O'Connor *et al.*, 1995 | 1. 7453-7464<br>2. 7860-7871<br>3. 35-46<br>4. 50-61 |

N = T, C, A or G. W = A or T.

# 1.6 MANIFESTATIONS OF HUMAN PAPILLOMAVIRUS INFECTION

Manifestations of HPV infection include latent infection (where there is no clinical or histological abnormality), subclinical infection (where there are no symptoms but the infection does cause cellular morphological changes) and the spectrum of clinical infections (where there are symptoms or the lesions are visible to the naked eye).

## 1.6.1 Benign Lesions

The most common benign clinical manifestations of HPV infection are skin warts (such as plantar and flat warts) and its genital manifestation, condylomata acuminata (Figure 1.10). Condylomata acuminata are small, fleshy, vascular growths resembling cauliflowers. They can be found on areas such as the penis and on or under the prepuce, or on the labia in the female. They can also be found on the perianal epithelium or anal canal of either sex. The majority of condylomata acuminata are caused by HPV types 6 and 11 which only rarely progress to invasive cancer (Schmauz and Owor, 1980, and Wieland and Pfister, 1997). Genital warts may persist, resolve spontaneously, and some frequently recur due to latent infection or small lesions that are missed during treatment. HPV is also responsible for respiratory papillomatosis, EV, and subclinical infection of sites other than the cervix, such as the vulva, vagina or penis. HPV can also infect the epithelium of the urethra, oesophagus, and eye.

a)  b)  c) 

Figure 1.10: a) Wart on the finger, b) Plantar wart on the sole of the foot, c) Anal warts

a) and b) Reproduced with permission from the National Skin Centre, Singapore at http://www.nsc.gov.sg. c) Reproduced with permission from S.T.D. Online at http://www.adraidtoask.com/STD/hpv.html.

### 1.6.2 Cytological Classification of Cervical Lesions

Cervical intraepithelial neoplasia (CIN), is a term that was first used by Ralph Richart, and encompasses all precursor lesions of cancer of the cervix (Barrasso and Guillemotonia, 1997). The Bethesda System (TBS) was introduced in an effort to standardise Pap smear technology. This system for reporting cervical cytologic disease (refer to the National Institutes of Health Consensus statement, 1997) reports the adequacy of the specimen, gives a descriptive diagnosis and does not use the visual stigmata of HPV as a major element of disease classification. Potential diagnosis includes benign cellular changes (due to infection or reactive inflammation) and different degrees of squamous and glandular abnormality. The current Bethesda system is reported to have 85% sensitivity and 90% to 90% specificity of detection (Apple *et al.*, 1995).

The hierarchy of squamous cell abnormalities (from Adimora and Quinlivan, 1995) is as follows:

1. Atypical squamous cells of undetermined significance (ASCUS)
2. Low grade squamous intraepithelial lesions (LG SIL) which include benign cellular changes suggestive of HPV infection and cervical intraepithelial neoplasia 1 (CIN 1), ie. mild dysplasia.
3. High grade squamous intraepithelial lesions (HG SIL) which include CIN 2, CIN 3/carcinoma *in situ*, ie. severe dysplasia.
4. Carcinoma.

### 1.6.2.1 Atypical Squamous Cells of Undetermined Significance (ASCUS)

ASCUS refers to the presence of abnormal-appearing squamous cells that do not meet the criteria for defined epithelial abnormalities (Kaufman, 1996). For example, the loss of normal nuclear/cytoplasmic ratio associated with nuclear enlargement. Nuclei may also appear hyperchromatic. Less than 5% of smears should be designated as ASCUS. Only 5% to 10% of ASCUS smears will harbour serious cervical disease (Manos *et al.*, 1999).

### 1.6.2.2 Low Grade Cervical Intraepithelial Neoplasia (CIN 1)

CIN 1 lesions are mostly confined to the transformation zone of the cervix (the area around the os of the cervix bounded by the original prepubertal epithelium and the present border of the endocervical glandular epithelium). CIN 1 cells display mild dysplasia, or a slightly altered patten of differentiation from normal keratinocytes. The basal layers constitute the lower third of the epithelium and cells containing large vacuoles surrounding dense nuclear chromatin, called koilocytes, are often seen in the suprabasal layers. These cells contain an increased nuclear to cytoplasmic ratio (nuclear atypia), and at least two of the following: nuclear enlargement, binucleation, irregular nuclear contours, hyperchromasia, and perinuclear halos. A wide clear zone of cytoplasmic degradation, devoid of glycogen can surround the nucleus. Koilocytotic atypia was first described in cervical smears by Koss and Durfee (1956) and has since been since associated with HPV infection, and in particular, viral maturation (Meisels *et al.*, 1976 and 1977). Marked koilocytotic atypia is the most reliable marker of CIN 1 for infection of high/intermediate risk HPVs (Ziol *et al.*,

1998), and as such, the microscopic manifestations of HPV infection are normally classified with CIN (Kiviat, 1996). Studies suggest that 10% to 15% of women with CIN 1 will subsequently develop CIN 2-3. At least 60% will have abnormal changes in their Pap smears that will regress to normal, and approximately 30% will have persistence of CIN 1 (Kiviat, 1996). The two year cumulative incidence of CIN 1/2 was 28% among women who were HPV positive and only 3% in those without detectable HPV DNA in one study (Koutsky *et al.,* 1992). Therefore the vast majority of infections are self-limited, although the exact mechanisms that lead to regression and clinical cure remain unknown.

#### 1.6.2.3 High Grade Cervical Intraepithelial Neoplasia (CIN 2 and 3)/Carcinoma *in situ*

The epithelium of CIN 2 and 3 lesions has an undifferentiated basal cell-like appearance and pronounced cellular disorganisation or dysplasia, although even a substantial proportion of these lesions will regress. Some will persist and a small proportion will progress. These lesions are diagnosed with the aid of the colposcope (see section 1.7.1.3), or microscope. Flat lesions (macules) need 5% acetic acid (aceto-whitening) for visualisation and can also be examined by use of the colposcope. CIN 2 and 3 cells display an increased nuclear to cytoplasmic ratio, nuclear pleomorphism and hyperchromasia. Koilocytotic changes disappear perhaps due to viral integration.

When the epithelial changes have not yet invaded the underlying stroma but have extended into the endocervical glands, these lesions are classified as carcinoma *in situ* or CIN 3. CIN 3 lesions include the whole epithelial thickness and contain disorganised, abnormally proliferating (dysplastic) basal cells with a large nuclear to cytoplasmic ratio, marker nuclear atypia, and a lack of normal differentiation and apoptosis (Lörincz, 1999).

#### 1.6.2.4 Carcinoma

Approximately 3% of women infected by HPV will develop cancer of the cervix during their lifetime. The mean period between a primary infection and cancer development is in the order of 20 to 50 years (zur Hausen, 1989). Generally cancers are caused by mutations, either acquired (chemical, radiation, viral) or genetic. The

key features of cervical cancers include their nuclear morphology, pattern of disturbed differentiation, and intraepithelial disorder. These changes in cell morphology help explain why in cervical cancers there is no viral particle production, as particle production requires differentiated epithelial cells, suggesting a parallel with the proliferative effect of HPV in the basal layer of warts. A comparison of the differences between some of the aspects of the virus life cycle in benign and malignant cervical disease is outlined in Table 1.5.

Macroscopically, cervical cancers may have a yellowish discolouration, obvious areas of tissue necrosis and large, erratically shaped blood vessels. The prognosis depends on the stage of the disease at diagnosis, as well as factors such as the histological type of the tumour, lymph node status and the patient's age. As the disease progresses, the malignant cells infiltrate the cervix and adjacent tissues. Metastasis to distant organs proceeds via lymph nodes. Death from cervical cancer results from these distant metastases, or uraemia, sepsis from the tumour, or haemorrhage following rupture from major blood vessels of the pelvis.

The phenomenon of rapid onset cervical cancer, defined as less than three years between normal cytological screening and development of invasive disease (Bain and Clocker, 1983, Figge *et al.,* 1970 and Benoit *et al.,* 1984) occurs, and is believed to result at least in part from inadequate screening (Janerich *et al.,* 1995). Alternatively it may result from false negative Pap smear tests. Nonetheless, it is likely that in a small percentage of cancers rapid-onset disease does actually exist (Eide *et al.,* 1987, Brand *et al.,* 1988, Hopkins and Morley, 1991, and Hildesheim *et al.,* 1999).

Table 1.5: A Comparison of Viral Characteristics in Benign versus Malignant Cervical Disease

| **Viral Characteristics** | **Benign Cervical Disease or Low Grade CIN** | **High Grade CIN and Malignant/Invasive Cervical Disease** |
|---|---|---|
| Examples of HPV types involved | 6, 11, 42, 44, 54, 16, 18, 30. | 16, 18, 31, 45. |
| L1/L2 genes | Expressed in upper terminally differentiated epithelial layers. | Genes often not present, and often not expressed when they are. mRNAs occasionally seen in cancers, which may be due to cellular promoters flanking integrated viral DNA. |
| E6/E7 genes | Few E6/E7 transcripts in basal layers and more in upper differentiated layers of lesions. | E6/E7 actively transcribed. Expression of E6/E7 transcripts throughout all cell layers of the undifferentiated epithelia. |
| E1/E2 genes | E2 binding to the URR inhibits early promoter activity in basal layers (E2 autoregulates its own expression). In upper epithelial layers, there is a change to late promoter usage and an increase in levels of E1/E2 proteins. | Integration of viral DNA often results in disruption of the E1/E2 genes and there is loss of repression of early gene transcription and replication ceases. Genes present in episomal isolates, but may not always be transcribed. |
| E4 gene | Proteins are abundant in low grade lesions in middle and upper epithelial layers. Is responsible for the appearance of koilocytes. | Integration often results in loss of E4 protein. However E4 is expressed in some cancers and is abundant in all cell layers when it is expressed. |
| URR | Low level activity | Always retained in cancers. High level activity. |
| Replication and production of virus particles | Occurs in upper epithelial layers – is dependent on differentiation of keratinocytes | No replication or viral production (loss of differentiated cells). |
| Physical state | Episomal | Mostly integrated (all of HPV 18), some episomal (eg HPV 16), and some with both. |
| Copy number | Variable copy number – low, up to approximately 100 in the basal layers, to thousands of copies in the upper epithelial layers. | Copy numbers probably more homogeneous - highest copy number for episomal virus. Lowest for integrated. Copy number for HPV 16 increases with increasing epithelial abnormality. |

*Some of the information used in this table came from Stoler *et al.*, 1992, and Fuchs and Pfister, 1994.

### 1.6.3 International Federation of Gynecologists and Obstetricians (FIGO) Staging System for Cervical Cancer

The criteria for staging cervical cancers is based on the FIGO staging system outlined in Table 1.6.

Table 1.6: General Guidelines for International Federation of Gynecologists and Obstetricians staging (Adapted from the National Institutes of Health Consensus Statement, 1997).

| Stage | Guidelines for Staging | Appropriate Management |
|---|---|---|
| 0 | Carcinoma *in situ*, intraepithelial carcinoma. | |
| I | Carcinoma strictly confined to the cervix. | Simple hysterectomy or cone biopsy |
| IA | Invasive carcinoma identified only microscopically. Invasion limited to measured stromal invasion with a maximum depth of 5mm and width of 7mm. | |
| IA1 | Measured invasion of stroma no greater than 3mm in depth and no wiser than 7mm. | |
| IA2 | Measured invasion of stroma greater than 3mm and no greater than 5mm in depth and no wiser than 7mm | Primary radical or modified radical hysterectomy or primary radiation therapy |
| IB | Kinchela lesions confined to the cervix or preclinical lesions greater than IA. | Radical hysterectomy with pelvic lymphadenectomy or radiation therapy |
| IB1 | Clinical lesions no greater than 4cm in size. | |
| IB2 | Clinical lesions greater than 4cm in size. | |
| II | Carcinoma extends beyond the cervix, but has not extended on the pelvic wall: carcinoma involves the vagina, but not as far as the lower third. | |
| IIa | No obvious parametrial involvement. | |
| IIb | Obvious parametrial involvement. | Primary radiation therapy using external beam radiation and brachytherapy |
| III | Carcinoma has extended on the pelvic wall; on rectal examination there is no cancer-free space between the tumour and the pelvic wall; the tumour involves the lower third of the vagina. | |
| IIIa | No extension on to the pelvic wall, but involvement of the lower third of the vagina. | |
| IIIb | Extension on to the pelvic wall. | |
| IV | Carcinoma has extended beyond the true pelvis or has clinically involved the mucosa of the bladder or rectum. | |
| IVa | Spread of the growth to adjacent organs. | |
| IVb | Spread to distant organs. | |

### 1.6.4 Classification of Invasive Cervical Tumours

Malignant neoplasias of the cervix are histologically divided into eight categories (Table 1.7). Squamous cell carcinomas (SCC) and adenocarcinomas represent the vast majority of cervical malignancies, accounting for greater than 90% and 5% respectively.

Table 1.7: Histological Classification of Malignant Cervical Cancers
(Adapted from Buckley and Fox, 1992)

- **Squamous cell carcinomas**
eg. keratinising/non-keratinising
well differentiated/poorly differentiated
- **Adenocarcinomas**
eg. endocervical adenocarcinoma
- **Mixed Tumours**
eg. adenosquamous carcinoma
- **Small cell carcinomas**
eg. neuroendocrine
- **Undifferentiated carcinoma**
- **Metastatic carcinoma**
- **Malignant melanoma**
- **Malignant non-epithelial tumours**
eg. sarcoma

### 1.6.5 Other Carcinomas Caused by Human Papillomavirus

Vulvar, vaginal, penile, and anal cancers, with an incidence rate only 1% to 2% of that of cervical cancer, often contain HPV DNA (Muñoz *et al.,* 1992). Buschke-Löwenstein tumours of the anogenital region for instance, contain HPV 6 or 11. HPV is also responsible for some SCCs of the oral cavity, tonsils, larynx, and perhaps nonmelanoma skin tumours (Berkhout *et al.,* 1995, Shamaniin *et al.,* 1996, D'Costa *et al.,* 1998, and Kook *et al.,* 1998). There is still debate over the degree of association of genital HPV types with cancer at nongenital body sites. In some cases, such as HPV 16 and periungual carcinoma, the evidence is strong (Moy *et al.,* 1989).

EV is a rare disorder, characterised by early onset of reddish, flat wart-like skin lesions, some of which will become premalignant and invasive SCCs, usually at sun-exposed sites. HPV 5 and 8 are the most frequent types found in these cancers.

Recurrent respiratory papillomatosis (RRP), also caused by HPV, has a bimodal age distribution, and is most commonly detected in children less than 5 years and in young adults. Occasionally it can be so severe and rapid that repeated occlusion of the airways necessitates frequent surgery to remove the growths. Although not always cancerous, RRP can become malignant.

## 1.7 CERVICAL CANCER PREVENTION

Primary prevention of cervical cancer is desirable and includes education directed towards adolescents and health care providers, encouraging measures to reduce HPV transmission such as delayed onset of sexual intercourse. Education is directed towards advertising campaigns and easy access to information. Information about HPV, warts, and cervical cancer is readily available on the internet. Examples of a few informative web pages include:

- The National Cervical Cancer Public Education Campaign (http://www.cervicalcancercampaign.org/),
- The AVSC International web page (http:/www.avsc.org/diseases/dhp.html).
- The National Cervical Cancer Coalition website (http://www.nccc-online.org/index.asp) – cervical cancer information centre.

Secondary prevention efforts include the development of antivirals, therapeutic vaccines to prevent progression of lesions, improving the sensitivity and specificity of current screening methods, and expanding education and screening programs to target under reached populations (NIH Consensus statement, 1997).

## 1.7.1 Detection of Cervical Lesions

### 1.7.1.1 Papanicolaou Smears and Cytology

Cervical cancer may be prevented by early detection of its precancer lesions, which are recognised by cytologic screening and confirmed by histologic diagnosis through colcoscopically-directed biopsy. In 1928 George N. Papanicolaou reported finding abnormal cells in vaginal smears of women with genital cancer. This observation was not followed any further for 10 years until Papanicoloau collaborated with gynaecologist Herbert Trout to establish prospective screening (Papanicoloau *et al.*, 1941). Eventually the medical world realised the importance of cytologic detection of cervical lesions. During the 1950's and 1960's a variety of studies provided the evidence of what we now know are lesions that are the precursors to cancer. Squamous cell carcinoma is ideal for screening because of the usually long preclinical phase.

The incidence of cervical cancer has steadily declined over the past 20 years in developed countries and this has been attributed to the efficiency of the Pap smear in detecting early disease (LaGuardia, 1993). Pap smear screening remains the best method of reducing the incidence and mortality of cervical cancer (NIH Consensus Statement, 1997). In several developed countries, Pap smear screening has resulted in a 70% or greater decline in mortality from cervical cancer in the last 50 years (Hakama *et al.*, 1991, and Parker *et al.*, 1996, and Piver, 1996).

It has been estimated in the USA that 5% of Pap smears are abnormal (Lörincz, 1999). It is recommended in Australia that a Pap smear is performed at two year intervals for women with no symptoms or history of abnormal cytology, from 18 to 20 years of age or 1 to 2 years after first sexual intercourse, whichever is later. This can cease at 70 years for women who have had normal smears within the last five years (NH&MRC, 1991).

Methods for specimen collection, preparation, and evaluation of Pap smears have changed little since its introduction in the 1940's. Although highly effective, Pap smears have a false-negative rate of between 5% and 50% (Koss, 1989 and 1993). Half of these errors are due to inadequate specimen sampling, and the other half to

failing to identify abnormal cells or interpret them accurately (Cirisano, 1999). More recently, the process has become automated to increase accuracy. This has involved the introduction of ThinPrep® and PAPNET and AutoPap 300QC. PAPNET, for example, is a computer-based Pap smear reading system used primarily to screen atypical Pap smears. It detects abnormal cells that may have been missed by conventional microscopic examination (Mango, 1996). Recognition of the association between HPV and cervical cancer led to investigations of the value of oncogenic HPV detection in conjunction with screening (see section 1.7.2.1).

### 1.7.1.2 The ThinPrep® Pap Test™

With a conventional Pap smear, a small portion of the collected sample of epithelial cells from the transformation zone of the cervix, taken with a spatula or cytobrush, is smeared onto a slide. The disadvantage is that the abnormal cells can be lost. The ThinPrep® Pap Test™ (Cytyc Corporation, USA), however, uses an automated slide preparation unit that produces a uniform thin-layer slide free of obscuring artefacts such as blood, mucous and inflammation. The sample is taken with a cytobrush or spatula-sampling device. With a traditional Pap smear, the sample is then smeared onto a slide. With Thin Prep the device is rinsed into a special vial with preservative and the vial is placed into the processor in the laboratory. This system has improved clarity and specimen adequacy over the conventional Pap system.

### 1.7.1.3 Colposcopy

The colposcope is a low-powered magnification device to identify lesions that require punch biopsy and to determine the extent of a lesion. The use of cytology alone provides 87% sensitivity for the detection of cervical lesions, whereas a combination of cytology and colposcopy may improve the sensitivity of screening to 98% (Burghardt *et al.,* 1993). The cervix is observed at 6- to 40-fold magnification through a colposcope with illuminated magnification. For treatment, a laser may be attached. For viewing lesions, 3% to 5% acetic acid is applied to the cervix. If the findings are abnormal, a specimen can be taken for histologic diagnosis; either a punch biopsy taken at the SCJ, or an endocervical curettage, or both. Endocervical curettage is performed with a sharp spoon-shaped or curved curette, which can be inserted through the narrow external os. Once positioned, a circumferential scaping

technique is used. At colpsocopy, intraepithelial lesions appear as abnormal findings within the transformation zone. They usually appear red (due to stromal hyperemia) and react to acetic acid (due to increased cellular and nuclear activity), and are iodine negative (absence of glycogen). The limits of the lesion are usually well defined as atypical epithelium is thicker than normal.

#### 1.7.1.4 The Acetic Acid Test

This test modifies the colour of the epithelium and allows the study of surface configuration. Epithelium with increased cell and nuclear density produces a colposcopic appearance of opacity and paleness, which is the basic characteristic of intraepithelial and invasive neoplasia. When acetic acid is applied, the lesion appears as an area of sharply delineated white or grey colour. The test allows clear distinction between squamous (normal and abnormal) and columnar tissue. The reason for the white appearance is not really known. The common hypothesis is that acetic acid induces a reversible coagulation of certain epithelial and stromal proteins (Barrasso and Guillemotonia, 1997).

### 1.7.2 Laboratory Identification of Human Papillomavirus Infection

No reliable serological test is available for detecting HPV infection. Present tests are still insensitive and fairly nonspecific for current HPV infection, leaving serology for research purposes only. It is difficult to prompt keratinocytes to terminally differentiate, which makes it difficult to produce HPV in culture, as its replication cycle is intimately associated with the different cycles of terminally differentiating epithelium. The signal transduction and gene expression pathways that are elicited during the differentiation process also regulate viral gene expression (Meyers, 1996). More recently, the organotypic 'raft' culture method has been developed that allows keratinocytes to stratify and express differentiation induced cytokeratins (Meyers, 1996). The addition of 12-o-tetradecanoyl phorbol-13-acetate (TPA) supports the differentiation of cells. The rafting method faithfully recapitulates squamous differentiation and has allowed the growth of some HPV types as mature viral particles, eg. HPV 16 and 31.

Because of the difficulty of culturing HPV, molecular methods are most commonly used for detection. Several different laboratory methods have been used for the detection of HPV. Some of these include (from LaGuardia, 1993):

- electronmicroscopy - for the detection of assembled virions
- immunocytochemistry - for the detection of viral antigens
- recombinant DNA techniques - for the detection of viral DNA or RNA
- polymerase chain reaction (PCR) - for the detection of viral DNA or RNA
- serological assays - for the detection of antibodies.

Methods for HPV detection developed following the successful cloning of individual HPV types and were initially based on hybridisation techniques. Recombinant DNA techniques such as Southern and dot blot, filter and tissue *in situ* hybridisations have all played major roles in establishing the association between HPVs and cervical cancer. Southern blot hybridisation represents the most specific method of HPV detection (which ranges in detection from 0.1 to 0.01 HPV genome copies per cell) (de Villiers, 1992). However this technique is time consuming, relatively labour intensive and costly in terms of reagents.

Dot and filter *in situ* hybridisation techniques are both cheaper, quicker and simpler procedures, but have very low sensitivity (one genome copy per cell in a sample of $10^5$ cells, and 50 genome copies per cell respectively). Additionally, with dot blot hybridisation, even under conditions of high stringency, background signals often complicate the interpretation of results. Tissue *in situ* hybridisation is relatively time consuming and with a detection limit of 20 to 100 copies per cell, it is the least sensitive of these traditional techniques (de Villiers, 1992, and Rose, 1993). PCR involves the exponential amplification of DNA and is the most sensitive methods of HPV detection available. PCR is relatively cheap and quick to perform and is currently the most commonly used method of HPV detection in the laboratory.

Immunological assays using synthetic oligonucleotides, fusion proteins and artificially produced VLPs have been used for the identification of viral structural antigens. However, they are still of limited value in the discrimination of individuals

with past or present HPV infections (zur Hausen and de Villiers, 1994). Specimens suitable for HPV testing include brushings, scrapes, lavages or biopsies from various anatomic sites. The methods that can be used to detect HPV have their own advantages and disadvantages, which are overviewed in Table 1.8.

In the USA, two HPV detection kits have been approved by the Food and Drug Administration (FDA) for clinical use. These include HPV Profile (Digene Diagnostics, Inc) which uses dot blot analysis, and the Hybrid Capture® System HPV DNA Assay (Digene Diagnostics, Inc) which uses RNA probes to identify HPV DNA. Target DNA combines with specific RNA probes creating RNA:DNA hybrids that are captured onto a solid phase coated with specific antibodies. The captured hybrids are detected with antibodies conjugated to alkaline phosphatase. The Hybrid Capture system yields the same prevalence of HPV in cervical smears as does PCR alone, suggesting its usefulness for routine HPV testing (Riethmuller *et al.*, 1999).

Table 1.8: Overview of the Advantages and Disadvantages of Methods for Detecting Human Papillomavirus Infection (Adapted from Zahm *et al.*, 1997)

| Method | Material | Advantages | Disadvantages |
|---|---|---|---|
| **Indirect Methods** | | | |
| Colposcopy | Cervix, vagina | Sensitive | Not specific, subjective, needs experience |
| Cytology | Smears | Koilocytosis and dyskeratosis are specific phenomena | Limited sensitivity |
| Histology | Tissue | Koilocytosis and dyskeratosis are specific phenomena | Limited sensitivity |
| **Direct Methods** | | | |
| Electronmicroscopy | Smears or tissue | Specific detection of virus particles | Laborious and costly, insensitive |
| Polymerase chain reaction | Smears or tissue | Fast, specific, and very sensitive | Issues of contamination |
| Hybridisation | Smears or tissue | Identify different HPV types, sensitive and specific, gold standard for clinical and research use | Laborious |
| **Serology** | | | |
| HPV antibody detection | Serum | No gynecologic examination necessary, natural history of HPV infection can be followed | Limited sensitivity, cross-reaction between response to different HPV types |
| Detection of antigen | Smears or tissue | Activity of infection can be followed, non-radioactive method | Low sensitivity, high concentrations of viral antigens necessary |

### 1.7.2.1 Routine Human Papillomavirus Testing

In recent years the value of routine HPV testing in conjunction with cytology has been under debate. There are arguments in favour of testing (Meijer *et al.,* 1998, and Manos *et al.,* 1999), as well as those against (Herrington and Wells, 1997, and Adam *et al.,* 1998). It is argued that HPV testing would increase the effectiveness of current screening procedures as it can decrease false-negative rates. A woman with a cytologically normal smear but with high risk HPV infection, has a greater risk of developing an abnormal smear in a subsequent two years than an HPV negative woman (Richart *et al.,* 1998), which is evidence for the use of HPV testing. HPV testing may be of particular use in the management of women who are diagnosed with ASCUS, or for confirmation of LG SIL (CIN 1 or 2), or for monitoring HPV cure. Five to 20% of patients with ASCUS or LG SIL have high grade CIN confirmed by subsequent colposcopy and biopsy (Wright *et al.,* 1995, Cox, 1996, and Kaufman, 1996). Some clinicians refer to colposcopy women with ASCUS smears to rule out underlying high grade disease. HPV testing could decrease this burden on colposcopists (Coutlée *et al.,* 1997). In comparison with women initially HPV negative, women who tested positive for HPV DNA in one study were 3.8 times more likely to have LG SIL subsequently diagnosed during follow up, and 12.7 times more likely to develop HG SIL (Liaw *et al.,* 1999). On the other hand, it has been argued that HPV testing may not be cost effective. The cost could be offset though if HPV testing could be used to identify only those low grade lesions that would develop into high grade lesions subsequently.

Routine testing is most commonly performed using the Hybrid Capture (HC) II kit. An effective approach of ASCUS triage is to perform the HPV test on the same specimen as the Pap smear, which is possible using the Thin Prep™ system to prepare thin layer Pap smears and then perform Hybrid Capture HPV testing from the residual medium if the cytological diagnosis is ASCUS.

# 1.8 MANAGEMENT OF CERVICAL PRECANCER AND CANCER

## 1.8.1 Historically

Although the causative agent of cervical cancer has only been known for a few decades, attempts to treat and manage the disease have been taking place for considerably longer. It was the accessibility of cervical lesions that encouraged the development of surgical intervention in the nineteenth century. An overview of past management and treatment of cervical cancer is presented in Table 1.9.

Table 1.9: Historical Overview of Management and Treatment of Cervical Precancer and Cancer

| YEAR | METHODS OF TREATMENT AND MANAGEMENT OF CERVICAL PRECANCER AND CANCER |
|---|---|
| 16th Century BC | Early attempts at chemotherapy - vagina instilled with fennel, garlic, fresh eggs, honey and wasps' dung (documented in the Ebers Papyrus) |
| 1812 | Vaginal hysterectomy first performed for cervical cancer by Paletta |
| 1825 | First abdominal hysterectomy performed by Lagensbeck (Meigs, 1954) |
| 1893 | Ries designed radical surgery to remove vagina, parametrium and broad ligament, the uterus, ovaries and clinically involved lymph nodes (Ries, 1895) |
| 1894 | First radical hysterectomy and bilateral pelvic lymphadectomy performed by Thring (Thring, 1895) |
| 1897 | Wertheim performed a similar operation to Thring, but is generally credited with this major advance due to his extensive work in developing the technique (Wertheim, 1912) |
| Late 1800's | Radiotherapy (using radium) was developed and became an alternative means of treatment (Kjellgren, 1992) |
| 1923 | Papanicolaou discovered the exfoliative nature of cervical cancer and stained cells for microscopic examination thus providing a method for the early diagnosis of cervical cancer (Papanicolaou and Traut, 1941) |
| 1927 | Colposcope, invented by Hinselmann, used to view subclinical infection which is detectable colposcopically after the application of a solution of 3% to 5% acetic acid that whitens the infected epithelium |
| 1940's | Introduction and use of sulpha drugs, antibiotics and nitrogen mustard gas (and decrease in use of radium) |
| 1950's | Nitrogen substitutes of folic acid, methotrexate and tetracycline employed. |
| 1950's onwards | Electrosurgical cautery and fulguration introduced for the physical destruction of CIN lesions - whereby an electrosurgical current is passed into tissues and the energy is focussed to vaporise them. |
| 1960's | Methotrexate and similar drugs used in intraarterial pelvic perfusions |
| 1970's to 1980's | Carbon dioxide and neodymium-yttrium aluminium garnet lasers introduced for treatment of CIN. Used for cutting, coagulating and vaporising tissues. |

### 1.8.2 Current Strategies

Cervical lesions are treated in order to prevent progression to invasive cancer and to decrease the risk of transmitting the virus to sexual partners. A 12mm deep incision is estimated to cure up to 90% of cervical lesions with endocervical canal involvement (Barrasso *et al.*, 1997). There is general agreement about treating HG SIL but no general agreement about treating LG SIL and ASCUS – either to treat or use cytologic follow-up at regular intervals only. Only a small proportion of low grade lesions will progress to HG SIL but as treatments can yield up to 95% cure rates after one session and persistent LG SIL contains contagious virus, there is certainly some value in the treatment of these lesions. Management of HG SIL includes colposcopy-directed cone biopsy and endocurvical curettage followed by conisation with scalpel, cautery, laser, or loop electrocautery excision (NIH Consensus statement, 1997).

Surgery, radiotherapy, and chemotherapy are all used, either individually, or in combination, to treat women with cervical cancer. Stage IA1 cervical cancer (refer to Table 1.6) is often successfully treated with a simple hysterectomy, or cone biopsy. Stage IA2 disease is treated by radical hysterectomy or radiation therapy. Patients should receive one treatment or the other, but not both, due to cost and morbidity of treatment. Women with more advanced, nonmetastatic disease (stages IB and above) should be treated with radiation. Recurrent cervical cancer confined to the pelvis should be treated with the modality not previously received (NIH Consensus statement, 1997).

Treatments that stimulate the immune system may be used alone or in association with surgical techniques. Some currently used immunological therapies (from Gross and Barrasso, 1997) include interferons (such as IFNα, and interleukin-2) and retinoids (such as Tretinoin). Chemotherapeutic treatment may be used in conjunction with radiotherapy and may prove more effective than standard therapy using radiation alone, eg. Cisplatin-based chemoradiation therapy (Rose and Lappas, 2000).

# 1.9 DEVELOPING STRATEGIES FOR THE TREATMENT OF CERVICAL LESIONS

The two main approaches that are being developed include vaccines and antiviral therapies. There is a small chance that gene therapy may become useful in the future.

## 1.9.1 Vaccines

It is likely that regression of HPV infections is mediated by immune mechanisms and that resistance to further infection is also immunologically based. Recent advances have involved directing the immune response to prevent infection, to virus-infected and virally transformed cells. There are two types of vaccine approaches, prophylactic and therapeutic.

### 1.9.1.1 Prophylactic Vaccines

Prophylactic vaccines would be useful before HPV infection to elicit neutralising antibody-mediated inhibition of cell surface receptor binding (Tindle, 1996). Prophylactic vaccination may circumvent constraints on immunity seen in natural infection and seek to establish a secretory IgA-mediated immunological barrier at the anogenital epithelium (Tindle, 1996). In eukaryotic cells, the HPV capsid protein, L1, organises itself into VLPs when it is expressed at high levels (Schiller and Okun, 1996). Prophylactic vaccine attempts so far have focussed on these VLPs made from L1 and/or L2 as they are effective virion surrogates and a source of conformationally correct epitopes (Murakami *et al.,* 1999). They may also serve as vehicles for the delivery of additional immunomodulatory components. VLP-based vaccines are being trialed because of the effectiveness of capsid proteins to prevent infection in several animal models. Immunisation is followed by an antibody response that neutralises viral particles and prevents infection. The problem is that protection is type-specific, due to type specificity of protective antibodies and divergence between HPV types (Roden *et al.,* 1996). Other possibilities for vaccines include peptide vaccines, recombinant viral vaccines (vaccinia or adenovirus), and naked DNA vaccines.

#### 1.9.1.2 Therapeutic Vaccines

Therapeutic vaccines are designed to stimulate or augment the host immune response to help eliminate infection or regress carcinoma. The usual CMI targets are the E6 and E7 proteins because they serve as tumour-specific antigens. Immunisation with the viral proteins E2, E6 and E7 has been attempted. For instance in animal models, E6 and E7 peptide vaccines induce a CTL response that is protective against subsequent challenge with E6 and E7 containing tumours (Feltkamp *et al.,* 1993). Similarly, specific CTLs to HPV 16 E7 can also be generated in humans using synthetic E7 peptides (Alexander *et al.,* 1996). These peptides are in use in phase I clinical trials (Steller and Schiller, 1996). Several other phase I clinical trials are also being conducted. These are reviewed by Murakami *et al.* (1999). One variation to a peptide vaccine is to fuse viral proteins to other sequences that might help direct the protein to particular cellular organelles.

Preparing an adequate vaccine is a large challenge. There is no animal model of HPV infection and the situation with animal PVs cannot be directly translated to HPVs. There are many different types of HPV-associated lesions which means a single vaccine is unlikely to provide adequate protection. As HPVs are not easily propagated in tissue culture systems, reagents are limited to recombinant expression vectors and peptide synthesis. The T cell response to a protein is limited to a few epitopes, dictated by the haplotype of the host (Gao *et al.,* 1994), which imposes restraints on the use of synthetic peptides as immunogens. Sequence variation among HPV types, and intratypic variation may also be sufficient to convert an immunogenic protein to a non-immunogenic variant (Gao *et al.,* 1994). Therefore an effective vaccine is not likely in the immediate future. Even if all the obstacles could be overcome, the cost of preparations would be huge and would not be affordable to those countries needing it most.

### 1.9.2 Antivirals

The aim of antiviral treatment is to provide viral-specific intervention. For instance antisense nucleic acids, which inhibit viral protein synthesis by binding to and inactivating specific viral RNAs. If an antiviral inhibits a function common to all

PVs it would have the major advantage of being potentially active against all HPV types (Lowy, 1997), however such an antiviral is yet to be identified.

### 1.9.3 Gene Therapy

The aim of gene therapy as a therapeutic agent is to modify the structure and/or the function of genes for the control and cure of a disease. E6 and E7 are the main targets. Oligonucleotides that bind to the DNA and prevent transcription are called triplex-forming oligonucleotides. Oligonucleotides can also be designed that bind to mRNA, called antisense oligonucleotides. In this approach antisense RNA molecules are hybridised with their respective complementary mRNA, forming untranslatable dsRNA molecules (Walder, 1988, and Cotten *et al.*, 1989). Another strategy is to replace oncogenic genes, called gene augmentation. There have been research efforts in these areas. For instance, an anti-E6 molecule has been developed which is able to direct RNase H activity on full length HPV 16 E6/E7 mRNA (Alvarez-Salas *et al.*, 1999). This caused a significant decrease in the growth rate of several cervical cancer cell lines by downregulating E6/E7 mRNA and increasing p53 levels. If E6 and E7 are downregulated, it is likely that the cellular response will be cell growth. Likewise oligonucleotides directed against E6 and E7 also have antiproliferative effects (Madrigal *et al.*, 1997). The major barrier for antisense therapy is delivery of the therapy to target cells.

# 1.10 INTRODUCTION TO THE STUDY

Almost all cervical cancers contain HPV sequences and the continued presence and expression of the E6 and E7 genes is essential for maintenance of the malignant phenotype. HPV type has come under scrutiny as a possible marker for tumour behaviour, although the results of published studies comparing HPV 16 and 18 are still somewhat controversial (see section 1.3.5). HPV 18 has greater immortalising ability than HPV 16 *in vitro* (Barbosa and Schlegel, 1989). However, both have similarly low immortalising abilities when expressed from the same SV40 early promoter (Romanczuk *et al.,* 1991). This indicates that any difference between the two viruses probably resides in the URR, which mediates E6/E7 expression through interactions with cellular TFs. The role of sequence variation in the URR, and in other regions of the viral genome, in explaining the variable oncogenic potential of HPV of the same genotype has therefore come under scrutiny.

HPV 16 is associated with most cervical cancers worldwide, but morbidity and mortality vary greatly in different countries and even amongst women of similar genetic background living in the same social environment. There are also many individual exceptions to the expected outcome for cancers assigned to the same group by clinical staging at the time of diagnosis. Analysis of HPV 16 isolates from patients throughout the world has demonstrated significant sequence diversity as compared with the prototype. PV genomes evolve through the accumulation of point mutations, deletions, insertions, and intragenomic recombination with diversity occurring mainly through point mutations (Bernard *et al.,* 1994). Although many studies on HPV 16 sequence variation have been conducted, most of these have described the geographical clustering of variants and only a few have sought to identify the functional effects of identified sequence changes. However, it is known from those studies that some sequence variations in TFBSs in the URR can affect the level of E6/E7 expression and that some amino acid changes can alter PV protein functions. These types of sequence variants may contribute to cervical cancer risk and to the behaviour of established cancers, and the identification of viral markers for increased carcinogenic potential may, in the long term, lead to a more discriminating molecular screening protocol in clinical procedures.

The overall aim of this study was to characterise HPV 16 variants from cervical cancers from Australia and the South Pacific. The initial aim was to examine sequence variation in the HPV 16 URR to determine the distribution of HPV 16 variant lineages in this region of the world. More importantly this would provide the basis for functional analysis of sequence variation in the URR in terms of effects on TF binding and E6/E7 promoter activity. A further objective was to determine the physical state of the virus in the cancer cells, since unlike other malignant HPV genotypes, integration is not always observed with HPV 16 and sequence variation in the URR leading to increased promoter activity may be necessary for malignant conversion of episomal virus. Examination of the sequences of the E2/E4 and E6 regions in conjunction with the data already obtained for the URR may provide additional evidence for the biological significance of sequence variation in coding regions.

# CHAPTER 2

# SEQUENCE VARIATION IN THE UPSTREAM REGULATORY REGION OF HUMAN PAPILLOMAVIRUS TYPE 16 CERVICAL CANCER ISOLATES

# 2.1 INTRODUCTION

HPV intratype variants are stable entities, and have 2% or less sequence variation over the entire coding region and up to 5% variation in the URR. HPV 16 is the predominant HPV type found in cervical cancers in most regions of the world, including Australia (Bosch *et al.,* 1995, and Rose and Thompson, unpublished data). Phylogenetic studies have been used to classify HPV 16 variants from most regions of the world on the basis of sequence differences in the URR and in the L1, L2, and E6 genes (Ho *et al.,* 1993a, Yamada *et al.,* 1995, and 1997, and Wheeler *et al.,* 1997). On this basis, five major variant lineages of HPV 16 have been identified; that is European (E), African 1 and 2 (Af 1 and 2), Asian (As) and Asian American (AA), in addition to some minor lineages. The URR is approximately as varied as the L2 gene, which is the most variable coding sequence in HPV 16 (Yamada *et al.,* 1995). However in integrated HPV isolates from cancers, the L2 gene is often not present, making the URR a better choice for nucleotide analysis of HPV 16 isolates from cancers.

The value of delineating intratypic sequence variation extends beyond tracing the evolution of HPV 16. For instance, analysis of HPV 16 variants has been used to differentiate between persistent versus acquired infection and to determine HPV transmission dynamics by identifying specific variants in sex partners (Xi *et al.,* 1993). In addition, evidence is emerging that variants may be significant determinants of clinical course, and they may also prove important to the effectiveness of potential vaccines.

Although there is considerable sequence information on HPV 16 isolates from many regions of the world, there is currently none available for isolates from Australia and the South Pacific. In this study, the entire URR from 34 cervical cancer isolates from Australia and New Caledonia was amplified by PCR and sequenced. The major HPV 16 variant lineages circulating in these regions were determined. Particular interest focussed on sequence variation involving *cis*-acting motifs that might affect p97 activity. Sequence information from the URR then formed the basis for

investigations described in Chapter 3 of the study where the functional effects of selected variant URRs were assessed.

# 2.2 METHODS

## 2.2.1 Study Population

Investigations were undertaken on fresh tumours from 34 patients who underwent treatment for cervical cancer at Royal Prince Alfred Hospital (RPAH), Sydney, Australia, between 1990 and 1996. Central Sydney Area Health Service Ethics Committee granted ethics approval for the study and all work was conducted according to the approved protocols. Twenty six of the patients were Australian residents living in the inner city region of Sydney or in regional NSW, and eight patients were from New Caledonia in the South Pacific, referred to RPAH for treatment. The cancers had tested positive for HPV 16 DNA by PCR in a previous study. Selection was otherwise based on the availability of sufficient material for the investigations and clinicopathological data (Table 2.1). The mean age of the patients in the study was 48.3 years (ranging from 27 years to 79 years). The CaSki cell line, which has approximately 600 copies of the HPV 16 viral genome per cell integrated into various chromosomes (Yee *et al.,* 1985, and Mincheva *et al.,* 1987), was analysed for comparison in addition to being used as a positive control for PCR. The CaSki URR was also sequenced to confirm the previously published sequence. The sequence of the HPV 16 reference (prototype) strain (HPV 16R), cloned from the cervical cancer of a German patient (Seedorf *et al.*, 1985, with corrections by Meyers *et al.*, 1995) provided an arbitrary reference.

## 2.2.2 Preparation of Clinical Isolates

### 2.2.2.1 Phenol Extraction of Nucleic Acids from Biopsy Material

Phenol extraction was performed using standard methods (Sambrook *et al.,* 1989).

1. Six hundred microlitres of cooled lysis buffer (50mM NaCl, 50mM Tris, pH 7.5, 0.5% SDS, and 10mM ethylenediaminetetraacetic acid [EDTA], sodium salt) was dispensed into tubes and placed on ice.
2. Tumour biopsy material was kept in cryovials in liquid nitrogen. The equivalent of 50mg of tissue was chopped off biopsy material. A negative control, consisting of HPV 16 negative cervical tissue, was always included.

Table 2.1: Clinicopathological Variables

| Patient | Country of Birth | Age | Clinical Stage | Tumour Classification | Lymph Node Status | Primary Therapy | Time of Follow-up (months) | Clinical Follow-up |
|---|---|---|---|---|---|---|---|---|
| B1 | Australia | 48 | Ib | adenocarcinoma | 0/23 | S/R | 20 | no recurrence |
| B2 | Australia | 57 | Ib | N/K SCC | 1/20 | S/R | 29 | no recurrence |
| C1 | Australia | 61 | IIb | N/K SCC | unknown | R | N/R | DOD |
| D1 | Australia | 36 | Ib | adenocarcinoma | 0/26 | S | - | unknown |
| D2 | New Caledonia | 58 | IIB | K SCC | 0/1 | S/C/R | 5 | no recurrence |
| D3 | Austria | 45 | Ib | N/K SCC | 2/40 | S/R | 29 | no recurrence |
| E1 | England | 34 | Ib | N/K SCC | 0/18 | S | 20 | no recurrence |
| G1 | England | 53 | IV | N/K SCC | unknown | S | 21 | recurrence (4 months post-op) - alive |
| H1 | Australia | 36 | Ib | N/K SCC | 0/20 | S | 30 | no recurrence |
| H2 | Australia | 57 | Ib | N/K SCC | 3/38 | S/R | 20 | no recurrence |
| H3 | Australia | 59 | Ib | K SCC | 0/29 | S/R | 39 | no recurrence |
| H4 | England | 38 | Ib | N/K SCC | 1/32 | S/R | N/R | recurrence (8 months post op) - DOD |
| J1 | Australia | 40 | Ib | N/K SCC | 2/32 | S/R | 33 | no recurrence |
| K1 | New Caledonia | 45 | Ib | N/K SCC | 0/41 | R | 9 | no recurrence |
| K2 | Australia | 49 | Ib | N/K SCC | 0/20 | S/R | 29 | no recurrence |
| K3 | Australia | 40 | Ib | N/K SCC | 0/22 | S | 24 | no recurrence |
| K4 | Korea | 45 | Ib | adenocarcinoma | 0/12 | S | 22 | no recurrence |
| K5 | unknown | 38 | Ib | N/K SCC | 2/32 | S/R | 29 | no recurrence |
| K6 | New Caledonia | 64 | IIIa | K SCC | unknown | R | N/R | recurrence (8 months) - DOD |
| L1 | New Caledonia | 27 | Ib | adenosquamous | 2/11 | S/R | N/R | recurrence - DOD |
| M1 | Australia | 65 | II | N/K SCC | 0 | S | 16 | no recurrence |
| M2 | New Caledonia | 58 | IIIb | N/K SCC | unknown | R | - | unknown |
| M3 | Australia | 79 | Ib | adenocarcinoma | 1/20 | S | 33 | recurrence (6 months post op)-AWD |
| M4 | New Caledonia | 33 | IIIb | K SCC | unknown | R | - | unknown |
| M5 | Australia | 49 | Ib | adenosquamous | 0/34 | S | 32 | no recurrence |
| M6 | Australia | 53 | Ib | K SCC | 0/27 | S/R | 21 | no recurrence |
| O1 | Australia | 71 | IIIb | K SCC | unknown | R | N/R | DOD |
| O2 | Australia | 49 | Ib | N/K SCC | 0/27 | S/C | 20 | no recurrence |
| P1 | Australia | 36 | IVb | N/K SCC | unknown | R | N/R | DOD |
| P2 | New Caledonia | 40 | IIIb | N/K SCC | unknown | S/R | - | unknown |
| R1 | Australia | 56 | IIa | K SCC | 2/20 | S/C/R | N/R | recurrence (8 months post op) - DOD |
| S1 | Australia | 54 | Ib | adenosquamous | 3/34 | S/R | 21 | no recurrence |
| S2 | Australia | 43 | Ib | K SCC | 1/22 | S/C | 29 | no recurrence |
| T1 | New Caledonia | 27 | IVb | N/K SCC | unknown | S/C/R | 14 | patient died |

N/K = non keratinising; K = keratinising; SCC = squamous cell carcinoma; S = surgery; R = radiotherapy; C = chemotherapy; N/R = not relevant; DOD = died of disease; AWD = alive with disease; "post op" = time post surgery.

3. Biopsy material was chopped into very small pieces, using a different scalpel blade for each specimen, and placed into the lysis buffer.
4. Two to 3µl of 20mg/ml proteinase K was added to each tube on ice.
5. Tubes were placed in a heating block at 37°C overnight to allow digestion of tissues.
6. To digested tissues, 300µl of Tris-buffered phenol, pH 7.9 (Amresco) was added and the contents of each tube mixed well.
7. Three hundred microlitres of chloroform:isoamyl alcohol (24:1) (Biolab Scientific Pty Ltd) was added to each tube and tube contents mixed by inversion for 1 minute.
8. The tubes were spun (in an MSE micro centaur bench centrifuge) for 3 minutes at 12,000 rpm.
9. The top layer from each tube was removed and transferred to a fresh tube. The material right at the interface containing protein was not removed.
10. Equal volumes (approximately 500µl to 600µl) of chloroform: isoamyl alcohol (24:1) was added to the tubes, which were then inverted for 1 minute.
11. The tubes were centrifuged for 3 minutes at 12,000 rpm.
12. The top layer was again removed and placed into another fresh tube.
13. One-tenth volume (40 to 60µl) of 3M sodium acetate, pH 5.2 was added.
14. Two volumes (800µl) of cold absolute ethanol were added and the tubes mixed by inversion to help precipitate the DNA.
15. The tubes were placed at -70°C for 2 hours or -20°C overnight to precipitate DNA.
16. The tubes were centrifuged for 10 minutes at 12,000 rpm at 4°C.
17. The ethanol was removed and the tubes placed into a heating block at 42°C for 30 to 45 minutes until the pellets were dry.
18. Fifty microlitres of TE (10mM Tris, and 0.1mM EDTA) was added and the tubes left to stand at room temperature for at least 45 minutes. Flicking was used to resuspend the nucleic acids. Samples to be processed within a short period were stored at 4°C or long term at -20°C.

Nucleic acid concentration was determined as described in Appendix A7.2.1.

### 2.2.3 The Polymerase Chain Reaction

PCR first appeared in 1985 as a method for the prenatal diagnosis of sickle cell anaemia (Saiki *et al.*, 1985). In 1988 a thermostable DNA polymerase was introduced (Saiki *et al.*, 1988), overcoming the necessity of adding fresh DNA polymerase after each denaturation cycle, allowing reactions to be performed at a stringent temperature, and greatly simplifying the procedure. Since 1985 PCR technology has been widely used in many areas of science and it has made possible large-scale epidemiological studies of the role of HPV in cervical cancer. PCR is currently the most sensitive (potentially one HPV genome copy can be detected in $10^6$ cells) and common method for detecting HPV DNA and the type of HPV in clinical specimens. It also forms the basis for other investigations such as sequence analysis and cloning. PCR as such was ideal for use in this study.

#### 2.2.3.1 Principle of the Polymerase Chain Reaction

PCR involves the exponential amplification of a small segment of DNA that lies between two regions of known sequence (Figure 2.1). Two oligonucleotides of approximately 20 bases in length are used as primers for a series of synthetic reactions catalysed by a thermostable DNA polymerase. These primers flank the region to be amplified. One primer has a sequence complementary to the positive strand of the region of DNA and the other is complementary to the negative strand. The template DNA is first denatured by heating in the presence of a large molar excess of each of the two oligonucleotides and the four deoxynucleoside triphosphates (dNTPs) in a buffer. The reaction mixture is then cooled to a temperature that allows the oligonucleotide primers to anneal to their target sequences. The annealed primers are extended by a DNA polymerase that adds complementary dNTPs encoded by the target sequence, resulting in the creation of new positive and negative strands. The newly synthesised strands themselves act as templates for the PCR primers, and repeated cycles of denaturation, primer annealing and extension result in the exponential amplification of a segment of double-stranded DNA with their ends defined by the distance between the primers. In theory, the PCR yield can be calculated from the formula: A $(2^n)$ = P, where A is the number of starting molecules, n is the number of amplification cycles and P is the final yield in moles or number of molecules. In practice this extent of efficiency is rarely achieved

with a typical amplification more in the order of $10^6$, after 30 to 40 cycles (Clewley, 1995). The potential to detect a single molecule of specific DNA in a sample of $10^6$ cells represents a substantial increase in sensitivity over recombinant technologies.

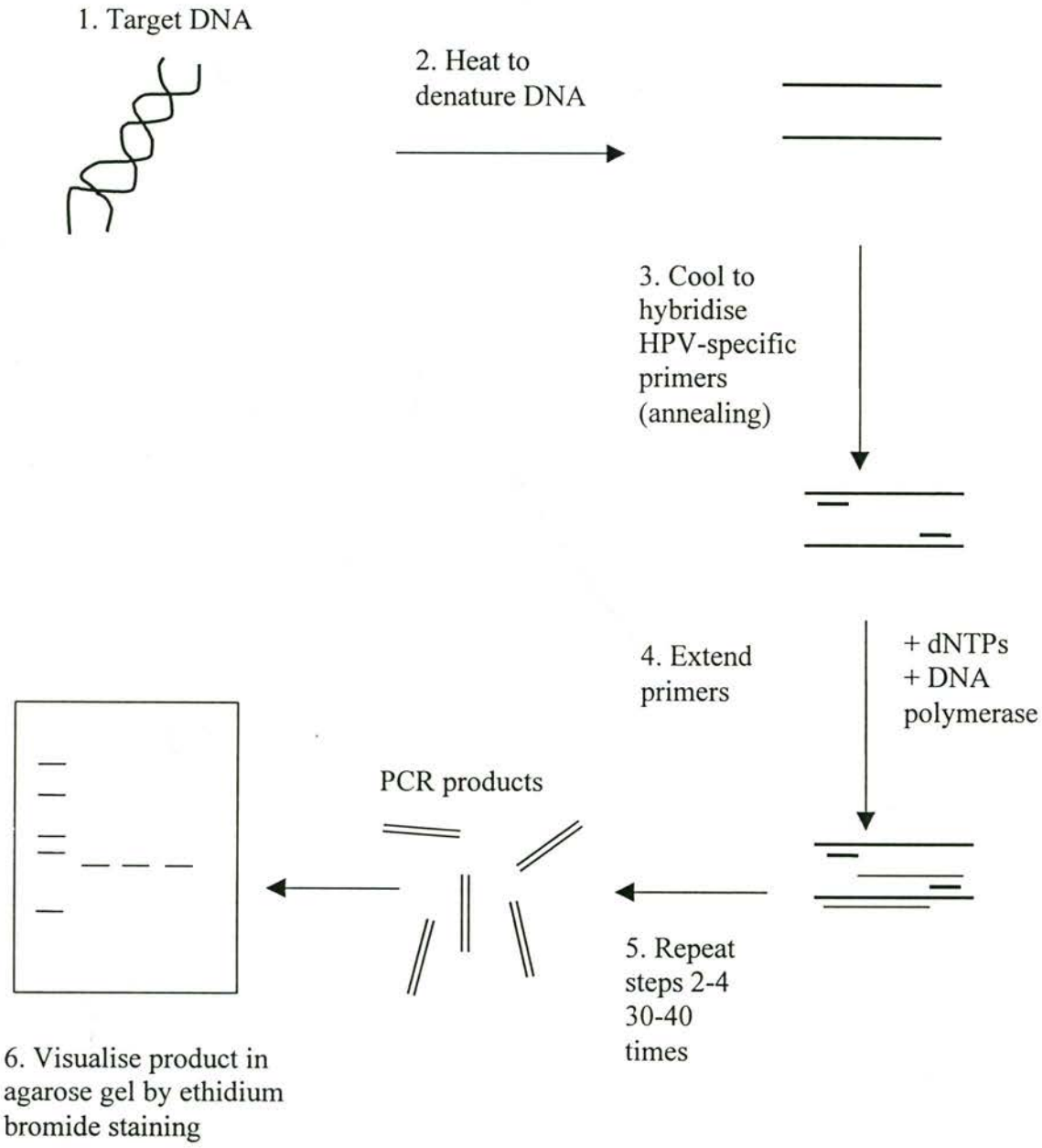



Figure 2.1: Principles of the polymerase chain reaction

(Adapted from Myint and Shaw, 1995)

Although identification of a band of predicted size following agarose gel electrophoresis is presumptive evidence of a positive result, subsequent hybridisation of the PCR product to a specific HPV probe, digestion with restriction endonucleases resulting in the production of specific fragment sizes, "nested PCR" or sequencing is required to verify results. For the most part in this study DNA sequencing was used.

#### 2.2.3.2 Precautions to Prevent Polymerase Chain Reaction Cross-Contamination

PCR has the advantages of extreme sensitivity, speed and simplicity. As a result, experimental conditions have to be very tightly controlled to minimise the chances of cross-contamination, leading to false positive results (de Villiers, 1992). False negative results, due to PCR inhibitors such as phenol, chloroform, or high concentrations of dNTPs (greater then 4mM), also have to be controlled for and PCR protocols require optimisation for each application. Examples of the procedures implemented to minimise the chances of cross-contamination include:

- A Class II cabinet (Email) dedicated to aliquot all of the PCR reaction components into sterile disposable plasticware, sufficient for small numbers of experiments. This area was also used for the preparation of PCR "cocktails" (which contained all of the reaction components except the template). The cabinet was cleaned with 70% alcohol and UV-irradiated for at least 20 minutes before and after use. A no DNA/RNA -20°C freezer in close proximity to this work station was used for the storage of all perishable PCR reagents on arrival from the supplier and all of the aliquoted PCR components ready for use. All reusable items of equipment used in the cabinet, for example, pipettes, marking pens, racks, ice containers, etc., were colour-coded and reserved for this purpose. Disposable items such as tubes and tips, were drawn from unused stocks stored in an area remote from that containing test samples.
- The sample DNA was added to the reaction mixes last and the tubes were capped after the addition of the DNA before proceeding to the next sample. All tubes containing DNA were microfuged before opening. Addition of DNA targets was performed in an area separate to that for preparing cocktails.

- A separate work station for the addition of the positive control template using pipettes, tips etc., dedicated for the purpose. This was performed after the assembly of the test reactions and sample targets had been completed.
- A thermal cycler located in a small separate area.
- A separate laboratory reserved for all manipulation involving PCR products, electrophoresis etc.
- Separate sets of protective clothing - a long sleeved, ankle-length coat and mop hat were worn in all work stations. The protective clothing was changed on a regular basis. Disposable latex gloves were worn throughout and changed frequently.
- No interchange of equipment, plasticware, pipettes, reagents, etc., between the different workstations.

## 2.2.4 Polymerase Chain Reaction Components

The following components are required for a PCR assay:

### 2.2.4.1 Deoxyribonucleoside Triphosphates (dNTPs)

2'-deoxycytidine-5'-triphosphate, trisodium (dCTP); 2'-deoxythymidine-5'-triphosphate, trisodium (dTTP); 2'deoxyadensine-5'-triphosphate, disodium (dATP); and 2'-deoxyguanosine-5'-triphosphate, trisodium (dGTP), were supplied at pH 7 and at a concentration of 100μmoles/ml by Biotech. Equal volumes of dCTP, dATP, dTTP, and dGTP were combined to give a 25mM stock concentration of each dNTP. The mix was aliquoted into 25μl volumes and stored at -70°C.

### 2.2.4.2 DNA Polymerase

*Tth Plus* DNA Polymerase is purified from the extreme thermophile *Thermus* by a method developed by Biotech International Ltd. It was supplied at 5.5units/μl (Biotech) and stored at -20°C in 20mM Tris-HCl pH 7.6, 300mM KCl, 1mM 2-mercaptoethanol and 50% glycerol (v/v) in the presence of 1mg/ml nuclease-free BSA.

*UlTma*™ DNA polymerase is encoded by a modified form of the *Thermotoga maritima* DNA polymerase gene that has been inserted into an *E.*coli host. The enzyme was supplied by Perkin Elmer at 6U/µl, stored in 20mM Tris-HCl, pH 8.0, 100mM KCl, 0.1mM EDTA, 1mM DTT, 0.5% Tween 20® (v/v) and 50% glycerol (v/v).

*Amplitaq*® DNA polymerase is a recombinant, thermostable DNA polymerase encoded by a modified form of the *Thermus aquaticus* DNA polymerase gene which has been inserted into an *E.coli* host. The enzyme was supplied by Perkin Elmer at 5U/µl, stored in 20mM Tris-HCl, pH 8.0, 100mM KCl, 0.1mM EDTA, 1mM DTT, 0.5% Tween 20® (v/v) and 50% glycerol (v/v) and 0.5% Nonidet P40®.

*Amplitaq Gold*™ DNA polymerase is derived from a recombinant, thermostable DNA polymerase encoded by a modified form of the *Thermus aquaticus* DNA polymerase gene which has been inserted into an *E.coli* host. *Amplitaq Gold*™ is provided in an inactive state and is activated by temperatures above 92°C which improve amplification of templates by lowering background and increasing amplification of specific products. The enzyme was supplied by Perkin Elmer at 5U/µl, stored in 20mM Tris-HCl, pH 9.0, 100mM KCl, 0.1mM EDTA, 1mM DTT, 0.5% Tween 20® (v/v) and 50% glycerol (v/v).

**2.2.4.3 Polymerase Chain Reaction Buffer**

The PCR buffer for *Tth Plus* DNA Polymerase (67mM Tris-HCl pH 8.8, 16.6 mM $(NH_4)_2SO_4$, 0.45% Triton-X and 200µg/ml gelatine) was supplied at ten times this concentration by Biotech.

10x *UlTma*™ buffer (100mM Tris-HCl, pH 8.8, 100mM KCl and 0.02% Tween 20® (v/v)) was supplied by Perkin Elmer at ten times this concentration. Tris-HCl ensures a limited pH, detergents such as Triton-X or Tween 20® help keep the polymerase in solution and gelatin improves enzyme stability.

*Amplitaq*® and *Amplitaq Gold*™ are supplied with 10x PCR buffer II (Perkin Elmer) containing 100mM Tris-HCl, pH 8.3 and 500mM KCl.

#### 2.2.4.4 Magnesium Chloride ($MgCl_2$)

A 25mM stock solution of $MgCl_2$ was supplied with each of the DNA polymerases. Magnesium ions are required for primer annealing, enzyme activity, and fidelity.

#### 2.2.4.5 Water

Filtered water of conductivity 18Ωohm (Water Purification System, Modulab™, Liquepure™) was obtained from a filter system and autoclaved. Alternatively 10ml vials of sterile, injectable water (Astra Pharmaceuticals) were used.

#### 2.2.4.6 Liquid Paraffin Oil

Most reaction mixtures that were amplified in hot-lid cyclers did not require the addition of paraffin oil. Tubes used in other thermal cyclers, or reactions containing only 25µl volumes, were overlayed with paraffin oil (Gilseal Homecare) to prevent evaporation of reaction components and to increase the rate of thermal equilibration.

#### 2.2.4.7 Thermal Cycler

A FTS-320 Corbett Research Australia thermal sequencer was used with tubes containing paraffin oil and a Corbett Research PC-960C cooled thermal cycler was used for tubes without oil (unless a tube only contained 25µl of reaction mix).

#### 2.2.4.8 Reaction Tubes

Reaction tubes were 0.5ml polypropylene tubes (Sarstedt), autoclaved before use.

#### 2.2.4.9 Primers

The five sets of type-specific primers used for amplification of the HPV 16 URR spanned nts 7010 to 123 representing the entire URR and part of the L1 gene (Table 2.2). Primers URR-1 and URR-2 (May *et al.*, 1994) amplified the entire URR and 146 nucleotides of the L1 gene. These primers were used in combination with primers URR-3, URR-4 and URR-5, specifically designed for this study in accordance with the recommendations published by Boehringer Mannheim (Appendix A7.2.2). Primers were synthesised by Gibco BRL (Life Technologies), Pacific Oligos or Sigma Aldrich. They were rehydrated upon arrival from the

supplier with TE (10mM Tris, and 1mM EDTA) to 1μg/μl, the concentration confirmed on the spectrophotometer, and stored at -70°C.

Table 2.2: Primers Used for Polymerase Chain Reaction of the Human Papillomavirus Type 16 Upstream Regulatory Region

| Primers | Genome Positions | Nucleotide Sequence (5'-3') | Product Size (bp) |
|---|---|---|---|
| URR-1* | 7010-7034 | *GGATCC*GACCTAGATCAGTTTCCTTTAGGAC | 1031 |
| URR-2* | 123-98 | *GGATCC*TCCTGTGGGTCCTGAAACATTGCAGT | |
| URR-3 | 7415-7434 | CTATATTTTGTAGCGCCAGG | 294 |
| URR-4 | 7703-7685 | CCTAACAGCGGTATGTAAG | |
| URR-5 | 7666-7684 | AATCACTATGCGCCAACGC | 363 |
| URR-2 | 123-98 | *GGATCC*TCCTGTGGGTCCTGAAACATTGCAT | |
| URR-3 | 7415-7434 | CTATATTTTGTAGCGCCAGG | 620 |
| URR-2 | 123-98 | *GGATCC*TCCTGTGGGTCCTGAAACATTGCAT | |
| URR-1 | 7010-7034 | *GGATCC*GACCTAGATCAGTTTCCTTTAGGAC | 699 |
| URR-4 | 7703-7685 | CCTAACAGCGGTATGTAAG | |

* Primers published by May *et al.*, 1994

Italics = *Bam*H I restriction sites on primers URR-1 and URR-2

## 2.2.5 Polymerase Chain Reaction Preparation

### 2.2.5.1 Cocktail Preparation

1. Cocktail preparation was performed in a nucleic acid free area. The appropriate cocktail reagents were transferred from -20°C to ice contained in a small esky.
2. Where feasible, bulk PCR cocktails were prepared which could be stored at -70°C for up to a week. The final volume of the cocktail depended on the number of samples to be tested. One extra mix was allowed for each ten tubes to allow for pipetting inaccuracies. One "no nucleic acid" negative control and one positive control (CaSki-derived nucleic acids) was included for each series of PCR reactions. The volume of water was adjusted accordingly for a final volume of 25μl, 50μl, or 100μl depending on the particular reaction.
3. Twenty microlitres, 45μl, or 90μl of cocktail was transferred into each 0.5ml tube and then one drop of paraffin oil was added to each tube if necessary to prevent evaporation during cycling.

4. The cabinet was wiped down with 70% alcohol and UV irradiated for at least 20 minutes.

### 2.2.5.2 Addition of Targets

Cocktails were transported to the workstation on ice to prevent activation of the DNA polymerase. Target volumes ranged from 5µl to 10µl depending on the size of the reaction. The operator dressed in a gown and cap. A clean aerosol resistant tip (ART) (Molecular Bio-Products) was used each time to add 5µl or 10µl of the appropriate samples into tubes containing the cocktail. Five microlitres or 10µl of water was added to the nucleic acid-free negative control tube. It was ensured that the samples being amplified were delivered into the PCR reaction mix below the oil, not into the oil.

### 2.2.5.3 Addition of Positive Control Nucleic Acids

Five microlitres of the positive control, 20pg/µl CaSki total cell nucleic acids (for a 25µl or 50µl reaction) or 10µl of 100ng/µl CaSki (for a 100µl reaction) was added to the reaction mixes at a separate workstation.

### 2.2.5.4 Thermal Cycling

Thermal cycling was initiated without delay to prevent DNA polymerase activation and non-specific amplification by the primers.

## 2.2.6 Optimisation of Polymerase Chain Reaction Protocols

The quality of PCR product is of critical importance to successful sequencing results; background can obscure real sequence data and shorten the length of read. It was necessary to optimise each PCR protocol to prevent non-specific binding of the primers and to increase the sensitivity of the assay. The concentration of dNTPs and DNA polymerase remained standard for each PCR. Optimisation was performed by titrating the concentration of primer (from 0.25µM to 0.75µM), against magnesium chloride (from 1.5mM to 3.5mM). The amount of target DNA was also varied from 200ng to 700ng, according to the amount of tumour in each of the samples. Finally, the annealing temperature was also varied (from 50°C to 60°C). To determine a

starting point for the annealing temperature, the following calculation was made: $T_m = 69.3 + 0.45(G+C)\% - 650/L$. Where $T_m$ is the melting temperature, and L the length of the oligonucleotide.

All PCR protocols, except that using primers URR-3 and URR-2, were optimised by the author during a previous study (Watts, 1996). The PCR protocol used for primers URR-3 and URR-2 was the same as that used for primers URR-1 and URR-2 except that 0.5mM less $MgCl_2$ was found to give optimal results by titration.

### 2.2.7 Polymerase Chain Reaction Protocols for Amplification of the Upstream Regulatory Region

Tables 2.3 to 2.7 represent the protocols optimised for each primer combination.

Table 2.3: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-1 and URR-2, or Primers URR-1 and URR-4

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 56.28µl | - |
| 10x Reaction Buffer | 10µl | 1x |
| 25mM $MgCl_2$ | 12µl | 3mM |
| Primer URR-1 (100ng/µl) | 5µl | 0.4µM |
| Primer URR-2 (100ng/µl) or Primer URR-4 (100ng/µl) | 5µl | 0.4µM |
| dNTPs | 1µl | 250µM |
| *Tth*+ or *Amplitaq Gold*™ DNA Polymerase | 0.72µl | 4 Units |
| | ____ | |
| | 90µl PCR Cocktail + 10µl Target DNA (100ng/µl) | |

Thermal Cycling Conditions:

| | |
|---|---|
| Cycle 1 | Step 1 94°C/4 minutes (Denaturation)<br>Step 2 57°C/1 minute (Annealing)<br>Step 3 72°C/1 minute (Extension) |
| Cycles 2-39 | Step 1 94°C/0.5 minutes (Denaturation)<br>Step 2 57°C/1 minute (Annealing)<br>Step 3 72°C/1 minute (Extension) |
| Cycle 40 | Step 1 94°C/0.5 minutes (Denaturation)<br>Step 2 57°C/1 minute (Annealing)<br>Step 3 72°C/10 minutes (Extension) |

Table 2.4: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-3 and URR-4

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 12.55µl | - |
| 10x Reaction Buffer | 2.5µl | 1x |
| 25mM $MgCl_2$ | 2.5µl | 2.5mM |
| Primer URR-3 (100ng/µl) | 1µl | 0.6µM |
| Primer URR-4 (100ng/µl) | 1µl | 0.6µM |
| dNTPs | 0.25µl | 250µM |
| *Tth*+ or *Amplitaq*® | 0.2µl | 1 Unit |
| DNA Polymerase | ——— | |
| | 20µl PCR Cocktail + 5µl Target DNA (50ng/µl) | |

Thermal Cycling Conditions:

| | |
|---|---|
| Cycle 1 | Step 1 94°C/4 minutes (Denaturation) |
| | Step 2 55°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycles 2-39 | Step 1 94°C/0.5 minutes (Denaturation) |
| | Step 2 55°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycle 40 | Step 1 94°C/0.5 minutes (Denaturation) |
| | Step 2 55°C/1 minute (Annealing) |
| | Step 3 72°C/10 minutes (Extension) |

Table 2.5: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-3 and URR-2

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 29.1µl | - |
| 10x Reaction Buffer | 5µl | 1x |
| 25mM $MgCl_2$ | 5µl | 2.5mM |
| Primer URR-3 (100ng/µl) | 2.5µl | 0.4µM |
| Primer URR-2 (100ng/µl) | 2.5µl | 0.4µM |
| dNTPs | 0.5µl | 250µM |
| *Tth*+ or *Amplitaq*® DNA Polymerase | 0.4µl | 2 Units |
| | ——— | |
| | 45µl PCR Cocktail + 5µl Target DNA (100ng/µl) | |

Thermal Cycling Conditions:

| | |
|---|---|
| Cycle 1 | Step 1 94°C/4 minutes (Denaturation) |
| | Step 2 57°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycles 2-39 | Step 1 94°C/0.5 minutes (Denaturation) |
| | Step 2 57°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycle 40 | Step 1 94°C/0.5 minutes (Denaturation) |
| | Step 2 57°C/1 minute (Annealing) |
| | Step 3 72°C/10 minutes (Extension) |

### 2.2.7.1 Hot Start Polymerase Chain Reaction

Hot start PCR (Tables 2.6 and 2.7) can reduce non-target primer amplification. It is accomplished by the manual addition of an essential reagent, in this case $MgCl_2$, to the reaction tube at elevated temperatures. This improves the specificity of the reaction and prevents the formation of primer dimers.

Table 2.6: Hot Start Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-5 and URR-2

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 13.05µl | - |
| 10x Reaction Buffer | 2.5µl | 1x |
| Primer URR-5 (100ng/µl) | 1µl | 0.4µM |
| Primer URR-2 (100ng/µl) | 1µl | 0.4µM |
| dNTPs | 0.25µl | 250µM |
| *Tth*+DNA Polymerase | 0.2µl | 1 Unit |
| | ——— | |
| | 20µl PCR Cocktail + 5µl Target DNA(50ng/µl) | |

Thermal Cycling Conditions:

Tubes sat in thermal cycler at 80°C for 5 minutes before adding 2µl of 25mM (2mM final) $MgCl_2$ and proceeding to cycling.

| | |
|---|---|
| Cycle 1 | Step 1 94°C/4 minutes (Denaturation) |
| | Step 2 57°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycles 2-39 | Step 1 94°C/0.5 minutes (Denaturation) |
| | Step 2 57°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycle 40 | Step 1 94°C/0.5 minutes (Denaturation) |
| | Step 2 57°C/1 minute (Annealing) |
| | Step 3 72°C/10 minutes (Extension) |

Table 2.7: Hot Start *UlTma*™ Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers URR-1 and URR-2

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 64.5µl | - |
| 10x Reaction Buffer | 10µl | 1x |
| Primer URR-1 (100ng/µl) | 4µl | 0.4µM |
| Primer URR-2 (100ng/µl) | 4µl | 0.4µM |
| dNTPs | 1µl | 1mM |
| *UlTma*™ DNA Polymerase | 0.5µl | 3 Units |
| | —— | |
| | 84µl PCR Cocktail + 10µl Target DNA(100ng/µl) | |

Thermal Cycling Conditions (as manufacturers recommendations):

Tubes sat in thermal cycler at 80°C for 5 minutes before adding 6µl of 25mM $MgCl_2$ and proceeding to cycling.

| | |
|---|---|
| Cycle 1 | Step 1 95°C/2 minutes (Denaturation) |
| Cycles 2-41 | Step 1 95°C/2 minutes (Denaturation) |
| | Step 2 54°C/2 minutes (Annealing) |
| | Step 3 72°C/2 minutes (Extension) |
| Cycle 42 | Step 1 72°C/7 minutes (Extension) |

## 2.2.8 Preparation of Polymerase Chain Reaction Products for Electrophoresis

Five microlitres of PCR product was added to a tube containing 10µl of x1.6 gel loading dye (0.08mg/ml bromophenol blue, 16% glycerol (v/v), and 42mM EDTA, pH 8.0).

### 2.2.8.1 Preparation of Size Markers

Many different markers are commercially available and the choice depends on the size of the bands of interest. Two different markers were chosen for this study. They were always prepared in bulk and stored at -20°C. For PCR involving primers URR-1 to URR-5, pUC19DNA/*Hpa*11 markers were used. When available, a 1kb DNA ladder (Life Technologies, Gibco BRL) was used with PCR products produced using primers URR-1 and URR-2. Fragment sizes for the markers are given in Appendix A7.3.

### 2.2.8.2 Gel Electrophoresis

The method for making agarose gels is given in Appendix A7.2.3.

1. Samples were loaded under the electrophoresis buffer (x1 TAE) into the wells of the gel and one well was left for the size marker.
2. The gel was electrophoresed at 140 volts/250mA for 40 minutes.
3. After completion of electrophoresis the gel was removed and eased into a solution of 150µl of 1mg/ml ethidium bromide in 100ml of x1 TAE buffer and stained for 15 minutes. The gel was then destained in distilled water for 15 minutes. Alternatively, 50µl of 1mg/ml ethidium bromide was added to the gel at the time of pouring.
4. The bands on the gel were visualised using a transilluminator (Ultraviolet Products Inc.) and photographed using polaroid film 667 in a polaroid Land camera with a 5" lens and a 16cm by 23cm (Cu 88-49) hood. The size of bands was estimated by comparison of their position on the gel relative to those of the size markers.

### 2.2.9 Polyethylene Glycol (PEG) Precipitation of Polymerase Chain Reaction Products for Sequencing

PCR products were purified for sequencing using polyethylene glycerol (PEG) precipitation using modifications (Rose *et al.*, 1997) of the methods published by Lis (1980). This was necessary to remove excess primers and dNTPs.

1. Duplicate tubes of PCR product were pooled, microfuged and any oil pipetted from the surface.
2. Forty five microlitres of PCR product was added to 45µl of x2 PEG 8000 mix (27% PEG 8000, 600mM sodium acetate, pH 5.2, and 6.5mM $MgCl_2$). Tubes were vortexed briefly to mix.
3. Tubes were allowed to stand at 4°C for 1 hour.
4. DNA was pelleted by centrifuging for 15 minutes at 12,000 rpm (in an MSE micro centaur centrifuge), ensuring that the tubes were all aligned in the same way, ie., all with hinges pointing outwards (for orientation as pellets could be small and difficult to see at this stage).
5. Supernatants were carefully removed with a pipette tip running down the side of each tube away from the pellet.
6. DNA was washed by the addition of 300µl of 95% ethanol. The tubes were spun for 15 minutes, making sure that they were again aligned the same way. The ethanol was carefully pipetted off.
7. The pellets were dried at 42°C using a heating block (thermal cycler) for about 30 minutes or longer depending on the amount of remaining ethanol.
8. Pellets were resuspended in approximately 15µl of sterile water (volume depending on the size of the pellet) and allowed to stand at room temperature for about 30 minutes, or until the pellets were completely dissolved. Tubes were stored at -20°C.
9. Two microlitres of resuspended DNA was run on a gel so that the concentration of the DNA could be roughly quantitated.

### 2.2.10 DNA Sequencing - Sanger Dideoxy Method

The method for DNA sequencing used in this study involves the enzymatic synthesis of labelled DNA using dideoxynucleotides that randomly terminate the elongating strand as demonstrated by Sanger *et al.* (1977). This is achieved by using DNA polymerase to extend a labelled DNA strand until a chain terminating 2'3'-dideoxynucleotide (ddNTP) is incorporated. These ddNTPS lack the 3'-OH necessary for chain elongation. The resulting labelled strands form a nested set of specifically terminated chains. These are separated according to size by electrophoresis on a high resolution denaturing gel.

The DNA once prepared using PEG precipitation was sequenced using dyedeoxy terminator chemistry (Appendix A7.4). Cycle sequencing was performed in a thermal sequencer using the ABI PRISM® Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit with *AmpliTaq*® DNA polymerase (Perkin Elmer) according to the manufacturer's instructions. Extension products were analysed on a 377 DNA sequencer (Perkin Elmer). Cycle sequencing reactions and gel analysis were performed by staff at the Children's Medical Research Institute, SUPAMAC (Sydney University Prince Alfred Macromolecular Analysis Centre) or AGRF (Australian Genome Research Facility) at the University of Queensland.

Forward or reverse primers were used for sequencing, depending on the quality of sequence data obtained with each of the primers. Pooling of PCR products, sequencing with a different primer and/or PCR product were strategies used to confirm sequence variations.

#### 2.2.10.1 Computer Analysis

Pileup (GCG) and Pretty (GCG) computer software available through the Australian National Genomic Information Service (ANGIS) were used to analyse the sequencing data. These programs aligned each of the sequences with the prototype and allowed sequences different from the prototype to be determined.

# 2.3 RESULTS

## 2.3.1 Polymerase Chain Reaction

### 2.3.1.1 Strategy

PCR was initially performed using primers URR-1 and URR-2, in order to obtain PCR products encompassing the entire URR. Occasionally it proved difficult to obtain this 1kb PCR product, especially where the DNA had not been recently extracted, or the sample contained only a very small amount of viral DNA. In these instances, PCRs using primer combinations URR-3 and URR-4, URR-5 and URR-2, URR-3 and URR-2, or URR-1 and URR-4 were used to span the 1kb segment.

### 2.3.1.2 Polymerase Chain Reaction Results

Single bands of expected size were obtained for all of the PCR protocols described. The yield of PCR product was variable, usually reflecting the differing concentrations of viral DNA present in samples and the quality of the DNA. Examples of PCR products obtained using primer combinations URR-1 and URR-4, and URR-1 and URR-2 are given in Figure 2.2.




Figure 2.2: Different size polymerase chain reaction products from the upstream regulatory region

Lanes 1 to 6 contain products produced using primers URR-1 and URR-4, and lanes 8 to 13 using primers URR-1 and URR-2. Lanes 5 and 12 contain positive controls (1μg of CaSki DNA) and lanes 6 and 13 contain negative controls (water). Lane 7 contains a pUC 19/*Hpa*II DNA marker.

### 2.3.2 Sequencing the Upstream Regulatory Region

Several variables influenced the quality of sequencing data: the quality and the amount of template, the primer used, and the expertise of the personnel performing the sequencing reaction. The antisense primer URR-2, in most instances yielded approximately 350 nucleotides of sequence data with 2% to 3% ambiguity codes and a good signal to noise ratio. However, sequencing with the sense primer, primer URR-1, proved less satisfactory. At best this primer produced up to 300 nucleotides of data with up to 10% ambiguity codes. The presence of palindromic sequences within this published primer (May *et al.,* 1994) may have accounted for the difficulty experienced. Where PCR products failed to sequence satisfactorily using primer URR-2, or uncertainties existed in the sequence data, PCR product was then sequenced using primer URR-5. This gave approximately 300 nucleotides of satisfactory data with 2% ambiguity codes. Sequencing using primer URR-4 was then performed to bridge the gap in the centre of the sequence data. The sequence data obtained using this primer were far superior to that of the other primers, consistently producing approximately 400 nucleotides of sequence data often with zero ambiguity codes. Examples of the electropherograms used to obtain sequence information for each of the primers is given in Appendix A7.5. The extent of sequence data obtained using each of the primers is represented in Figure 2.3.

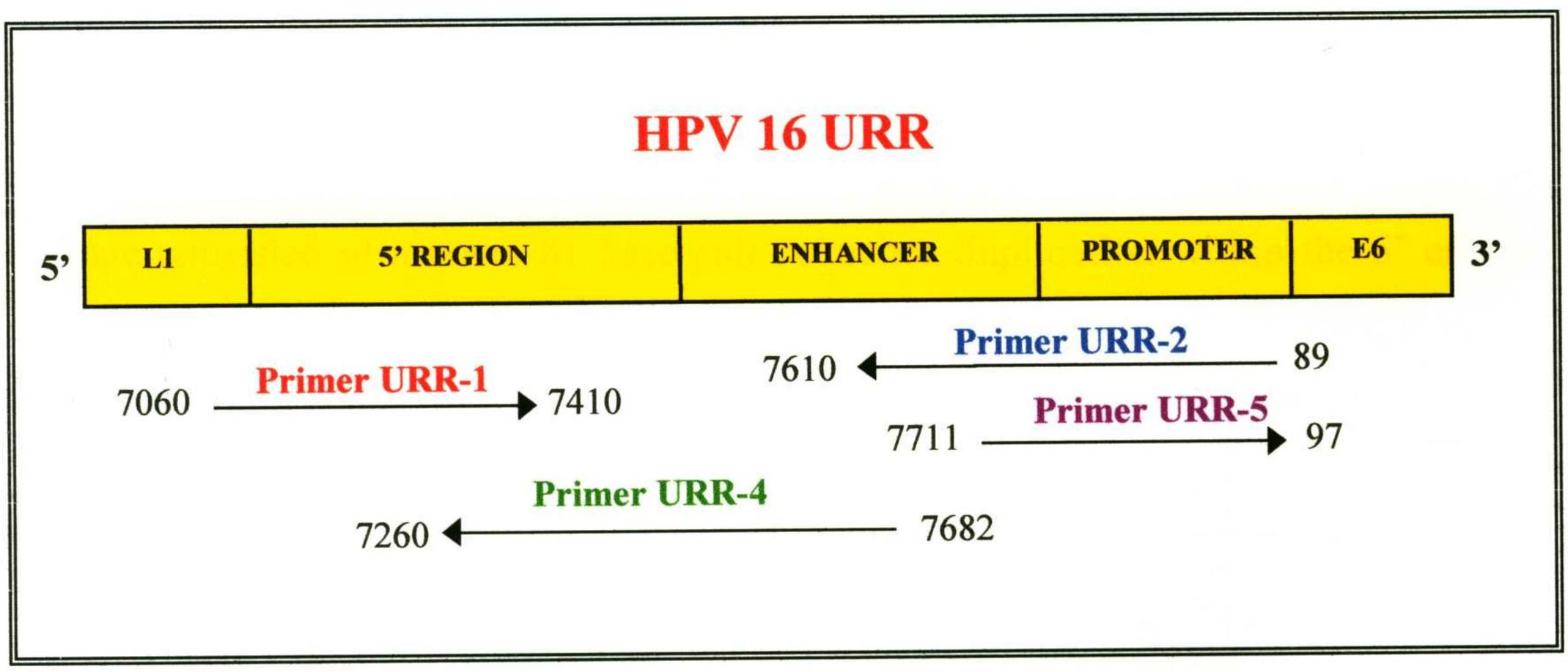



Figure 2.3: The human papillomavirus type 16 upstream regulatory region illustrating the extent of sequencing data obtained using each of the primers

Numbers represent nucleotide positions

### 2.3.2.1 Technical Considerations

In general, approximately 943 nucleotides of sequence were obtained from each of the isolates (from approximately nt 7060 to nt 97). DNA sequence over the primer sites was discarded, as it was identical to the sequence of the primer, making it impossible to determine underlying sequence variation in these regions. The CaSki sequence was compared with that already in ANGIS. Complete sequence alignments for the URR are given in Appendix A7.6.1.

### 2.3.2.2 Sequence Analysis

Overall, only one of the 34 isolates was identical to the prototype sequence over approximately 943 nucleotides sequenced (Figure 2.4). Between one and 15 single nucleotide variations were identified in the remaining 33 isolates. Overall, 44 different nucleotide variations and one single base insertion were identified. Of these, 19 impacted known TF or hormone binding sites. The percentage of multivariants identified (isolates with more than one nucleotide change) was 91% (31/34 isolates). Specific sequence variations were identified in the majority of the isolates, for example, a change from G to A at nt 7521 in 28 of the samples and a change from G to T at nt 7193 in the same 28 samples. Sequence variations in the CaSki URR were as previously reported (Xi *et al.,* 1993, and Meyers *et al.,* 1995b, and in ANGIS). There was no evidence of areas of hypervariability in this sample of URR sequences. The 3' region of the URR was found to be more highly conserved than the other regions.

A large-scale sequence change was identified in the O2 isolate. This large-scale change consisted of a novel 81 base pair sequence duplication within the 5' end of the enhancer, corresponding to nucleotides 7453 to 7534. The nucleotide sequence of the URR from O2 has been submitted to Genbank (accession number AF026034).

**Note:** Not all of the sequencing data on the URR were obtained during this study, including the identification of the large-scale duplication in O2. Most or the entire URR sequence of 24 of the 34 isolates was obtained during this study. The results given in Figure 2.4 are for the entire cohort of 34 isolates for completeness. In order to differentiate between those sequences obtained during this study and those obtained previously (Watts, 1996); the reader is referred to Appendix A7.6.1 for the alignment of sequences obtained only during this study.

Most of the variants were characteristic of the European (E) lineage (28/34 or 82%). Smaller proportions were Asian American (AA) and Asian (As) lineage isolates (2/34 or 6%, and 4/34 or 12% respectively). No African 1 or 2 lineage isolates were identified, or any minor lineage variants. As the major HPV 16 variant lineages have not been previously determined for Australia or any region of the South Pacific, a comparison of results with other regions of the world was made (Figure 2.5). Values used in Figure 2.5 were derived from a large multinational study conducted by Yamada *et al.* (1997). It is clear that variant patterns in Australia and the South Pacific are similar to those identified in Europe, although with larger proportions of As and smaller proportions of AA lineage variants. The proportion of As lineage variants in Australia is larger than all other regions of the world except in Asia itself. Only E variants were present in the New Caledonian patients. The majority of the E variants did not contain sequence variations in TFBSs (apart from the common variation at nt 7193 in a putative TEF-1 binding site), in comparison with the As and AA variant isolates which contained many variations from prototype in TFBSs.

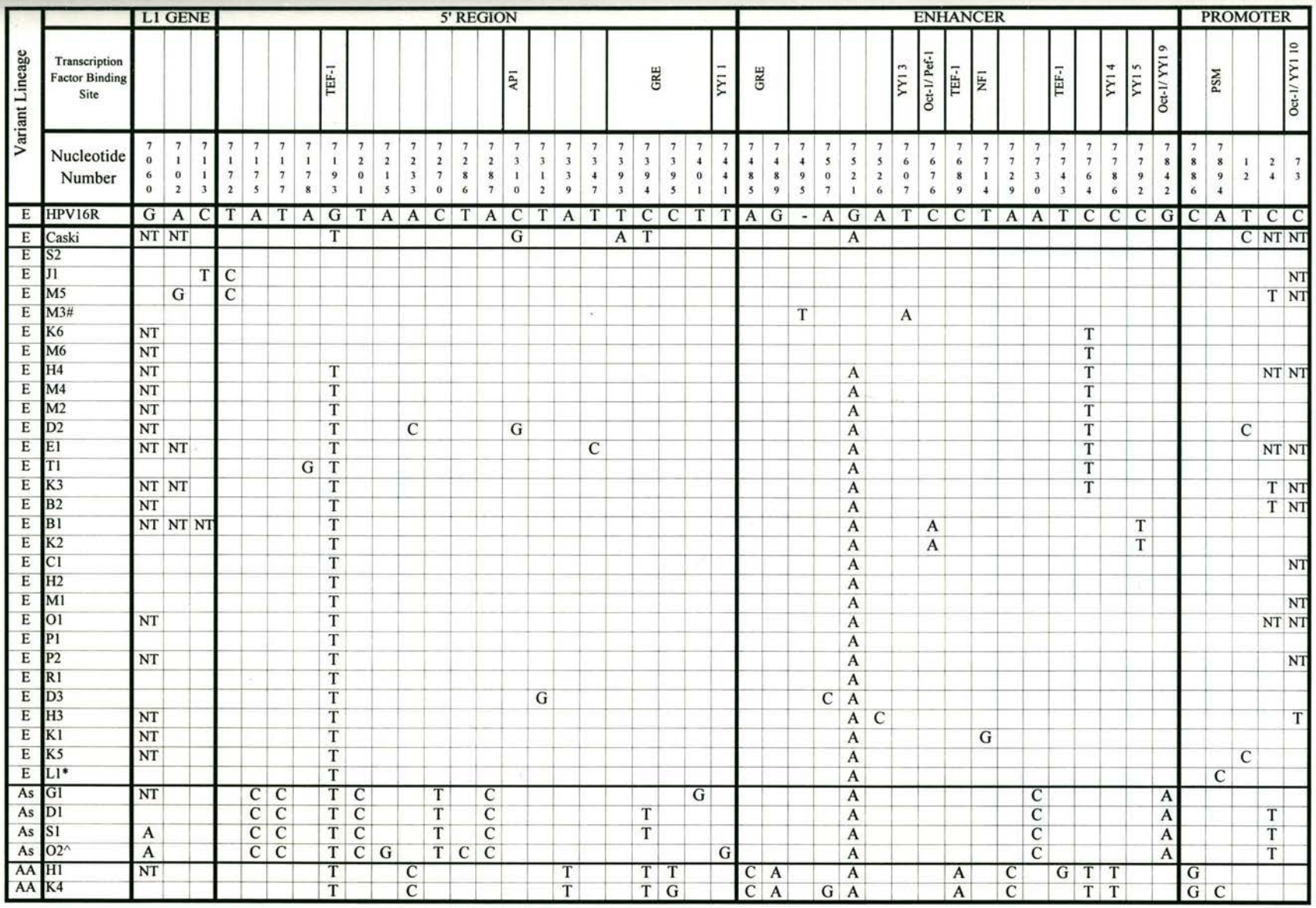



Figure 2.4: Sequence variation in the human papillomavirus type 16 upstream regulatory region

The location of sequence variations in TFBSs is indicated across the top; ^ duplication of nts 7453-7533 (81bp); # sequence unavailable between nts 7380-7440; * sequence unavailable between nts 7683-7710; E = European, As = Asian, AA = Asian American; and NT = not tested at this nucleotide

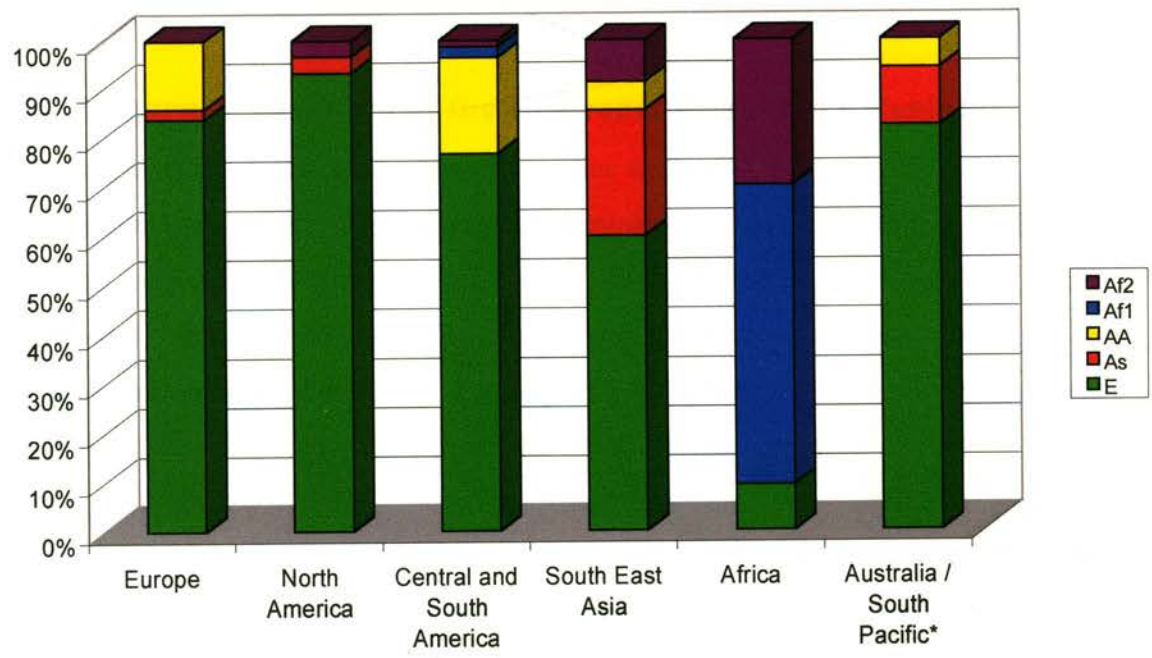



Figure 2.5: A comparison of human papillomavirus type 16 variant lineages in cervical carcinomas from different regions of the world

A comparison has been made with the results obtained during this study* (using isolates from Australia and the South Pacific) with previously published data from other regions of the world (Europe, North America, Central and South America, South East Asia, and Africa) (from Yamada *et al.*, 1997). Concept adapted from Yamada *et al.*, 1997.

The average age of the patients in this study for each of the lineages was 48.9 years for E variants, 48 years for As variants, and 40.5 years for AA variants. Patient age was considerably lower for AA variants, although only two patients were in this category and age may not be truly representative. No valid conclusions could be drawn between individual, and patterns of variations and clinicopathological variables in this relatively small and non-representative sample.

# 2.4 DISCUSSION

## 2.4.1 Sequence Variation in the Upstream Regulatory Region

This is one of the most comprehensive studies to date of the entire HPV 16 URR. There have been several multi-centre epidemiological studies examining HPV 16 variation from different geographical regions (Ho *et al.,* 1991 and 1993a, Chan *et al.,* 1992a, and Yamada *et al.,* 1997), however these studies have focussed primarily on the enhancer segment of the URR. The sequence information obtained in this study is the first for HPV 16 from Australia and the South Pacific region.

HPV 16 and 18 variants have a very slow rate of evolution (Chan *et al.,* 1995). The regions of greatest intratypic variability correspond to regions of greatest intertypic variability in the HPV genome (Stewart *et al.,* 1996). For instance, L2 is the most variable coding segment in HPV 16 and is approximately as varied as the URR (Yamada *et al.,* 1995). The amino terminal region of E6 is relatively varied and L1 is the most conserved (Wheeler and Icengole, 1995). The maximal evolutionary distance for the URR has been calculated to be 4.7% (Ho *et al.,* 1993a), which is similar to the 4.6% found in this study using a smaller number of isolates but a larger number of nucleotides (364 nts compared to 943 nts in this study). One major HPV variant seems to predominate over time in cervical infections, whereas minor variants appear to be more transient, or may be rare (van Belkum *et al.,* 1995, and Xi *et al.,* 1995). One of the largest studies so far conducted found only three of 411 invasive tumours containing two separate HPV 16 variants (Yamada *et al.,* 1997). Therefore, infection with more than one variant in cervical cancer is rare and the probability that it existed in the isolates from this study is low, although this was not directly tested.

Representative variants from all five major lineages are detected in populations worldwide, although specific prevalences differ by geography (Wheeler and Icenogle, 1995). These differences may be explained by founder effects and/or may reflect variant selection in ethnic groups. The majority of the isolates in this study belonged to the E lineage (28/34 or 82%). This is not surprising considering Australia was settled by persons of European descent more than 200 years ago.

Smaller proportions of the isolates belonged to the As (4/34 or 12%) and AA (2/34 or 6%) variant lineages. These results attest to more recent change in the composition of the Australian population, which now contains persons of many different ethnic and cultural backgrounds. In particular there is currently a large Asian population in Australia, especially in the Sydney region where much of the Australian sample was derived. Predictably most As variants are observed in South East Asia, while almost all AA variants are detected in Central and South America or Spain (Yamada *et al.,* 1997). A comparison of HPV 16 variants found in Australia and the South Pacific with those identified in other regions of the world revealed a similar composition to Europe, except for a greater proportion of As and a lower proportion of AA lineage variants.

In comparison with the Australian isolates, only E variants were identified in the eight isolates from New Caledonia. In the South Pacific, New Caledonia was taken over by French rule in 1853 and the current population is comprised of almost half Melanesian, one third European, and the rest minority groups. It is tempting to speculate that the French brought E variants of HPV to New Caledonia with them in the 19th century. To determine if small percentages of other variant lineages are represented in New Caledonia, a larger, more representative sample would have to be analysed.

The prototype HPV 16 URR was derived from a German tumour (Seedorf *et al.,* 1985) and is an artificial reference. It can be understood therefore why E lineage variants contain little sequence variation compared to the prototype, as evident from the results of this and other studies. The two common variant nucleotides 7193 and 7521, were present in most of the isolates in this study, suggesting that the prototype itself is the variant at these nucleotides. The As and AA isolates, on the other hand, contained sequences quite distinct from the prototype, perhaps reflecting the evolution of the virus on different continents or the presence of particular viral strains that have become more prevalent in certain regions of the world. Common variations were identified in multiple lineages and can therefore be considered biological entities. This has also been observed in other studies (for example in Ho *et al.,* 1991 and 1993a, Yamada et al., 1997, and Wheeler *et al.,* 1997). It is not known whether immunity to one HPV 16 variant can protect against infection with another

variant. Therefore HPV variants may prove important both to the risk of cervical neoplasia and to the design and effectiveness of vaccines (Wheeler *et al.,* 1997, and Yamada *et al.,* 1997). Caution should be taken however when determining the significance of measurements dependent on the actual prevalence of particular variants in a study population (Wheeler and Icenogle, 1995). If a single variant predominates in a population, the significance of detecting that variant is not high for the purposes of marking persistence or transmission.

In summary, in this chapter of the study sequence variation in the URR from HPV 16 positive cervical cancer isolates has been documented. E, As, and AA HPV 16 lineage variants were detected in the cohort from Australia and New Caledonia in the South Pacific, with E lineage variants the most prevalent. What is needed now is a better understanding of the biology of HPV intratype variation, which the remainder of this study aims to address.

# CHAPTER 3

# FUNCTION ANALYSIS OF SEQUENCE VARIATION WITHIN THE HUMAN PAPILLOMAVIRUS TYPE 16 UPSTREAM REGULATORY REGION

# 3.1 INTRODUCTION

Although there is now considerable data on sequence variation in the HPV 16 URR, its biological significance has not been well characterised. To some extent, this reflects the complexity of the situation, substitutions affect up to 5% of nucleotides and particular variant lineages cluster geographically. On current evidence at least some large-scale mutations and single nucleotide substitutions of the HPV 16 URR have biologic significance. Deletions or point mutations affecting binding of YY1 to its motifs in the HPV 16 URR (Dong *et al.*, 1994, and May *et al.*, 1994), and a duplication involving HPV 16 enhancer sequences and part of L1 (Hall *et al.*, 1997), may provide plausible explanations for malignant conversion in the absence of integration. Large-scale changes in the URR may also mediate malignant transformation of low risk HPV 6 and 11 (Rosen and Auborn, 1991, and DiLorenzo *et al.*, 1992). In addition, increased early promoter activity from HPV 6 due to duplication of enhancer segments has been reported (Kulke *et al.*, 1989), and single nucleotide variations in an Sp1 site in the HPV 18 URR increase promoter activity and may be biologically important (Rose *et al.*, 1998). Functional analysis of URR variants has not always revealed altered promoter activity however, and this may be due to the fact that particular variants analysed did not have sequence changes in TFBSs (Ho *et al.*, 1991).

Among the few epidemiological studies looking for associations between variants and clinical outcome, women and men with non-prototype-like HPV 16 variants were found to be more likely to develop either CIN 2/3 or anal carcinoma *in situ* respectively, than those infected with prototype-like variants (Xi *et al.*, 1997, and 1998). In one recent study, prototype-like variants, as well as variants with changes in the URR at nts 7521 and 7193, and AA variants were associated with more aggressive cervical disease, determined as rapid progression to carcinoma *in situ* (Kurvinen *et al.*, 2000). These findings have been contradicted by others who found similar variants in cervical smears from asymptomatic women, in CIN, and in cancers (Nindl *et al.*, 1999).

The aims of this part of the study were to determine if naturally occurring variant HPV 16 URRs identified in Chapter 2 have altered promoter activity, and to investigate associated mechanisms. Since it would have been impractical to examine all the variant URRs that were identified, six variants were selected for function analysis. The entire URRs containing the E6/E7 promoter were cloned upstream of a luciferase reporter gene, which was used as a surrogate for E6/E7 expression. Constructs were transiently transfected into HeLa and HT3 cells and luciferase activities compared with the prototype HPV 16 URR. For variants with enhanced promoter activity, attempts were made to identify the nucleotides responsible by performing PCR-mediated site-directed mutagenesis where specific nucleotides of interest were converted to those found in the prototype URR. Binding studies were also performed to determine the effects of sequence changes on the binding of corresponding TFs to their DNA recognition sequences.

# 3.2 METHODS

## 3.2.1 DNA Cloning

Recombinant DNA technology has in less than two decades become fundamental to the advances of medical science and molecular biology. DNA cloning began in the laboratories of Paul Berg, Herbert Boyer and Stanley Cohen in the early 1970's (Lehninger *et al.*, 1993). Cloning essentially involves separating a segment of DNA from a larger segment of DNA and attaching it to carrier DNA using DNA ligase, and then replicating this modified DNA thousands or millions of times. The result is the selective amplification of that particular gene or DNA segment. The methods used to accomplish this are known collectively as recombinant DNA technology or genetic engineering.

*Escherichia coli* (*E.coli*) was the first organism used for recombinant DNA work and is still the most common host cell of choice. *E.coli* has many advantages, its DNA metabolism, and many other biochemical processes, are well understood, and many naturally occurring cloning vectors such as plasmids associated with *E.coli* are well characterised. Effective techniques are also available for moving DNA, such as plasmids, circular DNA molecules that replicate independently from the host chromosome, from one bacterial cell to another (Lehninger *et al.*, 1993).

## 3.2.2 Selection of Human Papillomavirus Type 16 Upstream Regulatory Regions for Function Studies

The six variant URRs chosen for function studies from those identified in Chapter 2 of the study contained sequence variations either representative of those present in different lineages, or sequence variations of potential interest:

- Isolate O2, an As lineage variant, was chosen because of its interesting and unique 81bp duplication (from nts 7453 to 7534). This duplication involved a segment of the enhancer incorporating an E2 binding site as well as single GRE, TEF-1, Oct-1, and NF1 motifs. This isolate also carried many sequence

variations in TFBSs that were representative of those found in other As lineage isolates in the cohort.

- R1 was selected because the URR from this isolate was the most representative of the E lineage of all the isolates sequenced. It contained only two sequence changes at nts 7193 and 7521, which are common among isolates worldwide.
- Isolate K2 was chosen as one of the few E lineage variants with sequence changes in TFBS. It also contained the two common changes at nts 7193 and 7521. The two potentially interesting sequence variations in K2 fell within an overlapping Oct-1/PEF-1 site at nt 7676, and a YY1 binding site at nt 7792.
- Isolates K4 and H1, the two AA lineage variants, were selected because they contained many sequence changes in TFBSs compared to E isolates. K4 and H1 contained similar, although not identical sequences.
- During the latter stages of the study one other isolate, S1, more representative of the As lineage isolates than O2, was also cloned for investigations. The sequence of the URR in S1 was similar to O2, except for the large-scale duplication.

### 3.2.3 Cloning Strategy

The initial aim was to clone the selected URRs, which had been amplified by PCR using primers URR-1 and URR-2 (as described in section 2.2.7) into the multicloning site of the luciferase vector pALuc (kindly supplied by Ping Dong, China). The primers had *Bam*H I restriction sites located on their 5' ends to facilitate cloning. However after many unsuccessful attempts, this cloning strategy was abandoned. As a result, another vector, pCR®-Blunt (Invitrogen) was chosen as the initial cloning vector and entire HPV 16 URRs from blunt-ended PCR products were cloned into it. The URRs were then subcloned into pALuc using the *Bam*H I sites.

## 3.2.4 Cloning Upstream Regulatory Regions into pCR®-Blunt

### 3.2.4.1 Zero Blunt™ PCR Cloning

pCR®-Blunt (Invitrogen) is a cloning vector that allows selection of positive recombinants via disruption of the lethal gene *ccd*B fused to the carboxyl terminus of LacZα (Figure 3.1). The vector is provided linearised with blunt ends that allows cloning of PCR fragments without the need for restriction sites. Ligation of a blunt PCR fragment disrupts expression of the *lac*Zα-*ccd*B gene fusion product, allowing growth only of positive recombinants. The vector also contains the kanamycin resistance gene for selection in *E.coli*. The cloning efficiencies obtained using this vector are reported to be in the order of 80% to 90%, making this vector a good choice for the initial cloning of PCR products in this study. The H1, K2, K4, O2, R1, and S1 URRs were cloned into pCR®-Blunt to create pCR®-BluntH1, pCR®-BluntK2, pCR®-BluntK4, pCR®-BluntO2, pCR®-BluntR1, and pCR®-BluntS1.

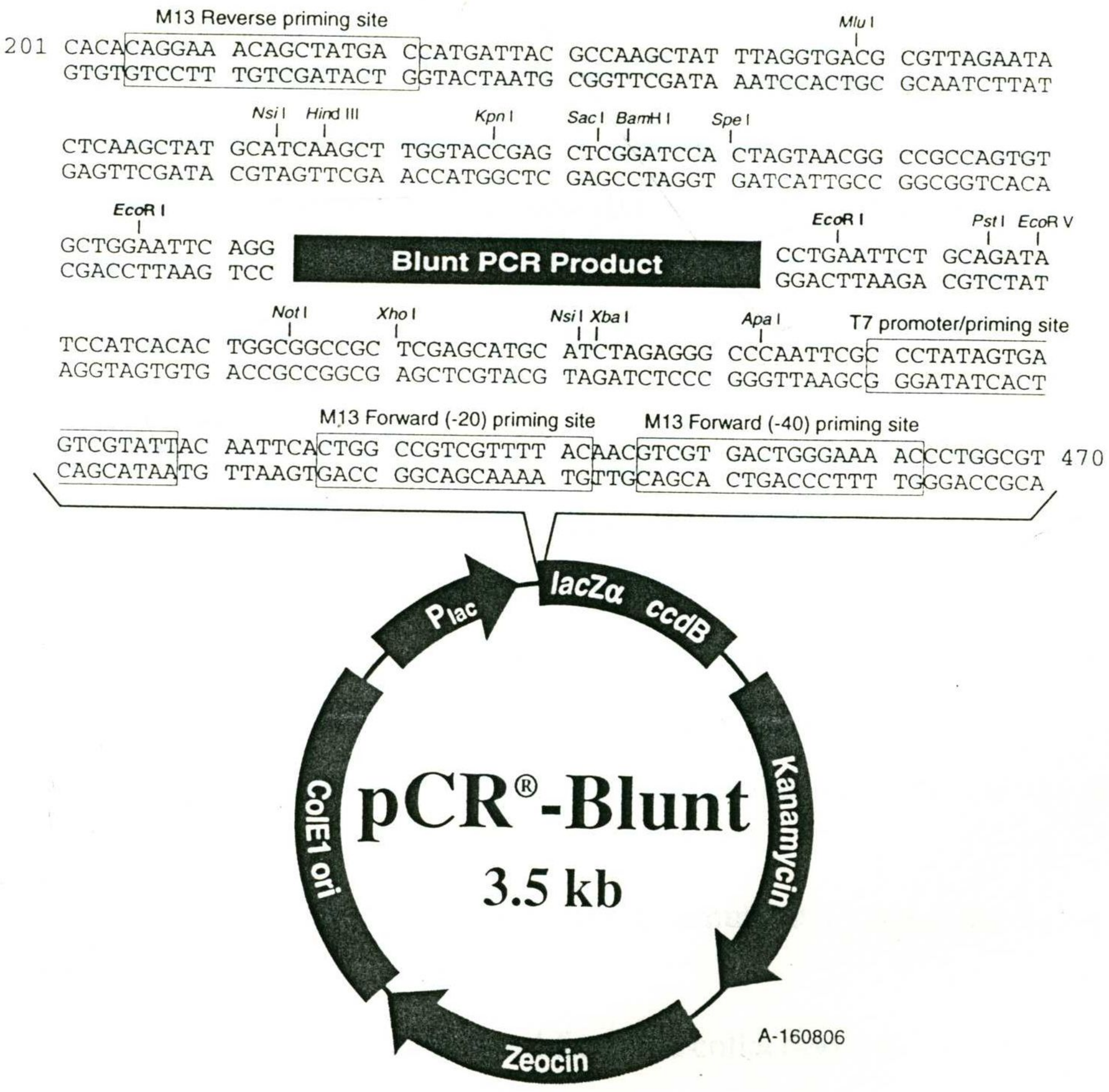



Figure 3.1: The cloning vector pCR®-Blunt showing the location of genes and the multicloning site

Restriction sites found only in the multicloning site are shown.

## 3.2.5 Preparation of Variant Upstream Regulatory Regions for Cloning into pCR®-Blunt

### 3.2.5.1 Production of Blunt-Ended Polymerase Chain Reaction Products

Ninety microlitres of PCR products (from PCR using primers URR-1 and URR-2) were blunt-ended by the addition of 5U of Klenow (Promega) and incubated at 30°C for 15 minutes. Klenow was inactivated by heating tubes to 70°C for a further 15 minutes. Approximately 16µl of x5 gel loading dye was added and the entire contents of the tubes electrophoresed on a 1% ultrapure agarose (Bio Rad) gel at 90V for 1 hour. Cytidine (1mmol/L) (Sigma) was added to the gel to prevent destruction of the DNA by UV transillumination (as recommended by Gründemann and Schömig, 1996). Bands corresponding to the blunt-ended PCR products were cut from the gel and the DNA purified by QIAquick gel extraction.

### 3.2.5.2 Purification of DNA by QIAquick Gel Extraction (QIAGEN®)

This procedure involves the solubilisation of agarose and the selective adsorption of nucleic acids to silicagel particles. This adsorption occurs quantitatively at high salt concentrations, whereas the elution of nucleic acids is accomplished in low salt. This procedure was performed as recommended by the manufacturer.

1. The DNA fragment of interest on the agarose gel was identified under UV illumination by comparing to a size marker and excised.
2. Three volumes of solubilisation buffer (QG) was added to 1 volume of gel.
3. The gel slice and buffer were incubated at 50°C for 10 minutes. The tube was mixed by flicking and inverting 2 to 3 times to help dissolve the gel.
4. A QIAquick spin column was placed into a 2ml collection tube and the sample loaded.
5. The column and tube were centrifuged for 1 minute at 12,000 rpm (in an MSE micro centaur centrifuge).
6. The flow-through fraction was drained from the collection tube and the QIAquick column returned to the same tube.
7. An additional 500µl of buffer QG was added and the tube recentrifuged, to remove any remaining traces of agarose.

8. To wash, 750µl of buffer PE was added to the QIAquick column and centrifuged for an additional minute after standing for 5 minutes at room temperature.
9. The QIAquick column was placed in a clean 1.5ml microcentrifuge tube.
10. The DNA was eluted by adding 30µl of TE (10mM Tris and 0.1mM EDTA, pH 8.0) or water or buffer EB (10mM Tris, pH 8.5), standing for 2 minutes and centrifuging at 12,000 rpm for 1 minute.
11. Two microlitres of the product was electrophoresed on a 2% Seakem™ LE agarose gel to roughly quantify the yield.

#### 3.2.5.3 Ligation

Ligation of DNA requires the formation of bonds between phosphate residues on the 5' termini of DNA and adjacent 3' hydroxyl moieties. This can be catalysed either by bacteriophage T4 DNA ligase or *E.coli* DNA ligase. T4 ligase is the more efficient of the two enzymes and was used in this study. For efficient blunt-ended PCR ligation it is recommended that a 10:1 to 100:1 molar insert-to-vector ratio be used in the ligation (Invitrogen). For a 10:1 molar ratio, approximately 70ng of PCR product were required. The following components were added together for each cloning reaction and incubated at 16°C for 1 hour: 5µl of approximately 15ng/µl gel purified PCR product, 2µl of water, 1µl of 10x ligation buffer (Promega), 1µl of 25ng/µl pCR®-Blunt, and 1µl of 4U/µl T4 ligase (Promega). Two microlitres of the reaction was used for transformation and the remainder stored at -20°C.

### 3.2.6 Transformation of Bacterial Cells

Plasmids can be introduced into host cells by the process known as transformation (Hanahan *et. al.*, 1995). Bacterial transformation can be induced using several methods. The most commonly used method is based on the observations of Mandel and Higa who showed that bacteria treated with ice-cold solutions of calcium chloride (and also several other divalent and multivalent cations), and then briefly heated at 42°C, could be transfected with bacteriophage λ DNA (Mandel and Higa, 1970). This treatment induces a transient state of 'competence' in the bacteria, during which they are able to take up DNA. This is due to the addition of the divalent cation which can interact with the phosphate backbone of DNA and increase the efficiency of transformation by promoting the association between DNA and the cell membrane

of the bacteria (Hanahan, 1983). This method was subsequently used to transform bacteria with plasmid DNA (Cohen *et al.*, 1972). The products of ligation with pCR®-Blunt were transformed into competent One Shot™ TOP 10 (Invitrogen) cells.

### 3.2.6.1 Transformation Using One Shot™ Competent Cells (Invitrogen)

Transformation was performed essentially as recommended by the manufacturer.

1. The vial of SOC medium (supplied in the kit, see Appendix A7.1) was warmed to 37°C.
2. On ice, 50µl of One Shot™ TOP 10 cells was thawed for each ligation/transformation reaction.
3. Two microlitres of 0.5M β-mercaptoethanol was added to each vial of competent cells and mixed by stirring with a pipette tip.
4. Two microlitres of each ligation reaction was pipetted directly into the competent cells and mixed by stirring with a pipette tip.
5. The vials were incubated on ice for 30 minutes.
6. Vials were then incubated for 45 seconds in a 42°C water bath without mixing or shaking.
7. The vials were removed from the water bath and quickly placed on ice for 2 minutes.
8. Two hundred and fifty microlitres of the pre-warmed SOC medium was added to each vial.
9. The vials were incubated with shaking at 37°C for 1 hour.
10. Twenty microlitres, 100µl and the remainder of the mix (approximately 180µl) were spread over LB agar plates containing 50µg/ml kanamycin (see Appendix A7.1).
11. The plates were inverted and placed in a 37°C incubator overnight.
12. The number of colonies on each plate was counted and a small number selected for further analysis.

### 3.2.7 Screening Bacterial Colonies for pCR®-Blunt Transformants

#### 3.2.7.1 Preparing Miniprep Plasmid DNA

Extraction of plasmids from *E.coli* was performed essentially according to alkaline lysis protocols recommended by Promega. Sodium dodecyl suphate (SDS) solubilises the phospholipid and protein components of the cell membrane, leading to lysis and release of the cell contents. Sodium hydroxide (NaOH) denatures the chromosomal and plasmid DNAs, as well as proteins. The lysis time allows maximum release of plasmid DNA from the cell without release of chromosomal DNA. Lysis is neutralised by the addition of acidic potassium acetate. The plasmid DNA remains in solution whilst the SDS precipitates and the proteins, chromosomal DNA and cellular debris become trapped in salt-detergent complexes.

1. Selected colonies were inoculated into 5ml of LB broth and shaken overnight at 2,000 rpm at 37°C (in a Ratek orbital mixer incubator). One and a half millilitres of each overnight culture were placed into tubes and centrifuged at 12,000 rpm (in an MSE micro centaur centrifuge) for 1 minute. The remainder of the overnight cultures was stored at 4°C.
2. Supernatant were removed by aspiration, leaving bacterial pellets as dry as possible.
3. Pellets were resuspended by vortexing in 100µl of ice-cold resuspension buffer (25mM Tris-HCl, pH 8.0, 10mM EDTA, pH 8.0, and 50mM glucose).
4. The tubes were incubated for 5 minutes at room temperature.
5. Two hundred microlitres of a freshly prepared lysis solution containing 0.2M NaOH and 1% SDS was added to each tube. The contents of the tubes were mixed by inversion and incubated on ice for 5 minutes.
6. One hundred and fifty microlitres of ice-cold potassium acetate, pH 4.8 (3M potassium, 5M acetate) was added to each tube, the contents of the tubes mixed by inversion for 10 seconds, and then incubated on ice for 5 minutes.
7. The tubes were centrifuged at 12,000 rpm for 5 minutes.
8. Supernatants were transferred to fresh tubes, avoiding the white precipitate, which contained cellular material.
9. DNase-free RNase A was added to each tube to a final concentration of 20µg/ml.
10. The tubes were incubated at 37°C for 20 minutes.

11. One volume (400µl) of TE-saturated phenol/chloroform was added to each tube, the tubes inverted for 1 minute and centrifuged at 12,000 rpm for 5 minutes.
12. The upper, aqueous phase was added to fresh tubes and one volume (400µl) of chloroform:isoamyl alcohol (24:1) added. The tubes were inverted for 1 minute and centrifuged at 12,000 rpm for 5 minutes.
13. The upper, aqueous phase was added to fresh tubes and 2.5 volumes (1ml) of ethanol added. The tubes were mixed and the DNA precipitated by placing the tubes on dry ice 5 minutes or at -70°C for approximately 2 hours.
14. The tubes were centrifuged at 12,000 rpm for 10 minutes and the pellets rinsed with 70% ethanol before being centrifuged for another 10 minutes.
15. The pellets were dried by inverting the tubes briefly and then placing in a heating block set at 42°C.
16. Dried pellets were dissolved in 10µl of sterile water. A yield of 1µg to 4µg of DNA was expected from 1.5ml of culture.

**3.2.7.2 Restriction Endonuclease Analysis of Miniprep Plasmid DNA**

Plasmid samples were analysed by digesting with *Bam*H I, which releases the HPV 16 URR. The following were added together for each reaction: 12.3µl of water, 2µl of 10x buffer, 0.5µl of 20U/µl *Bam*H I (New England Biolabs), 0.2µl of 10mg/ml BSA, and 5µl of miniprep plasmid DNA. Tubes were incubated at 37°C for 1.5 hours and reactions stopped by adding 5µl of x5 gel loading dye. Twenty microlitres of each reaction was electrophoresed on a 1% Seakem™ LE agarose gel at 100V for 50 minutes. The samples were always electrophoresed with a positive control plasmid containing the HPV 16 prototype URR, pALuc16, and photographed under UV transillumination.

### 3.2.8 Reporter Genes

Reporter genes are powerful tools for studying gene function and cellular events and are frequently used as indicators of transcriptional activity in cells. The analysis of *cis*-acting transcriptional elements is a common application of reporter genes. It allows the functional characterisation of promoter and enhancer elements because the expression of the reporter is correlated with the transcriptional activity of the reporter gene. Specifically, the reporter gene is joined to a promoter sequence in an expression vector that is transferred into cells. The cells are analysed for the reporter by either measuring the amount of mRNA, reporter protein, or the enzymatic activity of the reporter protein. The most commonly used systems in eukaryotic cells include chloramphenicol acetyltransferase (CAT), β-galactosidase, firefly luciferase and alkaline phosphatase (AP). Luciferase, a monomeric protein that does not require post-translational processing for enzymatic activity, was used in this study. The luciferase assay also provides a rapid and inexpensive method for monitoring promoter activity (de Wet *et al.*, 1987).

The URR was subcloned upstream of the *luc* gene, which encodes the enzyme firefly luciferase. The first 40 base pairs of the *luc* gene were removed during the construction of the plasmid, removing the promoter sequences necessary for transcription. As such, the E6/E7 promoter at the 3' end of the URR (p97) acted as a substitute promoter, and luciferase activity as a substitute for E6/E7 expression. The luciferase gene is derived from the coding sequence of the *luc* gene cloned from the firefly *Photinus pyralis* and the bioluminescent reaction catalysed is the oxidation of beetle luciferin with concomitant production of a photon (Figure 3.2), which can be measured in a luminometer. The light intensity is a measure of the rate of catalysis by luciferase and is directly proportional to luciferase concentration in the sample and directly correlates with the transcriptional activity of the cloned regulatory element. The luciferase activity of each construct in this study was compared to that of the HPV 16 prototype, pALuc16.

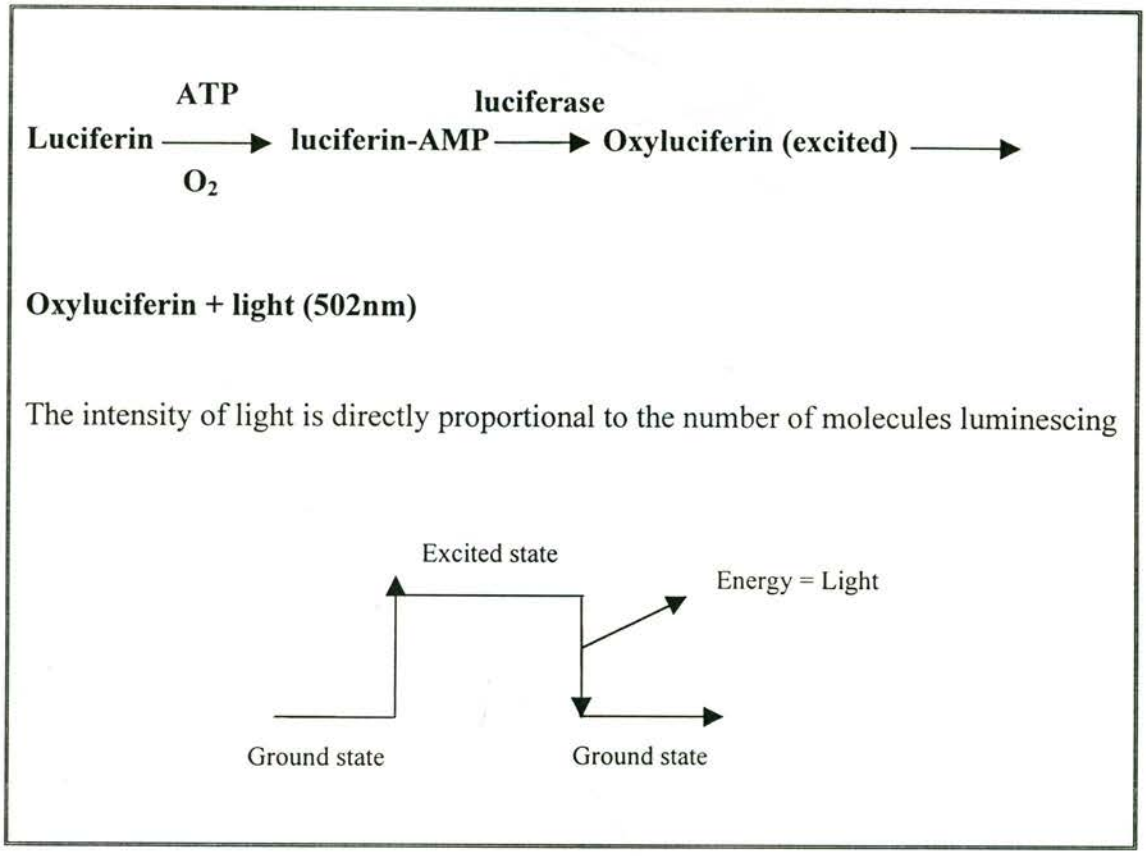



Figure 3.2: Creation of bioluminescence by luciferase

### 3.2.9 Subcloning Strategy

The HPV 16 URRs were released from pCR®-Blunt (pCR®-BluntH1, pCR®-BluntK2, pCR®-BluntK4, pCR®-BluntO2, pCR®-BluntR1, and pCR®-BluntS1) using *Bam*H I and subcloned into pALuc (Figure 3.3) to give pALucH1, pALucK2, pALucK4, pALucO2, pALucR1, and pALucS1.

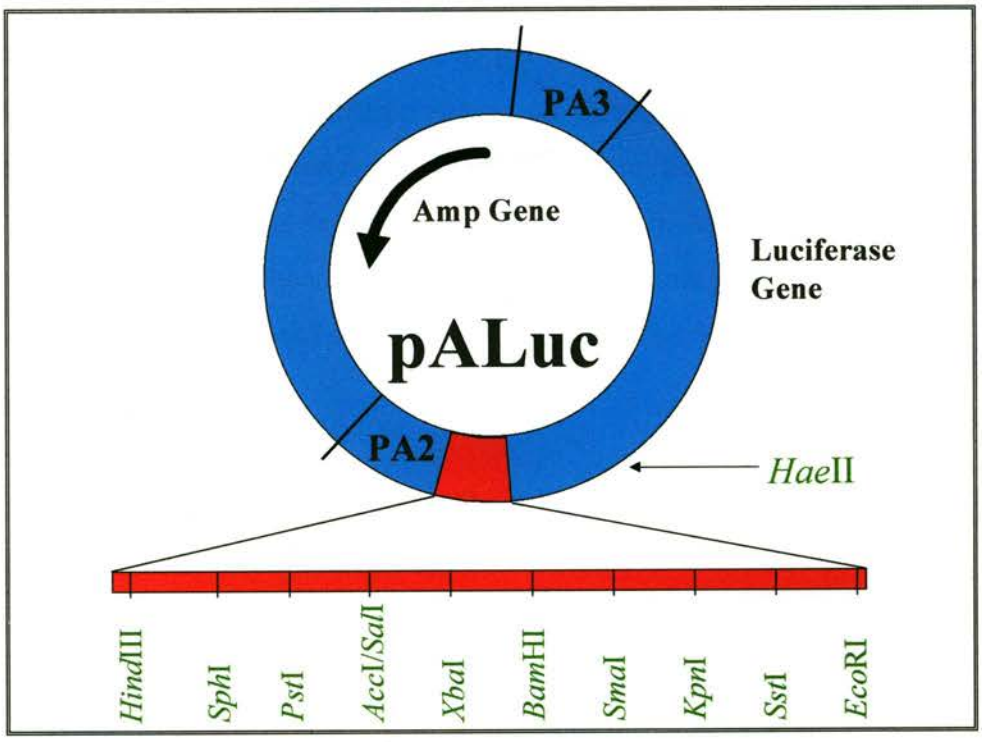



Figure 3.3: The cloning vector pALuc showing relevant genes and the multicloning site

The red area represents the multicloning site and restriction endonuclease sites found only in that region are shown; luc is the luciferase gene and Amp is the ampicillin gene.

### 3.2.9.1 Restriction Endonuclease Cleavage of pALuc

The following were added together and incubated at 37°C for 1 hour: 2.5µl of 2µg/µl pALuc, 38µl of water, 5µl of 10x buffer, 4µl of 20U/µl *Bam*H I, and 0.5µl of 10mg/ml BSA.

### 3.2.9.2 Dephosphorylation of pALuc

To minimise spontaneous reannealing, the sticky ends of the open pALuc were dephosphoylated.

1. One microlitre of 1U/µl sheep alkaline phosphatase (SAP) (Amersham) was added and the reaction incubated a further hour at 37°C.
2. One microlitre of 0.5M EDTA, pH 8.0 was added to the tube and heated at 65°C for 10 minutes to inactivate the phosphorylase.

To the dephosphorylated vector 12µl of x5 gel loading dye was added and the entire contents of the reaction mix electrophoresed on a 1% ultrapure agarose gel at 90V for 1 hour. The band of interest was excised from the gel and purified using QIAquick gel extraction (as explained in section 3.2.5.2). Alternatively, the digested product was purified by phenol extraction of 50µl of cutting mix (*Bam*H I digested and dephosphorylated pALuc) as described in section 2.2.2.1.

## 3.2.10 Preparation of Upstream Regulatory Regions for Subcloning into pALuc

Twenty microlitres of a bacterial suspension containing a pCR®-Blunt clone was added to 5ml of LB broth with 5µl of 50mg/ml kanamycin (Appendix A7.1), and incubated with shaking at 37°C overnight. To prepare for subcloning, miniprep plasmid DNA (as described in section 3.2.7.1), was digested with *Bam*H I by adding together: 31.5µl of water, 5µl of 10x buffer, 3µl of 20U/µl *Bam*H I, 0.5µl of 10mg/ml BSA, and 10µl (approximately 4µg) of miniprep plasmid DNA, and incubating at 37°C for 2 hours. The reactions were stopped by the addition of 10µl of x5 gel loading dye. The contents of the tubes were electrophoresed on a 1% the DNA purified by gel extraction (as described in section 3.2.5.2).

## 3.2.11 Subcloning

### 3.2.11.1 Ligation

Equimolar amounts of the vector and insert were added to ligase and buffer for each construct. For the individual HPV 16 URRs this comprised: 1µl of 10x ligation buffer, 2µl of *Bam*H I digested pALuc (approximately 0.3µg), 6.5µl of *Bam*H I digested HPV 16 URR, and 0.5µl of 4U/µl T4 ligase. A control consisting of vector only was also included and consisted of: 8µl of water, 1µl of 10x ligation buffer, 2µl of vector (pALuc, *Bam*H I cut), and 0.5µl of 3U/µl of T4 ligase. All ligation reactions were incubated overnight at 18°C.

## 3.2.12 Transformation

### 3.2.12.1 Preparation of XLI Blue Competent Cells for Transformation

pALuc constructs were transformed into XLI Blue *E.coli* cells (Sambrook *et al.*, 1989).

1. Five millilitres of LB broth was inoculated with 30µl of stock XLI blue cells and incubated at 37°C overnight with shaking at 180 rpm (in a Ratek orbital mixer incubator).
2. This culture was transferred to 50ml of LB broth and the bacteria allowed to grow at 37°C until the $OD_{600}$ was between 0.4 and 0.6 (approximately 2.5 hours).
3. After allowing the cells to stand on ice for 10 minutes, the culture was transferred to a 50ml polypropylene tube and centrifuged at 5,000 rpm (in a Beckman JA-14 fixed angle rotor) at 4°C for 12 minutes.
4. The supernatant was decanted and the tube quickly drained. The cells were gently resuspended in a 0.5 volume (25ml) of 50mM $CaCl_2$. The cells were allowed to stand on ice for 20 minutes.
5. The tube was centrifuged at 4,000 rpm at 4°C for 10 minutes. The supernatant was decanted and the cells gently resuspended in a 0.1 volume (5ml) of 50mM $CaCl_2$.
6. The cells were dispensed into 100µl volumes and left on ice for 30 to 60 minutes before transformation.

### 3.2.12.2 Transformation of XLI Blue *E.coli*

Two controls were processed along with the test; a control for the competency of cells (1ng of pCAT control plasmid [Promega]) and a vector only control.

1. The ligation mixes were added to 100µl aliquots of competent XLI Blue cells.
2. The tubes were allowed to stand on ice for 25 minutes, were placed in a heating block preheated to 42°C for exactly 90 seconds, and then transferred to ice for 2 minutes.
3. Four hundred microlitres of LB broth was added to each tube before incubating for 1 hour at 37°C.

4. For insert/vector and vector only controls, 30µl, 300µl and the remainder of the ligation mix (approximately 170µl) were spread evenly over LB plates containing 100µg/ml ampicillin. Similarly, 30µl of the competent cell control was delivered to a plate.
5. The plates were incubated overnight at 37°C.
6. The number of colonies on each plate were counted and the competency of cells estimated as follows: control plasmid = 1ng per 500µl of reaction mix, which is equal to 0.06µg of control plasmid per 30µl added to the plate. The competency of cells was measured using the formula: *x* colonies x 1000 / 0.06µg.

### 3.2.13 Screening Bacterial Colonies for pALuc Transformants

Since XLI cells are ampicillin sensitive, in theory all colonies on LB plates containing ampicillin should contain plasmids with the insert of interest. However, in practical terms, some colonies were expected to contain vector religated without the insert despite the use of SAP, particularly as a single cut restriction enzyme was used.

Several colonies of each sample were grown overnight with shaking at 37°C in 5ml volumes of LB broth plus 100µg/ml ampicillin. Constructs were extracted from bacteria and the plasmid DNA analysed by digesting with *Bam*H I (as described in sections 3.2.7.1 and 3.2.7.2). A positive control containing 1ng of pALuc16 was also included. Samples positive for a 1kb band on the gel, presumptively the inserted URR, were then digested with *Hae* II to determine if the URR was in the correct orientation.

There is only one *Hae* II restriction site in the HPV 16 URR at nt 7426, and several in the sequence of pALuc. It is possible to use *Hae* II to determine the orientation of clones by the expected sizes of digested product. All constructs produce fragments of 370bp and 2766bp. In addition, constructs containing the URR ligated in the correct orientation produce fragments of 701bp and 2388bp, and if in the incorrect orientation fragments of 514bp and 2575bp.

### 3.2.13.1 *Hae* II Digestion

Reactions were assembled for each construct containing the following: 13.3µl of water, 2µl of 10x buffer, 0.5µl of 20U/µl *Hae* II (New England Biolabs), 0.2µl of 10mg/ml BSA, and 4µl of miniprep plasmid DNA. Tubes were incubated at 37°C for 1.5 hours and reactions stopped by the addition of 5µl of x5 gel loading dye. Entire volumes were then electrophoresed on a 1% Seakem™ LE agarose gel for 1 hour at 100V. Clones containing the HPV 16 URR in the correct orientation were identified, preserved (as described in Appendix A7.2.4), and large-scale plasmid preparations made for further studies.

### 3.2.13.2 Large Scale Preparation of Bacterial Plasmids - QIAGEN Maxiprep Protocol (QIAGEN®)

QIAGEN plasmid purification protocols are based on modified alkaline lysis, followed by binding of plasmid DNA to QIAGEN anion-exchange resin under low salt and pH conditions. A medium-salt wash removes RNA, protein and other impurities. Plasmid DNA is eluted in a high-salt buffer and then concentrated and desalted by isopropanol precipitation. The QIAGEN maxiprep protocol was performed essentially to the instructions provided by the manufacturer.

1. Two hundred microlitres of a suspension of XLI Blue cells was inoculated into 1ml of LB broth containing 100ng/µl ampicillin and left shaking at 37°C for 5 hours.
2. This culture was added to 250ml of LB broth containing 100ng/µl ampicillin in a flask and left shaking for no longer than 16 hours at 37°C. The $OD_{600}$ of the LB broth was optimal between 1.5 and 2.5.
3. The turbid broth was poured into two 250ml centrifuge containers, which were spun at 10,000 rpm (using a Beckman JA-14 fixed angle rotor) for 12 minutes at 4°C and the resulting supernatant decanted.
4. The bacterial pellet was resuspended in 10ml of resuspension buffer P1 (50mM Tris-HCl, pH 8.0, 10mM EDTA, and 100µg/ml RNase). The mixture was transferred into two 50ml Falcon tubes.

5. To each Falcon tube 10ml of lysis buffer P2 (200mM NaOH and 1% SDS) was added. The tubes were mixed gently by inverting 4 to 6 times and then incubated at room temperature for no longer than 5 minutes.
6. 10ml of chilled neutralisation buffer P3 (3.0M potassium acetate, pH 5.5) was added and the tubes mixed by inverting 5 to 6 times, then incubated on ice for 15 to 20 minutes.
7. The tubes were inverted again and centrifuged at 13,000 rpm for 30 minutes at 4°C.
8. A QIAGEN tip was equilibrated by adding 10ml of equilibration buffer QBT (750mM NaCl, 50mM MOPS, pH 7.0, 15% ethanol and 0.15% Triton X-100) and allowing the column to empty by gravity flow.
9. The supernatant from step 7 was removed from the tube and added to the equilibrated QIAGEN tip.
10. The QIAGEN tip was washed with 30ml of wash buffer QC (1.0M NaCl, 50mM MOPS, pH 7.0 and 15% ethanol), then repeated.
11. The DNA was eluted by adding 15ml of elution buffer QF (1.25M NaCl, 50mM TrisHCl, pH 8.5, and 15% ethanol).
12. The DNA was precipitated with 0.7 volumes (10.5ml) of isopropanol (Sigma), equilibrated to room temperature, and centrifuged immediately at 10,000 rpm for 30 minutes at 4°C.
13. The isopropanol was carefully removed, the pellet washed with 5ml of ice cold 70% ethanol and the tubes centrifuged at the same temperature and speed for a further 15 minutes.
14. The ethanol was carefully removed and the pellet left to air dry, making sure that there were no further traces of ethanol.
15. The pellet was resuspended in 200µl of TE (10mM Tris and 0.1mM EDTA).
16. The concentration of the DNA was estimated on the spectrophotometer at $OD_{260nm}$ as described in Appendix A7.2.12, and then stored at -20°C.

### 3.2.14 pCH110 β-Galactosidase Vector

When measuring the effects of promoters and enhancers on gene expression, it is essential to include an internal control that will distinguish differences in the level of

promoter activity from differences in the efficiency of transfection or in the preparation of extracts. This is best achieved by co-transfecting cells with two plasmids – one that carries the construct under investigation and another that expresses an activity that can be assayed in the same cell extract. An enzyme frequently used for this is β-galactosidase, which is encoded by the *lacZ* gene in *E.coli* and catalyses the hydrolysis of β-galactoside sugars such as lactose to galactose and glucose. Extracts of mammalian cells contain low levels of endogenous β-galactosidase activity, and an increase of up to 100-fold can usually be detected during the course of a transfection (Sambrook *et al.,* 1989). β-galactosidase is an excellent reporter enzyme which can be assayed quickly and directly in cell extracts using a spectrophotometric assay. The pCH110 β-galactosidase expression vector (Amersham Pharmacia Biotech) was used in these experiments.

### 3.2.14.1 Preparation of pCH110

The β-galactosidase vector pCH110 (Figure 3.4) was transformed into *E.coli* and extracted in the same manner as the luciferase constructs (described in section 3.2.13.2).

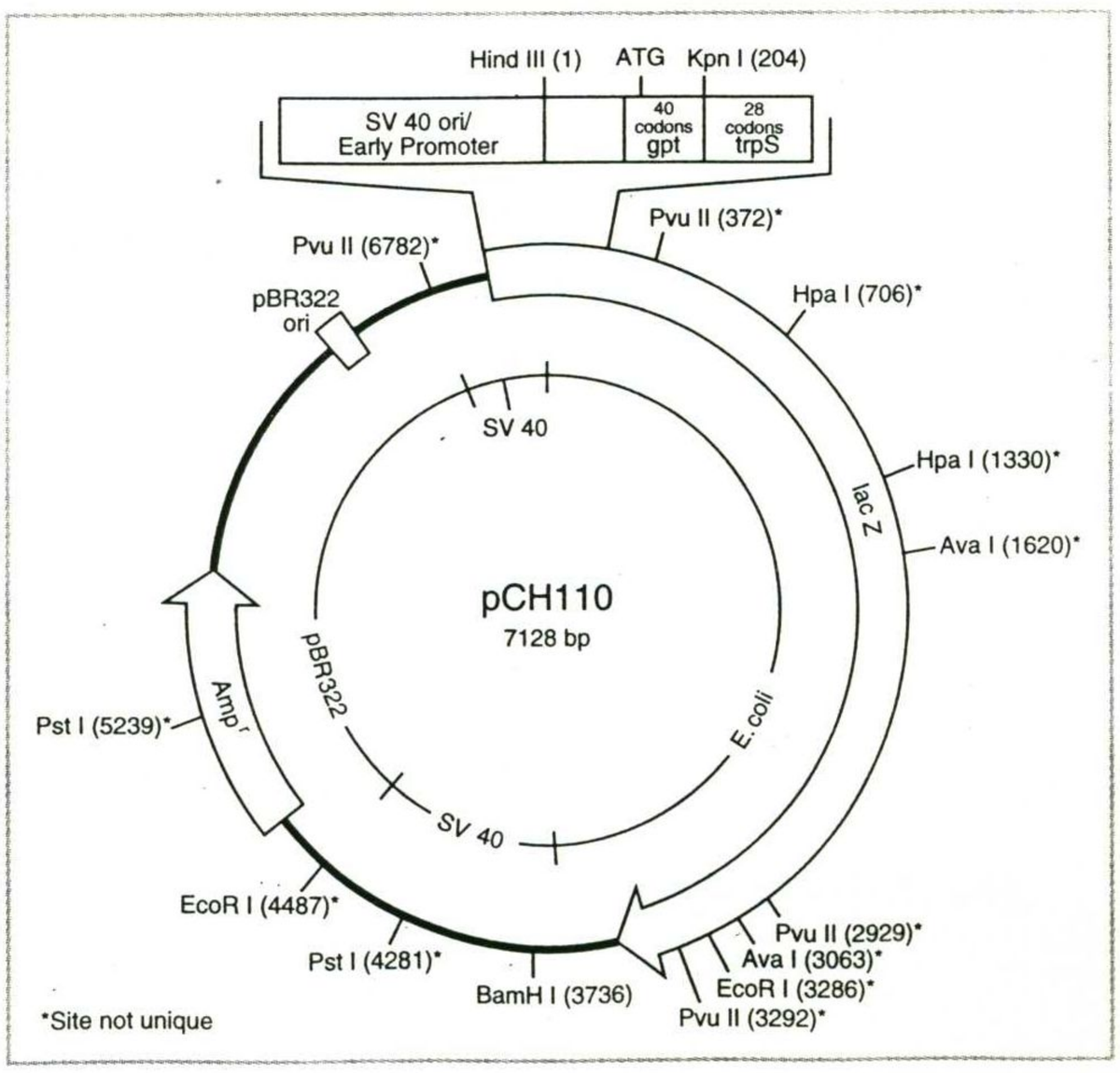



Figure 3.4: pCH110 β-galactosidase expression vector showing genes and restriction endonuclease sites (Amersham Pharmacia Biotech)

Amp = the ampicilline gene.

### 3.2.15 Transient Transfection of Mammalian Cells

#### 3.2.15.1 Strategy

Transfection is the process of introducing a segment of DNA into a eukaryotic cell. Different methods and reagents for transfection are available, for instance calcium phosphate coprecipitation, electroporation, DEAE-dextran (positively charged polymers), and cationic lipid reagents (liposomes). Calcium phosphate coprecipitation was the method selected for transient transfection experiments in this study as it worked well, was cost effective, and the method was well established in the laboratory. When nucleic acid is present as a calcium-phosphate coprecipitate, the uptake of DNA by cells in culture is markedly enhanced. The precipitant of calcium phosphate contains condensed DNA that adheres to cell membranes and enters the cytoplasm of cells by endocytosis. The DNA is then transferred to the nucleus. Increases in the efficiency of calcium phosphate coprecipitation have been achieved by incorporating additional steps such as glycerol shock following transfection. The success of transfection by calcium phosphate coprecipitation can be influenced by the volume and type of tube (eg. polypropylene), temperature, pH, speed of adding reagents, time of transfection, time of glycerol shock and percentage of $CO_2$ in the incubator.

In this study, transient transfections were performed in duplicate on cultures of human cervical carcinoma derived HeLa and HT3 cells. HeLa cells contain approximately 40 copies of HPV 18 (Schwarz *et al.*, 1985), whilst HT3 cells are HPV negative. Six micrograms of luciferase gene reporter plasmid and 1μg of pCMV β-galactosidase expression vector were used for each 6cm dish. Standard amounts of harvested cell lysates were used in assays to determine β-galactosidase activity and promoter activity was determined by calculating the luciferase/protein ratios adjusted to account for variations in transfection efficiency.

#### 3.2.15.2 Seeding Cells

Cell lines were maintained as described in Appendices A7.2.5 to A7.2.7. Transfections were performed in duplicate in 6cm dishes.

1. HT3 and/or HeLa cells were removed from flasks with trypsin (as outlined in Appendix A7.2.6) and transferred into sterile 10ml conical tubes.
2. The cells were washed with PBS and centrifuged for 5 minutes at 1,200 rpm (in an MSE large bench centrifuge). The supernatant was decanted and the cells resuspended.
3. The number of cells was estimated by counting in a Neubauer counting chamber (as described in Appendix A7.2.8).
4. DMEM plus 7% fetal calf serum (Appendix A7.1) was diluted to the required volume for the number of dishes to be seeded allowing $4x10^5$ cells in a volume of 4ml per dish, making sure the cell suspension was homogeneous (but without too much aeration that causes an undesirable rise in the pH of the medium).
5. Four millilitres of growth medium was dispensed into each dish, and the cells distributed evenly over the dish using backwards and forward motion.
6. The dishes were incubated in 5% $CO_2$ overnight.

### 3.2.15.3 Transient Transfection

Cells in dishes for transfection had to look healthy and be 60% to 70% confluent in order to proceed.

1. The growth media was changed 1 hour prior to transfection.
2. During this hour the transfection mix was prepared as follows. For duplicate tests 6μg of test plasmid, 1μg of pCH110, and water was added to a tube to a final volume of 450μl.
3. Forty five microlitres of 2.5M $CaCl_2$ was added to each tube just prior to each test. The tube was mixed once before dropping onto buffer. In a 14ml round bottom polypropylene Falcon tube, 450μl of x2 HBSS, pH 7.05 (280mM NaCl, 10mMKCl, 50mM HEPES, and 1.5mM $Na_2HPO_4$, pH 7.05) was added.
4. The $CaCl_2$/DNA mix was slowly dropped straight down on top of the HBSS and mixed once by flicking.
5. The tubes were left to stand for exactly 20 minutes at room temperature and then 300μl of the complexes were pipetted into each of the 6cm dishes of cells. The dishes were agitated gently by backwards and forward motion at least 16 times.
6. The dishes were incubated at 37°C with 5% $CO_2$ for 16 hours.

### 3.2.15.4 Glycerol Shock

Both the attrition rate of the cells, and the amount and quality of the precipitate were noted.

1. Depending on the number of dishes involved, 9ml of plain DMEM (no FCS or antibiotics) was pipetted into a 10ml tube, warmed to 37°C and when ready to proceed 1ml of sterile glycerol added and mixed gently.
2. Processing four dishes at a time, the medium was pipetted from each dish.
3. One millilitre of the glycerol/medium mix was added to each dish and allowed exactly 3 minutes contact with the cells, then quickly pipetted off.
4. The cells were gently washed with 5ml of warm PBS.
5. The PBS was removed and replaced with 4ml of medium containing 7% FCS and 100 IU/ml penicillin/40μg/ml gentamicin. The entire procedure had to take less than 10 minutes. The dishes were then incubated overnight at 37°C.

### 3.2.15.5 Harvesting Cellular Protein

The attrition rate of the cells was noted and PBS was cooled on ice. Reporter lysis x1 buffer was prepared by adding 4 volumes of water to 1 volume of reporter lysis x5 buffer (Promega).

1. The growth medium was removed from the cells. The cells were washed twice with PBS, with care taken not to dislodge any of the cells. As much of the final PBS wash as possible was removed.
2. Four hundred microlitres of reporter lysis x1 buffer was added to each 6cm dish. The dishes were rocked slowly several times to ensure the complete coverage of cells.
3. All areas of each dish were scraped using a 'rubber policeman', then each dish was tilted and the cell lysate scraped to the lower edge. The cell lysate from each dish was removed to a separate tube and placed on ice.
4. Tubes were freeze/thawed for two cycles (2 minutes each cycle) in a dry ice/ethanol bath and a 37°C water bath.
5. The tubes were then vortexed for 10 seconds, centrifuged for 2 minutes at room temperature and the supernatants transferred to new tubes.

6. Each extract was divided into two and one half stored immediately at -70°C.

### 3.2.15.6 Protein Estimation (Bio Rad Method)

The Bio Rad protein assay is a dye-binding assay in which a colour change in the dye occurs when different concentrations of protein are added. The absorbance maximum for an acidic solution of Coomassie® Brilliant Blue G-250 dye shifts from 465nm to 595nm when binding to protein occurs (Bio Rad). This dye binds to the basic and aromatic amino acid residues of proteins, especially arginine. The protein concentration was read as soon as possible after the cell protein was harvested.

1. The dye concentrate (kept at 4°C) was diluted with 4 volumes of sterile Liquipure water as required and filtered using Whatman no. 1 filter paper.
2. Diluted dye was aliquoted into 500µl volumes and 2.5µl of protein extract added. The tubes were mixed thoroughly and left to stand 10 minutes at room temperature.
3. Absorbances were read at 595nm in a spectrophotometer.
4. The following calculation was made: (30/OD) x 10, and the amount of protein standardised for the β-galactosidase assays, using approximately 30µl of protein extract.

### 3.2.15.7 β-Galactosidase Assay

To normalise luciferase activity to the β-galactosidase activity, the amount of protein in individual extracts was measured and the β-galactosidase assays then carried out using a standard amount of protein. The luciferase activity was then normalised to the β-galactosidase activity.

1. For 20 assays, o-nitrophenyl-B-D-galactopyranoside (ONPG) solution was made by adding 0.008g of ONPG to 2 ml of 100mM $Na_2HPO_4$, pH 7.5.
2. The cocktail was made as follows per tube: 200µl of 100mM $Na_2HPO_4$, 66µl of ONPG (as in step 1), and 3µl of Mg x100. The cocktail was dispensed into 270µl volumes.
3. Approximately 30µl of each extract was added to the cocktail (according to the previous calculation), made up to 30µl with water if necessary.

4. Tubes were incubated at 37°C for up to 2 hrs depending on the intensity of the colour. The reactions were stopped by adding 500µl of 1M $Na_2CO_3$ to each tube. Solutions were mixed well and the absorbances read using visible light at 420nm in the spectrophotometer (linear range 0.1 to 0.9).

### 3.2.15.8 Luciferase Assay

Luciferase assays were performed on harvested cell lysates as described by de Wet (1987). Findings were based on the results of at least three independent experiments.

1. Twenty microlitres of cell extracts were mixed with 100µl aliquots of room temperature luciferase assay reagent (Promega), one tube at a time.
2. Tubes were placed immediately into the luminometer (Fluoroskan Ascent FL, Labsystems) and the amount of light produced in 10 seconds measured.
3. The following calculations were made: Luciferase (raw data)/OD (protein) = R. R/OD (β-galactosidase) = final result.

### 3.2.16 Site-Directed Mutagenesis

Recombinant DNA technology has made it possible to generate mutations in cloned DNAs using a variety of chemical and enzymatic methods. PCR for instance, can be used to introduce restriction endonuclease sites or point mutations (Ausubel *et al.*, 1992). PCR-mediated site-directed mutagenesis was used in this study as a means of 'mutating' selected nucleotides to those found in the prototype in order to examine the effects of single base changes on promoter activity in the context of the entire URR from the particular variant. Figure 3.5 summarises the basic steps involved in site-directed mutagenesis.

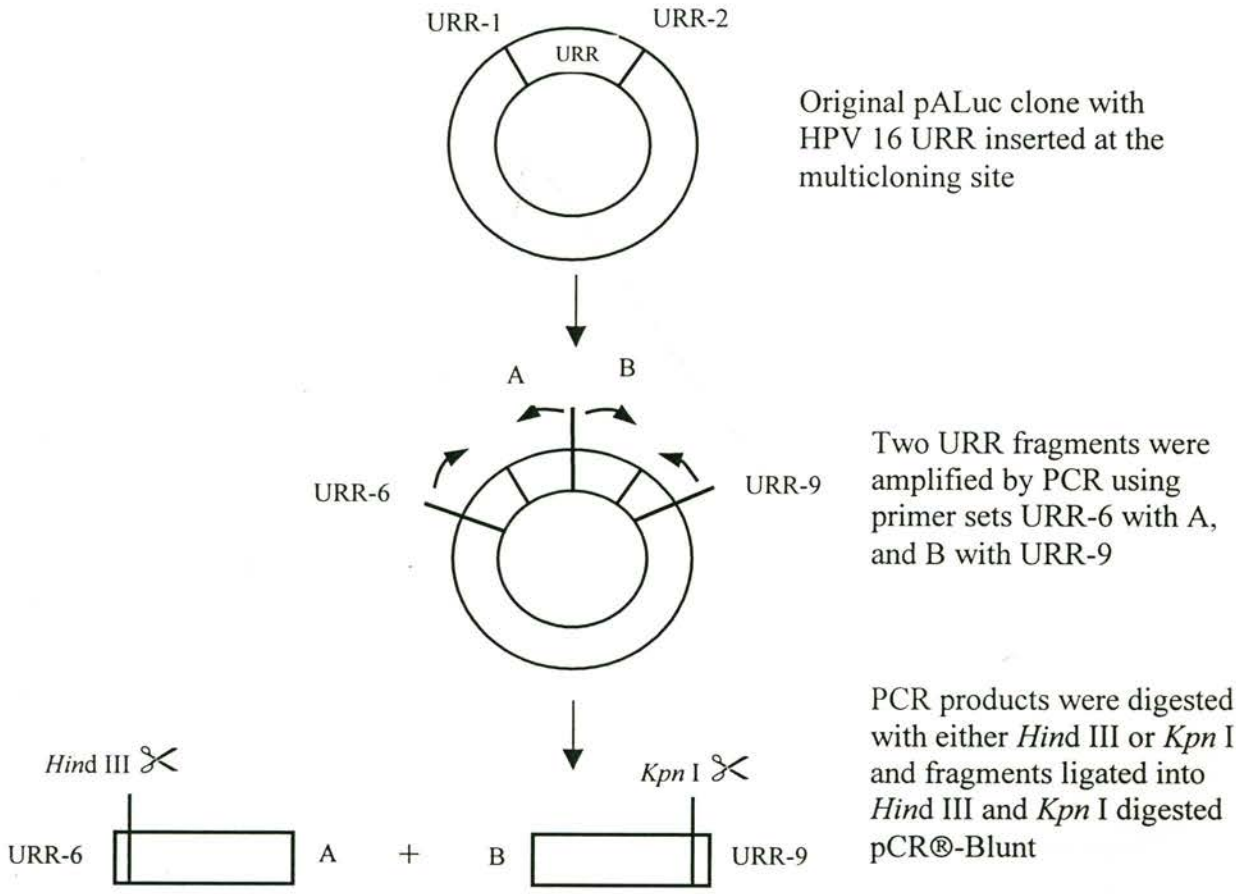



Figure 3.5: Introduction of a point mutation into the upstream regulatory region

Sites for two oligonucleotide primers, URR-6 and URR-9 (Table 3.2) flank the cloning site. In two separate PCRs, fragments upstream (URR-6/A) and downstream (B/URR-9) of the point mutation were amplified using the flanking primers (URR-6 or URR-9) and oligonucleotides containing the point mutation to be introduced (A and B – one of which contains the mutation to be introduced). PCR fragments were digested with *Hind* III or *Kpn* I, and then ligated and cloned into *Hind* III and *Kpn* I digested pCR®-Blunt. The URR was then subcloned into pALuc. The resultant plasmid contained an inserted fragment identical to the original inserted DNA except for the introduced point mutation.

### 3.2.17 Constructs Created by Site-Directed Mutagenesis

PCR-mediated site-directed mutagenesis was used to alter nucleotides of interest in the context of the URR from which they were identified. The resulting constructs and the changes induced are summarised in Table 3.1, whilst the protocol used for site-directed mutagenesis is given in Figure 3.6.

#### 3.2.17.1 Isolate K2

The original K2 URR was found to contain four nucleotide changes from prototype (refer to Figure 2.4). Two of these, at nts 7193 and 7521, are common in isolates worldwide, and were the only changes from prototype in R1. R1 did not have altered promoter activity, so these nucleotides were not investigated further. The two other variant nucleotides in K2, a C to A change at nt 7676 (in an overlapping Oct-1/PEF-1 binding site) and a C to T change at nt 7792 (in a YY1 binding site) were changed to prototype to create two new clones. One or both of these were hypothesised to be responsible for the increase in promoter activity observed for this isolate. These two nucleotide changes were also present in one other E lineage variant in the original cohort.

#### 3.2.17.2 Isolates K4 and H1

The K4 and H1 URRs were similar in sequence so it was interesting that promoter activities were upregulated by substantially different amounts. Both K4 and H1 contained many variant nucleotides when compared with the prototype. Several of these were common to both and were present in known TFBSs, including three variations in two GREs, and single changes in a TEF-1 and YY1 site. It was hypothesised that a change (C to T) at nt 7786 (in the YY1 binding site) increased the promoter activity of both K4 and H1 and this nucleotide was changed back to prototype to create new constructs for both isolates. It was further hypothesised that a nucleotide change (A to C) at nt 7894 (in the PSM) may have led to a further increase in promoter activity in K4, and that a TEF-1 variation in H1 (T to G) at nt 7743 may have led to decreased promoter activity. These nucleotides were changed back to prototype in the respective URRS to create new constructs. The nt 7786 change was present only in the two AA lineage variants in the cohort, whilst the nt 7894 and nt 7743 changes were present only in the isolates described.

Table 3.1: Nucleotides Altered by Site-Directed Mutagenesis and Resulting Constructs

| Isolate | Nucleotide Altered | TFBS/s Affected | Name of Construct |
|---|---|---|---|
| K2 | None | - | pALucK2 |
| | 7676 (C→A) | Oct-1/Pef-1 | pALuc16K2Oct-1P |
| | 7792 (C→T) | YY1 5 | pALuc16K2YY1P |
| K4 | None | - | pALucK4 |
| | 7786 (C→T) | YY1 4 | pALuc16K4YY1P |
| | 7894 (A→C) | PSM | pALuc16K4PSMP |
| H1 | None | - | pALucH1 |
| | 7786 (C→T) | YY1 4 | pALuc16H1YY1P |
| | 7743 (T→G) | TEF-1 | pALuc16H1TEF-1P |

P = Prototype TFBS Sequence

- Religate pCR®-Blunt (which was supplied from the manufacturer as a linearised blunt-ended plasmid)
- Transform into XLI Blue *E.coli* cells
- Grow up large quantities of pCR®-Blunt in *E.coli* and isolate
- Digest pCR®-Blunt with *Hin*d III and *Kpn* I
- Ligate digested pCR®-Blunt with PCR products from Figure 3.5
- Transform plasmid into TOP 10 *E.coli* cells
- Check for positive clones by PCR using primers URR-3 and URR-4
- Release inserted URR with induced point mutation using *Bam*H I (located on ends of original PCR primers URR-1 and URR-2)
- Ligate URR with point mutation to *Bam*H I digested pALuc
- Transform into XLI Blue *E.coli* cells
- Check for positive clones by PCR using primers URR-3 and URR-4
- Grow clone up in *E.coli*, isolate and sequence to check integrity
- Perform transient transfection and luciferase assays

Figure 3.6: Flow diagram protocol for site-directed mutagenesis

## 3.2.18 Polymerase Chain Reaction for Site-Directed Mutagenesis

### 3.2.18.1 Polymerase Chain Reaction Components

PCR reaction components were the same as listed in Chapter 2. The primers used for site-directed mutagenesis PCRs are given in Table 3.2.

Table 3.2: Polymerase Chain Reaction Primers Used for Site-Directed Mutagenesis

| Primers | Genome Positions | Nucleotide Sequence (5'-3') | Product Size (bp) | TFBS/s Affected |
|---|---|---|---|---|
| URR-6<br>URR-7 | pALuc5025-pALuc 5044<br>7664-7683 | TCCCCCTGAACCTGAAACAT<br>CGTTGGCGCATAGTGATTTA | 870 | Oct-1/ Pef-1 |
| URR-8<br>URR-9 | 7684-7703<br>pALuc92-pALuc109 | CCTTACATACCGCTGTTAGG<br>GTTTTTGGCGTCTTCCAT | 406 | |
| URR-6<br>URR-10 | pALuc5025-pALuc 5044<br>7764-7784 | TCCCCCTGAACCTGAAACAT<br>GTGACATTTAGTTGGCCTTAG | 971 | YY1 5 |
| URR-11<br>URR-9 | 7785-7806<br>pALuc92-pALuc109 | CCTAGTTCATACATGAACTGTG<br>GTTTTTGGCGTCTTCCAT | 305 | |
| URR-6<br>URR-12 | pALuc5025-pALuc 5044<br>7768-7792 | TCCCCCTGAACCTGAAACAT<br>AAACTAGGGTGACATTTAGTTGGCC | 968 | YY1 4 |
| URR-13<br>URR-9 | 7793-7818<br>pALuc92-pALuc109 | ATACATGAACTGTGTAAAGGTTAGTC<br>GTTTTTGGCGTCTTCCAT | 296 | |
| URR-6<br>URR-14 | pALuc5025-pALuc 5044<br>7848-7867 | TCCCCCTGAACCTGAAACAT<br>AAAACGGTTTGCACACACCC | 1024 | PSM |
| URR-15<br>URR-9 | 7868-7898<br>pALuc92-pALuc109 | GGGTTACACATTTACAAGGAACTTAT ATAAT<br>GTTTTTGGCGTCTTCCAT | 221 | |
| URR-6<br>URR-16 | pALuc5025-pALuc 5044<br>7693-7725 | TCCCCCTGAACCTGAAACAT<br>TAAAACAAGCCAAAAATATGTGCCTA ACAGCGG | 901 | TEF-1 |
| URR-17<br>URR-9 | 7726-7748<br>pALuc92-pALuc109 | ACTCACCTAATTGCATATTTGGC<br>GTTTTTGGCGTCTTCCAT | 363 | |

### 3.2.18.2 Phosphorylation of Primers

Phosphorylating primers on the 5' end was necessary for blunt-end cloning. Primers URR-7, -8, -10, -11, -12, -13, -14, -15, -16 and -17 were phosphorylated prior to use by adding the following per tube: 5µl of 1µg/µl primer, 1µl of 10mM ATP (Boehringer Mannheim), 1µl of 10x kinase buffer (Boehringer Mannheim), 1µl of 10 U/µl T4 polynucleotide kinase, and 2µl of water. Tubes were incubated at 37°C for 1 hour before the T4 polynucleotide kinase was inactivated by heating to 65°C for 10 minutes.

### 3.2.18.3 Optimised Polymerase Chain Reaction Protocol

To determine the optimal conditions for specific amplification, $MgCl_2$ concentration (from 1.5mM to 3.5mM) was titrated against annealing temperature (from 56°C to 62°C). PCRs were then carried out as described in Table 3.3. All products were amplified using the same protocol, but with different annealing temperatures. Products were electrophoresed on a 1% ultrapure agarose gel at 100V for 1 hour. The bands of interest were then cut from the gel and the gel slices purified as described in section 3.2.5.2. Two microlitres of each purified product was then electrophoresed at 120V for 30 minutes to roughly quantitate. Precautions to prevent PCR cross-contamination were as described in section 2.2.3.2.

Table 3.3: Polymerase Chain reaction Cocktail and Thermal Cycling Conditions for Site-Directed Mutagenesis

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 68.25µl | - |
| 10x Reaction Buffer | 10µl | 1x |
| 25mM $MgCl_2$ | 8µl | 2mM |
| Primer* (300ng/µl) | 1µl | ~0.5µM |
| Primer* (300ng/µl) | 1µl | ~0.5µM |
| dNTPs | 1µl | 250µM |
| *Amplitaq*® DNA polymerase | 0.75µl | 3.75 Units |
| | ____ | |
| | 90µl PCR Cocktail + 10µl Target Plasmid (1ng/µl) | |

Thermal Cycling Conditions:

| | |
|---|---|
| Cycle 1 | Step 1 94°C/1 minute |
| Cycles 2-31 | Step 1 94°C/1 minute |
| | Step 2 59-61°C/1 minute |
| | Step 3 72°C/1 minute |
| Cycle 32 | Step 1 25°C/1 minute |

* = Primer combinations as Table 3.2.

### 3.2.18.4 Restriction Endonuclease Digestion of Products

Purified PCR fragments were digested with the appropriate restriction endonuclease, either *Hin*d III (products from PCR reactions using primer sets containing primer URR-6), or *Kpn* I (products from PCR reactions using primer sets containing primer URR-9). For digestion with *Hin*d III the following were combined: 19µl of water, 5µl of 10x buffer, 1µl of 20U/µl *Hin*d III, and 25µl of PCR product. For digestion with *Kpn* I the following were combined: 17µl of water, 5µl of 10x buffer, 3µl of 8U/µl *Kpn* I, and 25µl of PCR product. Tubes were incubated at 37°C for 2 hours and the reactions stoped by adding 10µl of x5 gel loading dye. Products were electrophoresed and gel purified (as described in section 3.2.5.2). Pellets were resuspended in 30µl of TE (10mM Tris and 0.5mM EDTA, pH 7.6).

## 3.2.19 Preparation of pCR®-Blunt for Site-Directed Mutagenesis

For the purposes of these site-directed mutagenesis experiments, a cloning vector without blunt ends was required as it was necessary to have blunt-ended PCR products that would be ligated at the centre of each clone to induce the desired mutation (see Figure 3.5). If a blunt-ended cloning vector had been used in addition, this would have created too many blunt ends, and a high likelihood of the incorrect combination of molecules. As the pALuc primers (URR-6 and URR-9) extended the PCR products from the cloned URR into the multicloning site of pALuc, the end of each cloned PCR product contained several restriction sites. *Hin*d III and *Kpn* I were the restriction sites chosen for cloning back into pCR®-Blunt, as they were present in the multicloning site of pCR®-Blunt and pALuc. In order to digest pCR®-Blunt with these enzymes, religation of the vector was first necessary. pCR®-Blunt was religated in a reaction containing: 1µl of 25ng/µl pCR®-Blunt, 1µl of 10x ligation buffer, 1µl of 4U/µl T4 ligase, and 7µl of water. This reaction was incubated at 16°C for 1 hour before transformation into XLI Blue cells. A colony from an LB plate was selected and grown in 3ml of LB broth with 50µg/ml kanamycin for at least 5 hours. Then 30µl of culture was added to 20ml of LB broth with 50µg/ml kanamycin. The plasmid was isolated from an overnight culture by plasmid mini purification (using the QIAGEN® method that produces a cleaner product). The vector was digested

with *Hind* III and *Kpn* I, gel purified (section 3.2.5.2) and religated with the PCR fragments containing the point mutation of interest.

### 3.2.19.1 Plasmid Mini Purification Protocol (QIAGEN®)

Buffer components were the same as listed in section 3.2.13.2 for the Maxiprep protocol.

1. The bacterial cells from the overnight culture were harvested by spinning at 10,000 rpm (using a Beckman JA-14 fixed angle rotor).
2. The bacterial pellet was resuspended in 2ml of Buffer 1 (with RNase A added), leaving no cell clumps.
3. Two millilitres of Buffer P2 was added, mixed gently by inverting the tube 4 to 6 times, and incubated at room temperature for 5 minutes.
4. Two millilitres of chilled Buffer P3 was added and mixed immediately but gently by inverting 4 to 6 times, and incubated on ice for 5 minutes.
5. The tube was mixed again and then centrifuged at 14,000 rpm for 10 minutes. The supernatant was removed promptly.
6. Two QIAGEN-tip 20's were equilibrated by adding 1ml of Buffer QBT, and the columns allowed to empty by gravity flow.
7. The supernatant from step 5 was divided into two and added to the columns, which were allowed to empty by gravity flow.
8. The QIAGEN tips were washed four times with 2ml of Buffer QC.
9. The DNA was eluted with 1.6ml of Buffer QF.
10. The DNA was precipitated with 1.12ml of room temperature isopropanol and the tubes centrifuged immediately at 14,000 rpm for 30 minutes. The supernatant was carefully removed.
11. The DNA pellet was washed with 2ml of 70% ethanol, and then air-dried, before being redissolved in a suitable volume of TE (10mM Tris and 0.5mM EDTA, pH 7.6).

### 3.2.19.2 Ligation

As explained in section 3.2.5.3, a ratio of at least 10:1 molar insert-to-vector is required for efficient blunt-ended cloning. A slightly higher ratio was used for these experiments as three DNA segments were ligated together. The following

components were added together and incubated at 16°C for at least 1 hour or overnight: 3.5µl of each PCR product, 1µl of 10x ligation buffer, 1µl of 25ng/µl pCR®-Blunt (*Hin*d III and *Kpn* I digested), and 1µl of 4U/µl T4 ligase. The total volume of each reaction was used for transformation into TOP 10 cells.

### 3.2.20 Screening Bacterial Colonies for pCR®-Blunt Transformants

In order to determine if isolated plasmids resulting from transformation contained the insert of interest, PCR was performed using primers URR-3 and URR-4 (as described in section 2.2.7). As only clones containing the correctly aligned PCR fragments should amplify, a positive band of the correct size indicated that cloning was successful. The target used for this PCR protocol consisted of 5µl from 30µl of broth culture heated to 95°C for 5 minutes. Clones positive by PCR for HPV 16 insert were reconfirmed using the standard plasmid isolation procedure (section 3.2.7.1) and restriction endonuclease digestion (*Hin*d III and *Kpn* I). Bacteria containing clones were always preserved (Appendix A7.2.4). Constructs were then extracted (section 3.2.7.1), digested with both *Hin*d III and *Kpn* I, and gel purified (section 3.2.5.2) for subcloning.

### 3.2.21 Subcloning Upstream Regulatory Regions into pALuc

Subcloning reactions were set up as follows using equimolar vector-to-insert ratios: 2µl (approximately 0.3µg) of pALuc (*Hin*d III and *Kpn* I digested), 6.5µl of URR (*Hin*d III and *Kpn* I digested), 1µl of T4 10x buffer, and 0.5µl of 4U/µl T4 ligase. Tubes were incubated overnight at 18°C. Analysis of subclones was by miniprep (section 3.2.7.1) and *Bam*H I digestion (section 3.2.7.2).

### 3.2.22 Polymerase Chain Reaction for Mycoplasma

Mycoplasma belong to the class *Mollicutes*, and are major cell culture contaminants. Mycoplasma contamination of cell cultures results in very poor transfection efficiencies since mycoplasma attach themselves to the outside of cells preventing DNA uptake. As problems had been encountered over recent years in the department

with mycoplasma contamination, all cell cultures were routinely screened at regular intervals. Mycoplasma affect cell growth and morphology, can alter amino acid and nucleic acid metabolism, and immunological and biochemical properties, leading to unreliable experimental results (Dussurget and Roulland-Dussoix, 1994). Contamination can come from a variety of animal and human sources, but most commonly from other infected cultures or from laboratory personnel. They may also originate from products such as FCS. As mycoplasma are very small, they can pass through most filters used for cell culture.

Five mycoplasma species make up 95% of contaminations (Dussurget and Roulland-Dussoix, 1994). These include *Mycoplasma arginini*, *Acholeplasma laidlawii*, *Mycoplasma hyorhinis*, *Mycoplasma orale*, and *Mycoplasma fermentans*. As mycoplasma cannot be detected by visual inspection or light microscopy, a PCR-based detection method was used to detect any contaminated cultures when transfection efficiency was low or the cells did not appear healthy.

The cell extract used as target for this PCR was obtained by performing a standard phenol extraction (as described in section 2.2.2.1) using approximately one to two million cells, or by adding cells to lysis buffer (50mM Tris-HCl, pH 8.5, 1mM EDTA, pH 8.0, and 0.5% Tween 20, with 400µg/ml proteinase K) and incubating at 37°C for several hours. Thirty microlitres of lysate was then heated to 95°C for at least 5 minutes before adding 5µl as target to the PCR cocktail.

The genus-specific mycoplasma primers used (Table 3.4) are located in the mycoplasma 16S rRNA gene, and produce a single 717bp band for the five common mycoplasma contaminants, using the protocol in Table 3.5.

Table 3.4: Mycoplasma Polymerase Chain Reaction Primers

| Primer | Sequence (5'-3') |
|---|---|
| GPO-1* | ACT CCT ACG GGA GGC AGC AGT A |
| MGSO* | TGC ACC ATC TGT CAC TCT GTT AAC CTC |

*Taken from Dussurget and Roulland-Dussoix (1994)

Table 3.5: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Detection of Mycoplasma

| PCR Reaction: | Volume Added: | Final Concentration/ Amount: |
|---|---|---|
| Water | 30.1µl | - |
| 10x reaction buffer | 5µl | 1x |
| $MgCl_2$ (25mM) | 5µl | 2.5mM |
| GPO-1 (20pmol/µl) | 2µl | 0.8µM |
| MGSO (20pmol/µl) | 2µl | 0.8µM |
| dNTPs (25mM) | 0.4µl | 200µM |
| *Amplitaq Gold*™ DNA polymerase | 0.5µl | 2.5 Units |
| | 45µl + 5µl of 100ng/µl DNA target (500ng total). | |

Thermal Cycling conditions:

| | |
|---|---|
| Cycle 1 | Step 1 95°C/1 minute |
| Cycle 2 | Step 1 95°C/1 minute |
| | Step 2 57°C/1 minute |
| | Step 3 72°C/1 minute |
| Cycle 37 | Step 1 72°C/7 minutes |
| | Step 2 25°C/5 minutes |

Five microlitres of PCR product was electrophoresed on a 2% agarose gel to assess the mycoplasma status of the cell line. Results were compared with a positive control consisting of amplified DNA from a previously infected cell line if available. Positive cultures were treated for 7 to 10 days with 0.5µg/ml to 1µg/ml of Mycoplasma Removal Agent (ICN) in DMEM, and later retested to ensure that the treatment was effective. Mycoplasma Removal Agent contains a chemical derivative of the quinoline family of antibiotics and works by inhibiting Mycoplasma DNA gyrase. This method was successful after one or two treatments.

### 3.2.23 Electrophoretic Mobility Shift Assays (EMSAs)

The electrophoretic mobility shift assay (EMSA), or gel shift is a simple method for identifying and investigating sequence-specific DNA:protein binding. Important biological processes, such as the development of an adult organism from a single-celled zygote, and the maintenance of the developed structure and biochemical characteristics of tissues, are dependent on certain DNA binding TFs in different tissues. The spectrum of TFs present determines what genes may be expressed in a cell type. The expression of a single gene may be determined by a number of different TFs and the ratio of their concentrations may be vital, particularly if competing positive and negative factors are involved (Dent and Latchman, 1993). This is certainly the case with the transcription of the HPV E6 and E7 genes. Factors involved in transcription and DNA replication have been identified in all cell types examined by researchers and the EMSA is a powerful tool for the detection of these factors. EMSAs can be applied to the identification of both known and novel factors binding to a candidate DNA fragment and to the identification of variations in levels of known TFs in response to stimuli, eg. growth factors.

In an EMSA, radioactively labelled double-stranded oligonucleotides are incubated with cellular protein extracts or purified proteins. These complexes are then resolved on polyacrylamide gels. The DNA:protein complexes migrate more slowly than non-bound DNA (Figure 3.7). This method can be used to screen cellular extracts (such as HeLa nuclear extract), or *in vitro* expression systems for proteins binding to a known sequence, or as in this study, to examine the effect of sequence variation on DNA:protein binding. To minimise nonspecific DNA binding in crude nuclear extracts, the addition of a carrier DNA, such as poly (dI-dC) is necessary.

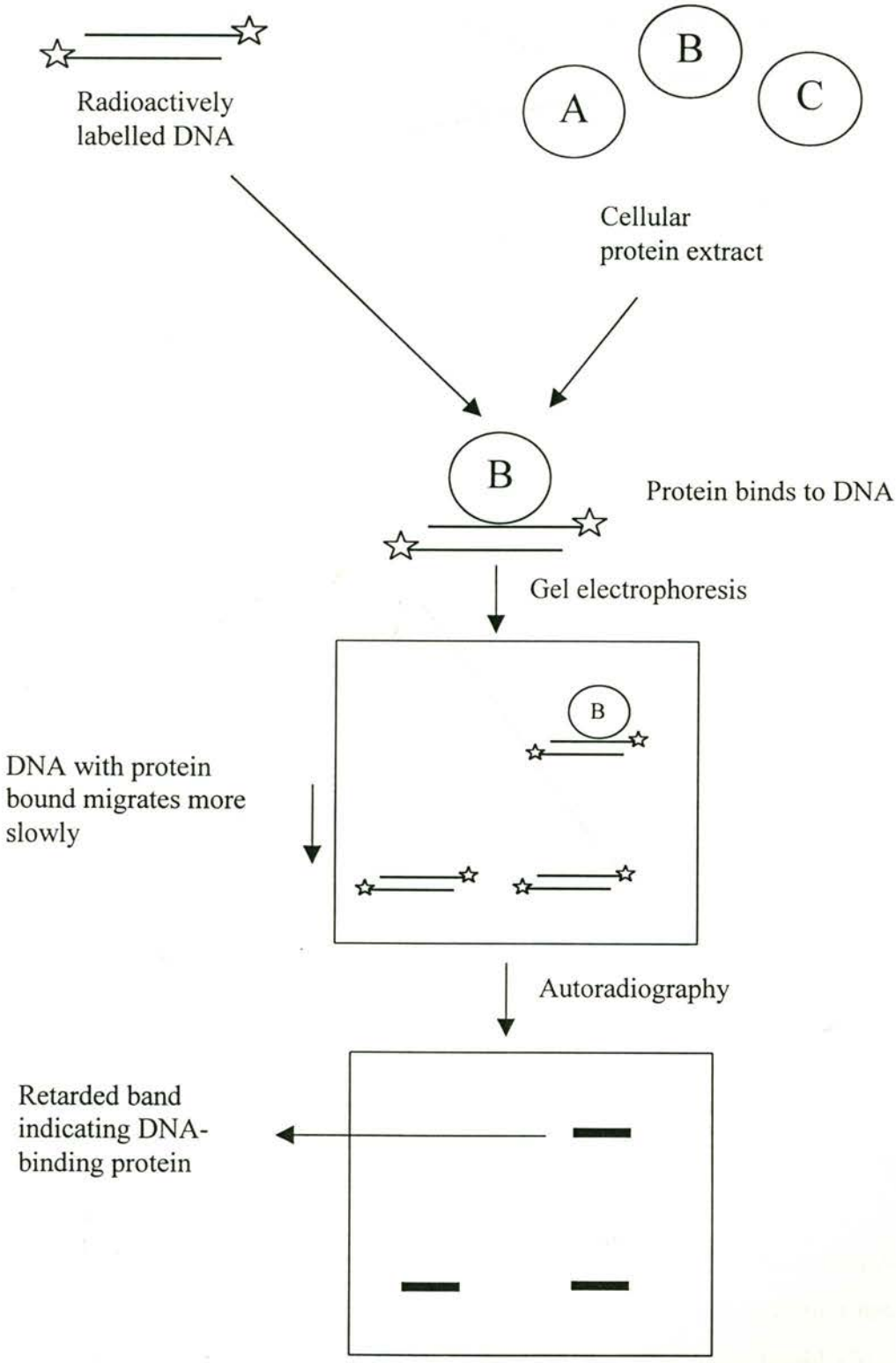



Figure 3.7: The electrophoretic mobility shift assay

(Adapted from Dent and Latchman, 1993)

The binding of cellular protein (B) to a radioactively labelled DNA causes it to move more slowly upon gel electrophoresis and results in the appearance of a retarded band upon autoradiography.

### 3.2.24 Strategy For Electrophoretic Mobility Shift Assays

EMSAs were performed to determine whether increased promoter activity associated with K2 (C to A at nt 7676 in an overlapping Oct-1/PEF-1 site and C to T at nt 7792 in YY1 5) and in K4 (A to C at nt 7894 in the PSM) resulted from altered DNA binding of associated TFs. In addition, a sequence change at nt 24 (C to T) found in six of the isolates in the original cohort including S1 and O2, located adjacent to an Sp1 site, was also examined to determine if binding of Sp1 was altered. HeLa nuclear extracts were used for these studies, as they have been used in other studies on the HPV URR.

#### 3.2.24.1 Preparation of HeLa Nuclear Extract

1. One million HeLa cells from a subconfluent culture were collected in 10ml of tris buffered saline (TBS). The cell suspension was centrifuged at 1,200 rpm (in an MSE large centrifuge) for 5 minutes. The pellet was resuspended in 1ml of TBS and transferred to a 1.5ml tube. The tube was microfuged briefly to pellet the cells.
2. The TBS was removed and the pellet was resuspended in 500µl of ice cold extraction buffer (20mM HEPES, pH 7.8, 10mM KCl, 0.1mM EDTA, pH 8.0, 1mM DTT, and 0.5mM PMSF) by gentle pipetting.
3. The tube was left on ice for 15 minutes, then 30µl of 10% nonidet NP-40 was added and the tube vortexed for 10 seconds.
4. The tube was microfuged for 30 seconds at 4°C (in a Sigma 1K15 refrigerated bench centrifuge) and the supernatant discarded. The pellet was resuspended in 50µl of extraction buffer with 400mM NaCl added. The tube was agitated vigorously for 15 minutes at 4°C.
5. The tube was microfuged at 13,000 rpm for 5 minutes at 4°C and the supernatant containing the nuclear extract was removed, the protein concentration measured (as described in section 3.2.15.6), and stored at -70°C or in liquid nitrogen in 10µl or 15µl aliquots.

### 3.2.24.2 Preparation of 5% Acrylamide Gel in Low Ionic x0.5 TBE Buffer

1. In a beaker, 5.6ml of 40% acylamide (Astral Scientific), 5ml of x5 TBE, 39.4ml of water, 200µl of 10% ammonium persulphate (APS) (Sigma), and 100µl of TEMED (Sigma), were added in that order. The solution was mixed gently and the gel poured immediately between glass plates with a 1.5mm spacer. A 20 well, well former was added and the gel left to polymerise for 1 hour.
2. The bottom spacer and well former were removed and the plates rinsed under a tap. The gel was clamped into a vertical gel electrophoresis unit (Life Technologies). The upper and lower chambers were filled with x0.5 TBE. The wells were rinsed out with a small amount of x0.5 TBE and the wells marked and numbered.
3. Three microlitres of x5 gel-loading dye was added to the end well and the gel pre-run for 20 minutes at 160V to equilibrate.

### 3.2.24.3 Oligonucleotides

The sequences of the oligonucleotides used for EMSAs are given in Table 3.6.

Table 3.6: Oligonucleotides Used for Electrophoretic Mobility Shift Assays

| Oligonucleotide | Nucleotide Sequence (5'-3') | Genome Positions (5'-3') | TFBS |
|---|---|---|---|
| AAVYY1<br>AAVYY1R | *CTAGA*TTGCGACATTTTGCG<br>*CTAGA*CGCAAAATGTCGCAA | - | AAV YY1 |
| YY1P* | *CTAGA*TTCATACATGAACTG | 7790-7804 | YY1 |
| YY1PR | *CTAGA*CAGTTCATGTATGAA | | |
| YY1M | *CTAGA*TT**T**ATACATGAACTG | | |
| YY1MR | *CTAGA*CAGTTCATGTATA**A**A | | |
| OCT1P^ | ATCACTATGCGCCAACGCCTTACATAC | 7667-7693 | Oct-1/Pef-1 |
| OCT1PR | GTATGTAAGGCGTTGGCGCATAGTGAT | | |
| OCT1M | ATCACTATG**A**GCCAACGCCTTACATAC | | |
| OCT1MR | GTATGTAAGGCGTTGGC**T**CATAGTGAT | | |
| PSMP# | *AGATCT*TATAATAATACTAAACTACAATAATTCAT | 7893-15 | PSM |
| PSMPR | *AGATCT*ATGAATTATTGTAGTTTAGTATTATTATA | | |
| PSMM | *AGATCT*T**C**TAATAATACTAAACTACAATAATTCAT | | |
| PSMMR | *AGATCT*ATGAATTATTGTAGTTTAGTATTATTA**GA** | | |
| Sp1P | *AAGCTT*AAAACTAAGGGCGTAACCG | 20-38 | Sp1 |
| Sp1PR | *AAGCTT*CGGTTACGCCCTTAGTTTT | | |
| Sp1M | *AAGCTT*AAAA**T**TAAGGGCGTAACCG | | |
| Sp1MR | *AAGCTT*CGGTTACGCCCTTA**A**TTTT | | |

P = prototype HPV 16 sequence; M = mutated HPV 16 sequence; R = reverse oligonucleotide

*Oligonucleotides published by O'Connor *et al.*, 1996

^ Oligonucleotides published by Sibbet *et al.*, 1995

#Oligonucleotides published by O'Connor *et al.*, 1998

Italics represent restriction endonuclease sites (*Xba* I for YY1, *Bgl* II for PSM, and *Hin*d III for Sp1)

Bold sequence indicates altered nucleotides

### 3.2.24.4 Labelling Oligonucleotides with dATP $\gamma$-$^{32}$P

All work involving radiation was performed behind a perspex screen. Gloves, a gown and a radiation monitoring badge were worn.

1. The oligonucleotides were rehydrated to 1μg/μl with TE (10mM Tris and 1mM EDTA, pH 7.6). Each oligonucleotide was then diluted to 100ng/μl in water.
2. The following were then added to a tube in order: 18.5μl of water, 1μl of the 100ng/μl 5' prototype oligonucleotide, 2.5μl of 10x kinase buffer, 1μl of 10U/μl T4 polynucleotide kinase, and 2μl of dATP $\gamma$-$^{32}$P (3000ci/mmol).
3. The tube was placed in a water bath at 37°C for 1 hour.

### 3.2.24.5 Annealing the Oligonucleotides

After incubating for 1 hour at 37°C, 1μl of the 100ng/μl 3' oligonucleotide was added and the oligonucleotides annealed at 65°C for 15 minutes. The tube was allowed to cool slowly over 2 hours to room temperature in a beaker of water from the 65°C water bath. The annealed oligonucleotides were stored at -20°C.

### 3.2.24.6 Purification of Radioactive Probes Using NAP 5 Columns

Labelled probes were purified from the free radioactive dATP by running through NAP 5 columns (Amersham Pharmacia Biotech) containing Sephadex G-25. Several millilitres of TE were run through the NAP 5 column and the column allowed to run dry. The radioactive sample was then added and allowed to run into the column. The sample was then chased with 475μl of TE and then 1ml of TE was added. Three-drop fractions were collected until the peak of the radioactively labelled probe came through.

### 3.2.24.7 Measuring Specific Activity of Probes

1. Three millilitres of scintillation fluid (Packard®) was dispensed into a scintillation container, 2μl of probe was added and the contents of the tube mixed.
2. The container was placed in the scintillation counter and the number of counts/μl/minute calculated.

## 3.2.25 Oct-1/PEF-1 Electrophoretic Mobility Shift Assay Protocol

### 3.2.25.1 Preparation of "Mix"

This protocol was adapted from that published by O'Connor *et al.* (1996). Per tube, the following mix was assembled:

- Three microlitres of poly [d(I-C)] (1μg/μl)
- Three microlitres of Buffer D (20mM HEPES, pH 7.9, 20% glycerol, 0.1M KCl, 0.2mM EDTA, 0.5mM PMSF, and 0.5mM DTT)
- Three microlitres of x4 SMK (16mM spermidine, 160mM KCl, 0.5mM $MgCl_2$, 0.2% NP-40, 4mM DTT, 4mM PMSF, 8μg BSA, and 20% glycerol)
- Water was then added to bring the final volume to 12μl.

### 3.2.25.2 Reaction Setup

For each tube on ice, 9μl of mix and then 1μl (approximately 2μg) of nuclear protein extract was added. Tubes were incubated on ice for 10 minutes. Then the probe was added (20,000cpm in 1μl) and the tubes incubated on ice for a further 30 minutes.

## 3.2.26 Oct-1, YY1, PSM and Sp1 Electrophoretic Mobility Shift Assay Protocol

### 3.2.26.1 Preparation of "Mix"

This protocol was adapted from that published by Steger *et al.* (1995). Per tube, the following was added:

- Ten microlitres of x2 binding buffer (500μl x2 binding buffer [24mM HEPES, pH 7.9, 20% glycerol, 1mM EDTA, pH 8.0, 8mM $MgCl_2$, 120mM KCl, and 0.2% Nonidet NP-40], plus 1μl of 1M DTT)
- Seven microlitres of mix (5.7μg/μl BSA, 11.4mM spermidine, 0.14μg/μl poly [d(I-C)], 0.018μg/μl calf thymus DNA, and 5.7mM PMSF), plus 1μl water if no competition or supershift was performed.

### 3.2.26.2 Reaction Setup

For each tube on ice, 10μl of x2 binding buffer was added, followed by 7μl of mix and then 1μl (approximately 2μg) of nuclear protein extract. The tubes were incubated on ice for 20 minutes. The probe was then added (20,000cpm in 1μl) and the tubes incubated on ice for a further 20 minutes. Unlabelled competition oligonucleotides were added after radioactive probe.

### 3.2.26.3 Competition Assays

Competition experiments were performed to determine if sequence variations in particular TFBSs had any effect on DNA:protein interaction. DNA binding proteins are often capable of binding to a series of variations on a basic consensus sequence. A higher affinity binding site will compete more of a particular protein from a labelled binding site, resulting in a reduction in the intensity of the retarded band and vice versa (Dent and Latchman, 1993). If a molar excess of an unrelated oligonucleotide is added, there is no difference in the intensity of the band corresponding to the specific complex. For competition experiments binding of protein extract to labelled prototype oligonucleotides (16P7676, 16P7792, 16P7894, or 16P24) was competed with either 200-, 100-, 50-, 25-, or 12.5-fold excess of unlabelled prototype (16P7676, 16P7792, 16P7894, or 16P24), or mutant (16M7676, 16M7792, 16M7894, or 16M24) oligonucleotides.

### 3.2.26.4 Supershifts

Identification of proteins within a DNA:protein complex can be made by including a specific polyclonal antibody in the binding reaction. Specific antibodies either prevent protein binding by binding to a vital site within the protein, resulting in the disappearance of the band of interest, or "shift" the band of interest relative to the control as a result of the increase in size of the complex. Supershifts were performed by adding 2μg of IgG polyclonal YY1 or Oct-1 antibodies (Santa Cruz) 10 minutes after protein, and incubating tubes an extra 20 to 40 minutes on ice.

## 3.2.27 Visualising and Measuring Complexes

### 3.2.27.1 Electrophoresis

The entire contents of each tube were added to an equilibrated acrylamide gel. The gel was electrophoresed at 160V for 1.45 to 2 hours.

### 3.2.27.2 Fixing and Exposing the Gel to X-Ray Film

One plate was prised off the gel and the gel was fixed for 15 minutes in 10% methanol/10% glacial acetic acid. A piece of 3M Whatmanns filter paper was cut to the size of the gel and then the gel was transferred to the filter paper. After removing the gel from the other glass plate, the gel was covered in plastic wrap and placed in a gel-dryer (Bio Rad) at 80°C for 1 hour. In a dark room, the dried gel was placed into an X-ray developing cassette containing a Dupont cronex Quanta III intensifying screen. A piece of Kodak X-Omat Blue XB-1 film was placed on top of the gel. The cassette was left 3 hours to overnight at -70°C. The autoradiograph was then developed using an automated facility at Royal Prince Alfred Hospital (RPAH), Sydney.

### 3.2.27.3 Measuring Intensity of Bands on Autoradiographs

In some cases bands on autoradiographs representing specific DNA:protein complexes, and the unbound label located at the bottom of each autoradiograph, were quantitated using a Bio Rad phosphoimager. This was performed by placing the dried gel on a radioactive phosphoimager screen (Bio Rad) for 10 minutes to expose and then placing the screen into the phosphoimager and using Molecular Analyst™ software to measure the intensity of selected bands. The percentage of bound protein was calculated by taking the volume for bound and dividing by bound + unbound.

# 3.3 RESULTS

## 3.3.1 Cloning Human Papillomavirus Type 16 Upstream Regulatory Regions

### 3.3.1.1 Technical Considerations

All six URRs were successfully cloned into pCR®-Blunt and then subcloned into pALuc to create pALucH1, pALucK2, pALucK4, pALucO2, pALucR1, and pALucS1. An example of pALuc clones containing a HPV 16 URR, shown as a band on the gel of approximately 1kb after *Bam*H I digestion, which releases the URR from pALuc, is given in Figure 3.8.




Figure 3.8: pALuc constructs digested with *Bam*H I to determine the presence of a successfully cloned upstream regulatory region

pALuc constructs retrieved from bacterial colonies and tested for the presence of the HPV 16 URR by digestion with *Bam*H I. Lanes 1, 3, 4, 6, 7, 9, 10, 11, and 12 contain positive clones with an insert of the correct size (a band at approximately 1kb). Lanes 2, 5, 8, and 13 contain incorrectly ligated constructs. Lane 14 contains the positive control (2.5µg of *Bam*H I digested pALuc16) and lane 15 a 1kb DNA Ladder.

Once positive clones were obtained for each of the six isolates, the orientation of the URR in the plasmid was checked. Due to the location of *Hae* II sites in pALuc and the HPV 16 URR, all *Hae* II digested clones produce bands on a gel of 370bp and 2766bp regardless of orientation of the inserted DNA (Figure 3.9). If the URR was

inserted in the correct orientation, two bands of 2388bp and 701bp were also visualised. If the clone was ligated in reverse orientation, two bands of 2575bp and 514bp were observed instead. Inspecting the bands at 514bp and 701bp (Figure 3.9) quite easily distinguished the two different types of inserted DNA.

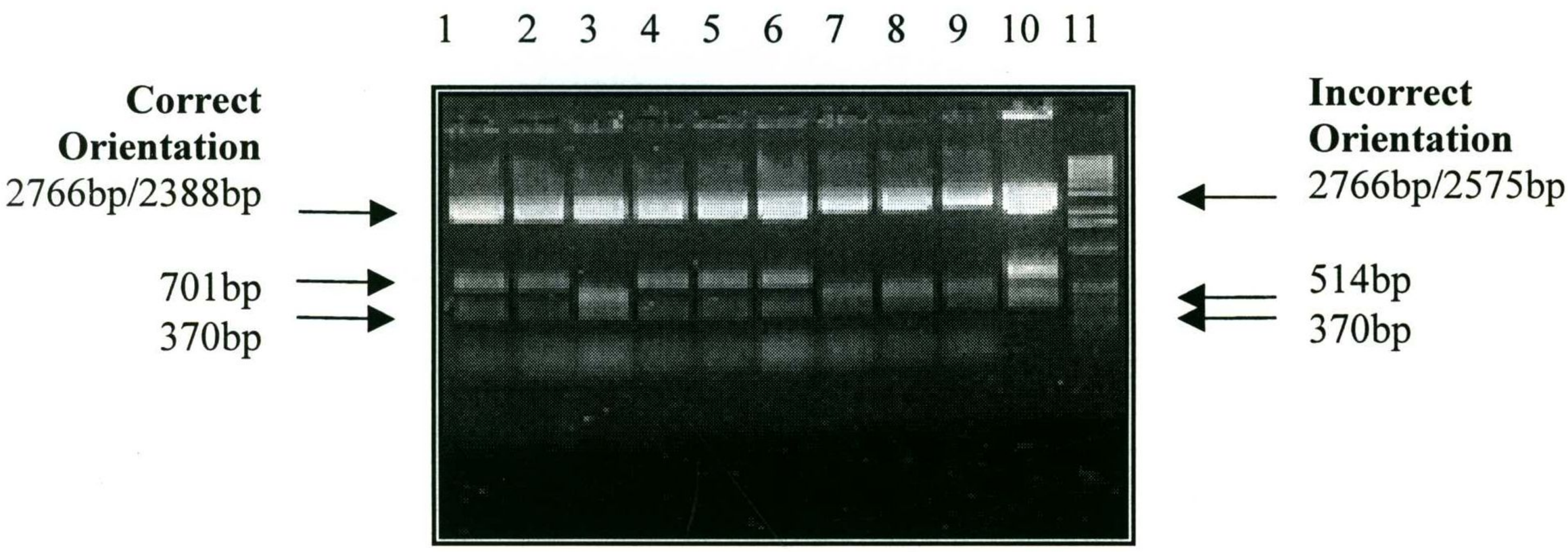



Figure 3.9: pALuc constructs digested with *Hae* II to determine which contain upstream regulatory regions cloned in the correct orientation

Lane 10 is the positive control (2.5µg of *Bam*H I digested pALuc16) with a band at 701bp, and clones containing an inserted URR in the correct orientation also have this band (lanes 1, 2, 4, 5 and 6). This is in comparison to clones that are incorrectly ligated which have a band at 514bp (lanes 3, 7, 8, and 9). Lane 11 contains a 1kb DNA Ladder.

All pALuc constructs were subjected to sequence analysis of the URR to ensure that sequences were the same as those reported in Chapter 2 of the study (Figure 2.4). No sequence anomalies were identified in the constructs.

The constructs containing the HPV 16 URR inserted in the correct orientation obtained for each of the six isolates, H1, K2, K4, O2, R1, and S1 are shown (Figure 3.10).

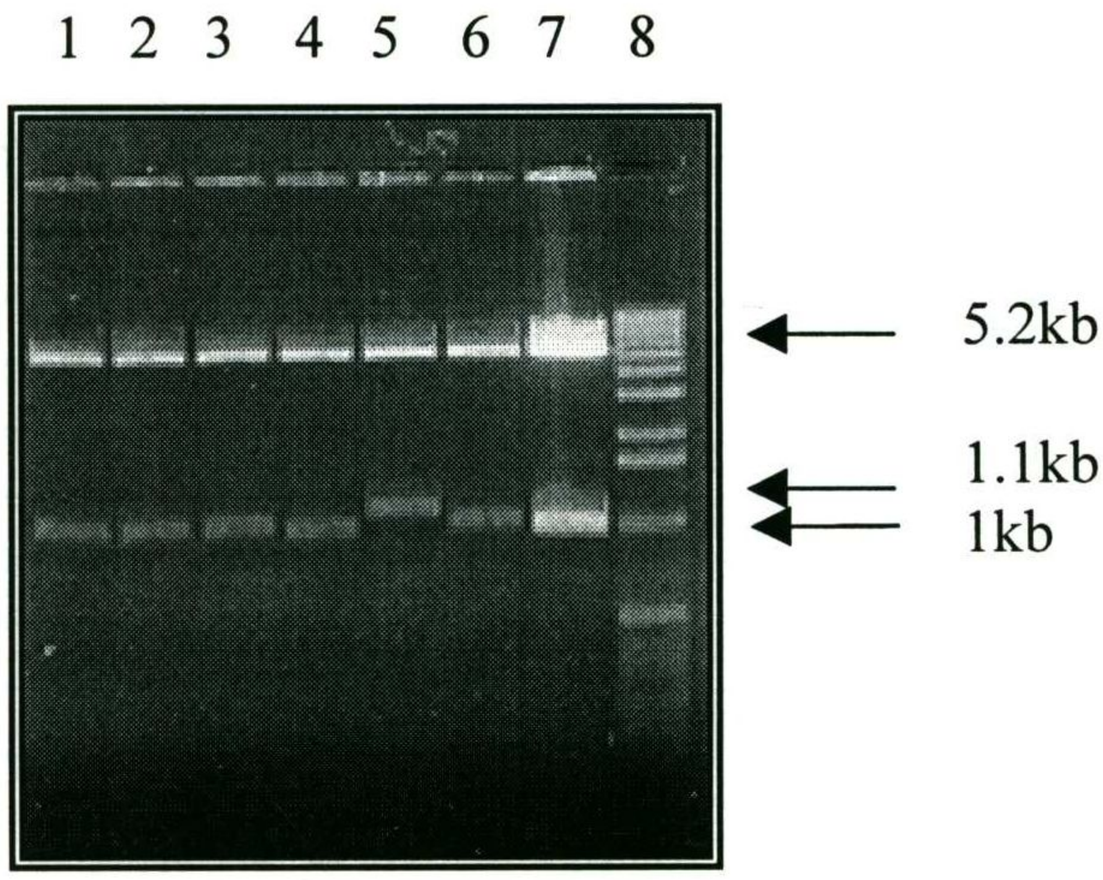



Figure 3.10: pALuc constructs pALucS1, pALucH1, pALucK2, pALucK4, pALucO2, and pALucR1, digested with *Bam*H I to release the upstream regulatory region

Lanes 1 to 6 contain the pALuc clones, lane 7 the positive control (2.5µg of pALuc16) and lane 8 contains a 1kb DNA Ladder. Note: pALucO2 contains an insert of larger size than the other constructs due to the large-scale duplication of 81bp (lane 5).

### 3.3.1.2 Luciferase Assays

The transfection efficiency of HeLa and HT3 cells depended predictably on several factors. The quality of the cells maintained in flasks before seeding into dishes was important. Flasks were best used at approximately 85% confluence. The distribution of the cells once plated also affected transfection, with an even surface of cells giving better results than "clumps" of cells. Approximately 70% plating density proved best for transfections.

Luciferase results in HT3 cells were similar to those in HeLa cells, and reproducible results were obtained across at least three assays. HPV 16 promoter activity was determined by calculating the luciferase/protein ratios adjusted to account for the variation in transfection efficiencies (Tables 3.7 to 3.9). In Figure 3.11 the luciferase

activities of the initial five variant URRs tested and the HPV 16 prototype URR (which served as reference and was set at a value of 1.0) are graphed. Standard deviations (lines extending from the bars) are also shown.

Table 3.7: Raw Luciferase Values in HeLa Cells*

| Isolate | pALuc16 | pALucH1 | pALucK2 | pALucK4 | pALucO2 | pALucR1 |
|---|---|---|---|---|---|---|
| Final Calculated Results | 1837401 | 1718234 | 17949751 | 8315172 | 4147541 | 2704392 |
| | 1920550 | 3844169 | 13252607 | 7202387 | 3149467 | 2267870 |
| | 1642751 | 3740111 | 13591810 | 8405512 | 3079392 | 2565987 |
| | 1984178 | | | | | |
| Average | 1846220 | 3100838 | 14931389 | 7974357 | 3458800 | 2512750 |
| SD# | 148363 | 1198500 | 2619474 | 670069.8 | 597495.4 | 223077.4 |

* Luciferase/OD for Protein = R. R/OD β-Galactosidase = Final Result

#SD = Standard Deviation

Table 3.8: Raw Luciferase Values in HT3 Cells*

| Isolate | pALuc16 | pALucH1 | pALucK2 | pALucK4 | pALucO2 | pALucR1 |
|---|---|---|---|---|---|---|
| Final Calculated Results | 1409299 | 3448923 | 11181613 | 7904911 | 965113 | 2646887 |
| | 1361339 | 3570528 | 13510283 | 5825797 | 1315201 | 2228988 |
| | 1181674 | 3131832 | 8553241 | 4924037 | 1165120 | 2364534 |
| | 841878 | | 7865152 | | 1139015 | |
| Average | 1198548 | 3383761 | 10277572 | 6218248 | 1146112 | 2413470 |
| SD# | 580430.3 | 226490.9 | 2585910 | 1528697 | 143484.5 | 213203.9 |

* Luciferase/OD for Protein = R. R/OD β-Galactosidase = Final Result

#SD = Standard Deviation

Table 3.9: Final Values for Constructs in Both HeLa and HT3

| Isolate | pALuc16 | pALucH1 | pALucK2 | pALucK4 | pALucO2 | pALucR1 |
|---|---|---|---|---|---|---|
| HeLa | 1.0 | 1.7 | 8.1 | 4.3 | 1.8 | 1.3 |
| SD | 0.08 | 0.65 | 1.42 | 0.36 | 0.3 | 0.12 |
| HT3 | 1.0 | 2.8 | 8.5 | 5.1 | 0.96 | 2.0 |
| SD | 0.48 | 0.18 | 2.1 | 1.25 | 0.12 | 0.18 |

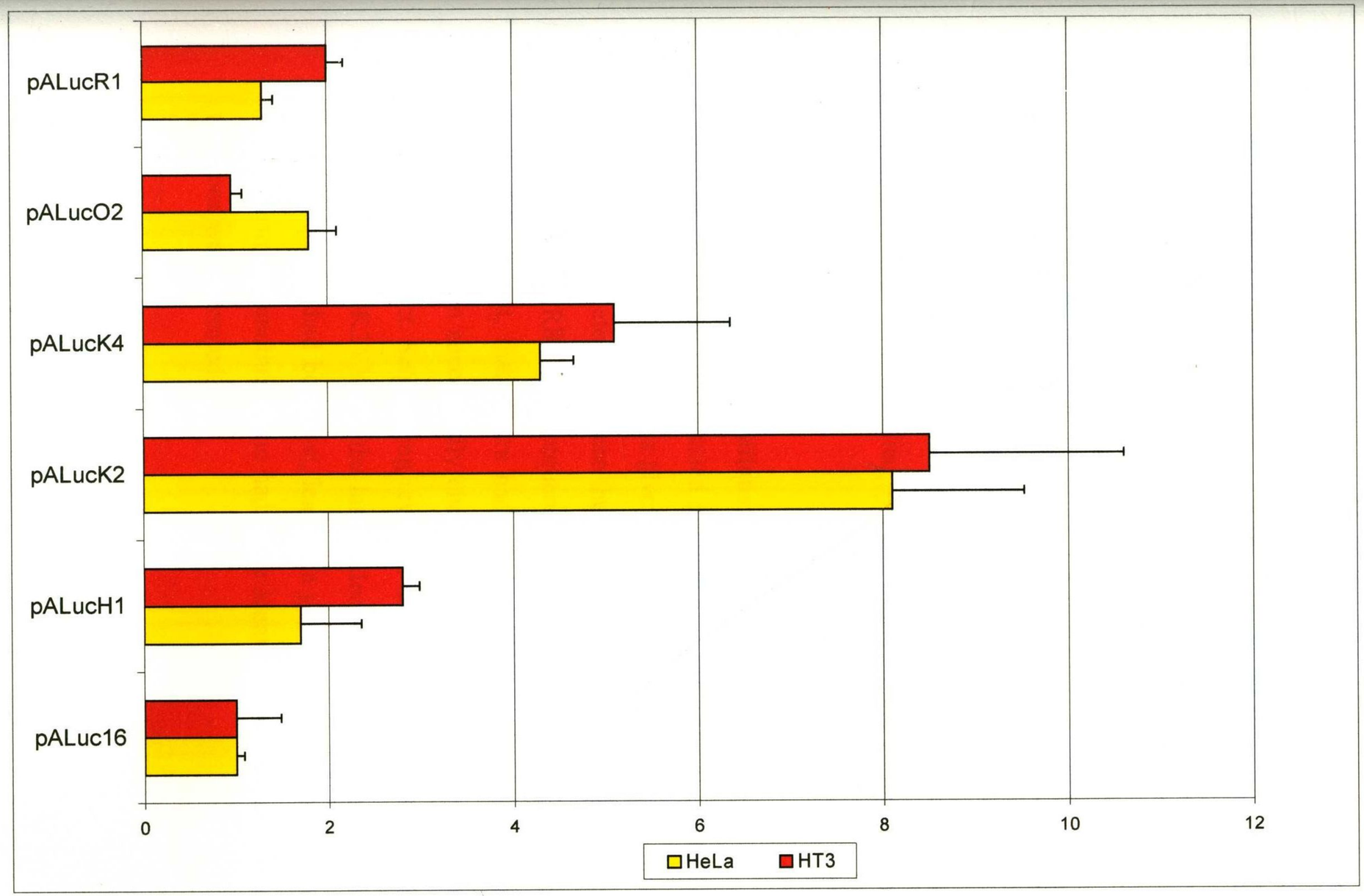



Figure 3.11: Luciferase activity under the control of the selected (H1, K2, K4, O2 and R1) human papillomavirus type 16 upstream regulatory regions

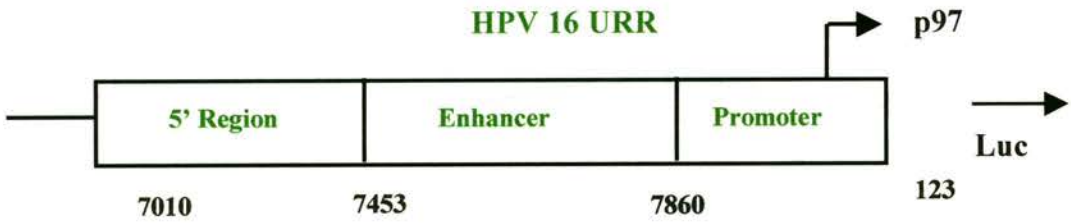
Luciferase activities mediated by all three variant URRs containing single nucleotide variations judged as potentially significant (pALucK2, pALucK4, and pALucH1), were upregulated in comparison with the prototype (8.1-8.5, 4.3-5.1 and 1.7-2.8 fold respectively). Those under the control of pALucO2 (which contained the large-scale duplication), and pALucR1, were comparable with the prototype. The luciferase activity of pALucS1, as determined during the second series of assays below, was also comparable to the prototype

## 3.3.2 Investigations to Identify Sequence Changes Mediating Increased Promoter Activity

A diagrammatic representation of the individual mutations created in the context of the TFBSs they were identified by PCR-mediated site-directed mutagenesis is given in Figure 3.12.

### 3.3.2.1 Technical Considerations

To create clones for site-directed mutagenesis, the original URRs were amplified in two segments. An example is given in Figure 3.13. The two products corresponding to the one complete URR were then ligated into *Hin*d II and *Kp*n I digested pCR®-Blunt. Cloned URRs were sequenced to ensure that the original sequence changes were still present, that others had not been introduced, that the nucleotide being mutated had been successfully changed, and that the innermost ligation site in the product was intact. Sequencing revealed that in two of the clones (pALuc16K2Oct-1P and pALuc16K2YY1P), the last nucleotide from the ligated end of the 5' or 3' primer sequence had been lost, leaving a gap in the sequence of one base pair. For those constructs another clone had to be identified and sequenced, and failing that, the clone was recreated.




**pALuc16K2Oct-1P** (C to A at nt 7676) Oct-1 7670 ACTATGCGC GCGCCAACGCC 7685 PEF-1

**pALuc16K2YY1P** (C to T at nt 7792) YY1 5 7790 TTCATACATGAACTG 7804

**pALuc16K4PSMP** (A to C at nt 7894) PSM 7893 TATAATAATACTAAACTACAATAAT 11

**pALuc16K4YY1P/ pALuc16H1YY1P** (C to T at nt 7786) YY1 4 7772 AACTAAATGTCACCC 7786

**pALuc16H1TEF-1P** (T to G at nt 7743) TEF-1 7737 TGCATATTTG 7746

Figure 3.12: Schematic representation of individual sequence mutations created in each of the K2, K4, and H1 constructs in the context of the transcription factor binding sites in which they were located

The sequence of each of the TFBSs Oct-1, PEF-1, YY1 5, PSM, YY1 4, and TEF-1 where sequence changes were mutated by site-directed mutagenesis to that found in the HPV 16 prototype are given. The prototype nucleotide created in each construct is given in red and beneath the name of each construct, the nucleotide change originally identified in the isolate is written.

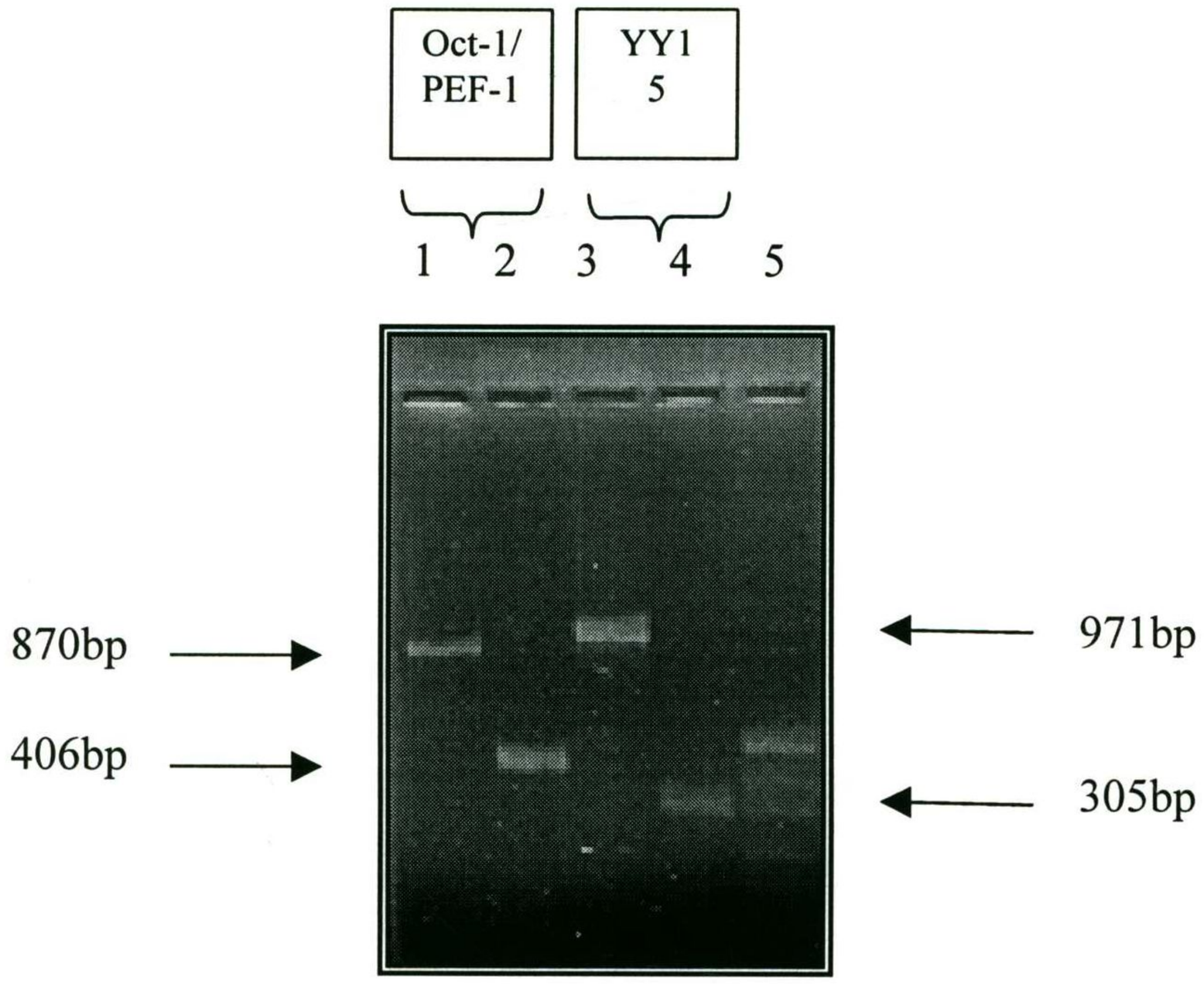



Figure 3.13: Example of polymerase chain reaction products used for site-directed mutagenesis cloning

Gel purified PCR product using K2 as template. The product in lane 1 was amplified using primers URR-6 and URR-7 (Oct-1/PEF-1 site). The product in lane 2 was amplified using primers URR-8 and URR-9 (Oct-1/PEF-1 site). The product in lane 3 was produced using primers URR-6 and URR-10 (YY1 5), and that in lane 4 using primers URR-11 and URR-9 (YY1 5). Lane 5 contains a pUC 19/*Hpa*II DNA marker. PCR products in lanes 1 and 2, and 3 and 4 correspond to complete K2 URRs, except that in lanes 1 and 2, nt 7676 in an overlapping Oct-1/PEF-1 site, has been changed to that found in the prototype. Similarly, in lanes 3 and 4, nt 7792 in a YY1 site has been changed to that found in the prototype.

All primer sets designed for site-directed mutagenesis produced bands of expected size, however the larger size PCRs, using primer URR-6, also produced a larger non-specific band. This extra product proved very difficult to eliminate even after optimisation of the PCR by titrating $MgCl_2$ and primer concentration, annealing temperature, and the amount of template against each other. In instances where a clone with this non-specific band was the only one identified, this extra product was carefully cut around during gel purification, so that it did not subclone into pALuc.

### 3.3.2.2 Luciferase Assays Using Constructs Created by Site-Directed Mutagenesis

Luciferase assay values for each of the constructs are given in Tables 3.10 and 3.11. These values were calculated in the same manner as described in section 3.3.1.2. Figure 3.14 represents the final luciferase activities in HeLa cells for the constructs created by site-directed mutagenesis and for isolate S1 (which was cloned during this set of experiments), standardised to the HPV 16 prototype.

It was interesting to note that luciferase values for each of the original URRs, K2, K4 and H1, were higher during site-directed mutagenesis experiments than in the original luciferase assays. This probably resulted from the use of a different luminometer (Fluoroskan Ascent FL [Labsystems] as opposed to a Lumat Berthoid, LB9501 in the first series of experiments). The changes in final values did not affect the order in which the constructs were upregulated or the proportional differences between them. It was not possible to obtain three complete sets of luciferase assay results using the new constructs in HT3 cells as well as in HeLa. Transfections using the constructs and HT3 cells were attempted several times but the transfection efficiency was consistently low. Mycoplasma contamination of the HT3 cells was suspected, however PCR for mycoplasma was negative and the cells looked healthy (whereas HT3s that were previously infected with mycoplasma showed a marked cytopathic effect). Original luciferase results, however, showed the same trends in both HeLa and HT3 (Figure 3.11). In two sets of results obtained using HT3s and the site-directed mutagenesis constructs (where the transfection efficiency was low) the trend was the same as that found in HeLa.

Table 3.10: Luciferase Values in HeLa Cells for Constructs Created by Site-Directed Mutagenesis

| | **pALuc16** | **pALucK2** | **pALuc16K2 Oct-1P** | **pALuc16K2 YY1P** | **pALucK4** | **pALuc16K4 YY1P** | **pALuc16K4 PSMP** | **pALucH1** | **pALuc16H1 YY1P** | **pALuc16H1 TEF-1P** | **pALucS1** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.1204 | 1.2567 | 0.753 | 0.18 | 0.7933 | 0.877 | 0.627 | 0.446 | 0.34 | 0.418 | 0.095 |
| | 0.0869 | 1.068 | 0.79 | 0.2 | 0.8036 | 0.783 | 0.381 | 0.336 | 0.338 | 0.249 | 0.149 |
| | 0.0976 | 1.1375 | 0.571 | 0.187 | 0.6184 | 0.7328 | 0.4064 | 0.2957 | 0.398 | 0.282 | 0.09 |
| Average | 0.101633 | 1.154067 | 0.704667 | 0.189 | 0.738433 | 0.7976 | 0.471467 | 0.359233 | 0.358667 | 0.316333 | 0.111333 |
| SD | 0.01711 | 0.095435 | 0.117228 | 0.010149 | 0.104079 | 0.0732 | 0.135293 | 0.077797 | 0.034078 | 0.089579 | 0.032716 |

Table 3.11: Final Values in HeLa Cells for Site-Directed Mutagenesis Constructs

| | **pALuc16** | **pALucK2** | **pALuc16K2 Oct-1P** | **pALuc16K2 YY1P** | **pALucK4** | **pALuc16K4 YY1P** | **pALuc16K4P SMP** | **pALucH1** | **pALuc16H1 YY1P** | **pALuc16H1 TEF-1P** | **pALucS1** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HeLa | 1 | 11.3 | 6.9 | 1.9 | 7.3 | 7.9 | 4.6 | 3.5 | 3.5 | 3.1 | 1.1 |
| SD | 0.16 | 0.9 | 1.1 | 0.1 | 1 | 0.7 | 1.3 | 0.75 | 0.3 | 0.9 | 0.3 |

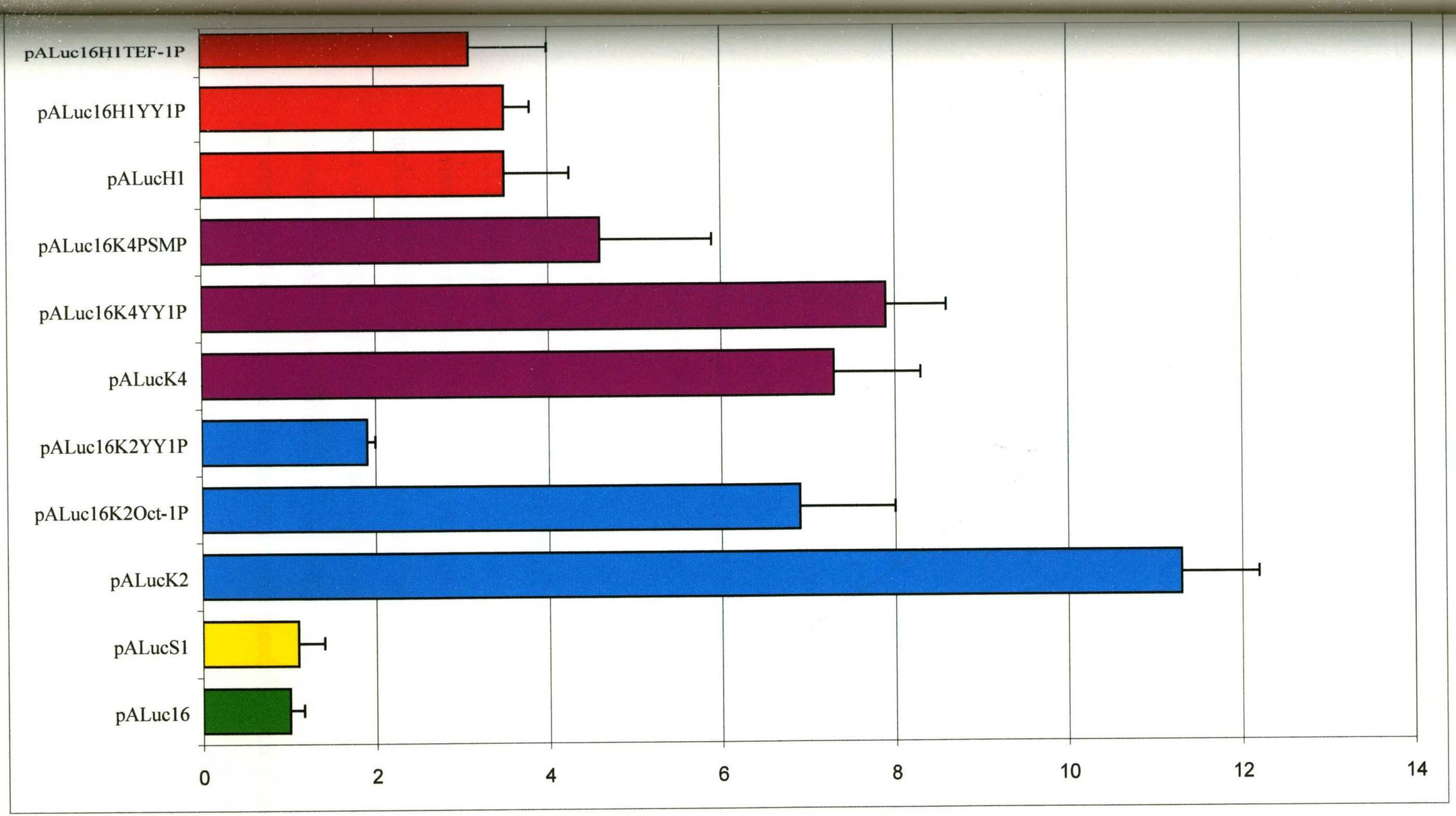



Figure 3.14: Expression of luciferase under the control of human papillomavirus type 16 upstream regulatory regions in HeLa cells.

The promoter activity of the URRs derived from the H1, K2, K4, and S1 isolates was assessed in the context of the prototype URR which was taken as one (green bar). The promoter activity of H1, K2 and K4 altered by site-directed mutagenesis are designated by bars the same colour as the original constructs.

It was found that nucleotide changes at 7676 and 7792 both contributed to the overall increase in promoter activity associated with the K2 URR. Changing the sequence variation at nt 7792, at the periphery of a YY1 binding site back to that found in the prototype, had the greatest effect, decreasing promoter activity by approximately 9-fold (pALucK2 versus pALuc16K2YY1P). However, the variation at nt 7676, in an overlapping Oct-1/PEF-1 site, also contributed several-fold (pALucK4 versus pALuc16K4Oct-1P).

The possibility that the change at nt 7786 at the periphery of another YY1 site accounted for the elevation in promoter activity seen for K4 and H1 was excluded because reversion to the prototype had no effect on promoter activity in either case (pALucK4 versus pALuc16K4YY1P and pALucH1 versus pALuc16H1YY1P). The promoter activity of pALuc16H1TEF-1P was comparable with that of pALucH1 indicating that the variation in the TEF-1 site in H1 played no part in altering H1 promoter activity.

The sequence variation at nt 7894 in the PSM accounted in part for the upregulation of the K4 promoter. This sequence variation reduced the promoter activity of K4 from 7.3-fold to 4.6-fold when the nucleotide was changed to that found in the prototype (pALucK4 versus pALuc16K4PSMP), and explained at least in part the difference in promoter activity between H1 and K4.

## 3.3.3 Electrophoretic Mobility Shift Assays (EMSAs)

### 3.3.3.1 Technical Considerations

Several factors affected the quality of EMSA results. The first was the quality of the nuclear protein extract. Since protein degrades easily, the proteinase inhibitor PMSF was used, protein extracts were handled completely on ice, and were stored in liquid nitrogen. Reaction conditions for probe and protein (such as time and temperature of incubation) affected results and had to be optimised for each of the probes used.

### 3.3.3.2 Nucleotide 7792 at the Periphery of a YY1 Binding Site (nts 7790-7804)

The protocol of O'Connor *et al.* (1996) which was initially used for the YY1 EMSAs, yielded unsatisfactory results with poor resolution, despite many attempts at optimisation. Superior results were obtained using a protocol published by Steger *et al.* (1995).

The AAV P5-60 YY1 site is well characterised and is often used as a control in YY1 EMSAs. YY1 antibody specifically bound YY1 using the AAV probe, resulting in a supershifted band. YY1 antibody also prevented YY1 binding to the HPV YY1 site, but did not produce a supershifted band (Figures 3.15). Competition assays for YY1 using non-radioactively labelled prototype (16P7792) and mutant (16M7792) oligonucleotides, showed that the variation at nt 7792 did not reduce YY1 protein binding affinity (Figure 3.15). In fact, the mutant oligonucleotide appeared to bind YY1 more strongly, as seen by better competition, and this result was reproducible across several experiments. Bands corresponding to YY1 binding and free probe were quantitated on a phosphoimager, and the percentage of labelled oligonucleotide bound to YY1 determined (Figure 3.16). These results substantiated the finding that excess unlabelled mutant probe prevented binding of the labelled prototype probe more effectively than the prototype itself.

YY1 EMSAs were performed using both HT3 and HeLa nuclear extracts and yielded comparable results The amount of YY1 protein in each of these cell lines also appeared to be comparable on the basis of the intensity of the band on the gel.

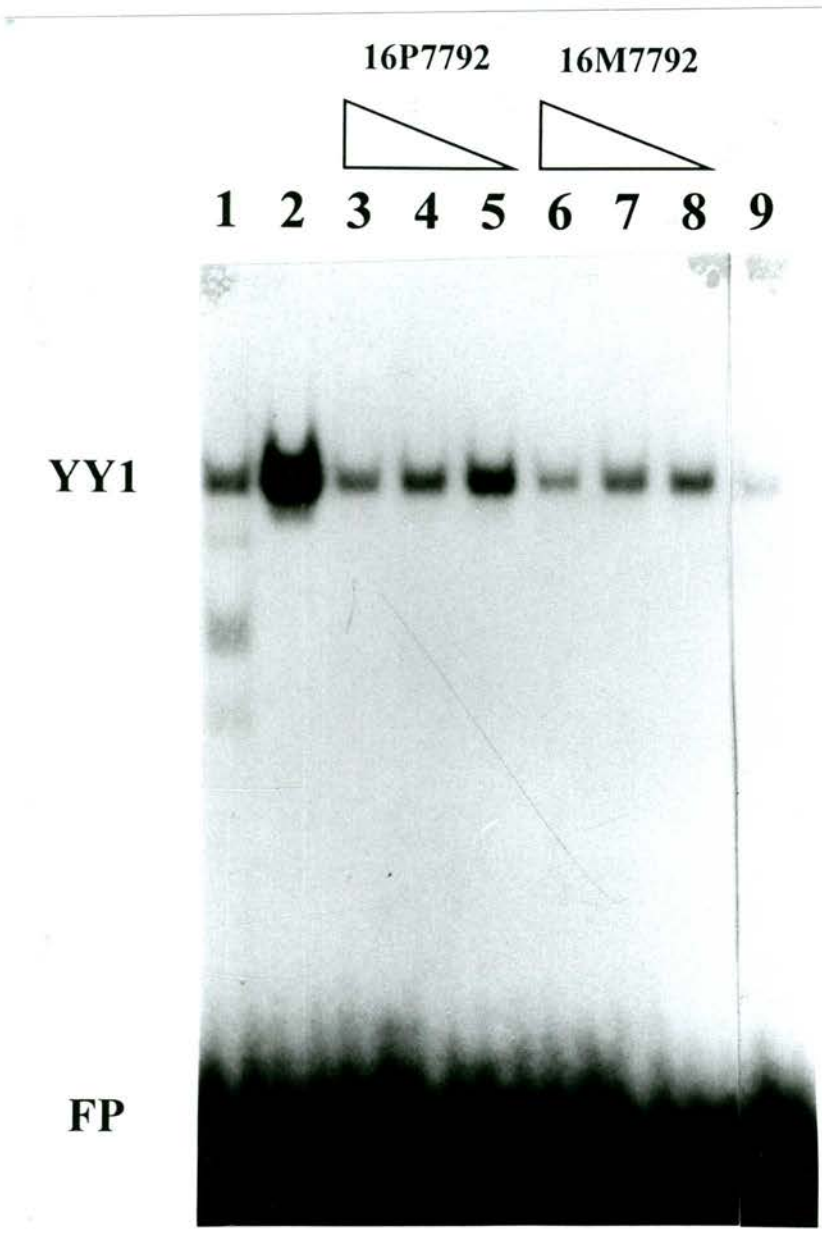



Figure 3.15: Electrophoretic mobility shift assay autoradiograph showing the binding of HeLa nuclear extract to $^{32}$P-labelled prototype YY1 oligonucleotide (nts 7790-7804)

AAV P5-60 YY1 (lane 1) and HPV 16 prototype oligonucleotide nts 7790-7804 not competed (lane 2), or competed with 200- (lanes 3 and 6), 100- (lanes 4 and 7), and 50-fold excess (lanes 5 and 8) of unlabelled prototype oligonucleotides (16P7792) with a C at nt 7792 (lanes 3-5) or unlabelled mutant oligonucleotide (16M7792) with a T at nt 7792 (lanes 6-8). Confirmation that the band represents YY1 binding is provided in lane 9 where addition of polyclonal YY1 antibody has almost completely abolished binding. FP = free probe.

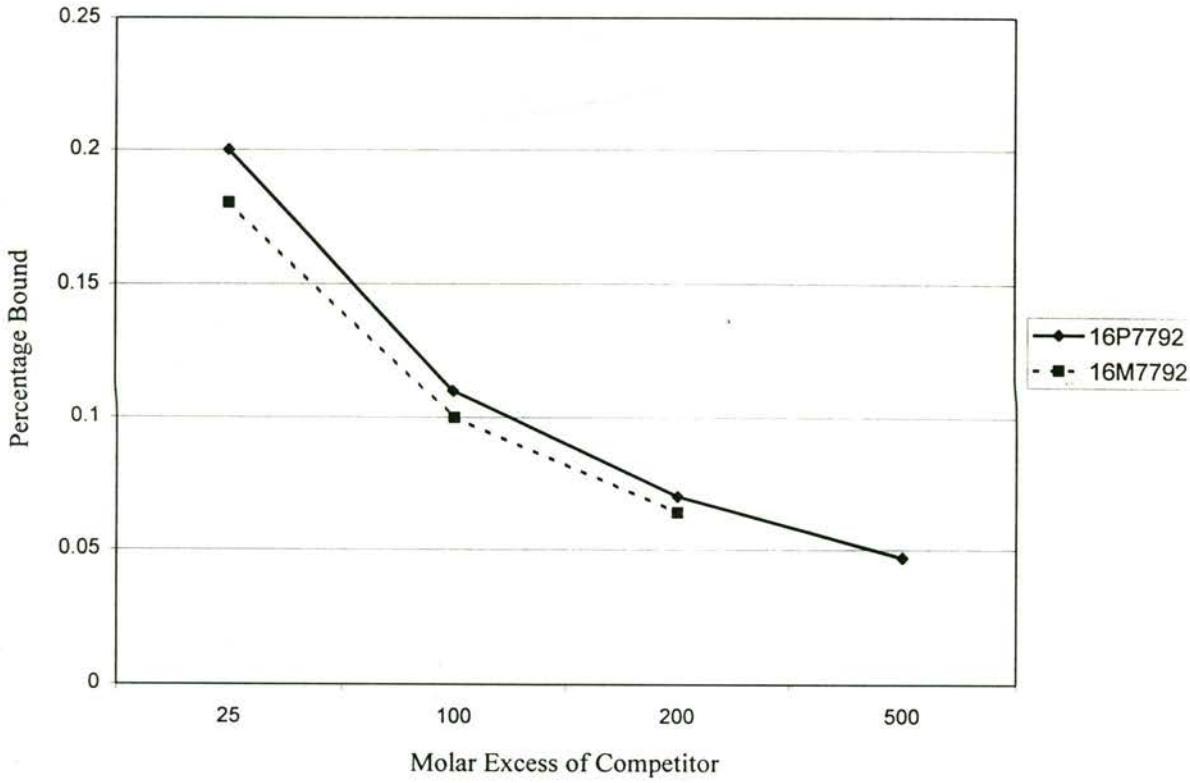



Figure 3.16: Percentage of bound YY1 to unlabelled prototype (16P7792) or mutant (16M7792) oligonucleotides derived from Molecular Analyst™ software on the phosphoimager

The percentage bound refers to the amount bound divided by bound plus unbound. The percentage of YY1 bound to the probe decreases with increasing amount of unlabelled competitor. Note the mutant oligonucleotides have slightly less YY1 protein bound for all concentrations of excess probe.

The YY1 binding site sequence variation at nt 7786 in K4 and H1 was also investigated by EMSAs, despite the fact that there was no effect on promoter activity. This variation had no effect on the binding affinity of YY1 with the unlabelled prototype oligonucleotide (16P7786) and mutant oligonucleotide (16M7786) competing equally efficiently (data not shown).

### 3.3.3.3 Nucleotide 7676 in an Overlapping Oct-1/PEF-1 Binding Site (nts 7670-7678 and 7675-7685 Respectively)

Two different EMSA protocols were used with the Oct-1/PEF-1 oligonucleotide. Using the protocol published by Steger *et al.* (1995) only one band was observed on the autoradiograph. This band was removed completely using specific polyclonal Oct-1 antibody (Figure 3.17). This protocol did not allow binding of PEF-1 protein to the probe. Using the method described by O'Connor *et al.* (1996) however, two bands were observed close together. Supershifting with Oct-1 polyclonal antibody showed that the top band corresponded to Oct-1 binding. It was assumed that the lower band represented binding to PEF-1, as it remained after supershifting (not shown), although this was not directly tested. An antibody corresponding to PEF-1 is not available, making it difficult to prove the real identity of the complex. Although a band corresponding to PEF-1 binding was not present using the protocol of Steger *et al.* (1995), the complex representing Oct-1 using this protocol had better resolution on the autoradiograph and the results of competition assays using this protocol are shown instead (Figure 3.17). Neither Oct-1 or PEF-1 binding was noticeably affected by the variation at nt 7676.

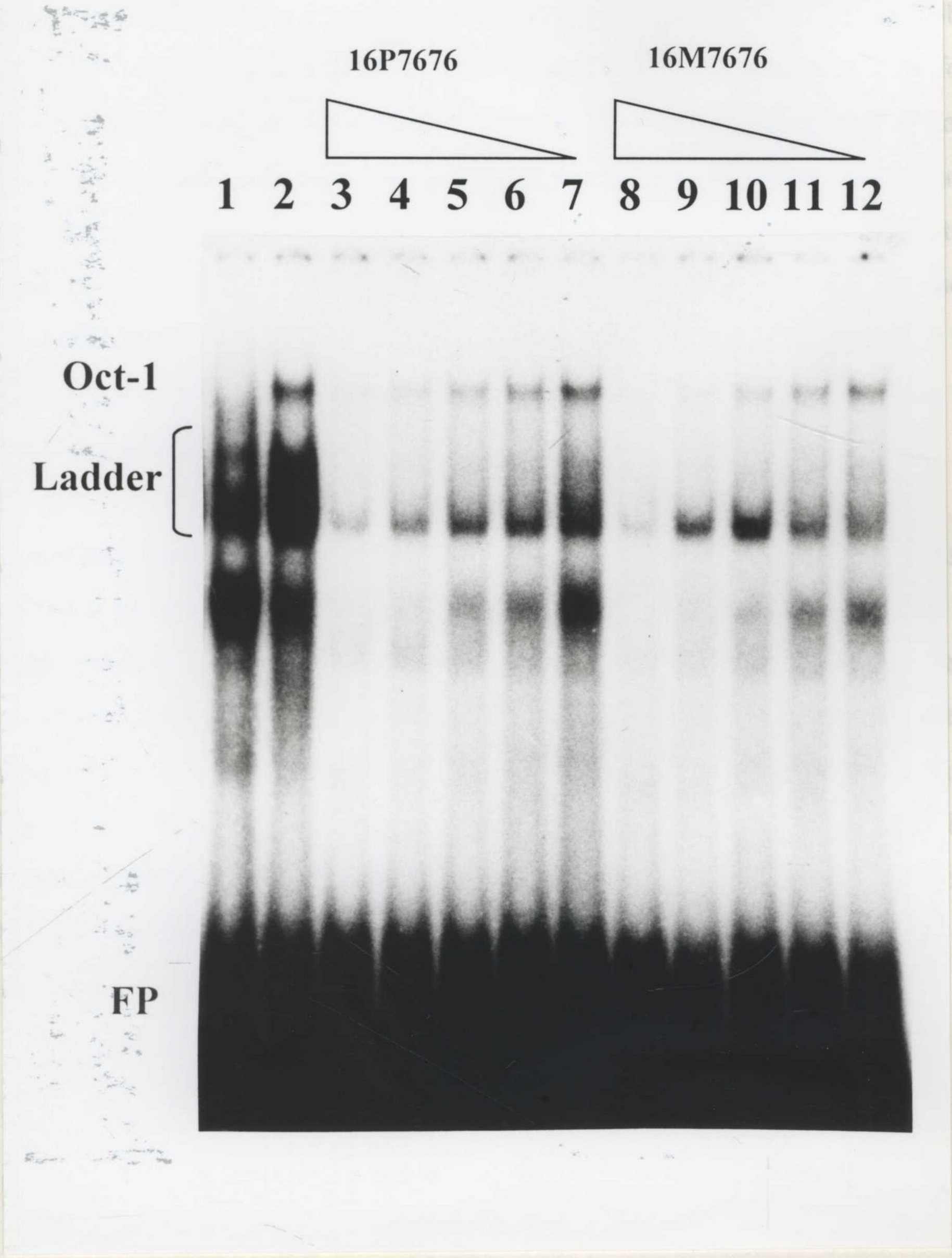



Figure 3.17: Electrophoretic mobility shift assay autoradiograph showing the binding of HeLa nuclear extract to $^{32}$P-labelled prototype Oct-1/PEF-1 oligonucleotide (nts 7667-7693).

Confirmation that the upper band in lane 2 represents Oct-1 binding is provided in lane 1 by the addition of polyclonal Oct-1 antibody, which has abolished binding of the labelled prototype. The effect of the variation at nt 7676 on binding to Oct-1 and the ladder of faster-migrating complexes is shown by competition with 200- (lanes 3 and 8), 100- (lanes 4 and 9), 50- (lanes 5 and 10), 25- (lanes 6 and 11) and 12.5-fold excess (lanes 7 and 12) of unlabelled prototype (16P7676) with a C at nt 7676 (lanes 3-7), or unlabelled mutant oligonucleotide (16M7676) with an A at nt 7676 (lanes 8-12). FP = free probe.

### 3.3.3.4 Nucleotide 7894 in the PSM (nts 7893-7901 and 3-11)

Using the protocol of Steger *et al.* (1995), two complexes, C1 and C2, were observed using the PSM probe (Figure 3.18). These represent the monomer and dimer forms of the PSM-BP respectively (O'Connor *et al.,* 1998). There were also other bands on the autoradiograph representing either non-specific or other uncharacterised complexes. An antibody was not available at the time the experiments were performed, as the identity of the PSM-BP as CDP was not then known. To show that the two complexes, C1 and C2, were not non-specifically bound, competition with an unrelated oligonucleotide (Oct-1) was performed. The Oct-1 oligonucleotide did not prevent the binding of C1 or C2 (lane 10 in Figure 3.18).

The sequence variation at nt 7894 in K4 is located in the first of the two parts of the motifs that make up the PSM (at nts 7893-7901). Both the prototype and mutant oligonucleotides progressively competed out the binding of C1 the same amount. However, the same molar excess of prototype oligonucleotide completely prevented binding of C2, whilst the mutant oligonucleotide did not (in particular a comparison of lane 5 with lane 9 shows that C2 is clearly evident in lane 9, but not in lane 5). This demonstrates that the mutant oligonucleotide does not bind to the dimer form of CDP (or C2 in Figure 3.18) as well as the prototype.

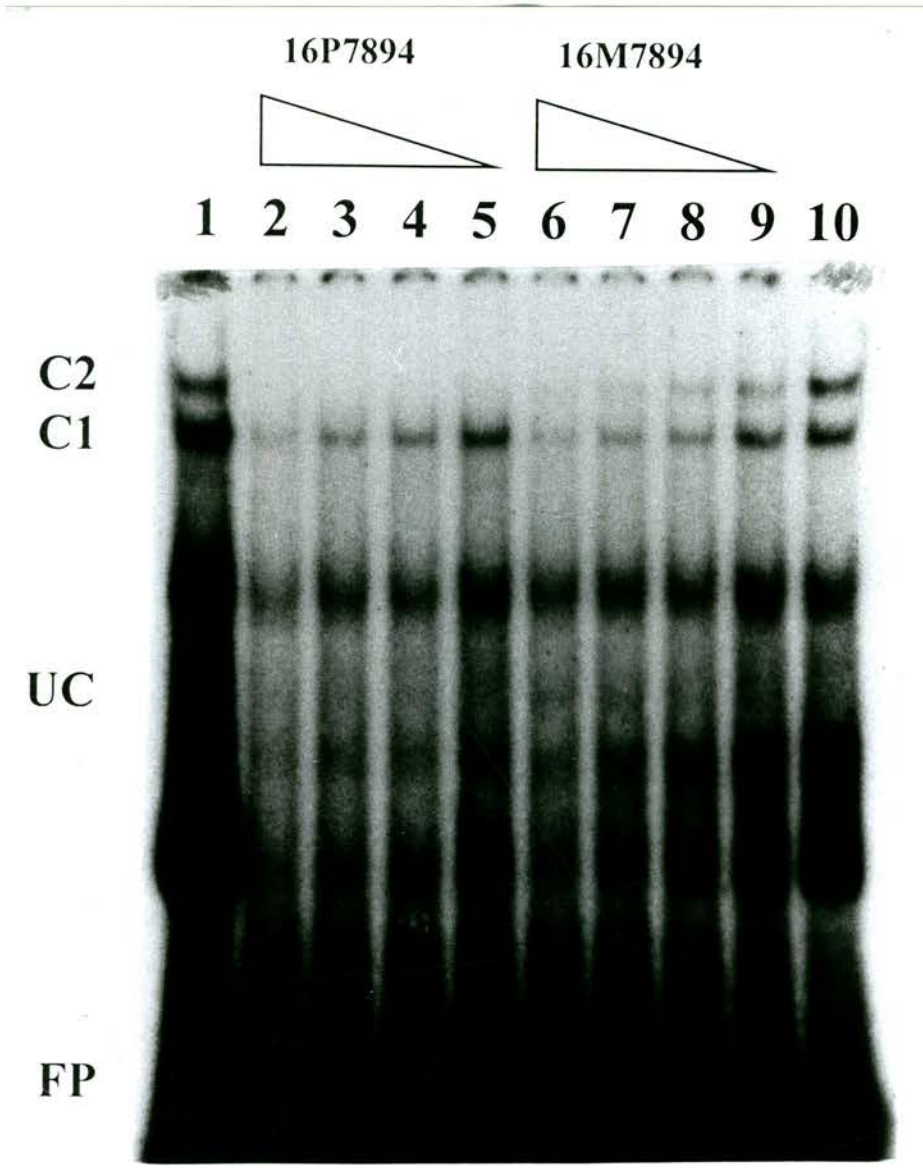



Figure 3.18: Electrophoretic mobility shift assay autoradiograph showing the binding of HeLa nuclear extract to $^{32}$P-labelled prototype PSM oligonucleotide (nts 7893-15)

Binding resulted in two complexes, C1 and C2, thought to represent the monomer and dimer forms of the PSM-BP respectively (lane 1). The effect of the variation at nt 7894 is shown by competition with 200- (lanes 2 and 6), 100- (lanes 3 and 7), 50- (lanes 4 and 8), and 25-fold excess (lanes 5 and 9) of unlabelled prototype (16P7894) with an A at nt 7894 (lanes 2-5), or unlabelled mutant oligonucleotide (16M7894) with a C at nt 7894 (lanes 6-9). Specificity of C1 and C2 is provided in lane 10 by competition with 200-fold excess of an unrelated unlabelled oligonucleotide (16P7676), where binding of both forms of the PSM-BP has been largely unaffected. UC = uncharacterised bands; FP = free probe.

### 3.3.3.5 Nucleotide 24 Adjacent to an Sp1 Binding Site (nts 28-33)

Sequence variation adjacent to an Sp1 site at nt 24 was investigated because similar Sp1 binding site mutations in the HPV 18 URR have been reported to be important in terms of increased promoter activity and more efficient binding to Sp1 motifs (Rose *et al.*, 1998). Specific binding of Sp1 to the oligonucleotide was confirmed using purified Sp1 protein. The sequence variation did not affect Sp1 binding (Figure 3.19).

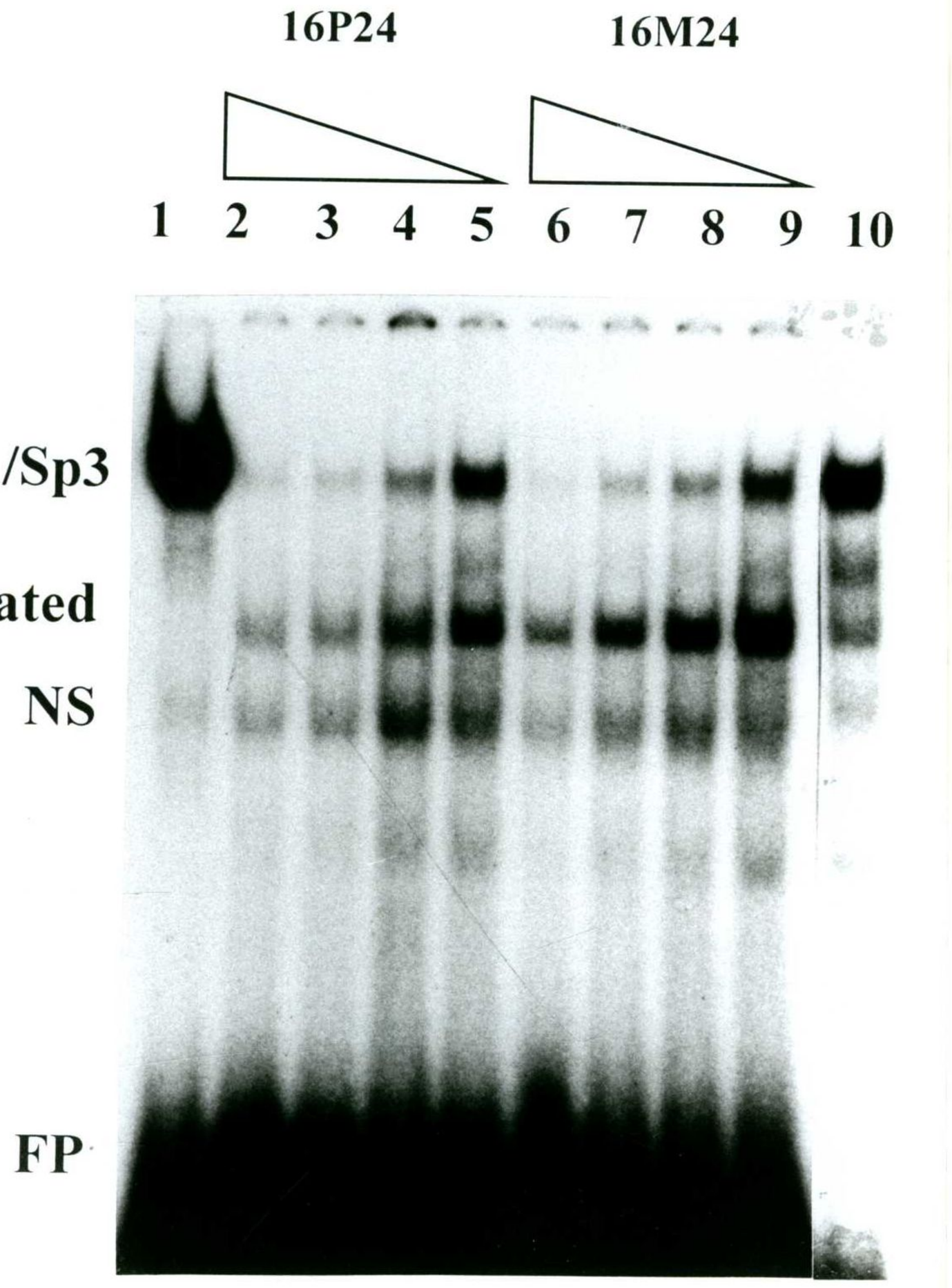



Figure 3.19: Electrophoretic mobility shift assay autoradiograph showing the binding of HeLa nuclear extract to $^{32}$P-labelled prototype Sp1 oligonucleotide (nts 20-38)

Binding resulted in two complexes, Sp1/Sp3 and Sp3 truncated, not competed (lane 10), or competed with 200- (lanes 2 and 6), 100- (lanes 3 and 7), 50- (lanes 4 and 8), or 25-fold excess (lanes 5 and 9) of unlabelled prototype oligonucleotide (16P24) with a C at nt 24 (lanes 2-5) or unlabelled mutant oligonucleotide (16M24) with a T at nt 24 (lanes 6-9). Confirmation that the band represents Sp1 binding is provided in lane 1 where purified Sp1 protein is used.

NS = nonspecific bands: FP = free probe.

# 3.4 DISCUSSION

## 3.4.1 The Potential Significance of Sequence Variation in the Upstream Regulatory Region

In this study, interest was centred on HPV 16 URRs with sequence variation in known TFBSs as it is possible that sequence differences that markedly alter the level of E6/E7 expression might equate to differences in the behaviour of cancers associated with the same HPV type. The importance of the level of E6/E7 expression is highlighted by studies showing that when the E7 oncogenes from high risk HPV types (16, 18 or 33) are placed under the control of strong promoters, markedly increased expression of E7 occurs and this results in increased transformation of cells (Watanabe *et al.,* 1992, and Liu *et al.,* 1995). Increased oncogene expression from variant URRs in episomal viral isolates may also be the functional equivalent of integration, overcoming the inhibitory effects of the viral E2 protein.

During this study, HeLa cells, which do not produce E2 protein, were used for performing function assays; therefore the results that were obtained were separable from any effects of E2. None of the isolates examined in this study by functional analysis contained sequence variation in E2 binding sites, suggesting that any effect of E2 on the promoter was not likely to be different for each of the variants, particularly as comparisons were made for each isolate with the HPV 16 prototype. A study by Veress and colleagues (1999) for example, demonstrated that when the E2 gene is cotransfected with several different HPV 16 URRs, the transactivation effect of E2 on each promoter was similar. However, any real effect of E2 protein on the promoter activity of each construct in this study cannot be completely discounted. Isolate O2 did contain an extra E2 binding site, and it is unknown what effect this may have had on promoter activity. Interestingly, O2 also contained other duplicated TFBSs, and these did not affect overall promoter activity. In any case the potential effects of E2 protein could have been tested using a cell line known to express the HPV 16 E2 protein, or by coexpressing E2 in cells such as HeLa. It must be noted however that even when the E2 gene is present in cervical cancers, it may not always be expressed (Sherman *et al.,* 1992, Stoler *et al.,* 1992, Böhm *et al.,* 1993, Fuchs and Pfister, 1994, and Daniel *et al.,* 1995), and to make any real conclusions about the

effects of the E2 protein on promoter activity its expression in each of the isolates would have to be determined. The effects of sequence variation were also examined in the context of the entire URR, as opposed to only the enhancer region used in some earlier studies.

Initial attention was focussed on assessing the effects of the O2 URR, containing the large-scale duplication, on promoter activity. The duplication contained a segment of the enhancer including a cluster of TFBSs, and an E2BS. The isolate also contained a series of single nucleotide changes typical of the As lineage. The promoter activity of O2, as well as that of the other As lineage isolate S1, was not substantially different from the E prototype, suggesting that the large-scale duplication in O2 probably did not affect promoter activity. Other large-scale duplications and deletions of the HPV 16 URR have been reported, with deletions being more common. A 213bp duplication of the URR (including most of the region duplicated in O2) was identified in an isolate from Bowenoid papulosis of the vulva (Fang *et al.,* 1993). The authors believed that the duplication was unlikely to alter promoter activity, even though it contained regulatory elements, although this was not tested. Transforming capacity was found to be dependent on the presence of duplicated HPV 16 URR sequences (larger than, but including the same segment duplicated in O2) in another study (Hall *et al.,* 1997). Mechanisms by which the reported duplication altered transforming activity are unknown, however it may be that duplication of particular TFBSs or single nucleotide changes in the isolate were necessary for transformation. Similarly, when the HPV 11 URR is duplicated, it is able to transform baby rat kidney (BRK) cells in conjunction with *ras* (Rosen and Auborn, 1991). This is significant because the low risk HPV 11 is rarely associated with malignancies. Another low risk HPV type with a duplicated URR, HPV 6a, has also been identified in malignancies (DiLorenzo *et al.,* 1992), and this type of anomaly can lead to higher promoter activity (Grassman *et al.,* 1996a). This evidence supports a role for some mutant genomes in the pathogenesis of malignancy. On the other hand, some large deletion mutants of the HPV 16 URR (from nts 7622-7816) have reportedly led to a large decrease in promoter activity (Grassman *et al.,* 1996b). Interestingly, different length deletants had markedly different promoter activities in the Grassman study, illustrating that the complex interplay between positive and negative factors would

ultimately determine the effect of the large-scale change. It is possible that large-scale sequence changes of this nature may also be secondary events during long-term viral persistence.

It was interesting to find that luciferase activity controlled by three of the five URRs containing only single base substitutions was substantially upregulated in comparison with the European prototype. The greatest effect was an 11-fold increase associated with the E lineage variant K2 that contained only four variant nucleotides, two of which are common among HPV 16 isolates worldwide. The sequence variation at nt 7792, located in YY1 5 (as designated by O'Connor *et al.,* 1996), was largely responsible for the increase in promoter activity of K2. Single base substitutions and deletion of YY1 binding sites were identified in cervical cancer isolates by Dong *et al.* (1994), and May *et al.* (1994). In particular, sequence variation at nt 7799 identified in isolates from both of those studies (within the same YY1 binding site, but designated as nt 7796 in both studies using an uncorrected HPV 16 nucleotide sequence), had a similar effect on promoter activity to the nucleotide change at 7792. May *et al.* (1994) demonstrated that the nucleotide change at 7799 abolished YY1 binding, which in turn released the promoter from YY1-mediated repression (as discussed in section 1.5.10.2). In addition, the deletion of a portion of the HPV 16 URR (nts 7756 to 7795) involving several YY1 binding sites including part of YY1 5, from the cervical cell line SiHa, has been postulated as the mechanism leading to increased levels of E7 mRNA and protein per cell per HPV DNA copy compared with CaSki (Kozuka *et al.,* 2000). It was hypothesised that the change at nt 7792 identified in this study mediated increased promoter activity by preventing YY1 binding in line with the previous observations. It was therefore surprising to find that the nt 7792 variation did not lead to decreased DNA:protein binding. This may be explained by its location outside the core binding element of YY1. The mechanism/s underlying upregulated promoter activity due to sequence variation at nt 7792 therefore remain to be determined. Several possibilities exist. As there are many other suggested functions of YY1, the variation could affect its activities by means other then preventing activation by AP1. For instance, several YY1 sites exist in the HPV 16 URR that are not involved in repression of AP1 activity, suggesting that they function by other means. DNA bending has been suggested as a possible function of YY1 in other systems (Natesan and Gilman, 1993, and Kim and Shapiro,

1996), although this concept is not universally accepted (Galvin and Shi, 1997). It is also possible that the nt 7792 variation in K2 could affect the binding of an unknown transcription or other factor. Approaches used to determine the cellular factors that interact with the HPV 16 URR have included DNase I hypersensitive site (HSS) analysis, footprinting studies using SiHa or HeLa nuclear extracts, and EMSAs using oligonucleotides corresponding to the HSS and footprint sites (Gloss *et al.,* 1989a, and Sibbet and Campo, 1990). These analyses showed that up to 23 sites were protected and therefore had cellular factors bound. However, there were no reported footprints corresponding to nt 7792, incorporating the YY1 site now known to exist, so it is possible that another factor will bind to the site under different conditions.

New TFBSs are still being identified in the HPV 16 URR, although the rate of identification has slowed over the last few years. For instance, a second Sp1 site overlapping YY1 9 was recently reported (Dong and Pfister, 1999). Point mutations in the YY1 site that prevented YY1 binding increased Sp1 binding and therefore p97 activity, hence illustrating another mechanism of YY1-mediated repression now known to exist in the HPV 16 URR. Sequence variation within the newly identified Sp1 site was not identified in this study. All four As lineage variants contained a nucleotide change in YY1 9 at nt 7842 however. In the O2 and S1 URRs this nucleotide change appears not to have had any significant effect on promoter activity though, even considering its location adjacent to the central important 'CAT' sequence in the YY1 binding motif. However the possibility remains that any effect may have been masked by other sequence changes in the O2 or S1 URRs.

The other sequence change investigated in K2, that at nt 7676, also contributed to increased promoter activity. This is believed to be the first report of naturally occurring sequence variation at this nucleotide, located within a site that Cuthill and colleagues initially reported as binding a factor they designated PEF-1 (nts 7675-7685) (Cuthill *et al.,* 1993). The same group more recently showed Oct-1 binds to an overlapping site (nts 7670-7678), and in addition noted a ladder of faster-migrating complexes binding to the same oligonucleotide, the components of which remain largely uncharacterised (Sibbet *et al.,* 1995). Sibbet and colleagues proposed that Oct-1 binding to the site in HPV 16 downregulates the promoter whilst PEF-1 upregulates it. The group further studied the overlapping site by creating a C to G

sequence change at nt 7676, as opposed to the naturally occurring C to A change identified in this study. The C to G change in the Sibbet study had little effect on either Oct-1 or PEF-1 binding and only increased promoter activity slightly in the spontaneously immortalised keratinocyte cell line, HaCaT (by less than 1.5-fold). It was therefore interesting that the variation at the same nucleotide in K2 affected promoter activity as much as 4-fold. These differences may relate to the cell types used in the investigation (HaCaT as compared with HeLa in this study). Repression of promoters by TFs can be cell type specific due to differences in the transcriptional environment, with the HPV 16 enhancer being more strongly repressed in HeLa cells as compared with HaCat cells (Sailaja *et al.,* 1999). In agreement with Sibbet and colleagues, sequence variation at nt 7676 in this study was not observed to have any real effect on the binding of either Oct-1 or PEF-1, which does not explain the effects of this change on promoter activity. The nucleotide change is not located in the critical 'CCAAC' element of PEF-1 (Cuthill *et al.,* 1993) which explains why it did not affect PEF-1 binding.

Although the change at nt 7676 in the study by Sibbet and coworkers did not significantly affect either Oct-1 or PEF-1 binding, it did enhance the ladder of faster-migrating complexes. It was suggested that the factors forming the faster-migrating ladder play a role in enhancer function, although a much smaller role than either Oct-1 or PEF-1. The increase in promoter activity observed in K2 may relate to these factors, but this could not be confirmed as the results of the binding experiments (using two different protocols) were inconclusive for the factors forming the ladder. Clarification of the mechanisms by which this sequence change affects promoter activity may come with the development of binding protocols allowing a clearer delineation of the ladder and ultimately characterisation of the factors involved.

The other two isolates exhibiting upregulated promoter activity, K4 and H1, both displayed many sequence variations typical of the AA lineage. It has recently been suggested that AA variants may have enhanced oncogenic activity (Veress *et al.,* 1999, Kammer *et al.,* 2000, and Kozuka *et al.,* 2000), and the results of this study are consistent with that suggestion. Six of the 13 substitutions common to K4 and H1 fell within known TFBSs: two GREs, a YY1 and two TEF-1 motifs. It was hypothesised either that K4 contained sequence variation leading to upregulated promoter activity,

and/or that H1 contained an additional variation which downregulated promoter activity. Investigations here showed that a TEF-1 variation (nt 7743), found only in H1, did not play a part in altering promoter activity, suggesting that a nucleotide change in K4 upregulated promoter activity.

Sequence variation in the recently identified PSM in K4 largely accounted for the difference in promoter activity between K4 and H1, and this report is the first indicating the importance of naturally occurring PSM sequence variants. The PSM consists of two copies of a 9bp repeated sequence separated by 7bp, overlapping the origin of replication. Both copies of the 9bp motif are required for full transcriptional repression. The PSM is bound by a dimer of the PSM-BP (CDP/Cut), a transcriptional repressor that binds with limited sequence specificity to purine-thymidine rich sequences (Pattison *et al.,* 1997). The sequence variation at nt 7894 identified in this study, an A to C change, alters the PSM to a sequence less like consensus. O'Connor *et al.* suggested that complexes they designated as C1 and C2 on EMSAs represent binding of a monomer and dimer of the PSM-BP, the CDP/Cut protein respectively (O'Connor *et al.,* 1998 and 2000). Both of these complexes were observed in the binding studies reported here. In addition, other uncharacterised complexes were also observed on the autoradiograph, which is not surprising since the probe, covering the genomic region over the origin of replication was 35 bases in length. The single nucleotide variation identified in this study affected C2, or what is thought to be CDP dimer formation, and led to a reduction in the level of promoter repression. The findings for the PSM from this study are important in building on the observations of others that variations in TFBSs, in particular negative regulators, may play important roles in over-expression of the viral oncogenes and hence, in tumour progression. Mutated HPV 16 and 18 URRs from oral cancer cell lines with sequence variations in binding sites for the negative regulator NF-IL6, have similarly been proposed as a cause for increased promoter activity (Chen *et al.,* 1997).

The sequence change at nt 7786 in another of the important YY1 sites, YY1 4, which is relatively common in As and AA isolates, was initially considered a likely candidate leading to upregulated promoter activity of both K4 and H1. However it was found not to be responsible for increasing promoter activity and did not affect the binding of YY1 protein. This finding is in agreement with the recent report of

Veress *et al.* (1999), who examined the functional effect of this change in the context of the prototype URR, whereas in this study the assessment was made in the context of the other changes in which it occurred naturally. This finding also helps to illustrate that not all naturally occurring sequence variations in the five important HPV 16 YY1 binding sites alter promoter activity. This change at nt 7786 was postulated as an explanation for upregulated promoter activity in two episomal cervical cancer isolates in a recent study (Park *et al.,* 1999), although referred to as nt 7785 in that study. Park *et al.* did not perform site-directed mutagenesis or EMSAs to determine the cause of upregulated promoter activity in their isolates. Although the function assays reported here were not carried out using the same cell line as the Park study (C33A versus HeLa), it is likely that the nt 7786 variation was not responsible for the upregulated promoter activity they observed. The isolates in the Park study contained similar sequences to K4 and H1 (eight of the same variations). It was deemed likely that one of these eight other changes, such as in a TEF-1 binding site or a GRE, was responsible for the upregulated promoter activity seen for K4, H1 and their isolates.

Sequence variation common to both K4 and H1 in a TEF-1 motif at nt 7689 results in a site less like consensus, and played no part in promoter upregulation in the study by Veress *et al.* (1999). Chan *et al.* (1989) induced a point mutation in one of the GRE sites in the HPV 16 URR that eliminated palindromicity and consequently reduced response to dexamethasone. On the other hand it was shown that mutations of GREs such that they resemble consensus motifs, lead to increased promoter activity due to increased hormone response and increased DNA/protein binding (Mittal *et al.,* 1993, and Khare *et al.,* 1996). The four GRE variants in K4 and H1 do not appear to create GRE sequences closer to consensus and therefore seemed unlikely to play a role in increased promoter activity. The underlying basis for upregulated promoter activity associated with K4 and H1 was therefore not determined.

A very recently published report by Kammer *et al.* (2000) revealed the nucleotide change in AA variants leading to upregulated promoter activity. This group demonstrated that North American 1 (NA1) variants in addition to AA variants had increased promoter activity in comparison with the E prototype, and the region responsible for enhanced transcription of both lineage variants was localised to the 3'

end of the URR, as in AA variants in the study by Veress *et al.* (1999). Further investigations showed that an A to C change at nt 7729, common to both AA and NA1 variants, was largely responsible for the enhanced promoter activity of both types of variants. As this nucleotide change was also present in K4 and H1, the two AA variants in this study, it is probable that this sequence change was responsible in part for upregulated promoter activity, with the nucleotide change in the PSM in K4 explaining the difference in promoter activity between them. The nt 7729 change was not considered to be of interest in this study when site-directed mutagenesis experiments were conducted as no known TFBSs included nt 7729. The identity of the factor/s binding to nt 7729 remains unknown and there were no DNA:protein binding differences in EMSAs using prototype and mutant oligonucleotides in the study by Kammer *et al.* (2000). Of particular interest, the AA variant was associated with higher promoter activity than the NA1 variant in the Kammer study, and since the AA variant contained the nt 7894 change in the PSM as described here, it is likely that this explains the higher promoter activity.

Sequence variation in positive regulators are also likely to play a part in promoter upregulation. For instance, two Sp1 motif sequence variants in HPV 18 cervical cancer isolates upregulated E6/E7 promoter activity four- to six-fold in one study (Rose *et al.,* 1998). This resulted from Sp1 binding more efficiently to the mutant binding site. The changes (to a G in each case) brought the aberrant Sp1 motif closer to the high affinity-binding consensus Sp1 motif. It was suggested that stronger Sp1 binding may decrease the ability of the E2 protein to repress the promoter. In this study a C to T sequence variation at nt 24 adjacent to an Sp1 site was identified in several of the isolates. In line with expectations, this change did not affect Sp1 binding in EMSAs and was deemed as unlikely to affect promoter activity, although this was not directly tested. This correlates with the fact that the URRs containing this variation had promoter activity comparable with the prototype (O2 and S1) and that the C to T change moves the Sp1 site further away from a high affinity consensus binding site.

In summary, the functional implications of sequence variation in six selected natural HPV 16 URR variants have been investigated in this part of the study. Sequence variation in TFBSs was found to be responsible for the upregulated promoter activity

exhibited by two of the URRs, and the binding of a negative regulator to the PSM was shown to be altered due to a sequence change. This is the first report of the functional importance of naturally occurring PSM sequence variations. However some of the nucleotide changes responsible for increased promoter activity did not affect the binding of their corresponding TFs, and it remains to be determined how they mediated promoter upregulation. Novel factors may bind to those sites and play important roles in regulation by the HPV 16 URR. This study provides further evidence that URRs from particular variant lineages, such as those from the AA lineage, may generally have enhanced promoter activity compared with URRs from other lineages. Others have suggested that sequence variation in the URR from E lineage variants may not be a common strategy to promote increased transcription of the oncogenes (Kurvinen *et al.,* 2000). While the majority of E lineage variants in this study probably have promoter activity comparable with the prototype, this study has shown that even single nucleotide changes in TFBSs in E variants, although not very common, can have a substantial effect on promoter activity.

# CHAPTER 4

# DETERMINATION OF THE PHYSICAL STATE OF HUMAN PAPILLOMAVIRUS TYPE 16 DNA IN CERVICAL CANCERS AND SEQUENCE ANALYSIS OF THE E2 AND E4 GENES

# 4.1 INTRODUCTION

The precise mechanisms of HPV-induced oncogenesis are not well understood, although malignant conversion is often associated with integration of the virus into the host chromosome and cessation of viral synthesis (Jeon *et al.,* 1995). Cells with integrated HPV 16 DNA have a selective growth advantage over those that are episomal, which may result from disruption of an mRNA instability element in the 3' untranslated region of the E6/E7 mRNA, leading to increased stability of E6 and E7 transcripts (Kleiner *et al.,* 1986, Jeon and Lambert, 1995 and Jeon *et al.,* 1995).

Integration of the viral genome frequently results in disruption of the E1 and/or E2 ORFs, while the E6 and E7 genes and the URR are always retained (Choo *et al.,* 1987, 1988, and Wilczynski *et al.,* 1988). The loss of functional E2 protein releases E2-mediated repression of the E6/E7 promoter, leading to an increase in expression of the E6 and E7 proteins, which is a critical event during tumourigenesis. Although it appears that all HPV 18, 31 and 35, and many HPV 16, positive cervical cancers carry integrated HPV DNA, between 9.8% and 27% of HPV 16 positive cervical cancers carry only episomal HPV DNA (Table 1.3). Small proportions of HPV 16 cancers contain both integrated and episomal DNA. As discussed in Chapter 3, malignant conversion in the absence of integration has been attributed both to disruption of YY1 binding sites in the URR and to large-scale URR duplications. Since several isolates in this study exhibited upregulated promoter activity due to sequence variation in TFBSs, and another contained a large-scale duplication in the URR, a major aim was to relate results from function studies to the physical state of the virus. The physical state of all 34 cancer isolates was investigated both by Southern hybridisation and PCR of the E2 gene. In addition, in isolates where the E2 region was found to be disrupted, the general location of the break within E2 was determined.

A further objective was to catalogue sequence variation within the complete E2 gene from 13 of the isolates, in addition to the E4 gene (which is contained within the coding sequence for E2). The E2 gene derived from three isolates containing benign HPV 16 infection was sequenced for comparison. On current evidence, the sequence

of the E2 gene appears to be relatively stable, although different HPV 16 lineages tend to have different subsets of E2 sequence changes. There is some epidemiological and molecular evidence that sequence changes in E2 may be biologically important (Hecht *et al.,* 1995, and Casas *et al.,* 1999). In one report, mutations in the E2 region increased immortalising capacity *in vitro* (Romanczuk and Howley, 1992). The potential importance of amino acid changes in the E2 and E4 proteins identified in this study are discussed.

# 4.2 METHODS

## 4.2.1 Determining the Physical State of Human Papillomavirus DNA

There are several methods for determining the physical state of viral DNA including Southern blot hybridisation (Cullen *et al.,* 1991, Pirami *et al.,* 1997, and Corden *et al.,* 1999), 2D gel electrophoresis, *in situ* hybridisation (Vernon *et al.,* 1997), PCR of the E2 gene (Das *et al.,* 1992a, and Chen et al., 1994b), and RT-PCR of mRNA (Daniel *et al.,* 1995, and Klaes *et al.,* 1999). In this study, the physical state of HPV 16 DNA was determined both by PCR of the E2 region and Southern hybridisation.

## 4.2.2 Southern Blot Hybridisation

DNA hybridisation was first reported in 1961 (Hall and Spiegelman, 1961), but it was not until 1975 that it was combined with a method to transfer or blot electrophoretically separated DNA fragments from a gel. This method is now known as "Southern blotting" after its inventor (Southern, 1975). The principle of DNA hybridisation is based on the ability of two complementary DNA strands to form a double helix which is held together by hydrogen bonding. By labelling one strand, which is used as a probe during hybridisation, the complementary strand can be identified with great specificity. DNA extracted from tissue samples is treated with a panel of restriction endonucleases, with the choice of enzyme/s dependent on the number of recognition sites in the viral DNA. It is desirable to choose a panel of enzymes with at least one which does not cut the viral DNA and one which cuts only once. The DNA is electrophoresed on an agarose gel before being "blotted" onto a nitrocellulose or nylon membrane. Membranes are then incubated with a radioactively labelled probe, in this case, the entire HPV 16 genome and complexes are visualised by autoradiography.

If HPV DNA in a tumour is episomal, cellular DNA digested with the enzyme that does not have a recognition site in the HPV genome, will contain several bands representing different forms of circular HPV DNA upon hybridisation with a viral DNA probe: open circular, supercoiled and multimeric forms. These different forms electrophorese at different rates on an agarose gel. The sample containing the

enzyme that digests HPV DNA singularly, on the other hand, will contain a band at approximately 8kb, because circular viral genomes have become linearised. Any multimeric copies are also converted to 8kb. If the HPV DNA in a tumour is integrated, digestion of cellular DNA with the enzyme that does not recognise HPV results in a smear of viral DNA that migrates with the bulk of the cellular DNA. Integrated DNA digested with the single cut enzyme will display either many bands, often of high molecular weight DNAs, or a smear. If the viral DNA is both integrated and episomal, then a complex array of smears and bands at 8kb is visualised.

#### 4.2.2.1 Clinical Isolates

Nucleic acids were extracted from the 34 patient tumours and CaSki cells as described in section 2.2.2.1.

#### 4.2.2.2 Restriction Endonuclease Digestion of DNA

Several restriction endonucleases were trialed for digesting cellular nucleic acids from cervical cancers including *Hin*d III (non-cut), *Bgl* II (non-cut), *Hin*c II (cuts at nt 3211 in E2), *Bam*H I (cuts at nt 6152 in L1), *Hae* II (cuts at nt 1314 in E1 and nt 7430 in the URR), and *Eco*R I (cuts at nt 7457 in the URR and nt 6820 in L1). *Bgl* II and *Bam*H I were selected for analysis of all isolates, and total cellular DNAs were digested with *Bam*H I and *Bgl* II in separate reactions containing 10µg of total nucleic acids, 2µl of 10U/µl *Bgl* II (Amersham Pharmacia Biotech) or 1µl of 20U/µl *Bam*H I (New England Biolabs), 2µl of the corresponding 10x buffer, 0.2µl of 10mg/ml BSA for reactions containing *Bam*H I, and water to a final volume of 20µl. Reactions were incubated at 37°C for 2 hours, before adding 5µl of x5 gel loading dye. Samples were then electrophoresed on a long 0.7% agarose gel overnight at 20V.

#### 4.2.2.3 Rapid Alkaline Transfer

This was performed essentially according to the methods first published by Reed and Mann (1985).

1. The gel, once electrophoresed, was trimmed and orientation notches were made.

2. The gel was depurinated in 0.25M HCl until the blue dye line went yellow (approximately 5 to 10 minutes) to break up large pieces of DNA to allow for easier transfer.
3. The HCl was tipped off and the gel washed.
4. The pH was then neutralised by the addition of 0.4M NaOH.
5. A perspex bridge was set up in a shallow dish containing enough 0.4M NaOH to bring the level of NaOH to half way up the bridge. Two long pieces of 3MM filter paper were cut to fit exactly over the perspex bridge and hang over either side as wicks into the NaOH. Another four pieces of filter paper were cut the same size as the gel (as shown in Figure 4.1).
6. The long pieces of filter paper were moistened in 0.4M NaOH and placed on the bridge one at a time. The sheets were rolled carefully to remove any air bubbles. This procedure was repeated with the two sheets of filter paper the same size as the gel.
7. The gel was placed well side down onto the filter paper and rolled lightly to remove any air bubbles.
8. A piece of positively charged nylon Genescreen Plus® membrane (NEN™ Life Science Products) was cut 2mm larger than the gel in all dimensions. This was placed in distilled water for 2 minutes, and then placed with the concave side onto the gel and rolled gently to remove any air bubbles. The membrane was not moved once it was in place.
9. One of the last pieces of filter paper was moistened by dipping in 0.4M NaOH and placed onto the membrane. This was likewise rolled to remove air bubbles.
10. The last piece of dry filter paper was also placed onto the gel.
11. Five to 8cm of paper towel was placed onto the top of the filter paper.
12. On top of this was placed a weight of approximately 500 grams. The set up apparatus was left at room temperature overnight.
13. At the end of the incubation period the paper towels and filter paper were removed, and the gel and membrane inverted. The position of the wells was marked on the membrane.
14. The membrane was washed in x2 SSC for 20 to 30 minutes at room temperature to remove any agarose on the membrane.

15. The membrane was then placed between 3MM filter paper and dried in an oven at 42°C for a few hours, then stored in a desiccator at room temperature until ready for hybridisation.

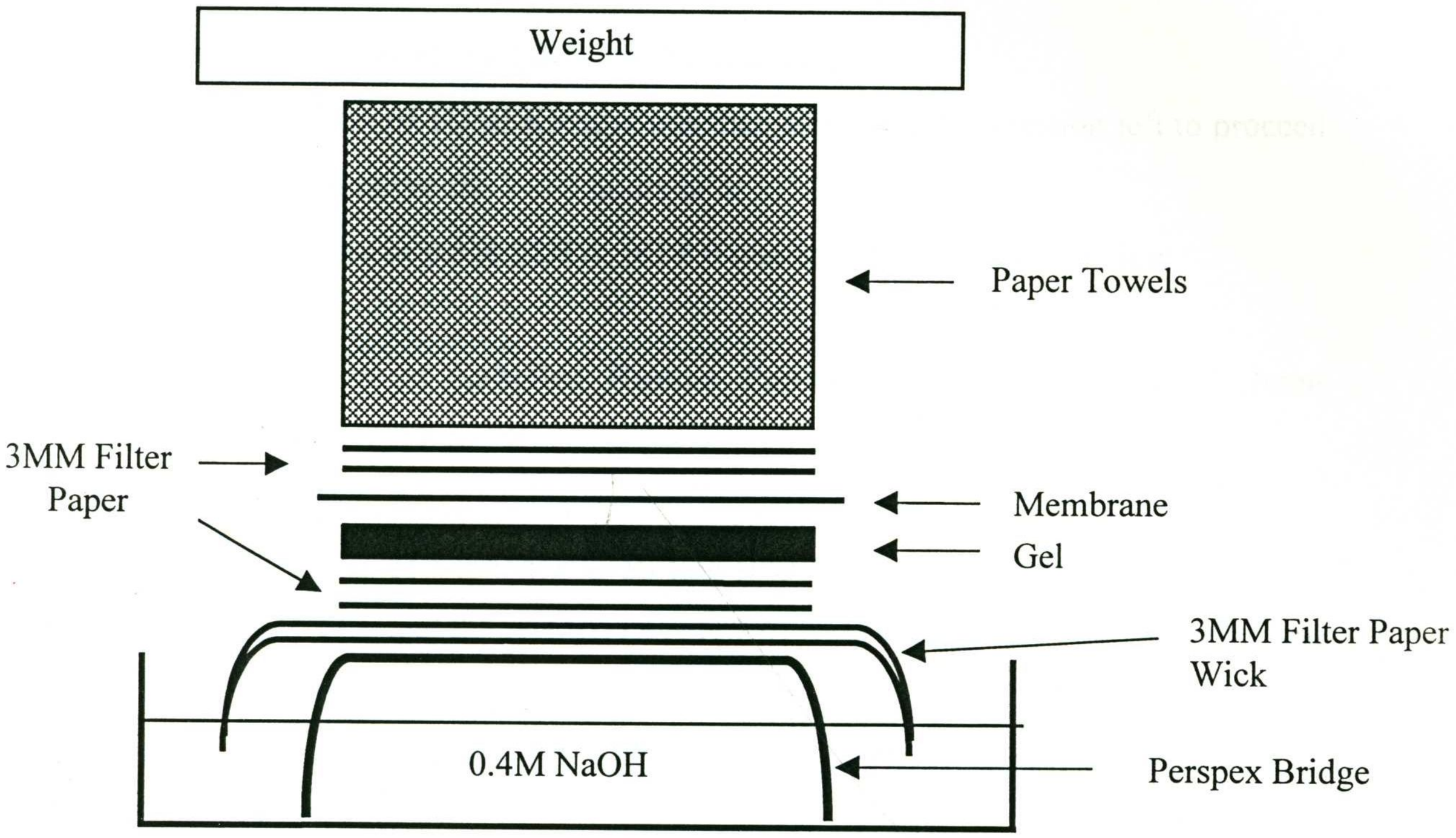



Figure 4.1: Diagrammatic representation of the Southern hybridisation apparatus

### 4.2.2.4 Preparing Human Papillomavirus Type 16 Probe DNA

The entire HPV genome was released from pBR322 using *Bam*H I and gel-purified as described in section 3.2.5.2.

### 4.2.2.5 Labelling the Probe Using a Multiprime DNA Labelling Kit (Amersham)

Labelling DNA by random priming involves annealing short oligonucleotides to denatured DNA and using these as primers. One of the dNTPs is labelled so the label is incorporated into the newly synthesised strand (Alphey and Parry, 1995). All work involving radiation was performed behind a perspex screen. Gloves, a gown and a radiation monitoring badge were worn.

1. Twenty two nanograms of HPV 16 DNA was added to water to a final volume of 23µl.
2. The DNA was boiled for 5 minutes in a water bath and then cooled on ice.
3. Ten microlitres of Multiprime buffer and 5µl of bovine serum albumin from the Multiprime DNA Labelling Kit were added to the DNA.
4. Then 6µl of 3000ci/mmol dCTP $\alpha$-$^{32}$P was added.
5. Two microlitres of 1U/µl Klenow was then added and the reaction left to proceed at room temperature for 3 to 5 hours or at 37°C for 30 minutes.
6. Five microlitres of 0.5M EDTA, pH 8.0 was added to stop the reaction.

#### 4.2.2.6 Separation of Labelled Human Papillomavirus Type 16 DNA from Unincorporated dCTP $\alpha$-$^{32}$P

Making the column:

1. One or 2 grams of Sephadex G-50 fine (Pharmacia) and approximately 25ml of TE (10mM Tris, and 0.5mM EDTA, pH 7.6) were mixed together in a beaker and left to stand overnight at 4°C.
2. Prior to use, the TE-Sephadex suspension was degassed for at least 20 minutes in a desiccator under vacuum.
3. Approximately 5ml of TE were run through a column (Pharmacia).
4. The TE-Sephadex suspension was slowly pipetted into the column with the tap open, and this process was repeated until the height of Sephadex in the column was 1cm below the top.
5. A further 2ml of TE was run through the column and the flow halted when the TE was just above the Sephadex. The column was not allowed to run dry.

Using the Column:

1. The labelled probe was added to the top of the column and allowed to run into the Sephadex.
2. A little TE was added and allowed to flow into the Sephadex.
3. The reservoir was filled with TE.
4. Approximately 20 drops were discarded and then three drop fractions (90µl) were collected for approximately 20 tubes.

5. There were two peaks of radiation. The first peak represented labelled HPV 16 DNA. The second peak, representing the unincorporated dCTP $\alpha$-$^{32}P$, was discarded.
6. The three tubes containing the labelled HPV 16 DNA were pooled and the specific activity was measured.
7. The probe was stored at -20°C in a lead container until required.

### 4.2.2.7 Measuring Specific Activity of the Probe

1. Three millilitres of scintillation fluid (Packard®) was dispensed into a scintillation container, 2µl of probe was added and the contents of the container mixed.
2. The container was placed in the scintillation counter and the number of counts/µl/minute was calculated and the volume corresponding to 5 x $10^6$ counts was used per hybridisation block.

### 4.2.2.8 DNA Prehybridisation

Each perspex block, which took two membranes, required 20ml of prehybridisation solution.

1. Prehybridisation solution for one block was prepared by adding 1ml of blotto (10% skim milk and 0.2% sodium azide), 5ml of x20 SSC, 2ml of 10% SDS, and 12ml of water. This solution was heated to 62°C for approximately 20 minutes.
2. Two millilitres of 3mg/ml calf thymus DNA per block was placed into a separate container, boiled for 6 minutes, then placed in ice. This was then added to the prehybridiation solution and kept at 62°C.
3. Using a perspex hybridisation block, the dry membrane/s were inserted into the block and warm prehybridisation solution was added.
4. The block/s were incubated at 62°C for a minimum of 4 hours.

### 4.2.2.9 DNA Hybridisation

1. The blotto and calf thymus DNA solution was made as for prehybridisation.
2. The HPV 16 DNA probe was boiled at 95°C for 5 minutes and then cooled on ice, before it was added to the blotto solution.

3. The prehybridisation solution was tipped from the block and the hybridisation solution immediately pipetted in.
4. The block/s were incubated at 62°C for 20 hours.

#### 4.2.2.10 Washing Membranes and Autoradiography

A low stringency wash consisting of x2 SSC and 0.1% SDS, and a high stringency wash consisting of x0.1 SSC and 0.1% SDS, were made up. Both wash solutions were warmed to 62°C prior to use.

1. Membranes were placed into the first pre-warmed low stringency washing solution and placed in a shaker bath at 62°C for 15 minutes.
2. The first low stringency wash was discarded and replaced with a second low stringency wash for a further 15 minutes in the shaker bath at 62°C.
3. This was repeated for the high stringency washes for time intervals of 15 and 30 minutes.
4. Membranes were removed from the wash solution and blotted lightly with 3MM filter paper.
5. Membranes were placed in cling wrap making sure that there were no folds in the plastic over the membrane.
6. In a dark room, the cling wrapped membranes were placed with an X-ray film (Kodak X-Omat Blue XB-1) into a developing cassette containing a Dupont Cronex Quanta III intensifying screen.
7. The cassette was placed into a -70°C freezer for a period of 1 to 3 days.
8. To develop the film, the cassette was taken out of the freezer and left 30 minutes to defrost before developing using the automated facility at RPAH, Sydney.

## 4.2.3 Polymerase Chain Reaction of the E2 Gene for Physical State Determination

### 4.2.3.1 Strategy

Results obtained by Southern hybridisation can be complemented by the more sensitive PCR method. Integration of HPV 16 DNA often occurs with a break in the E1 and/or E2 ORFs, resulting in deletion of part or the whole of these regions. PCR of the entire E2 gene (1097bp) has been used for determining the physical state of HPV DNA in several studies (Das *et al.*, 1992a, and Chen *et al.*, 1994b). If the E2 gene does not amplify, this is presumptive evidence of integration of the viral genome, providing the quality of the DNA has been controlled for. Small proportions of isolates containing integrated HPV DNA are also likely to contain a nondisrupted E2 region. PCR then, does not always give a definitive answer for physical state, particularly if integrated HPV consists of more than one copy of the HPV genome integrated in head to tail repeats.

### 4.2.3.2 Clinical Isolates

Nucleic acids were extracted from the 34 cancer isolates (as described in section 2.2.2.1).

### 4.2.3.3 Primers

Three sets of primers were used to amplify the E2 gene (from nts 2756 to 3853). The first set (Das *et al.*, 1992a) was used to amplify the entire E2 gene (E2-1 and E2-2). The second (E2-3 and E2-2) and third (E2-1 and E2-4) sets of primers were designed for this study to amplify the 3' or 5' regions of the E2 ORF respectively (Table 4.1). Samples producing negative results using primers E2-1 and E2-2 were then amplified using the other two sets of primers to determine which parts of the E2 gene were present and to assess the general location of disruption in E2. CaSki contains HPV 16 DNA integrated in at least 11 chromosomal sites (Mincheva *et al.*, 1987). It contains all HPV ORFs, including E2 (Baker *et al.*, 1987, and Smits *et al.*, 1991), and was used as a positive control for E2 PCR. Other controls and precautions to prevent PCR cross-contamination were as described in sections 2.2.5.2 and 2.2.3.2. PCR of the E6 gene (as described in section 5.2.2) was used as a control for the quality of HPV DNA in each sample and was performed at the same time as PCR of the E2 gene.

Table 4.1: Primers Used for Polymerase Chain Reaction of the Human Papillomavirus Type 16 E2 Gene

| Primers | Genome Positions | Nucleotide Sequence (5'-3') | Product Size (bp) |
|---|---|---|---|
| E2-1*<br>E2-2* | 2735-2753<br>3873-3854 | AGGACGAGGACAAGGAAAA<br>GGATGCAGTATCAAGATTTG | 1139 |
| E2-3<br>E2-2* | 3028-3040<br>3873-3854 | GTGGACATTACAAGACGTTAGCC<br>GGATGCAGTATCAAGATTTG | 814 |
| E2-1*<br>E2-4 | 2735-2753<br>3341-3360 | AGGACGAGGACAAGGAAAA<br>CTGCTAAACACAGATGTAGG | 606 |

*Primers published by Das *et al.*, 1992a

## 4.2.4 Optimised E2 Polymerase Chain Reaction Protocols

Although the primers used for PCR of the E2 gene have been published (Das *et al.*, 1992a), the protocol used for amplification had to be optimised. Primer concentration (from 0.25μM to 0.75μM), magnesium chloride concentration (from 1.5mM to 3.5mM), and annealing temperature (from 55°C to 65°C) were titrated against each other for each primer set. Final optimised PCR protocols for E2-1/E2-2, E2-3/E2-2, and E2-1/E2-4 are given (Tables 4.2, 4.3 and 4.4 respectively).

Table 4.2: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers E2-1 and E2-2 (Full Length E2 Gene Product)

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 29.1µl | - |
| 10x Reaction Buffer | 5µl | 1x |
| 25mM $MgCl_2$ | 7µl | 3.5mM |
| Primer E2-1 (100ng/µl) | 1.5µl | 0.5µM |
| Primer E2-2 (100ng/µl) | 1.5µl | 0.5µM |
| dNTPs | 0.5µl | 250µM |
| *Amplitaq*® DNA Polymerase | 0.4µl | 2 Units |
| | 45µl PCR Cocktail + 5µl Target DNA (100ng/µl) | |

Thermal Cycling Conditions:

| | |
|---|---|
| Cycle 1 | Step 1 94°C/5 minutes (Denaturation) |
| | Step 2 64°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycles 2-39 | Step 1 94°C/30 seconds (Denaturation) |
| | Step 2 64°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycle 40 | Step 3 72°C/7 minutes (Extension) |

Table 4.3: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers E2-3 and E2-2

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 26.1µl | - |
| 10x Reaction Buffer | 5µl | 1x |
| 25mM $MgCl_2$ | 7µl | 3.5mM |
| Primer E2-3 (100ng/µl) | 3µl | 0.5µM |
| Primer E2-2 (100ng/µl) | 3µl | 0.5µM |
| dNTPs | 0.5µl | 250µM |
| *Amplitaq*® | 0.4µl | 2 Units |
| DNA Polymerase | ____ | |
| | 45µl PCR Cocktail + 5µl Target DNA (100ng/µl) | |

Thermal Cycling Conditions as Table 4.2.

Table 4.4: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers E2-1 and E2-4

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 13.55µl | - |
| 10x Reaction Buffer | 2.5µl | 1x |
| 25mM $MgCl_2$ | 2.5µl | 2.5mM |
| Primer E2-1 (100ng/µl) | 0.5µl | 0.3µM |
| Primer E2-4 (100ng/µl) | 0.5µl | 0.3µM |
| dNTPs | 0.25µl | 250µM |
| *Amplitaq*® | 0.2µl | 1 Unit |
| DNA Polymerase | ____ | |
| | 20µl PCR Cocktail + 5µl Target DNA (50ng/µl) | |

Thermal Cycling Conditions as Table 4.2.

### 4.2.5 Analysis of Full Length E2 Polymerase Chain Reaction Products by *Hinc* II Digestion

PCR products corresponding to the entire E2 gene (PCR using primers E2-1 and E2-2) were validated by digestion with *Hinc* II. *Hinc* II digests the full length E2 gene PCR product at nt 3211, producing fragments of 662bp and 475bp. For *Hinc* II digestion, the following were added per reaction: 3.3µl of water, 0.5µl of 10U/µl *Hinc* II (New England Biolabs), 1µl of 10x New England Biolabs buffer 3, 0.2µl of 10mg/ml BSA, and 5µl of PCR product. Tubes were incubated for 1.5 hours at 37°C and reactions stopped by the addition of 5µl of x5 gel loading dye and the products electrophoresed on a 2% agarose gel.

### 4.2.6 Sites of Disruption in E2 Genes

For isolates containing a disrupted E2 region (negative with primers E2-1 and E2-2), the two smaller E2 PCRs were used to determine the presence or absence of smaller regions of the E2 gene. Primer E2-3, located downstream of primer E2-1, was used in combination with primer E2-2 to determine the presence of the 3' region of E2 (as shown diagrammatically in Figure 4.2). Primer E2-4 was also used in combination with primer E2-1 to determine the presence of the 5' region of the E2 gene.

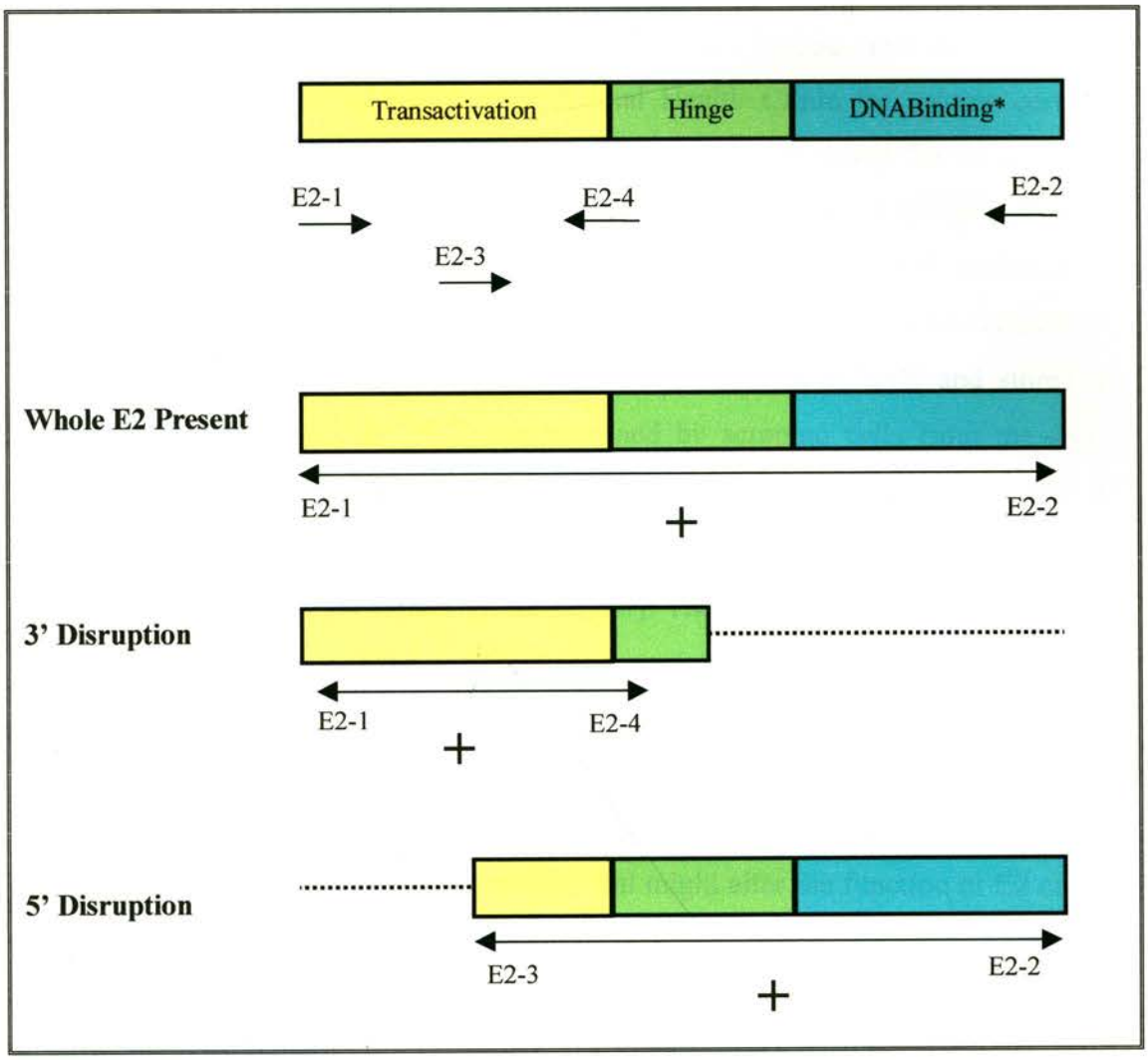



Figure 4.2: Location of primers for amplification of the E2 gene and corresponding products that can be amplified when segments of E2 are disrupted.

*DNA binding and dimerisation domain

## 4.2.7 Polymerase Chain Reaction of the E2 and E4 Genes for Sequence Analysis

### 4.2.7.1 Strategy

Thirteen cancer isolates selected for sequence analysis of the E2 gene were grouped into two categories. The first included isolates containing only integrated HPV DNA and an intact E2 ORF (M2 and H3). The second category contained episomal isolates from each lineage represented in the cohort: E (J1, C1, K1, H2, K2, K6, M6, and T1), AA (K4) and As (G1, and S1). Primers E2-1 and E2-2 were used to amplify the E2 gene from these isolates for sequence analysis.

A third category comprising three isolates (PRE262, PRE250, and PRE259) from patients presenting at Sydney Hospital Sexual Health Clinic for routine cervical screening in 1996 with benign HPV 16 infection, were analysed for comparison. These selected isolates had been shown to contain HPV 16 in the absence of other HPV types by strip typing (Roche), and there was no cytological evidence of dysplasia. Otherwise, selection was based on the availability of sufficient material for the investigations. The cells were fixed onto slides with CytoFix® and stored at room temperature. DNA for PCR was obtained by scraping cells from the glass slides into 110µl of lysis buffer for PCR (50mM Tris, pH 8.5, 1mM EDTA, pH 8.0, and 0.5% Tween 20®), plus 2µl of 20mg/ml proteinase K, and digesting the cells overnight at 37°C (not performed by the author). The crude extracts were then used directly for PCR. Primers E2-2 and E2-3 were used to amplify the E2 gene from the cervical smear isolates.

The aim was to identify sequence variation present in cancer but not in noncancer isolates, and to identify sequence variations that might alter the function of E2 or E4 in either integrated or episomal isolates.

#### 4.2.7.2 DNA Sequencing

Electrophoresis, PEG precipitation and DNA sequencing of the E2 and E4 genes were performed as described in Chapter 2. Sequence analysis was performed in both directions using primers E2-1 and E2-2. Procedures to ensure the validity of sequence data were as described in section 2.2.10. Sequences were aligned using Pileup (GCG) and Pretty (GCG) software from ANGIS as described. A sequence error in the E2 gene of HPV 16R at nt 2926 (correct nt a 'G', Meissner, 1997) was corrected.

#### 4.2.7.3 E2 and E4 Protein Analysis

Nucleotide variations leading to amino acid changes in the E2 and E4 genes were determined using Translate and Eclustalw software available through ANGIS.

# 4.3 RESULTS

## 4.3.1 The Physical State of Human Papillomavirus DNA by Southern Blot Hybridisation

### 4.3.1.1 Technical Considerations

Southern blot hybridisation results were affected both by the quality of nucleic acids and the amount of HPV DNA in an isolate. A nucleic acid extract that was PCR positive was not always positive by Southern hybridisation. It is probable that the number of HPV copies per cell was below the level of threshold for detection by Southern in the samples where no signal was observed on the autoradiograph.

### 4.3.1.2 Results by Southern Hybridisation

By Southern hybrisdisation alone, 7 of the 34 isolates (21%) were found to contain integrated HPV 16 DNA, ten (29%) contained episomal HPV DNA, four (12%) contained integrated and episomal HPV DNA, and the physical state of 13 of the isolates (38%) could not be determined. Episomal HPV DNA on average produced stronger signals on autoradiographs compared to isolates with integrated viral DNA. This suggests that episomal HPV was generally present in much higher copy numbers, although the possibility that these results are accounted for by variable amounts of tumour tissue in samples cannot be discounted. The AA and As lineage isolates appeared to have high copy numbers (which was probably related to the fact that the majority contained episomal HPV DNA), whilst the E lineage isolates had variable copy numbers. An example of a typical agarose gel containing digested cellular nucleic acids before alkaline transfer is given in Figure 4.3. A typical autoradiograph showing the difference between episomal and integrated HPV DNA is presented in Figure 4.4 and the results for physical state are given in Table 4.5.

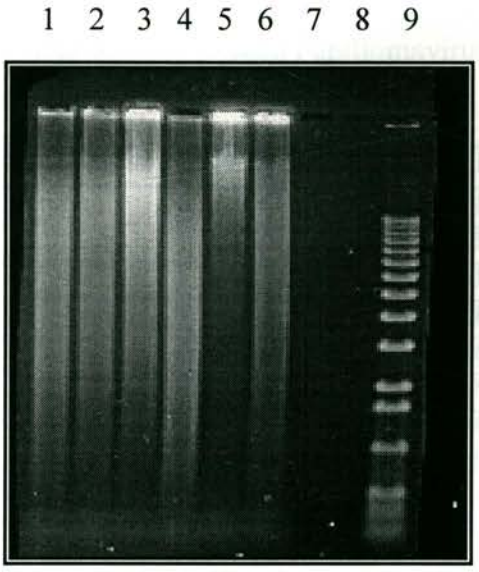



Figure 4.3: An example of cellular nucleic acids digested with *Bgl* II or *Bam*H I, prior to alkaline transfer for Southern hybridisation

Lanes 1 to 6 contain 10µg of cellular nucleic acids digested with *Bgl* II or *Bam*H I. Lanes 1, 3, and 5 contain *Bgl* II digested cellular nucleic acids from three different isolates, whilst lanes 2, 4, and 6 contain *Bam*H I digested cellular nucleic acids from the same three isolates. Lanes 7 and 8 contain 125pg each of HPV 16 control DNA (linearised whole genome), and lane 9 a 1kb DNA Ladder.

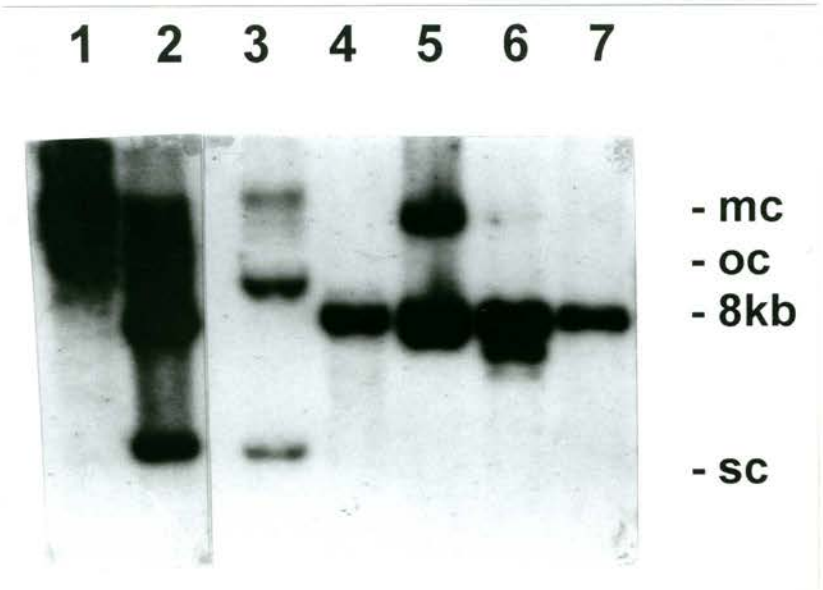



Figure 4.4: Southern hybridisation demonstrating integrated and episomal forms of human papillomavirus DNA

Lanes 1 and 2 represent *Bgl* II and *Bam*H I digested DNA from one isolate containing integrated and episomal DNA; lanes 3 and 4, and 5 and 6 represent episomal DNA from two isolates digested with the same restriction endonucleases, and lane 7 contains the positive control (linearised HPV 16 genome). mc = multimeric circles, oc = open cut circles, and sc = supercoils.

Table 4.5: The Physical State of Human Papillomavirus DNA in Each Isolate

| Isolate | Southern hybridisation | E2 PCR | E6 PCR* | Final result |
|---|---|---|---|---|
| CaSki | Integrated | + | + | Integrated |
| B1 | No Result | - | - | No Result |
| B2 | Integrated | - | + | Integrated |
| C1 | Episomal | + | + | Episomal |
| D1 | No Result | - | - | No Result |
| D2 | No Result | - | - | No Result |
| D3 | Not Tested | - | + | Integrated |
| E1 | No Result | + | + | No Result |
| G1 | Integrated + Episomal | + | + | Integrated + Episomal |
| H1 | No Result | - | - | No Result |
| H2 | Episomal | + | + | Episomal |
| H3 | Integrated | + | + | Integrated |
| H4 | No Result | - | + | Integrated |
| J1 | Episomal | + | + | Episomal |
| K1 | Integrated + Episomal | + | + | Integrated + Episomal |
| K2 | Episomal | + | + | Episomal |
| K3 | Integrated | + | + | Integrated |
| K4 | Episomal | + | + | Episomal |
| K5 | No Result | - | - | No Result |
| K6 | Episomal | + | + | Episomal |
| L1 | No Result | - | - | No Result |
| M1 | Not Tested | + | + | No Result |
| M2 | Integrated | + | + | Integrated |
| M3 | No Result | + | + | No Result |
| M4 | Integrated | - | + | Integrated |
| M5 | Integrated | - | + | Integrated |
| M6 | Episomal | + | + | Episomal |
| O1 | Episomal | + | + | Episomal |
| O2 | Episomal | + | + | Episomal |
| P1 | No Result | - | + | Integrated |
| P2 | No Result | - | + | Integrated |
| R1 | Integrated | - | + | Integrated |
| S1 | Integrated + Episomal | + | + | Integrated + Episomal |
| S2 | Integrated + Episomal | + | + | Integrated + Episomal |
| T1 | Episomal | + | + | Episomal |

*As all isolates were known to contain HPV 16, but not necessarily the HPV 16 E2 gene, PCR of the E6 gene was performed as a positive control to test for the integrity of nucleic acids.

## 4.3.2 The Physical State of Human Papillomavirus DNA by Polymerase Chain Reaction of the E2 Gene

### 4.3.2.1 Sensitivity Assay for the E2 Gene Polymerase Chain Reaction

As the results of PCR using primers E2-1 and E2-2 were used to help determine physical state, and results were recorded on a positive/negative basis, it was necessary to determine the sensitivity of the PCR assay. Sensitivity was assessed by amplifying serial dilutions of 1μg of CaSki nucleic acids (from 1 in 10 to 1 in $10^9$), shown in Figure 4.5.

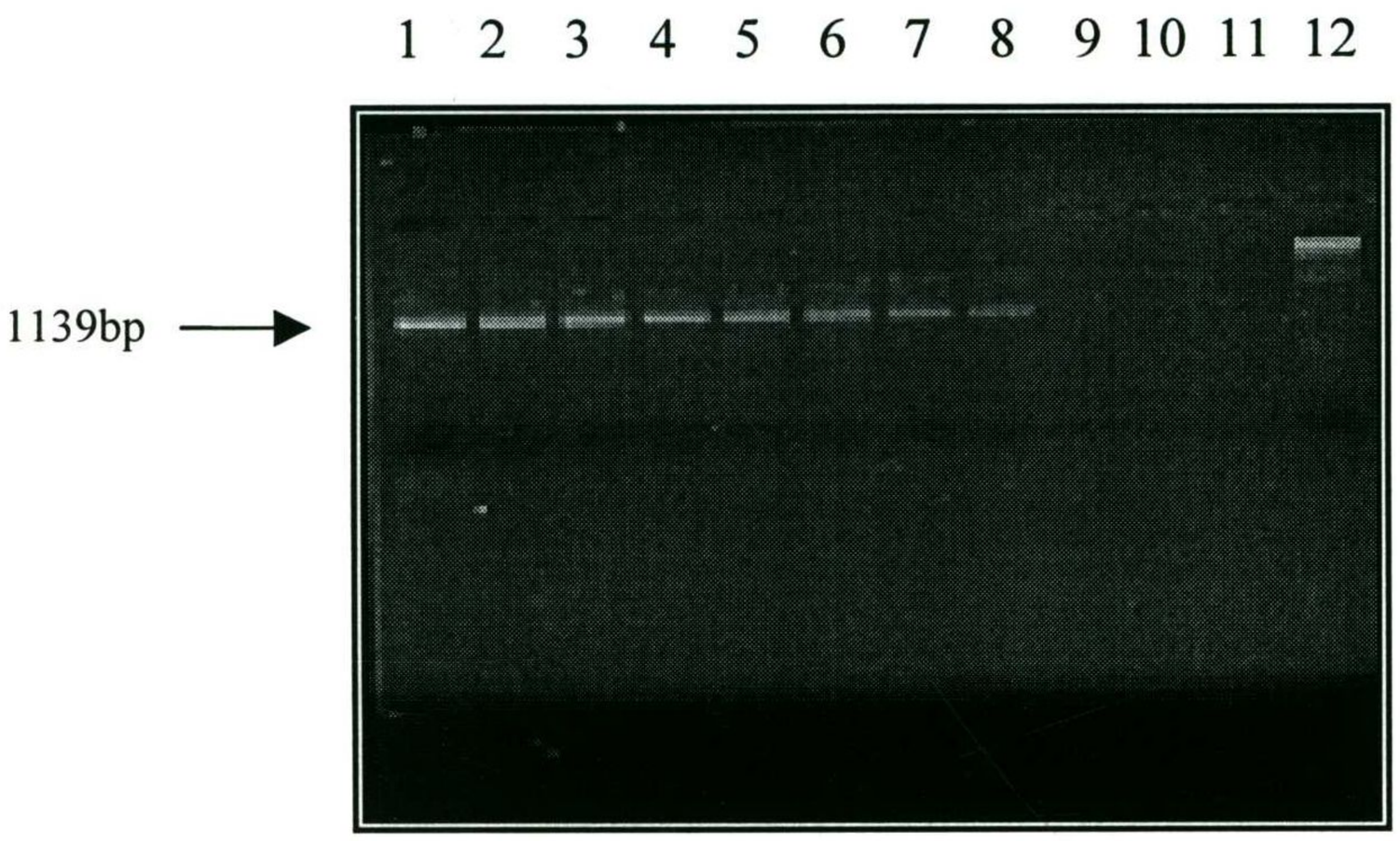



Figure 4.5: Sensitivity assay for E2 polymerase chain reaction (using primers E2-1 and E2-2)

Lanes 1 to 10 contain serial dilutions of CaSki DNA ranging from 1μg (lane 1) to 0.001pg (lane 10). Lane 11 contains the negative control (water) and lane 12 a 1kb DNA Ladder. The end point was determined when a visible band of expected size was no longer present on an agarose gel. In this case, the end point of the assay was at tube 8 (0.1pg of CaSki nucleic acids).

One human cell contains an average of 10pg of DNA (8pg to 17pg range) and 25pg of RNA (20pg to 40pg range), the actual values varying slightly depending on what part of the cell cycle the cell is in. On this basis, the total amount of nucleic acids in a CaSki cell was considered to be equal to 35pg. One CaSki cell also contains approximately 600 copies of HPV DNA (Yee *et al.,* 1985). Using this information and the concentration of total nucleic acids, the sensitivity of the E2 PCR in terms of HPV copy numbers was determined to be less than 10 HPV copies.

#### 4.3.2.2 Results for E2 Polymerase Chain Reaction

Overall, 28 of the original 34 isolates contained amplifiable HPV DNA (determined by PCR of the E6 gene) and of these, 20 (71%) were PCR positive for the entire E2 gene (Table 4.5). The six isolates where no results were obtained contained DNA that was degraded or there was no remaining target to carry out the investigations. As intratypic variation in the E2 gene is very small (sequences vary by only a few percent) it is not likely that sequence variations prevented primer annealing. Where E2 PCR was positive and no result was obtained by Southern hybridisation, it was not possible to determine the physical state of the HPV DNA. Where sufficient tumour DNA was available for testing, all isolates that were identified as episomal by Southern were also positive for full length E2 by PCR. For four of the isolates where no Southern result was obtained, E2 PCR was negative when E6 PCR was positive, indicating that the isolates most probably contained integrated HPV DNA, bringing the number of integrated isolates to 11. Three (27%) of the 11 integrated isolates contained the full length E2 gene (H3, K3, and M2).

#### 4.3.2.3 Restriction Endonuclease Analysis of Polymerase Chain Reaction Products

Cleavage of PCR products produced using primers E2-1 and E2-2 by *Hinc* II resulted in two products corresponding to the expected sizes of 476bp and 662bp for each of the isolates, except K4, where three bands were observed. One of these bands corresponded to the 476bp product, and two bands of approximately 400bp and 200bp were also present. Subsequent sequence analysis revealed that K4 contained sequence variation at nt 3566 that created an extra *Hinc* II restriction site, resulting in the breakage of the 662bp product into the two products observed.

### 4.3.3 Overall Results for Physical State

Overall, using the results from both Southern and PCR of the E2 gene, 11 of the 34 isolates (32%) contained integrated HPV 16 DNA; 10 contained episomal (29%), and four (12%) contained both integrated and episomal HPV 16 DNA. For nine isolates (26%) physical state could not be resolved by either Southern or PCR of the E2 gene. The overall results are represented diagrammatically in Figure 4.6. If only

the isolates where a result for physical state could be obtained are taken into account, the proportion in each category are 11/25 (44%) containing integrated HPV 16 DNA, 10/25 containing episomal (40%), and 4/25 (16%) containing both integrated and episomal HPV 16 DNA. Of the nine isolates for which physical state was unresolved there was either no result by Southern hybridisation and no sample remaining for testing by E2 PCR (six isolates), or the E2 PCR result was positive with no Southern result (three isolates). Interestingly, the physical state of four of five adenocarcinomas could not be determined.

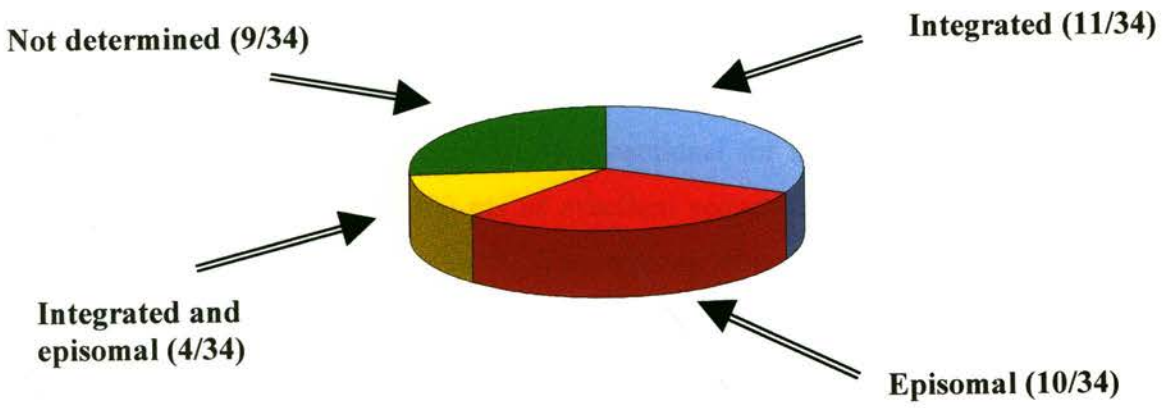



Figure 4.6: The physical state of human papillomavirus type 16 DNA in 34 cervical cancer isolates from Australia and the South Pacific

### 4.3.4 Site of Disruption of E2 in Integrated Papillomavirus Genomes

Eight of the 28 cancer isolates carried integrated HPV 16 DNA without a full-length E2 gene (negative for PCR with primers E2-1 and E2-2). Only two of the eight were positive for PCR using primers E2-2 and E2-3 and five were positive for PCR using primers E2-1 and E2-4 (Table 4.6). From these results it was concluded that of the eight isolates: two lacked an entire E2 ORF, one contained the entire E2 ORF but discontinuously, one was disrupted in the 5' region (the sequence between nt 3028 and nt 3873 was present), and four were disrupted in the 3' region (the sequence between nt 2753 and nt 3341 was present).

Table 4.6: Amplification of Integrated Human Papillomavirus Type 16 DNA for Different Regions of the E2 Gene

| Isolate | E2-1/E2-4 | E2-3/E2-2 | Comment |
|---|---|---|---|
| B2 | + | + | Contains entire E2 sequence but not present continuously |
| D3 | + | - | 3' breakage |
| H4 | - | - | Whole E2 missing |
| M4 | + | - | 3' breakage |
| M5 | - | - | Whole E2 missing |
| P1 | + | - | 3' breakage |
| P2 | - | + | 5' breakage |
| R1 | + | - | 3' breakage |

### 4.3.5 Sequence Variation in the E2 and E4 Genes

The E2 and E4 coding regions from 13 of the cancer isolates were sequenced. Both primers, E2-1 and E2-2, proved to be exceptional for sequencing, each producing approximately 600 nts to 800 nts of excellent sequence data with zero ambiguity codes (examples in Appendix A7.5). The sequence obtained from each of the primers overlapped by several hundred nucleotides. Sequence variation in the E2 and E4 genes for the selected cancer isolates is given in Figure 4.7. The complete nucleotide sequence alignments for E2 and E4 are presented in Appendix A7.6.2.

Thirty nucleotide changes were identified in the E2 gene overall, which is equivalent to 2.7% sequence variation over approximately 1080 nts sequenced. Patterns of sequence variation were lineage specific. As observed for the other regions of the HPV 16 genome sequenced during this study, the E lineage isolates varied least from the prototype, whilst the As and AA isolates contained the largest numbers of nucleotide changes. The two As lineage variants sequenced contained nine nucleotide variations each, whilst the AA lineage variant contained 19. Several of the nucleotide changes identified in the As lineage were also present in the AA lineage. A nucleotide change at 3410 (C to T) was the most common sequence alteration, present in 10 of the 13 isolates (77%) and in all variant lineages represented. The second most common nucleotide change, a C to A at nt 3684 was present in four isolates. In the transactivation region of E2, 12 nucleotide changes were identified, whilst five were identified in the hinge region, and 13 in the DNA binding region.

| | | Transactivation Region | | | | | | | | | | | | Hinge Region | | | | | DNA Binding Region | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variant Lineage | Isolate | 2828 | 2860 | 2938 | 3159 | 3182 | 3205 | 3224 | 3249 | 3362 | 3377 | 3384 | 3410 | 3422 | 3449 | 3516 | 3517 | 3524 | 3538 | 3566 | 3567 | 3575 | 3603 | 3664 | 3684 | 3694 | 3706 | 3778 | 3787 | 3798 | 3805 | |
| E | HPV16R | G | C | A | C | G | T | T | G | A | C | T | C | C | G | C | T | T | A | T | T | T | G | T | C | T | T | G | C | C | T | |
| E | J1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| E | M2 | | | | | | | | | | | | T | | | | | | | | | | | | | | | | | | | |
| E | C1 | | | | | | | | | | | | T | | | | | | | | | | | | | | | | | | | |
| E | K1 | | | | | | | | | | | | T | | | | | | | | | | | | | | | | | | | |
| E | H2 | | | | | | | | | | | | T | | | | | | | | | | | | | | | | | | | |
| E | K2 | | | | | | | | | | | | T | | | | | | | | | | | | | | | | | | | |
| E | H3 | | | | | | | | | | | | T | T | | | | | | | | | A | | | | | | | | | |
| E | K6 | | | | | | A | | | | | | | | | | | | | | | G | | | | | | | | | | |
| E | M6 | | | | | | A | | | | | | | | | | | | | | | G | | C | | | | | T | | |
| E | T1 | | | G | | | | | | | | C | T | | | | | | | | | | | | A | | | | | | | |
| As | G1 | | | | A | | | | A | | | C | T | | A | | | C | | | G | | | | A | | | | A | | | |
| As | S1 | A | | | A | | | | A | | | C | T | | A | | | C | | | | | | | A | | | | A | | | |
| AA | K4 | | A | | A | A | | A | A | G | G | | T | | A | A | C | | C | G | | | | | A | A | C | T | A | | G | |

Figure 4.7: Sequence variation in the human papillomavirus type 16 E2 gene from cervical cancer isolates

Note: the E2 ORF encompasses nts 2756 to 3853, whilst the E4 ORF encompasses nts 3357 to 3617.

### 4.3.5.1 Amino Acid Analysis of the E2 Protein

Overall, 30 nucleotide changes were identified in the E2 gene resulting in 21 nonsynonymous amino acid substitutions, and nine synonymous changes (Table 4.7). Generally, amino acid changes were located across the entire E2 protein (Figure 4.8). Eight of the nine nucleotide changes in one of the As lineage variants, S1, led to amino acid changes, while in the AA variant K4, 14 of the 19 nucleotide changes led to amino acid changes. Nonsynonymous changes were greatest in the transactivation region (10 from 12), intermediate in the DNA binding region (eight from 13), and lowest in the hinge region (three from five).

The common nucleotide change at 3410 leads to an amino acid change (proline to serine). None of the amino acid changes in the E2 gene were found in amino acids conserved across different types of HPV (based on amino acid alignments for SuperGroup A). Several amino acids changes were to others that were common at those positions in related HPV sequences. No frameshift mutations were identified and no nucleotide changes conferred a premature stop codon.

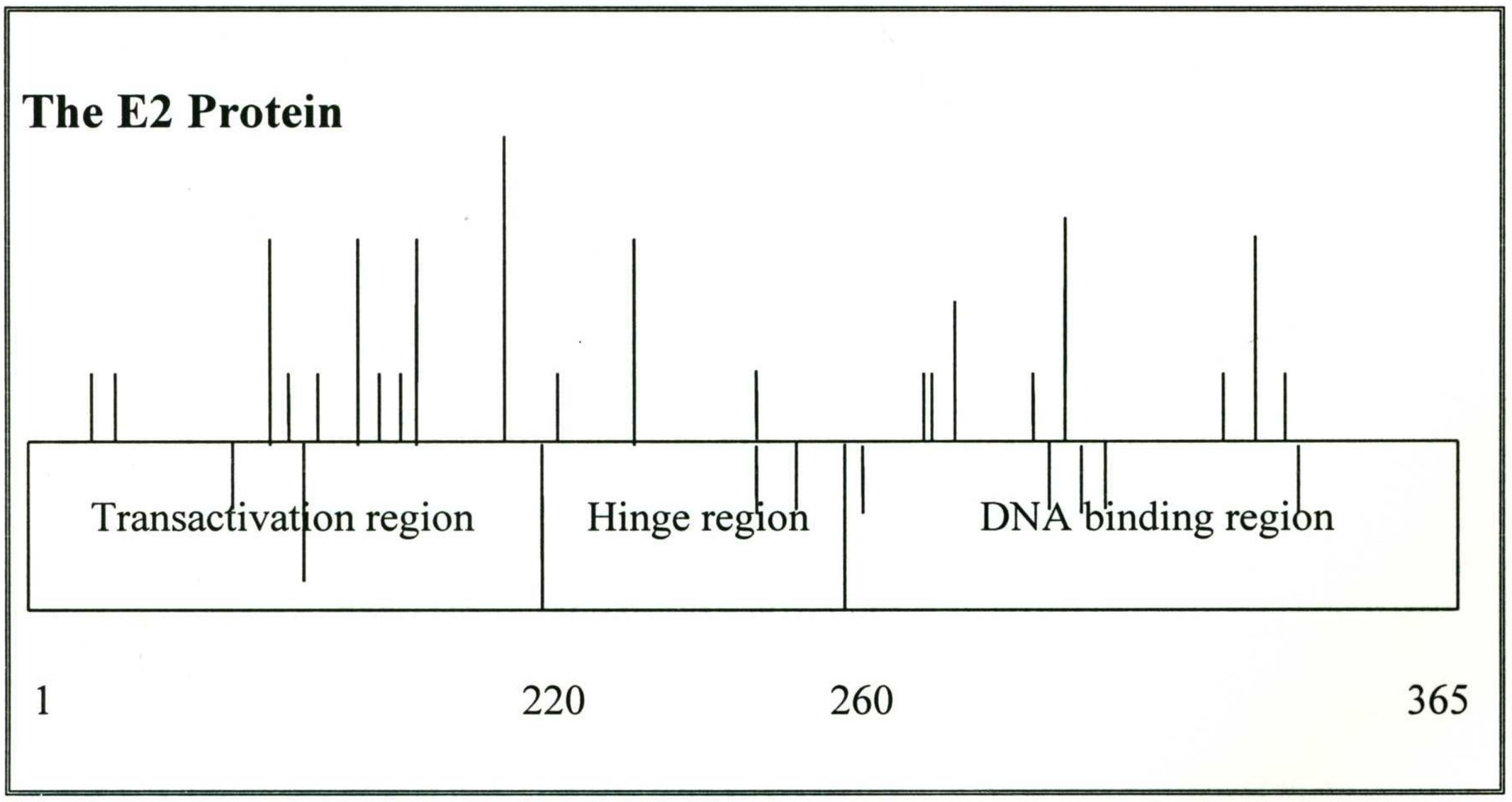



Figure 4.8: Distribution of amino acid changes in the E2 protein

Lines above the protein represent nucleic acid changes that led to amino acid changes (nonsynonymous changes). Lines below the protein represent nucleic acid changes that did not lead to amino acid changes (synonymous changes). The height of the line indicates the number of times that change was found, the larger the line, the more times that change was found among the isolates. The shortest line represents a change identified in only one isolate.

Table 4.7: Amino Acid Changes in the E2 Protein

| Region | Nucleotide | Nucleotide Change | Amino Acid Change | Amino Acid Position |
|---|---|---|---|---|
| Transactivation Region | 2828 | **G**AC to **A**AC | D to N | 25 |
| | 2860 | CA**C** to CA**A** | H to Q | 35 |
| | 2938 | AC**A** to AC**G** | No Change | 107 |
| | 3159 | **A**CA to **A**A**A** | T to K | 135 |
| | 3182 | **G**CA to **A**CA | A to T | 143 |
| | 3205 | GG**T** to GG**A** | No Change | 150 |
| | 3224 | **T**TA to **A**TA | L to I | 157 |
| | 3249 | C**G**A to C**A**A | R to Q | 165 |
| | 3362 | **A**AC to **G**AC | N to D | 203 |
| | 3377 | **C**CT to **G**CT | P to A | 208 |
| | 3384 | A**T**T to A**C**T | I to T | 210 |
| | 3410 | **C**CC to **T**CC | P to S | 219 |
| Hinge Region | 3422 | **C**AT to **T**AT | H to Y | 223 |
| | 3449 | **G**AA to **A**AA | E to K | 232 |
| | 3516 | A**C**T to A**A**T | T to N | 254 |
| | 3517 | ACT to AC**C** | No Change | 254 |
| | 3524 | **T**TG to **C**TG | No Change | 257 |
| DNA Binding Region | 3538 | TC**A** to TC**C** | No Change | 261 |
| | 3566 | **T**TT to **G**TT | F to V | 271 |
| | 3567 | T**T**T to T**G**T | F to C | 271 |
| | 3575 | **T**CA to **G**CA | S to A | 274 |
| | 3603 | A**G**T to A**A**T | S to N | 283 |
| | 3664 | TA**T** to TA**C** | No Change | 303 |
| | 3684 | **A**CA to **A**A**A** | T to K | 310 |
| | 3694 | AC**T** to AC**A** | No Change | 313 |
| | 3706 | TC**T** to TC**C** | No Change | 317 |
| | 3778 | TG**G** to TG**T** | W to C | 341 |
| | 3787 | GA**C** to GA**A** | D to E | 344 |
| | 3798 | **T**CT to **T**T**T** | S to F | 348 |
| | 3805 | GT**T** to GT**G** | No Change | 350 |

Nucleotide changes are in bold.

### 4.3.5.2 Amino Acid Analysis of the E4 Protein

The HPV 16 E4 gene is entirely contained within the E2 gene, although translated in a different frame. Nucleotide changes in the E4 gene leading to amino acid changes are presented in Table 4.8.

Table 4.8: Amino Acid Changes in the E4 Protein

| Nucleotide | Nucleotide Change | Amino Acid Change | Amino Acid Position |
|---|---|---|---|
| 3362 | GC**A** to GC**G** | No Change | 2 |
| 3377 | CT**C** to CT**G** | No Change | 7 |
| 3384 | **T**TA to **C**TA | No Change | 10 |
| 3410 | AA**C** to AA**T** | No Change | 18 |
| 3422 | CC**C** to CC**T** | No Change | 22 |
| 3449 | CC**G** to CC**A** | No Change | 31 |
| 3516/3517* | **CT**A to **AC**A | L to T | 54 |
| 3524 | TG**T** to TG**C** | No Change | 56 |
| 3538 | CA**G** to CC**G** | Q to P | 61 |
| 3566 | CA**T** to CA**G** | H to Q | 70 |
| 3567 | **T**TA to **G**TA | L to V | 71 |
| 3575 | GC**T** to GC**G** | No Change | 73 |
| 3603 | **G**TA to **A**TA | V to I | 83 |

* Individual nucleotide changes led to different amino acid changes, however as they were present in the same isolate they effectively led only to one amino acid change.
Nucleotide changes are in bold.

Fourteen nucleotide changes were identified in the E4 gene resulting in five nonsynonymous amino acid substitutions, and nine synonymous changes. In the corresponding region of E2 (all of the E2 hinge region and small parts of the transactivation and DNA binding domains), 11 nonsynonymous amino acid changes were identified compared with five, so that nonsynonymous substitutions were more common in E2 than E4. None of the nucleotide changes in the amino terminal region of E4 led to amino acid changes (Figure 4.9). Three of the five amino acid changes in the E4 protein were observed only in K4. The other two amino acid changes were identified in an As lineage variant (G1) and an E lineage variant (H3). The amino acid change in G1 was located towards the carboxyl terminal of E4 (amino acid 70) and created a new start codon. Amino acid alignments for the E2 and E4 proteins are given in Appendix A7.6.2.

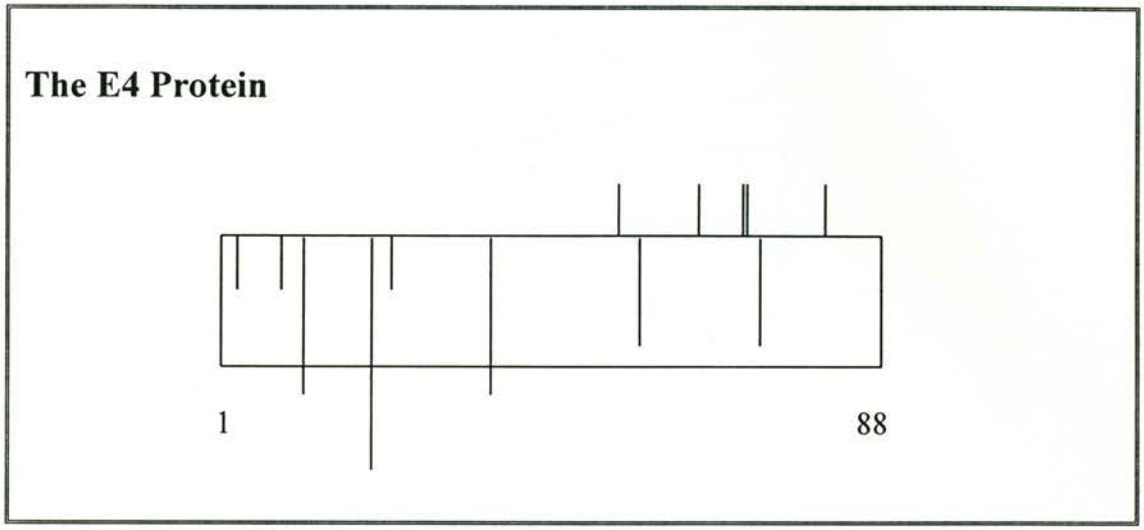



Figure 4.9: Distribution of amino acid changes in the E4 protein

Lines above the protein represent nucleic acid changes that led to amino acid changes (nonsynonymous changes). Lines below the protein represent nucleic acid changes that did not lead to amino acid changes (synonymous changes). The height of the line indicates the number of times that change was found, the larger the line, the more times that change was found among the isolates. The shortest line represents a change identified in only one isolate.

**4.3.5.3 Sequence Analysis of Noncancer Human Papillomavirus Type 16 Isolates**

PCR using primers E2-1 and E2-2 and the noncancer isolates, PRE262, PRE250, and PRE259 failed. This suggested either that nucleic acids extracted from samples fixed on slides were not suitable for amplification of the 1139bp segment or that the PCR was not optimal for crude cellular extracts. Disruption of the 5' region of the E2 gene due to integration of HPV was considered very unlikely, as the isolates were not derived from high grade CIN or cancer. Primer E2-3 when used in combination with primer E2-2 produces a smaller PCR product of 814bp, lacking the first 325bp of the E2 gene. PCR using these primers yielded satisfactory results for isolates PRE250 and PRE262, but the product for PRE259 contained non-specific bands and was not analysed further. Sequence data using primer E2-2 was obtained for PRE250 and PRE262 (Table 4.9). Nucleotide and amino acid alignments for E2 and E4 derived from the noncancer isolates are given in Appendix A7.6.2.

Table 4.9: Sequence Variation in the Human Papillomavirus Type 16 E2 Gene from Noncancer Isolates

| Variant Lineage | Isolate | 3244 | 3384 | 3410 | 3684 |
|---|---|---|---|---|---|
| E | HPV16R | A | T | C | C |
| E | PRE250 | | | T | |
| As | PRE262 | C | C | T | A |

Isolate PRE250 contained only the common nucleotide variation at 3410, which was also the case with five of the 13 cancer isolates, all of which were E lineage variants (or five of 10 E lineage cancer isolates). PRE262 contained four nucleotide variations, three of which (at nts 3384, 3410, and 3684) were present in both As and one E lineage cancer isolates. The other sequence change in PRE262, that at nt 3244, was unique to the isolate, although it did not result in a change in amino acid sequence at 163 (GG**A** to GG**C**) in the E2 protein. Nucleotide changes at 3384, 3410, and 3684 resulted in amino acid changes in the E2 protein, but not in the E4 protein (see Tables 4.7 and 4.8). To determine the variant lineages to which the noncancer HPV 16 variants belonged, the enhancer of the URR was sequenced (nts 7457 to 7810) using primer URR-3 (as described in Chapter 2). Sequence alignments for the URR of these isolates are given in Appendix A7.6.2. From the sequence changes in each URR (Table 4.10), it is likely that isolates PRE262 and PRE250 are E lineage variants, whilst PRE259 probably belongs to the As lineage.

Table 4.10: Sequence Variation in the Human Papillomavirus Type 16 Upstream Regulatory Region from Noncancer Isolates

| Isolate | Variant Lineage | Sequence Variation (nt) | |
|---|---|---|---|
| | | 7521 | 7730 |
| HPV 16R | E | G | A |
| PRE250 | E | A | |
| PRE262 | E | A | |
| PRE259 | As? | A | C |

# 4.4 DISCUSSION

## 4.4.1 The Physical State of Human Papillomavirus Type 16 DNA in Cervical Cancers

Woodworth *et al.* (1989) published some of the earliest indications that integration of the PV genome is important for immortalisation, and that the URR, and the E6 and E7 genes are invariably retained. The expression of HPV 16 E6 and E7 genes can be identified in cervical carcinoma cell lines and tumours irrespective of the physical state of the viral DNA. The proportion of integrated isolates identified in this study (44%) was lower than that identified in most other studies of HPV 16 cervical cancer isolates (53% to 100%; refer to Table 1.3). It is likely that most of the isolates where physical state could not be resolved (nine isolates) contained integrated HPV DNA, since evidence suggests that HPV copy number is related to the physical state of the viral genome, with episomal DNA associated with the highest copy numbers, integrated DNA with the lowest, and both forms with the intermediate range (Vernon *et al.,* 1998). The proportion of episomal isolates identified (40%) was higher than that in most studies (9.8% to 27%; refer to Table 1.3). However the proportion of episomal isolates in this cohort is probably an over representation because of likely under-detection of integrated HPV 16. The proportion of cancers containing both integrated and episomal HPV 16 DNA (16%) was within the limits of published data (4.75% to 22%; refer to Table 1.3). It is possible that variations in the physical state of HPV DNA in different studies reflect differences in ethnic groups, geographical locations, or the presence of different viral variants.

Failure to detect the entire E2 ORF was taken as a presumptive indication of integration of HPV DNA in this study. However, some reports suggest that E1 or E2 deletions are not confined exclusively to integrated HPV 16, or to invasive cancers (Wagatsuma *et al.,* 1990, Cullen *et al.,* 1991, and Krajinovic *et al.,* 1993). It is conceivable that deletion of the E2 gene may occur independently of the integration event. However all episomal isolates identified in this study contained the entire E2 ORF, findings consistent with those of another study where E2 mRNA was identified in all cancers in which the viral DNA was episomal or episomal and integrated (Park *et al.,* 1997).

In this study, it was of particular interest that K2 and K4, the two isolates with substantially upregulated URR activity, carried HPV DNA episomally. It was not possible to determine the physical state of HPV 16 DNA in H1, which also exhibited upregulated promoter activity. Of the isolates tested in function studies that had promoter activity comparable with the prototype, R1 contained integrated HPV and S1 contained integrated and episomal HPV. These results support the work of Dong *et al.* (1994) where HPV 16 URRs exhibiting upregulated promoter activity were found in episomal viral isolates. In that study, the HPV 16 URR from four out of seven cervical cancer isolates with purely episomal viral DNA had deletion of, or sequence variation in, YY1 binding sites. Sequence variations in YY1 sites were not very common in this cohort, but were present in cancers containing episomal viral DNA, such as in K2 and K4. However, the binding of YY1 to these sites (YY1 4 and 5) was not altered. SiHa, a cell line containing integrated HPV 16 DNA, has a 38bp deletion that removes one YY1 site, so this type of variation may not be restricted to episomal cancers and it has been suggested that these changes may occur as advanced dysplasia develops (Pfister, 1996). In this study, sequence variation in another motif important for negative regulation of the HPV 16 URR, the PSM, mediated upregulated promoter activity in the episomal isolate K4. These results substantiate the important observation that alterations in TFBSs play a significant role in over-expression of the viral oncogenes in episomal virus and therefore may influence the rate of tumour progression or the aggression of established tumours.

Interestingly, a comparison of URR sequence variation patterns with viral physical state reveals that isolates containing only the nucleotide change at 7764 from prototype URR sequence (K6 and M6) were episomal. Conversely, isolates containing sequence changes at nts 7193, 7521, and 7764 (H4, M4, and M2) were integrated. However, of those isolates containing sequence variation only at nts 7193 and 7521, a mixture of viral physical states was observed. While nt 7193 falls within a proposed TEF-1 binding site, nts 7521 and 7764 do not fall within known TFBSs and it is not clear whether these sequence variation patterns are associated with viral physical state; alternatively this may just be an incidental finding.

Although the O2 URR, which contained the large-scale duplication, did not display increased promoter activity compared to the prototype, the HPV DNA in O2 was carried episomally. An episomal genome in a study by Hall *et al.* (1997) containing a similar but larger duplication of the URR, exhibited transforming activity comparable with the HPV 16 prototype. Transforming capacity was found to be dependent on the presence of the duplicated URR sequences though, suggesting that duplications may also achieve the functional equivalence of integration. Although transformation assays were not conducted in this study, it is tempting to speculate that the O2 URR might likewise exhibit transforming ability dependent on its duplication.

There are two distinct forms of integrated HPV DNA; either insertion of single genome/s, usually resulting in disruption of the E1 or E2 ORF, or integration involving multiple copies of HPV in head-to-tail tandem repeats, derived from a multicopy concatemer episome. In the latter form of integration, only the viral copy flanking the cellular DNA is interrupted in the E1 or E2 region and the internal copies have intact E1 and E2 ORFs (Wagatsuma *et al.,* 1990, and Jeon and Lambert, 1995). Although not investigated directly, it is likely that this type of viral integration occurred in the integrated isolates in this study containing the entire E2 ORF, namely H3, K3, and M2. Integrated HPV DNA poses some interesting questions regarding alterations to HPV gene expression. It is evident that HPV promoters, once integrated, may be under altered regulation due to controls exerted by cellular DNA. In one study of HPV 16 positive tumours there was no difference in the variety of mRNAs or splicing patterns between integrated or episomal HPV 16 (Sherman *et al.,* 1992). However in the cell line SiHa, the early region polyadenylation signal in the integrated HPV 16 DNA is no longer downstream of the early viral promoter (Baker *et al.,* 1987). In this instance HPV 16 sequences must be spliced to RNA from adjacent human DNA, and transcription must use cellular polyadenyation sites, which may have played a role in malignant progression (Baker *et al.,* 1987). In CaSki cells, all HPV 16 ORFs are present, but there is no evidence of transcription of the late region. CaSki HPV 16 DNA is arranged in tandem arrays. However only the terminal HPV 16 copies in the large tandem repeats may be transcriptionally active, which is perhaps due to methylation of integrated HPV DNA. As in SiHa, CaSki may not utilise the viral polyadenylation site (Baker *et al.,* 1987), which is important

because the level of polyadenylation controls 'translatability' which is a mechanism for the control of gene expression. In addition, transient expression of the E2 protein in cervical carcinoma cell lines such as HeLa causes the cells to become significantly growth-inhibited (von Knebel Doeberitz *et al.*, 1991, and Desaintes *et al.*, 1997), however other cell lines such as C41 are only marginally growth-inhibited. This further illustrates that different cellular control elements can dominate over viral control elements in the transcriptional regulation of integrated virus.

### 4.4.2 Disruption of the E2 Gene

Unlike HPV 16, in HPV 18, 31, and 35 positive cervical tumours, the viral DNA always appears to integrate (Cullen *et al.*, 1991, Berumen *et al.*, 1994, Park *et al.*, 1997, and Pirami *et al.*, 1997). HPV 18 seems to integrate at a breakpoint in the E1/E2 boundary producing an approximately 4000bp integrant, with loss of part or all of the E2 ORF, all of the E4, E5, L2, and part of the L1 ORFs (Corden *et al.*, 1999). HPV 18 integration also appears to occur at a preferential site within the host genome (Couturier *et al.*, 1991, and Corden *et al.*, 1999). The fact that HPV 18 invariably integrates may be related to its greater transforming efficiency *in vitro*, and probably has something to do with the greater frequency of complete loss of E2 protein and greater deregulation of E6/E7 transcription (Cullen *et al.*, 1991). The total region of integrated viral DNA appears to be much less conserved for HPV 16 (particularly in the L1 and E1 regions). As in other oncogenic HPV types, a minimal conserved region of viral integration of 2,745bp has been identified for HPV 16 that encompasses the URR and the E6 and E7 genes (Chen *et al.*, 1994b). A general deletion domain of 1,465bp was defined in that study between nts 1417 and 2881, covering most of the E1 gene including the 3' end as well as 60bp at the 5' end of the E2 gene. In other studies on HPV 16, various deleted lengths of the viral genome were found in integrated isolates, with preferential loss of the E1 and E2 ORFs (Awady *et al.*, 1987, Choo *et al.*, 1987 and 1988, Wilczynski *et al.*, 1988, Vernon *et al.*, 1997, and Kalantari *et al.*, 1998). Kalantari *et al.* (1998) reported that 23% of HPV 16 positive tumours could not be amplified with E1 primers and 29% could not be amplified with E2 primers. Similarly, 29% of isolates (for which amplifiable

DNA was available) could not be amplified with primers for the entire E2 gene in this study.

In the Kalantari study (1998), five patterns of disruption were identified involving all of the E2 gene or subsections of the 5', central, or 3' end. When the E1 or E2 genes were discontinuous, 5' disruption was the most common scenario. Disruption of the 5' region of E2 was also commonly observed in at least one other study (Chen *et al.*, 1994b). The 5' sense primer used for amplification of the E2 gene in this study was located upstream of this reported break in the 5' region of E2 (Chen *et al.*, 1994b) and may provide an explanation for the fact that integrated HPV DNA was not amplifiable using this primer. However, PCR using primer E2-3, which is located further downstream, was expected to be positive if only a small proportion of the 5' region of E2 was not present in the cancer isolates. Interestingly, 5' disruption was identified in only one of the patient isolates, and four of the eight isolates without an entire E2 gene were found to be disrupted in the 3' region of E2 instead. The 5' region of the HPV 16 E2 protein contains replicative and E1 interaction functions, so disruption of the E2 gene leads to the cessation of viral replication functions. As there may be a switch between viral transcription and replication, disruption may lead to permanent transcription of the viral genome, greatly assisting malignant progression.

Some clinical associations have been made with the physical state of the virus. In some studies cervical carcinoma patients whose isolates contain disrupted E1/E2 have reportedly had poorer disease-free survival then patients without such disruptions (Vernon *et al.*, 1997, and Kalantari *et al.*, 1998). In other surveys, women with stage Ib cancer were reported to have a shorter disease-free survival period if their tumours contained integrated HPV than if they contained episomal, or episomal and integrated DNA (Unger *et al.*, 1995, and Vernon *et al.*, 1997). However, in the Vernon study there was no appreciable influence of E2 status on disease-free survival. This may be explained by methylation of integrated HPV DNA that reduces or abolishes transcription of intact E2 DNA, resulting in a lack of E2 function. This is consistent with reports that sometimes when the E2 gene is intact in cancer isolates, there is little evidence of mRNAs encoding the protein (Sherman *et al.*, 1992, Stoler *et al.*, 1992, Böhm *et al.*, 1993, Fuchs and Pfister, 1994, and Daniel *et*

*al.,* 1995). It is not known whether the E2 gene was transcribed in H3, K3, and M2, the three isolates containing integrated HPV DNA and an intact E2 gene in this study.

### 4.4.3 The Sequence of the E2 and E4 Genes

Nucleotide changes identified in the E2 and E4 genes in this study were often lineage specific and there were no obvious differences in the patterns of nucleotide variation between the E2 genes from integrated and episomal isolates. The sequence of the E2 region from noncancer isolates was also similar to the corresponding region in the E lineage from cancer isolates, although the sample size was too small to permit any valid conclusions. Across the E2 region from the cancer isolates, 2.7% of nucleotides varied and E2 multivariants made up 54% of the cohort (seven of the 13 isolates containing more than one nucleotide change). Nucleotide changes and nonsynonymous amino acid substitutions were greater in E2 than in E4. These nonsynonymous substitutions, or amino acid replacing changes, may provide information about negative or positive selection pressures. On the other hand, synonymous amino acid substitutions or 'silent' mutations may provide a measure of genetic variability under minimal selection pressure. These results support those of Eriksson *et al.* (1999) who suggested that E4 might be subject to negative selection pressure that does not affect the corresponding region of E2. Eriksson and colleagues also found no amino acid variations in the amino terminal of the E4 protein, which is responsible for associating with tonofilaments resulting in disruption of the cytokeratin filament network (Rogel-Gaillard *et al.,* 1992, and Robert *et al.,* 1994). Likewise, no amino acid changes in the amino terminal region of E4 were identified in this study.

### 4.4.4 The Significance of E2 Sequence Variants

The loss of E2 probably enhances transformation, reducing the number of time-dependent events required for oncogenesis. However it is also possible that episomal virus acquires mutation/s in the E1 and/or E2 ORFs that have little or no effect on viral replication, but result in the commitment of cells to a deregulated proliferation

state similar to that found in cells with integration-disrupted E1/E2 (Chen *et al.*, 1994b). "The E2 protein is a critical regulator of transcription and replication and any changes in its function are likely to have important consequences for the life cycle of the virus as well as the clinical manifestations of the infection" (Kovelman *et al.*, 1999).

Although a few recent studies have considered the biological importance of sequence variation in the E2 gene, there is much less information currently available for E2 than for the URR or for E6 and E7. One of few epidemiological studies so far performed for HPV 18 positive cervical lesions identified alterations in the sequence of E2 with potential to affect transforming ability (Hecht *et al.*, 1995). A HPV 16 E2 variant with a truncated carboxyl terminal DNA binding domain has also been isolated from normal keratinocytes that was unable to immortalise cells (Storey *et al.*, 1992). Investigations on HPV 6 E2 proteins from condylomata acuminata isolates demonstrated that amino acid changes in the three protein domains did not affect its transcriptional activation function (Kovelman *et al.*, 1999). *In vitro* studies on the HPV 16 E2 protein have been performed to show that mutation of the E2 protein, particularly in the transactivation region, can increase immortalisation capacity (Romanczuk and Howley, 1992). Similarly, Sakai *et al.* (1996) identified single amino acids in engineered HPV 16 E2 mutants at the amino terminal of the transactivation region that affected transactivation activity and were likely to be defective in important protein-protein interactions. None of these particular amino acid changes were identified in this study of naturally occurring sequence variants, although amino acid changes in the transactivation region of the E2 protein were relatively common in isolates from the As and AA lineage variants. The functional importance of sequence variations in the hinge region of the E2 protein is less clear, although they could potentially affect the three dimensional relation between the transactivation and DNA binding domains (Graham and Herrington, 2000). Of interest, in the study by Graham and Herrington there was no relationship between the presence of E2 mutations and the histological diagnosis of either CIN 3 or carcinoma.

HPV 16 AA variants from cervical cancers contain E2 sequences quite distinct from E lineage variants (Casas *et al.*, 1999, and Veress *et al.*, 1999). The sequence of the

E2 gene was obtained for one of only two AA lineage variants identified in this study (K4), so comparisons cannot be made with these studies, however it should be noted that the sequence of K4 was quite distinct from the E lineage variants examined, both in terms of numbers of nucleotide changes and distribution across the three domains of the E2 gene. It is evident that E2 is under evolutionary selection pressure and AA variants may also be functionally different from variants of the other HPV 16 lineages. It was suggested by Casas *et al.* (1999) that E2 variation found in the AA lineage may be an alternative mechanism for deregulation of oncogene expression, since the AA lineage variants in their study were associated with a high viral copy number and retention of the E1/E2 genes. Retention of E1 and/or E2 has also been correlated with high copy number in several other studies (Berumen *et al.*, 1994, and Hecht *et al.*, 1995). Veress *et al.* (1999) speculated that genetic changes in E2 might lead to upregulation of HPV transcription in episomal HPV genomes with intact and transcriptionally active E2 regions. However their co-transfection studies, determining p97 activity in the presence of the E2 protein, showed that there was no difference between E and AA variants, suggesting that sequence variation in E2 did not affect p97 activity to any great extent. This is consistent with the fact that none of the amino acid changes in E2 were in positions conserved among HPVs (Veress *et al.*, 1999).

A T helper cell response was recently identified against the carboxyl terminal domain of the HPV 16 E2 protein (Bontkes *et al.*, 1999b). This response frequently occurs at the time of virus clearance, but whether this response is responsible for viral clearance is not known. E2 protein variants could have structural differences in immunological epitopes that affect responses such as this, which assist viral persistence. In at least one study, HPV 16 and 18 E2 variants occurred more frequently in individuals with certain HLA DR/DQ haplotypes (Terry *et al.*, 1997). As the majority of nucleotide changes identified in this study led to amino acid changes in the E2 protein, it is conceivable that the variants identified could differ immunologically.

In summary, the physical state of HPV DNA in each of the cervical cancer isolates, and the presence of the E2 ORF were investigated in this part of the study. Isolates with upregulated promoter activity and that with large-scale change in the URR were

found to carry only episomal HPV DNA, suggesting that these types of sequence alterations may assist progression to malignancy when the E2 gene is translated. Sequence variation in the putatively translated E2 and E4 proteins potentially affecting their functions were also identified. Sequence variation in other coding regions of the viral genome may also play a part in malignant progression or tumour behaviour, particularly if the E2 protein is present in a cancer. Alterations to the sequence of the E6 oncogene and possible functional consequences are therefore presented in the next chapter of this study.

# CHAPTER 5

# SEQUENCE ANALYSIS OF THE HUMAN PAPILLOMAVIRUS TYPE 16 E6 ONCOGENE

# 5.1 INTRODUCTION

The oncogenic potential of HPVs is determined by the activity of their E6 and E7 proteins. The E6 protein, in a complex with the cellular protein E6-AP, targets p53 for degradation via the ubiquitin-mediated pathway, which circumvents p53-mediated growth arrest. This viral function is aimed at controlling the activities of p53 to assist the successful replication of HPV. However, failure of the virus to balance p53 levels results in genetic instability, which can ultimately lead to immortalisation and tumourigenesis. The E6 protein also binds E6BP, which appears to be a vitamin D receptor-associated factor. Most E6 function studies have been performed using the HPV 16 prototype strain, and although the sequence of the E6 gene is relatively stable, there is both epidemiological and molecular evidence for sequence variation imparting different biological properties. Studies have shown that mutations in the amino terminus, zinc fingers, or central region of E6 can severely reduce its ability to interact with p53 (Foster *et al.,* 1994). Conrad Stöppler *et al.* (1996) also demonstrated that naturally occurring variants of HPV 16 E6, particularly AA lineage variants with sequence changes predominately at the amino terminus, differ biologically and biochemically. E6 variants, including those carrying a nucleotide change at 350 (T to G), have also been associated with persistence and progression of CIN (Londesborough *et al.,* 1996, and Zhebe *et al.,* 1998b). However, the significance of the nt 350 variant (350G) remains controversial, with suggestions that it may have high oncogenicity in some populations but low oncogenicity in others (Zhebe *et al.,* 1998a). It has also been suggested that a T at nt 350 (350T) in association with the codon 72 Arg/Arg p53 protein genotype may be a risk factor for progression to cervical cancer (Brady *et al.,* 1999, and van Duin *et al.,* 2000). Single amino acid changes in the E6 protein may also be sufficient to convert an immunogenic protein to a non-immunogenic variant, thereby assisting the virus to subvert the immune response. For instance, an association has been made between a transition at nt 131 (A to G) in the HPV 16 E6 gene, and HLA-B7 with poorer outcome for cervical cancer patients (Ellis *et al.,* 1995 and 1997).

The aims of this part of the study were to catalogue nucleotide and amino acid variations within the E6 gene of the 34 HPV 16 positive cervical cancer isolates. The

potential importance of sequence changes on host immune response and the function of the E6 protein is discussed.

# 5.2 METHODS

## 5.2.1 Clinical Isolates

Nucleic acids were obtained from 27 of the 34 fresh tumour samples by phenol extraction as described in section 2.2.2.1. There was insufficient sample for analysis of the E6 gene in the remaining seven isolates.

## 5.2.2 Polymerase Chain Reaction of the E6 Gene

Primers designated E6-1 and E6-2 and published by Wheeler *et al.* (1997) were used to amplify the entire E6 gene product (nts 104 to 560) and the 5' region of the E7 gene (nts 562-640) (Table 5.1). The PCR protocol (Table 5.2) was identical to that optimised previously for paraffin-embedded tissue extracts (Stephen *et al.*, 2000) except that the annealing temperature was raised from 52°C to 58°C. PCR controls were as described in sections 2.2.5.2 and 2.2.5.3, and precautions to minimise PCR cross-contamination as described in section 2.2.3.2.

Table 5.1: Primers for Polymerase Chain Reaction of the E6 Gene

| Primers | Genome Positions | Nucleotide Sequence (5'-3') | Product Size (bp) |
|---|---|---|---|
| E6-1# | 56-72 | ACCGGTTAGTATAAAAG | 584 |
| E6-2 | 640-625 | GCTCATAACAGTAGAG | |

# Primers published by Wheeler *et al.*, 1997

## 5.2.3 DNA Sequencing

Electrophoresis, PEG precipitation, DNA sequencing, and computer analysis of sequence data were performed as described in Chapter 2. Procedures to ensure the validity of the sequence data were as described in section 2.2.10.

Table 5.2: Polymerase Chain Reaction Cocktail and Thermal Cycling Conditions for Primers E6-1 and E6-2 (E6 Gene Product)

| Per Reaction: | Volume Added: | Final Concentration/Amount: |
|---|---|---|
| Water | 30.1µl | - |
| 10x Reaction Buffer | 5µl | 1x |
| 25mM $MgCl_2$ | 5µl | 2mM |
| Primer E6-1 (200ng/µl) | 2µl | 1.2µM |
| Primer E6-2 (200ng/µl) | 2µl | 1.2µM |
| dNTPs | 0.5µl | 250µM |
| *Amplitaq*® DNA Polymerase | 0.4µl | 2 Units |
| | ____ | |
| | 45µl PCR Cocktail + 5µl Target DNA (100ng/µl) | |

Thermal Cycling Conditions:

| | |
|---|---|
| Cycle 1 | Step 1 94°C/4 minutes (Denaturation) |
| | Step 2 58°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycles 2-39 | Step 1 94°C/0.5 minutes (Denaturation) |
| | Step 2 58°C/1 minute (Annealing) |
| | Step 3 72°C/1 minute (Extension) |
| Cycle 40 | Step 1 94°C/0.5 minutes (Denaturation) |
| | Step 2 58°C/1 minute (Annealing) |
| | Step 3 72°C/10 minutes (Extension) |

### 5.2.4 E6 Protein Analysis

Nucleotide variations leading to amino acid changes were determined using the programs Translate and Eclustalw in ANGIS.

To determine the possible consequences of amino acid changes in terms of function of the E6 protein, it may be important to know if the change results in a shift in the 'type' of amino acid "R" group, or side chain, eg. a hydrophobic residue to an acidic one. Alterations in amino acids can have an effect on hydrophilicity, the probability that the amino acid will be on the surface of the protein (surface probability), flexibility, and antigenic index. Each of these parameters was examined using PeptideStructure software in ANGIS, which makes secondary structure predictions for a peptide sequence.

# 5.3 RESULTS

## 5.3.1 E6 Polymerase Chain Reaction

### 5.3.1.1 General Polymerase Chain Reaction Results

Single bands of expected size were obtained for 26 of the 27 isolates (an example is given in Figure 5.1). The remaining isolate, G1, contained two E6 PCR products, the expected product of 584bp and another of approximately 700bp. The band on the gel corresponding to the expected size PCR product was cut from the gel (as described in section 3.2.5.2) and sequenced. It was not possible to sequence the product corresponding to the larger sized band, as this band on the gel was faint and there was insufficient material for sequence analysis. Attempts to amplify the larger band using a nick of gel containing amplified DNA as target were unsuccessful.

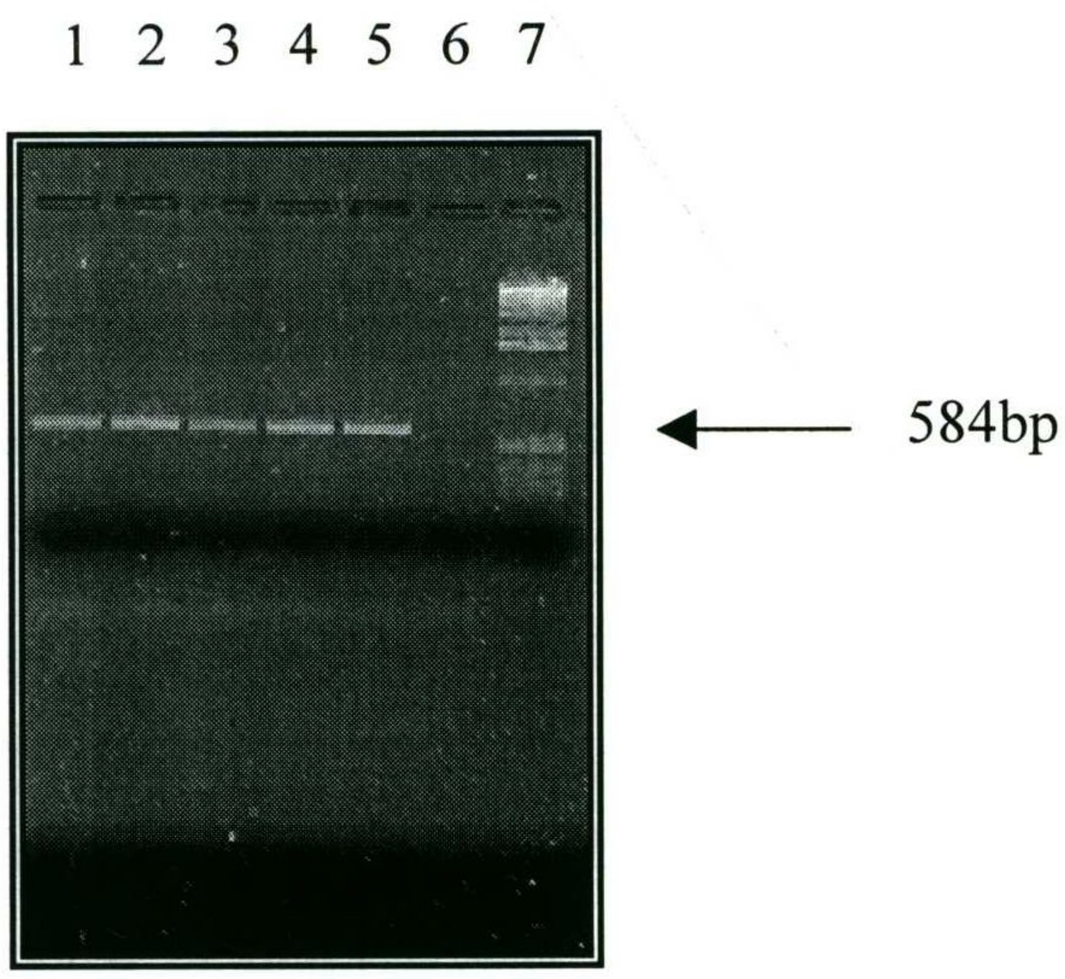



Figure 5.1: An example of E6 polymerase chain reaction products

Lanes 1 to 4 contain products from patient nucleic acids. Lane 5 contains the CaSki positive control (100pg), lane 6 is the negative control (water) and lane 7 a 1kb DNA Ladder.

## 5.3.2 Sequence Variation in the E6 Gene

The sequence of the E6 gene (from approximately nts 73 to 615) was determined for the 27 isolates (Figure 5.2). Primer E6-2 was used for sequence analysis and

routinely provided between 400 and 500 nucleotides of good quality DNA sequence (an example is given in Appendix A7.5). The complete nucleotide sequence alignment for E6 is given in Appendix A7.6.3.

Overall, 20 nucleotide changes were identified at 19 positions in the E6 gene, representing approximately 3.3% sequence variation over the length sequenced. Sequence variation was usually related to the variant lineage of the isolate, as discussed for URR sequence analysis. Six of the isolates (22%) contained the prototype E6 sequence and another five (19%) differed from prototype by containing only the commonly reported T to G change at nt 350. These were designated as E-subclass variants 350T and 350G respectively. The remaining 12 E lineage variants (44%) contained other nucleotide changes in conjunction with or without the T to G at nt 350, and been classified as 'other E variants' (Figure 5.2). Overall, 11 of the 27 isolates (41%) contained the T to G change at nt 350, the most commonly identified sequence change overall. No nt 350 variants were identified in As lineage variants. A sequence variation at nt 172 (A to C) was the second most common variation, present in five E lineage variants. All isolates with the G variation at nt 350 also contained the two common URR changes at nts 7193 and 7521 (Chapter 2). However not all isolates with the changes at nts 7193 and 7521 contained a change at nt 350.

The only AA lineage isolate K4, contained the greatest number of nucleotide variations, eight in total, similar to the trend observed in the URR and the E2 gene. The only isolate with prototype URR sequence, S2, also contained prototype E6 sequence. The percentage of multivariants in the E6 gene was 33% (9 isolates), which is considerably lower than found for the URR (91%). No nucleotide changes were located in the small region of the E7 gene (78bp) that was also sequenced.

| Variant Lineage | Nucleotide Number | 145 | 164 | 167 | 172 | 176 | 178 | 183 | 234 | 257 | 286 | 289 | 334 | 335 | 350 | 362 | 378 | 385 | 505 | 532 | E6 Subtype* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E | HPV16R | G | A | A | A | G | T | T | A | A | T | A | A | C | T | A | A | A | T | A | |
| E | S2 | | | | | | | | | | | | | | | | | | | | Prototype |
| E | J1 | | | | | | | | | | | | | | | | | | | | |
| E | K6 | NT | NT | | | | | | | | | | | | | | | | | | |
| E | E1 | NT | NT | | | | | | | | | | | | | | | | | | |
| E | D3 | NT | NT | | | | | | | | | | | | | | | | | | |
| E | B2 | NT | NT | | | | | | | | | | | | | | | | | | |
| E | C1 | NT | NT | | | | | | | | | | | | G | | | | | | 350G |
| E | H2 | | | | | | | | | | | | | | G | | | | | | |
| E | P1 | NT | NT | | | | | | | | | | | | G | | | | | | |
| E | P2 | | | | | | | | | | | | | | G | | | | | | |
| E | R1 | NT | NT | | | | | | | | | | | | G | | | | | | |
| E | H4 | A | | | C | | | | | | | | | | G | | | | | | Other E Variants |
| E | M2 | NT | NT | | C | A | | | | | | | | | G | | | | | | |
| E | K3 | | | | | | | | | G | | | | | G | | | | | | |
| E | K2 | NT | NT | | C | | | | | | | | | | G | | | | | | |
| E | M1 | NT | NT | C | | | | | | | | | | | G | | | | | | |
| E | K1 | NT | NT | | C | A | | | | | | | | | | | | | | | |
| E | K5 | | | | | | | | | | | | | | | | | | C | | |
| E | T1 | | | | | | | | | | | | | | | | | | C | | |
| E | M4 | NT | NT | | C | | | | G | | | | | | | | | | | | |
| E | M5 | NT | NT | NT | NT | NT | NT | NT | | | | | | | | | T | | | | |
| E | M6 | | C | | | | | | | | | | | | | | | | C | | |
| E | H3 | | | | | | | | | | | | | | | G | | | | | |
| As | G1 | NT | NT | NT | | | | | | | | | G | | | | | G | | | As |
| As | S1 | | | | | | G | | | | | | | | | | | | | | |
| As | O2 | | | | | | G | | | | | | | | | | | | | | |
| AA | K4 | T | | | | | | G | | | A | G | | T | G | | | | C | G | AA |

Figure 5.2: Sequence variation in the human papillomavirus type 16 E6 gene

* E lineage variants are divided into those that contain the prototype sequence (prototype), contain the common nucleotide change at 350 (350G), or contain other variations either in conjunction with 350G or without (other E variants).

NT = not tested at this nucleotide.

### 5.3.2.1 Amino Acid Analysis

Of the 20 nucleotide changes identified in the E6 gene, 12 resulted in nonsynonymous amino acid substitutions and eight were synonymous (Table 5.3). The majority of amino acid changes occurred at the amino terminus of the E6 protein, with nine occurring in the general location of the E6-AP binding region and six in the general location of the E6BP binding region (the regions indicated in Figure 5.3 are only approximate locations). No variant amino acids were located in the zinc fingers (CXXC), or in the p53 binding region (Figure 5.3). No frameshift mutations were observed and no nucleotide changes conferred a premature stop codon. The amino acid alignment for the E6 protein is given in Appendix A7.6.3.

Table 5.3: Amino Acid Changes in the E6 Protein

| Nucleotide | Nucleotide Change | Amino Acid Change | Amino Acid Position |
|---|---|---|---|
| 145 | CA**G** to CA**T** | Gln (Q) to His (H) | 14 |
| 145 | CA**G** to CA**A** | No Change | 14 |
| 164 | **A**CA to **C**CA | Thr (T) to Pro (P) | 21 |
| 167 | **A**CT to **T**CT | Thr (T) to Ser (S) | 22 |
| 172 | ATA to ATC | No Change | 23 |
| 176 | **G**AT to **A**AT | Asp (D) to Asn (N) | 25 |
| 178 | GAT to GA**G** | Asp (D) to Glu (E) | 25 |
| 183 | ATA to A**G**A | Ile (I) to Arg (R) | 27 |
| 234 | GAC to G**G**C | Asp (D) to Gly (G) | 44 |
| 257 | ATA to **G**TA | Ile (I) to Val (V) | 52 |
| 286 | GCT to GC**A** | No Change | 61 |
| 289 | GTA to GT**G** | No Change | 62 |
| 334 | AGA to AG**G** | No Change | 77 |
| 335 | **C**AT to **T**AT | His (H) to Tyr (Y) | 78 |
| 350 | **T**TG to **G**TG | Leu (L) to Val (V) | 83 |
| 362 | **A**CA to **G**CA | Thr (T) to Ala (A) | 87 |
| 378 | TAC to TT**C** | Tyr (Y) to Phe (F) | 92 |
| 385 | AAA to AA**G** | No Change | 94 |
| 505 | GGT to GG**C** | No Change | 134 |
| 532 | TCA to TC**G** | No Change | 143 |

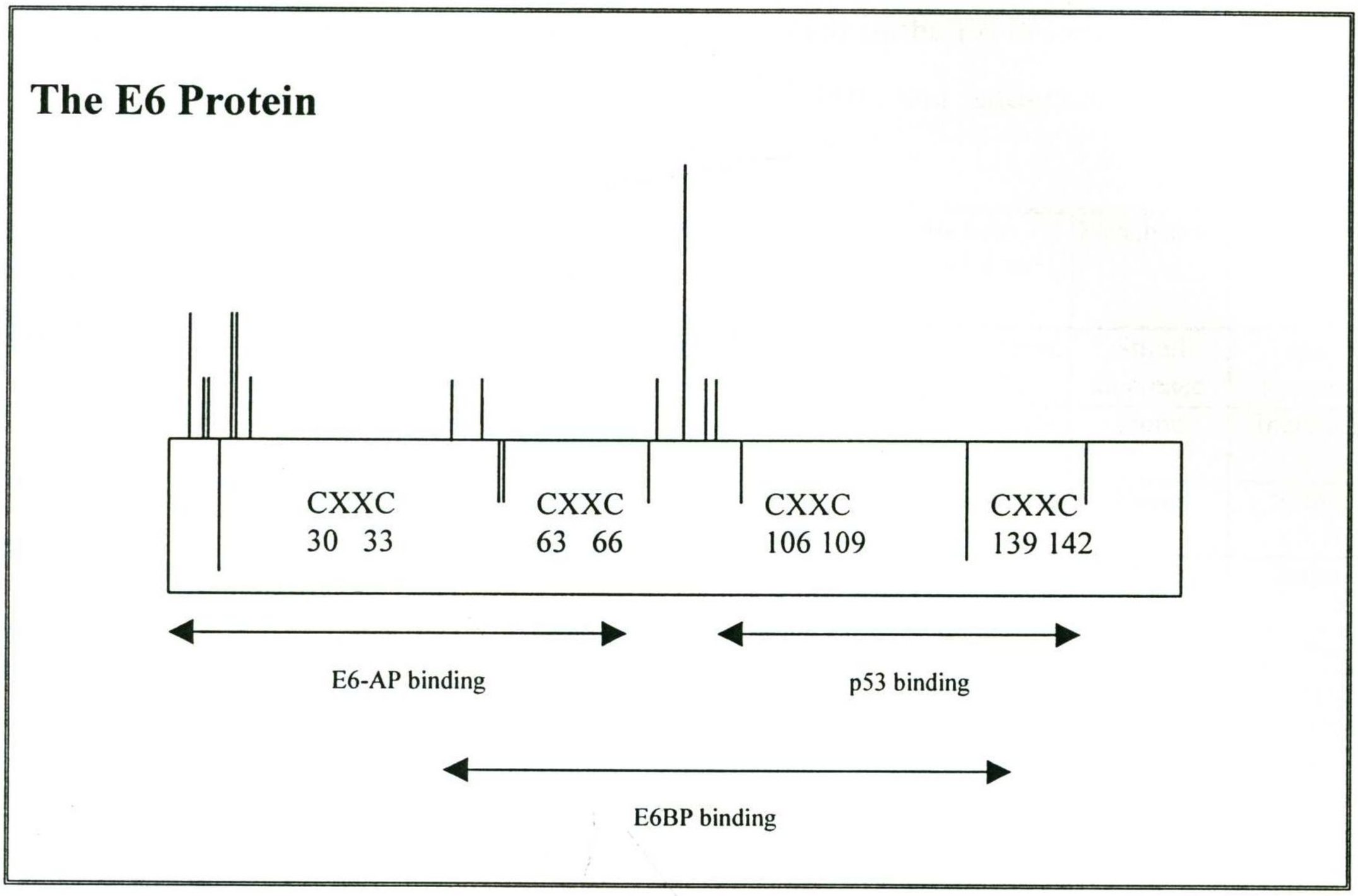



Figure 5.3: Distribution of amino acid changes in the E6 protein

CXXC represents the zinc fingers (C=cysteine, and X=any amino acid) and the numbers beneath are the corresponding amino acid positions of the zinc fingers. Lines above the protein represent nucleic acid changes that led to amino acid changes (nonsynonymous changes). Lines below the protein represent nucleic acid changes that did not lead to amino acid changes (synonymous changes). The height of the line indicates the number of times that change was found, the larger the line, the more times that change was found among the isolates. The shortest line represents a change identified in only one isolate. The approximate positions of the protein required for different functions are indicated.

The effect of each amino acid change on hydrophilicity, surface probability, flexibility and antigenic index is generalised in Table 5.4. In general, it was found that substitution of a hydrophilic amino acid for a non-hydrophilic amino acid resulted in decreased hydrophilicity. The change from a hydrophilic residue to another hydrophilic residue either made no difference, or caused a slightly increased or decreased hydrophilicity. Several of the amino acid changes identified, such as those at positions 14 and 27 (which were present only in K4) resulted in large changes to the antigenic index. Of the eight nucleotide variations in the E6 gene from K4, four led to amino acid changes. All four affected hydrophilicity. The nucleotide change at 350 (L83V) did not result in a change in antigenic index. None of the amino acid changes identified in the E6 protein altered flexibility substantially.

Table 5.4: Significance of Changes in Amino Acids in the E6 Protein in Terms of Hydrophilicity, Surface Probability, Flexibility, and Antigenic Index

| Amino Acid Position | Amino Acid | Type of Side Chain Change | Hydrophilicity# | Surface Probability* | Flexibility^ | Antigenic Index% |
|---|---|---|---|---|---|---|
| 14 | Q to H | Hydrocarbon (non-polar) to basic (positive, charged) | Small decrease | Small decrease (both <1) | Small decrease | Large decrease |
| 21 | T to P | Hydroxyl (polar, uncharged) to non-polar | Increase | Small increase (both <1) | Same | Increase |
| 22 | T to S | No change | Small increase | Small decrease (both >1) | Same | Same |
| 25 | D to N | Acidic (negative, charged) to amide (polar, uncharged) | Same | Same | Same | Same |
| 25 | D to E | No change | Same | Same | Same | Same |
| 27 | I to R | Hydrocarbon (non-polar) to basic (positive, charged) | Large increase | Increase (both <1) | Small increase | Large increase |
| 44 | D to G | Acidic (negative, charged) to hydrocarbon (non-polar) | Large decrease | Decrease (both <1) | Same | Same |
| 52 | I to V | No change | Increase | Small increase (both <1) | Same | Same |
| 78 | H to Y | Basic (positive, charged) to carbon ring (aromatic) | Decrease | Increase (both >1) | Same | Same |
| 83 | L to V | No change | Decrease | Same | Same | Same |
| 87 | T to A | Hydroxyl (polar, uncharged) to hydrocarbon non-polar) | Large decrease | Decrease (>1 to <1) | Decrease | Decrease |
| 92 | Y to F | No change | Large decrease | Large decrease (Both >1) | Same | Decrease |

#A hydrophilic protein is a 'water-loving' protein. Hydrophilic compounds are therefore very water-soluble. Hydrophilic amino acids typically have polar or charged side chains and are normally found near the surface of water-soluble proteins. Values can be used to determine membrane-spanning segments (highly hydrophilic), or regions that are likely to be exposed on the surface of the protein (hydrophilic domains), and therefore potentially antigenic. All hydrophilicity values in this study were determined using the Kyte-Doolittle scale (Kyte and Doolittle, 1982).

*Surface probability values refer to an increased probability of the amino acid being on the surface of the protein. Values were calculated according to the Emini method (Emini *et al.*, 1985, and Janin and Wodak, 1978).

^Protein flexibility was determined according to the Karplus-Schulz method.

%The antigenic index is a method for predicting B-cell epitopes, regions characterised by high levels of hydrophilicity. Antigenic index takes into account hydrophilicity, flexibility and surface probability, and was determined according to the Jameson-Wolf method (Jameson and Wolf, 1988).

# 5.4 DISCUSSION

## 5.4.1 The Sequence of the E6 Gene

Sequence variation in the HPV 16 E6 genome from most regions of the world has been analysed by PCR and/or DNA sequence analysis or dot blot hybridisation. The E6 sequences reported in this study, however, are the first from Australia and the South Pacific region. The percentage of variable E6 nucleotide positions at 3.3%, was the same as that identified in a study of 408 isolates from 22 countries (Yamada *et al.,* 1997). In one published study 43% of HPV 16 positive cervical cancer isolates from Europe contained the HPV 16 E6 prototype sequence (Nindl *et al.,* 1999), a higher proportion than the 22% identified here. In the study by Nindl and colleagues, an additional 38% of the cancer isolates were associated with only one nucleotide change, that at 350 (T to G). Overall however, 51% of E lineage variants in that study contained a G at nt 350, individually or in combination with other nucleotide changes. In this study, 350G was the only nucleotide change identified in 19% of isolates, whilst 37% of E lineage isolates (and 41% of the isolates overall) contained the G at nt 350 either alone or in combination with other nucleotide changes. Therefore 350G seems less prevalent in Australian and South Pacific tumours than in those from Europe, however it was more prevalent than in Chinese tumours, where 350G was only present in 10% of isolates in a recent survey (Stephen *et al.,* 2000).

In this study, 33% of the E6 isolates were multivariants, in comparison with only 10% of a patient group from the UK, 16% from German patients, 25% from Swedish and 61% from Italian patients (Zhebe *et al.,* 1998a and b, Brady *et al.,* 1999, and Nindl *et al.,* 1999). It is likely that these differences relate somewhat to the proportions of each variant lineage present in the population studied. The AA variant, K4, had changes at nts 145, 286, 289, 335, and 532, which were identified in all AA variants in a study by Yamada *et al.* (1995). A comparison of sequence variation in the E6 gene revealed no correlation of individual changes or patterns of mutation with HPV 16 physical state.

### 5.4.2 Potential Functional Significance of E6 Variation

The zinc finger domains of E6 are preceded by a hydrophilic amino terminal domain and are separated by a hydrophobic domain. Amino acid changes in this study were generally located before or between the zinc finger domains and some appeared to affect hydrophilicity, such as those at positions 21, 27, and 44. It is possible that changes in these amino acid residues affect the structure of the E6 protein by altering the location of amino acids from the interior to the surface of the protein, or vice versa, which may alter interaction with E6-AP or E6BP. The binding of E6 to p53 requires amino acids 106 to115, and amino acids 132 to 141 are required for p53 degradation and for abrogation of p53 transactivation (Sherman *et al.,* 1997). The ability of HPV 16 E6 to bind and target p53 for degradation *in vitro* is reportedly sensitive to mutations or small deletions (Foster *et al.,* 1994), however no sequence changes in these important domains were identified in isolates in this study. Most of the amino acid variations were located in the amino half of the protein, which is the region important for E6-AP binding. It is conceivable that amino acid changes in the E6-AP binding domain may alter the association of E6 with p53 through E6-AP, although this remains to be tested. Amino acid changes at the extreme amino terminus of E6 have also been shown to be important for different abilities of E6 variants to cooperate with E7 to mediate keratinocyte immortalisation (Conrad Stöppler *et al.,* 1996). In the study by Conrad Stöppler and colleagues, the E6 protein from an AA lineage variant exhibited higher p53 degradation activity and formed more differentiation resistant colonies in primary cells compared with E and Af2 lineage variants. Amino acid changes at 145, 335, and 350 were present in the AA variant analysed in that study, amino acid changes also present in K4. Enhanced p53 degradation by their variant may however have been due to an amino acid change at 138, in the domain required for degradation, an amino acid change not present in the only AA variant analysed here, K4. Either way, this evidence further suggests that AA variants may have enhanced oncogenic potential, which may be due in the E6 protein to subtle sequence changes.

No amino acid changes were identified in the zinc fingers in this study, as in other studies of HPV 16 cancer isolates. This might be expected, as the zinc fingers of E6 are evolutionally conserved structures and disruption of any one results in defective

function (Kanda *et al.,* 1991, and Sherman and Schlegel, 1996). Grassman and colleagues reported that HPV 6 E6 isolates from malignancies contained only a few sequence changes (Grassman *et al.,* 1996b). In contrast, HPV 6 E6 isolates from benign tumours were found to have a high degree of diversity, some of the changes were identified in important functional domains of the E6 protein such as the zinc fingers.

The possible biological importance of the T to G transition at nt 350 (L83V), the most common nucleotide change identified in the E6 gene in this study, has come under considerable scrutiny. There is some evidence to suggest that 350G may be a high risk factor for cervical cancer in some populations, such as the Swedish as it is more commonly associated with invasive cervical carcinoma than CIN, and a low risk factor in others such as the Italians, as it is more commonly associated with CIN than cancer (Zhebe *et al.,* 1998a). However, in a German population the variant was found to be highly prevalent in both low grade CIN and invasive cervical cancer and it was suggested that variants, such as 350G, are not associated with an increased risk of cervical cancer (Nindl *et al.,* 1999).

It is possible that the 350G variant has different transforming potential in squamous and glandular cervical lesions, with some evidence suggesting that 350G has greater oncogenicity in squamous lesions (Andersson *et al.,* 2000). The conflicting findings on 350G may also be due to genetic differences in the populations studied. Van Duin *et al.* (2000) found no association of nt 350 variants with virus persistence or development of cervical cancer in patients from The Netherlands, but suggested that the combination of a T at nt 350 with p53 Arg/Arg at codon 72, confers an increased risk of transition of CIN 3 to invasive cervical cancer. This common p53 polymorphism in cells results in either an arginine (Arg) or proline (Pro) at codon 72. Importantly, this polymorphism lies in a region of p53 necessary for induction of apoptosis (Walker and Levine, 1996, and Sakamuro *et al.,* 1997). Biochemical analysis of p53Pro and p53Arg showed that p53Pro is twice as strong a transcription activator than p53Arg (Thomas *et al.*, 1999a), but that p53Arg is twice as effective in preventing immortalisation of primary rodent cells. Therefore, p53Arg may be more effective in inducing apoptosis (Thomas *et al.,* 1999b). In another study from the UK, the 350 polymorphism did not correlate with HLA genotype or clinical

outcome, although there was also a trend for an increased frequency of p53Arg homozygotes among 350T carcinoma patients (Brady *et al.,* 1999). Storey *et al.* (1998a) also reported a significant over representation of the codon 72 Arg/Arg p53 genotype in cervical cancer patients in the UK. Individuals homozygous for p53Arg were approximately seven times more susceptible to HPV-associated tumourigenesis than heterozygotes. However other surveys have not substantiated these findings (Lanham *et al.,* 1998, Minaguchi *et al.,* 1998, Rosenthal *et al.,* 1998, Storey *et al.,* 1998b, and Tachezy *et al.,* 1999).

CIN progression and/or regression suggest immune function influences tumour development. It is possible that variants of HPV 16 may interact differentially with host genetic factors such as HLA type, altering disease course. Multiple different HLA factors reportedly influence HPV infection and subsequent disease progression (Bontkes *et al,* 1998). Although HPV variants probably lead to new immunogenic properties, Bontkes and colleagues did not find any significant correlation between variant incidence and disease progression, or with viral persistence and HLA type. However, an association of a nucleotide change at 131 (A to G) in E6 with HLA-B7 positive women with poor clinical outcome has been linked to evasion of the host immune response (Ellis *et al.,* 1995 and 1997). Brady *et al.* (1999) reported a frequency of 131G of only 9.6%, however 78% of those patients were HLA-B7 positive. Patients with HLA-B7 also had a slightly reduced survival time. A significantly lower frequency of antibody responses to VLPs has been observed in patients carrying 131G (Ellis *et al.,* 1997, and Etherington *et al.,* 1999). HLA-B7 was also identified as a risk factor for cervical disease in a study conducted in the US (Hildesheim *et al.,* 1998). However recent studies found no major role for the E6 variant in the pathogenesis of HPV 16-related disease in England (Luxton *et al.,* 2000), or association of the variant with any HLA type even when the variant was common in a population (Etherington *et al.,* 1999). In 12 isolates from this study where the sequence at nt 131 was available, there were no variations from the prototype, however the presence of the nt 131 variant in the other isolates cannot be excluded.

In general, the T-proliferative antibody response in cervical cancer patients is poor, however T and B cell epitopes in the HPV 16 E6 protein have been identified

(Bartholomew *et al.,* 1994, Dillner *et al.,* 1995, Ressing *et al.,* 1995, and Azoury-Ziadeh *et al.,* 1999). When E6 and E7 proteins are processed and presented at the cell surface, they are likely to be tumour rejection antigens. Indeed, lymphoproliferative responses to specific HPV 16 E6 and E7 peptides have been associated with clearance of HPV infection and regression of CIN, and lack of response to E6 and E7 sequences correlates with persistent HPV infection and disease (Kadish *et al.,* 1997). In particular, it has been determined that amino acids 1 to 30 of HPV 16 E6 are important for these functions (Kadish *et al.,* 1997), the region of E6 where the majority of amino acid changes were identified in this study. It is possible that E6 sequence variation could affect cytotoxic T-lymphocyte responses (Kast *et al.,* 1994, and Ellis *et al.,* 1995), although there is also evidence that E6 variants may not define a lack of antibody response in some patients (Nindl *et al.,* 2000). Mice immunised with recombinant vaccinia virus expressing E6 produce specific antibodies, as well as proliferative and cytotoxic T lymphocyte responses (Gao *et al.,* 1994). T cell determinants have also been identified using synthetic peptides to stimulate peripheral blood mononuclear cells (Strang *et al.,* 1990), and by characterising endogenously processed peptides in cells infected with recombinant vaccinia virus expressing E6 (Bartholomew *et al.,* 1994). Amino acid changes in E6 at positions 21, 22, 25, 44, 52, 87, and 92 are located in B or T cell epitopes identified in the previously mentioned studies (Strang *et al.,* 1990, Bartholomew *et al.,* 1994, and Gao *et al.,* 1994). These amino acid changes were identified in 10 of the 27 isolates for which the sequence of the E6 gene was obtained.

Changes in hydrophilicity and antigenic index due to amino acids variations were also large, further indicating the potential for alterations to immune response. In particular, in the AA variant K4, all four amino acid changes affected hydrophilicity, with two also leading to large changes in antigenic index. This suggests that AA variant E6 proteins may differ immunologically from those of the other lineages. Finally, it is possible that even synonymous sequence changes in E6 may correlate with clinical aggressiveness (Alvarez-Salas *et al.,* 1995). In the Alvarez-Salas study, aggressive tumours were heterozygous at nt 511, containing both a C and a T. The authors suggested that aggressiveness may have been due to genomic instability, and that mutations were introduced post infection.

In summary, in this chapter of the study the sequence of the E6 gene was determined for 27 of the original 34 isolates. The potential for some of the sequence variations identified in this study to alter E6 protein function or immune response was suggested.

# CHAPTER 6

# GENERAL DISCUSSION AND FUTURE DIRECTIONS

# 6.1 GENERAL DISCUSSION

PVs are quantitatively the most important group of viruses associated with benign and malignant neoplasia in humans. However the majority of HPV infections, even with high risk types, don't usually result in cancer, therefore other factors must be involved in malignant progression. Some suggested host-related and environmental factors include co-carcinogens, genetic predisposition (such as HLA genotype), p53 polymorphisms, and immunosuppression. Viral factors such as persistence of infection and high viral load also appear to be important (Koutsky, *et al.,* 1992, Munoz and Bosch, 1992, and Critchlow *et al.,* 1995). An increasing body of literature is providing evidence for the role of intratypic variants in malignant progression and the behaviour of established tumours.

Circumstances that lead to increased levels of E6 and E7 proteins may represent important events in the onset of tumourigenesis. This may result from integration of the virus and loss of negative regulation by the viral E2 protein, or by sequence changes in the URR affecting TF binding, or because of increased viral copy number expressing E6/E7. If sequence variation in coding regions leads to a lack of specific immune recognition, then the transforming effect of the virus on the host cell is likely to be compounded. It is also possible that sequence variation in the E2 protein alters its ability to repress p97 or affects the likelihood of viral integration and hence transformation. Sequence variation in the viral oncoproteins may also enhance their transformation functions. It is important to note that this study of cancer isolates examined the end point in a long process that began with benign infection, and ended in invasive cervical carcinoma.

HPV 16 sequences are diverse and suited to the study of molecular evolution, which is most likely a continuous process involving the stepwise accumulation of mutations (Ho *et al.,* 1993a). The investigations described in this study represent a comprehensive analysis of HPV 16 cervical cancer isolates from Australia and the South Pacific region: the first data from this region of the world. Intratypic variation in the entire URR and the E6, E4, and E2 genes has been described. No genetic recombination was obvious between viruses of different HPV 16 lineages. From

approximately 2,600 HPV 16 nucleotide positions sequenced in the viral URR, E2/E4 and E6 genes from the 34 isolates, 3.6% of nucleotides varied overall. This is consistent with the results obtained in a study of 29 cervical cell samples, predominantly of normal cytology, where 3,690 HPV 16 nucleotides were sequenced (Yamada *et al.,* 1995). Sequence changes in HPV 16 have been reported to be greatest in the URR, followed by E6 and then E7 (van Duin *et al.,* 2000). Similarly, the number of nucleotide changes in this study was greatest in the URR, followed by E2, E6 and then E4. Proportions of nonsynonymous substitutions were also greatest in E2, then E6, and E4 respectively. Multivariants, on the other hand, were greatest in the URR, followed by E6 and then E2. Nucleotide and amino acid variation occurring more frequently in some regions of PV genomes than others likely reflects functional constraints on the more conserved regions (Wheeler and Icengole, 1995), such as E4. Collections of nucleotide substitutions in 'hotspots' could also be the result of selective immune pressures. It has been suggested that nucleotide alterations in one region of HPV 16 are associated with changes in other regions of the genome (Chan *et al.,* 1992). However, some investigators have identified HPV-subclass variants (such as the E subclass variant G131) by sequencing the E6 gene, variants that were not identified by nucleotide changes in the URR (Yamada *et al.,* 1995, and Wheeler *et al.,* 1997). In this study, the sequence of the E6 and E2 genes was relatively lineage-specific, as was the URR, but sequence changes in the URR did not always predict sequence changes in the E6 or E2 genes.

Regulation of the HPV 16 oncogene promoter, p97, is a complex process involving transcriptional repression and activation. Activators provide the capacity to express viral proteins in an appropriate host cell type, while repressors may provide a means of timing E6 and E7 expression to when it is most appropriate. The relationship of HPV with the cell involves the use of numerous DNA binding sites for cellular factors and small sequence differences in binding sites may have a large cumulative effect on the virus, particularly if they act in concert. The results of this study support the hypothesis that sequence variation in the URR from cervical cancer isolates may be functionally important. Evidence has been provided for mechanisms of malignant conversion in the absence of integration of the viral genome. Specifically, naturally occurring sequence variation in the recently identified PSM from an episomal AA lineage variant was found to affect the binding of the PSM-BP (CDP) to DNA,

leading to derepression of the oncogene promoter. This sequence change at nt 7894 has been identified in an AA lineage variant in at least one other study (Kammer *et al.,* 2000), and as such may be a relatively common mechanism for enhanced promoter activity from the URR of AA lineage variants. In addition, sequence variation in motifs known to bind YY1 and Oct-1/PEF-1 was identified in another episomal isolate that resulted in substantially elevated promoter activity (at nts 7792 and 7676 respectively). The corresponding TFs did not display altered binding characteristics however, suggesting that as yet unidentified factors may bind to the motifs. Interestingly, an isolate with a novel large-scale duplication of the URR was also found episomally, suggesting that the large-scale change may have been necessary for malignant conversion even though promoter activity was not upregulated in comparison with the prototype.

Even small sequence changes in coding regions may have a large effect on the function of the viral proteins. The HPV 16 prototype L1 protein, for example, does not self-assemble into VLPs and this may be due to a single amino acid change (Kirnbauer *et al.,* 1992 and 1993). The identification of variants associated with persistent HPV infection, high grade CIN, or invasive cancer may also prove important to the rational design of diagnostics, therapeutics, and vaccine strategies. Although functional analysis was not performed on the E2, E4 and E6 proteins sequenced during this study, it is possible that at least some of the sequence changes identified have biological importance, particularly in terms of altering immunological epitopes and perhaps the functions of the proteins. For instance, amino acid changes in the transactivation region of the E2 protein were relatively common in As and AA lineage isolates. Some amino acid changes in the E6 protein were also found to affect hydrophilicity and antigenic index, particularly in an AA lineage variant. In addition, 10 E6 variants contained amino acid changes in known B and T cell epitopes.

Ten isolates were found to contain episomal HPV 16 DNA in this study, and large-scale duplication or upregulated p97 activity was identified in three that were examined. Of the seven remaining episomal isolates, none contained sequence variation in known TFBSs, apart from the common sequence variation at nt 7193 in a possible TEF-1 binding site. It is likely that those isolates have promoter activity

comparable with the HPV 16 prototype. Integration of HPV is likely to reduce the number of steps required for transformation and hence the time from HPV infection to invasive cancer, so it is possible that episomal HPV variants generally take longer than their integrated counterparts to cause cancer. Sequence variation in the URR leading to increased promoter activity may be equivalent to integration though if it increases the likelihood of the development of invasive cervical carcinoma, and decreases the time required from infection to cancer. It is possible that amino acid changes in the E2 and E6 proteins identified in several of those seven episomal isolates may also have influenced transformation. However, it should also be noted that host factors, such as immune response, probably have some influence on the course of disease.

In conclusion, the results of this study strengthen earlier evidence that HPV 16 URR variants may be biologically significant in determining the malignant potential of infection or the behaviour of established tumours, through their influence on the levels of expression of the E6 and E7 oncogenes. The results also indicate that variants from certain HPV 16 lineages, such as those from the AA lineage, may be functionally distinct from variants of other lineages, in terms of enhanced early promoter activity and possibly altered viral protein functions. In particular, new evidence for sequence changes in the URR of episomal variants that may have been involved in malignant transformation by individual variants was provided. Sequence changes in HPV proteins that may affect carcinogenic potential were also identified. However "defining a variant by a single nucleotide change clearly does not provide all of the information needed to understand the potential selective advantages of a virus variant in human disease" (Bontkes *et al,* 1998), and determining the exact roles played by viral variants will clearly require further research.

# 6.2 FUTURE DIRECTIONS

Although new mechanisms likely to be responsible for malignant transformation in episomal cervical cancer isolates were identified, reconstructing the entire story was beyond the scope of this study. Further investigations using different EMSA protocols and specific antibodies will be necessary to identify the protein factor/s mediating increased promoter activity in the URR of K2 at nts 7676 and 7792. It is likely that unidentified cellular factors bind to these sites and play important roles in p97 regulation. Transformation assays (co-transfecting plasmids containing the HPV 16 genome from O2 and activated *ras* into primary cells) would also help resolve whether the large-scale duplication in O2 aided malignant conversion.

Sequence variation in the viral E5 and E7 genes is also worthy of further investigation. HPV 16 E5 variants have been associated with high transcriptional activity and development of neoplasia (Bible *et al.,* 2000), and although some E7 variants have shown similar transforming potential to prototype E7 (Fujinaga *et al.,* 1994), others may be functionally important (Edmonds and Vausden, 1989, Sang and Barbosa, 1992, and Song *et al.,* 1997).

It has been debated whether sequence variation such as that in the URR leading to increased promoter activity is present at the initiation of cervical infection, or whether changes occur during infection, particularly at the border of dysplasia and malignancy. Limited studies have addressed these issues. The results of one study suggested that the sequence of viral genomes does not change during the course of infection, although the study group was very small (Hu *et al.,* 1999). There is also a need to better define the relevance of HPV variants to transmissibility, and host factors (eg. presenting HLA molecules). A link between viral load and progression of cervical lesions was recently made (Josefsson *et al.,* 2000, and Ylitalo *et al.,* 2000). New technologies now allow accurate quantitation of virus and open the way for studies relating viral copy number to progression or regression and sequence variation.

Importantly, conclusions drawn from investigations carried out in submerged monolayer cultures of cervical cancer derived cell lines may not accurately reflect what occurs in stratified differentiating epithelium *in vivo*. Experiments have shown that early promoter activity is significantly higher in tumourigenic cells compared to immortalised or primary cells, using the same sequences (Chen *et al.,* 1997, and Kook *et al.,* 1998). However, little is known of cellular TF/URR interactions that underpin E6/E7 regulation in stratified epithelium because until recently there has been no suitable *in vitro* model. There are many examples of overlapping TFBSs in the URR and TFs are members of multigene families that bind to the same motif but differ functionally, complicating the situation. The recent development of the organotypic "raft" culture system which faithfully reproduces cellular differentiation *in vitro* has opened the way for studies to determine how differential expression of TFs and competition for binding during cellular differentiation might affect the level of E6/E7 transcription. This would provide a more accurate determination of the functional effects of natural sequence variation in the URR. Preliminary studies in raft cultures have confirmed that the activities of the HPV 11 and 18 promoters in differentiated epithelia are distinct from those in submerged monolayer cultures of cancer cells or primary keratincocytes (Nuss Parker *et al.,* 1997, and Zhao *et al.,* 1997). However, HPV 16 URR/TF interactions and the effects of naturally occurring sequence variations on promoter activity in rafts have yet to be investigated.

Finally, it is not yet clear whether sequence variation is directly responsible for progression of precancer lesions, and/or increased clinical aggression of established tumours. The findings from this study could not be related to clinicopathological variables because the group was non-representative and not large enough to permit statistical analysis. With regard to noncancer isolates, most published sequences have come from isolates with unspecified clinical, histological or cytological correlations, and therefore the relationship between specific nucleotide changes and propensity for progression versus regression cannot be ascertained by comparative analysis. Large multicentre studies carried out on material from patients with well-defined clinical associations will help clarify the situation. If specific nucleotide changes associated with an increased risk of cervical disease progression could be identified, assays for these sequence changes could be used in conjunction with cytology in cervical screening programs, or to determine prognosis for cervical cancer patients.

# CHAPTER 7

# APPENDICES

# APPENDIX 7.1

## SOLUTIONS

*Sterilisation by autoclaving was performed at 15lb for 15 minutes at 121 ℃.*
*Sterilisation by filtration was performed with a 0.2μm pore filter (Sartorius) attached to a syringe.*
*Water was filtered in a water purification system (Modulab™, Liquepure™) using a conductivity of 18Ωohm.*
*If not specified, solutions were stored at room temperature.*

**25mg/ml Ampicillin**

A stock of 25mg/ml ampicillin, sodium salt (Sigma) was made by dissolving 1g of ampicillin powder in 40ml of water. Filtered aliquots of 5ml were stored at -20°C.

**Blotto (10% skim milk and 0.2% sodium azide)**

For 100ml of blotto, 10g of skim milk powder and 0.2g of sodium azide were made up to 100ml with water. Blotto was stored at 4°C.

**Buffers for EMSAs:**

**Buffer D for EMSA (20mM HEPES, pH 7.9, 20% glycerol, 0.1M KCl, 0.2mM EDTA, 0.5mM PMSF, and 0.5mM DTT)**

For 1ml of buffer D, the following were combined: 20μl of 1M HEPES, pH 7.9, 200μl of glycerol, 100μl of 1M KCl, 0.4μl of 0.5M EDTA, pH 8.0, 50μl of 10mM PMSF, 0.5μl of 1M DTT, and 629.1μl of water. This buffer was made just prior to use due to the short half-life of protease inhibitors.

**x4 SMK for EMSA (16mM Spermidine, 160mM KCl, 0.5mM $MgCl_2$, 0.2% Nonidet NP-40, 4mM DTT, 4mM PMSF, 8μg BSA, and 20% glycerol)**

For 100μl of x4 SMK, the following were combined: 16μl of 100mM Spermidine, 6.5μl of 2.5M KCl, 0.5μl of 100mM $MgCl_2$, 2μl of 10% Nonidet NP-40, 4μl of 100mM DTT, 4μl of 100mM PMSF, 16μl of 50mg/ml BSA, 40μl of 50% glycerol, and 11μl of water. This buffer was made just prior to use due to the short half-life of protease inhibitors.

**x2 Binding Buffer for EMSA (24mM HEPES, pH 7.9, 20% glycerol, 1mM EDTA, pH 8.0, 8mM $MgCl_2$, 120mM KCl, and 0.2% Nonidet NP-40)**

For 20ml of x2 binding buffer, the following were combined: 480μl of 1M HEPES, pH 7.9, 4ml of glycerol, 40μl of 500mM EDTA, pH 8.0, 160μl of 1M $MgCl_2$, 2.4ml of 1M KCl, and 40μl of Nonidet NP-40. The volume was adjusted to 20ml with water. Aliquots of 1ml were stored at -20°C.

**EMSA 'Mix' (5.7μg/μl BSA, 11.4mM Spermidine, 0.14μg/μl Poly [d(I-C)], 0.018μg/μl Calf Thymus DNA, and 5.7mM PMSF)**

For 7μl of EMSA mix (ie. per tube), the following were combined: 4μl of 10mg/ml BSA, 0.8μl of 100mM Spermidine, 1μl of poly [d(I-C)], 0.25μl of 0.5mg/ml calf thymus DNA, 0.4μl of 100mM PMSF, and 0.55μl of water. This mix was made just prior to use due to the short half-life of protease inhibitors.

**2.5M Calcium Chloride**

For 20ml, 7.349g of $CaCl_2$ were dissolved in water, the volume adjusted to 20ml and filtered. 1ml aliquots of 2.5M $CaCl_2$ were stored at -20°C. For 50mM $CaCl_2$ needed for making competent cells, 1ml of 2.5M $CaCl_2$ was made up to 50ml with water.

**Chloroform:Isoamyl Alcohol (24:1)**

Twenty four millilitres of chloroform (BDH Chemicals) were combined with 1ml of isoamyl alcohol (BDH Chemicals).

**Digestion Buffer (50mM Tris-HCl, pH 8.5, 1mM EDTA, pH 8.0, and 0.5% Tween 20®) for Paraffin-Embedded Tissues**

For 20ml of digestion buffer the following were combined: 1ml of Tris-HCl, pH 8.5, 40μl of 0.5M EDTA, pH 8.0, and 100μl of Tween 20® (Sigma). The volume was adjusted to 20ml with water and digestion buffer stored at 4°C short term, or at -20°C long term.

**0.5M Disodium Ethylenediaminetetraacetate (EDTA) pH 8.0**

To 800ml of water, 186.1g of EDTA (Sigma) were added. The solution was stirred vigorously on a magnetic stirrer and the pH adjusted to 8.0 with NaOH (approximately 20g of NaOH pellets). The volume was adjusted to 1L with water and the solution autoclaved.

**1M Dithiotheitol (DTT)**

One gram of DTT (Amersham Pharmacia Biotech) was dissolved in 6.5ml of 0.01M sodium acetate, pH 5.2. DTT was filtered and stored in 1ml aliquots at -20°C.

**Dulbecco's Modified Eagle's Medium (DMEM)**

For 1L, the contents of one pack of powdered DMEM containing D-glucose, L-glutamine and sodium pyruvate (Life Technologies, Gibco BRL), and 3.7g of $NaHCO_3$ (Sigma) were dissolved in approximately 900ml of water. The pH was adjusted to 7.0 with a small amount of concentrated HCl and the volume adjusted to 1L. The medium was filtered immediately (using compressed nitrogen gas) through a sterilised stainless steel filter holder containing a prefilter and a 0.22μm filter into dry-heat sterilised 500ml bottles. A 48 hour sterility test was performed by incubating a small amount of the medium. This base media was stored for up to 4 months at 4°C as required. Five to 25ml of fetal calf serum (CSL), 5ml of combined stock penicillin (100 IU/ml) and gentamicin (5mg/ml) was added

to each 500ml of medium before use. When the medium was used after 6 weeks storage, 10ml of 2.92% w/v glutamine (Life Technologies, Gibco BRL) was added per 500ml of medium.

**10mg/ml Ethidium Bromide**

Making sure that gloves and a mask were worn, one gram of ethidium bromide (Sigma) was added to 100ml of water. The powder was dissolved after several minutes agitation on a magnetic stirrer. The solution was stored in a foil-covered container at 4°C.

**x5 Gel Loading Dye**

For 10.5ml of x5 gel loading dye, 5.3ml of glycerol (Sigma), 2.5mg of bromophenol blue (Biorad) and 2.7ml of 0.5M EDTA, pH 8.0 were combined. For running PCR reactions on an agarose gel, 10µl of 1 in 3.1 diluted gel loading dye was added to 5µl of PCR product. Once PCR product was added this became x1 gel loading dye.

**x2 HEPES Buffered Saline Solution (HBSS), pH 7.05 (280mM NaCl, 10mMKCl, 50mM HEPES, 1.5mM $Na_2HPO_4$, pH to 7.05)**

For 100ml of x2 HBSS the following were combined: 5.6ml of 5M NaCl, 1.0ml of 1M KCl, 1.5ml of 100mM $Na_2HPO_4$, and 1.192g of HEPES. The volume was adjusted to 100ml with water and the solution filtered. x2 HBSS was stored in 10ml aliquots at –20°C.

**1M HEPES, pH 7.9**

For 10ml of 1M HEPES, 2.383g of HEPES were added to water and the pH adjusted to 7.9. The volume was adjusted to 10ml, the solution filtered and stored at -20°C in 1ml aliquots.

**Luria-Bertani (LB) Broth, pH 7.5**

For 1L of LB broth, 10g of bacto-tryptone, 5g of bacto-yeast extract and 5g of NaCl were added to 800ml of water. The pH was adjusted to 7.5 with concentrated NaOH, the volume adjusted to 1L, and the solution autoclaved. 100µg/ml ampicillin was added after cooling the media to approximately 55°C if required.

**Luria-Bertani (LB) Plates, pH 7.5**

A 1L solution of LB broth was made and 15g of agar added. The pH was adjusted to 7.5 with concentrated NaOH and the solution autoclaved. After allowing the medium to cool to approximately 55°C, 100µg/ml ampicillin or 50µg/ml kanamycin was added. Approximately 30ml of medium were poured into 100mm petri dishes. These plates were left to dry and then stored at 4°C for up to one month.

**Lysis Buffer (50mM NaCl, 50mM Tris, pH 7.5, 0.5% SDS, and 10mM EDTA, pH 8.0) for DNA Extraction from Fresh Tissue**

For 20ml of lysis buffer the following were combined: 200µl of 5M NaCl, 1ml of 1M Tris, pH 7.5, 1ml of 10% SDS, and 400µl of 0.5M EDTA, pH 8.0. The volume was adjusted to 20ml with water and the lysis buffer stored at 4°C.

**Lysis Buffer (0.2M NaOH and 1% SDS) for Minipreps**

For 10ml of lysis buffer the following were combined: 400µl of 5M NaOH and 1ml of 10% SDS. The volume was adjusted to 10ml with water. Lysis buffer was made as required.

**Lysis Buffer (50mM Tris, pH 8.5, 1mM EDTA, pH 8.0, and 0.5% Tween 20®) for Polymerase Chain Reaction**

For 20ml of lysis buffer, 1ml of 1M Tris, pH 8.5, 40µl of 0.5M EDTA, pH 8.0, and 0.5ml of 20% Tween 20® were made up to 20ml with water.

**1M Magnesium Chloride**

For 20ml, 8.412g of $MgCl_2.6H_20$ were dissolved in 10ml of water. The volume was adjusted with water and the solution autoclaved.

**Membrane Washing Solution for Southern Hybridisation**

For 1L of low stringency wash (x2 SSC and 0.1% SDS), 100ml of x20 SSC and 10ml of SDS were added to water up to 1L.

For 1L of high stringency wash (x0.1 SSC and 0.1% SDS), 5ml of x20 SSC and 10ml of 10% SDS were added to water up to 1L.

**x100 Mg (0.1M $MgCl_2$ and 4.5M β-mercaptoethanol)**

For 10ml of x100 Mg, 1ml of 1M $MgCl_2$ and 3.146ml of β-mercaptoethanol were added to water up to 10ml. x100 Mg was stored at 4°C.

**Nuclear Protein Extraction Buffer (20mM HEPES, pH 7.8, 10mM KCl, 0.1mM EDTA, 1mM DTT, and 0.5mM phenylmethylsulfonyl fluoride [PMSF])**

For 10ml of nuclear protein extraction buffer the following were combined: 200µl of 1M HEPES, 100µl of 1M KCl, 2µl of 0.5M EDTA, 10µl of 1M DTT, 500µl of 10mM PMSF, and 9.2ml of water.

For nuclear protein extraction buffer with 5M NaCl, 400µl of 5M NaCl was added to 4.6ml of nuclear protein extraction buffer (as above).

**x100 Penicillin/Gentamicin**

A 100x stock solution of penicillin/gentamicin was prepared by dissolving 1g (2,000,000 IU) of benzylpenicillin (sodium salt for injection, CSL) and 80mg of gentamicin (80mg/2ml, Astra

Pharmaceuticals) in 200ml of sterile water. This solution was filtered and stored in 5ml aliquots at -20°. For use, 5ml of stock solution was added to 500ml of cell culture medium.
Final concentrations: 100 IU/ml penicillin, 40μg/ml gentamicin.

**10mM Pheylmethylsulfonyl Fluoride (PMSF)**

To make 1.74mg/ml (10mM) PMSF (Sigma), 8.7mg of PMSF were added to isopropanol up to 5ml whilst wearing a face mask. PMSF was aliquoted into 1ml lots and stored at -20°C.

**Phosphate-Buffered Saline (PBS)**

For 5L of PBS, 40g of NaCl, 5.7g of $Na_2HPO_4$, 1g of KCl, and 1g of $KH_2PO_4$ were dissolved in water. The volume was adjusted to 5L and the solution autoclaved.

**1μg/μl Poly-Deoxy-Inosinic-Deoxy-Cytidylic Acid (Poly [d(I-C)])**

Ten OD units were added to 500μl of water to make 1μg/μl of poly [d(I-C)]. 1 OD unit of poly [d(I-C)] is equal to 50μg. Aliquots of 100μl were stored at -20°C.

**x2 Polyethylene Glycol 8000 (PEG)**

Forty grams of stock PEG were added to water, the volume adjusted to 100ml and the solution autoclaved. 3.338ml of 40% stock PEG (v/v), 1ml of 3M sodium acetate, pH 5.2, 32.5μl of 1M $MgCl_2$ and 629.5μl of water were added together and aliqouted into 1ml volumes.

**Potassium Acetate for Alkaline Lysis, pH 4.8**

To 60ml of 5M potassium acetate, pH 4.8, 11.5ml of glacial acetic acid (Ajax Chemicals) and 28.5ml of water were added. This solution was 3M with respect to potassium and 5M with respect to acetate and was stored at 4°C.

**1M Potassium Chloride**

For 20ml, 1.49g of KCl were dissolved in water, the volume adjusted to 20ml and the solution autoclaved.

**Prehybridisation/Hybridisation Solution for Southern**

Prehybridisation/hybridisation solutions were made just prior to use. Per block, 1ml of blotto, 5ml of x20 SSC, 2ml of 10% SDS, and 12ml of water were combined. The solution was dissolved at 62°C.

**20mg/ml Proteinase K**

A 20mg/ml stock proteinase K solution was made by dissolving 25mg of proteinase K (Boehringer Mannheim) in 1.25ml of water. The solution was filtered, aliquoted into 50μl volumes and stored at -20°C.

**Resuspension Buffer (25mM Tris-HCl, pH 8.0, 10mM EDTA, pH 8.0, and 50mM glucose) for Minipreps**

For 100ml of resuspension buffer the following were combined: 2.5ml of Tris-HCl, pH 8.0, 2ml of 0.5M EDTA, pH 8.0, and 5ml of 1M glucose. The volume was adjusted to 100ml with water. Resuspension buffer was stored at 4°C.

**RNase A (DNase-free)**

Ten milligrams of RNase A (Boehringer Mannheim) were dissolved in 1ml of 10mM Tris-HCl, pH 7.4, 15mM NaCl, and then heated to 100°C for 15 minutes and cooled slowly to room temperature. 50µl aliquots were stored at -20°C.

**Sephadex G-50 Fine**

One to 2g of Sephadex G-50 fine (Pharmacia) was added to 25ml of TE (10mM Tris, and 1mM EDTA). The Sephadex solution was left overnight at 4°C and degassed for at least 20 minutes to remove any air bubbles before use.

**SOC Medium (2% tryptone, 0.5% yeast extract, 0.05% NaCl, 2.5mM KCl, 10mM $MgCl_2$, and 20mM glucose)**

For 1L of SOC medium the following were combined: 20g of typtone, 5g of yeast extract, 0.5g of NaCl, and 950ml of water. 10ml of 250mM KCl (1.86g KCl in 100ml of water) was added to the solution and the pH was adjusted to 7.5 with concentrated NaOH. The volume was adjusted to 1L with water and the solution autoclaved. After cooling to approximately 55°C, 10ml of sterile 1M $MgCl_2$ and 7.2ml of 50% glucose was added. SOC medium was stored at 4°C.

**3M Sodium Acetate, pH 5.2**

For 25ml, 10.2g of sodium acetate were dissolved in 12.5ml of water using heat. The pH was adjusted to 5.2 with glacial acetic acid and the volume brought up to 25ml with water. The solution was autoclaved.

**5M Sodium Chloride**

For 50ml of 5M NaCl, 14.6g of NaCl were dissolved in water, the volume adjusted to 50ml and the solution autoclaved.

**x20 Sodium Citrate Sodium Chloride (SSC)**

For 1L, 175.3g of NaCl and 88.2g of sodium citrate were dissolved in 800ml of water. The pH was adjusted to 7.0 with concentrated NaOH. The volume was adjusted to 1L with water and the solution autoclaved.

**10% Sodium Dodecyl Sulphate (SDS)**

One hundred grams of SDS (Sigma) were added to 900ml of water and the SDS solution heated to 68°C to assist dissolution. The pH was adjusted to 7.2 with concentrated HCl and the volume adjusted to 1L with water.

**100mM Sodium Hydrogen Phosphate, pH 7.5**

For 100ml, 1.42g of $Na_2HPO_4$ was dissolved in 100ml of water. The pH was adjusted to 7.05 and the solution autoclaved.

**1M Tris, pH 7.5**

For 1L, 121.1g of Tris base were dissolved in 800ml of water. The pH was adjusted to 7.5 by adding approximately 65ml of concentrated HCl. The volume of the solution was adjusted to 1L with water and autoclaved.

**TE (10mM Tris and 1mM EDTA), pH 7.6**

For 100ml of TE, 1ml of 1M Tris and 200µl of 0.5M EDTA were added to water and the volume adjusted to 100ml. The pH was adjusted to 7.6 and the solution autoclaved.

**TE (10mM Tris and 0.5mM EDTA), pH 7.6**

For 100ml of TE, 1ml of 1M Tris and 100µl of 0.5M EDTA were added to water and the volume adjusted to 100ml. The pH was adjusted to 7.6 and the solution autoclaved.

**x5 Tris Acetate Electrophoresis (TAE) Buffer**

For 1L of buffer, 242g of Tris base (Sigma) were added to 57ml glacial acetic acid and 100ml of 0.5M EDTA, pH 8.0. The volume was adjusted to 1L with water and the solution autoclaved.

**x5 Tris Borate Electrophoresis (TBE) Buffer**

For 1L of buffer, 54g of Tris base were added to 27.5g of boric acid (ICN) and 20ml of 0.5M EDTA, pH 8.0. The volume was adjusted to 1L with water and the solution autoclaved.

**Tris-Buffered Saline (TBS)**

Eight grams of NaCl, 0.2g of KCl, and 3g of Tris base were dissolved in 800ml of water. The pH was adjusted to 7.4 with HCl, and the volume adjusted to 1L with water. The solution was autoclaved.

**0.25% Trypsin/0.25% EDTA**

Trypsin/EDTA solution was made by combining: 1.6g of NaCl, 0.08g of KCl, 0.2g of dextrose (D-glucose), 0.12g of $NaHCO_3$, 0.5g of trypsin 1:250 (Difco), 0.04g of $Na_2EDTA$, and 0.01g of phenol red in 200ml of water by vigorous stirring. After allowing the solution to stand for 30 minutes at 4°C, the solution was stirred again for 5 to 10 minutes. The pH was adjusted to 7.8 with 1M NaOH (trypsin is not active at a pH less than 7.6). The solution was filtered and stored in 20ml aliquots at -20°C.

# APPENDIX 7.2

## TECHNIQUES

### A7.2.1 Nucleic Acid Quantitation

The ratio between the optical density (OD) of samples at 260nm and 280nm ($OD_{260}/OD_{280}$) provides an estimate of the purity of nucleic acids provided the value lies between 1.8 and 2.0. The OD at 260nm measures the DNA in the sample, whilst the OD at 280nm measures protein. Contamination with protein or phenol produces ratios significantly less than these, in which case accurate quantitation of nucleic acids is not possible.

An OD of one, at a wavelength of 260nm and a slit width of 0.2, corresponds to approximately 50μg/ml of double-stranded DNA, approximately 40μg/ml of single-stranded DNA and RNA, and approximately 30μg/ml of single-stranded oligonucleotides. Two microlitres of each nucleic acid to be quantitated was added to 48μl of water (a dilution of 1 in 25). The absorbance was read at 260nm and 280nm in a microcuvette under UV using a spectrophotometer (Beckman DU®640). The concentration of the sample was determined as follows:

$Absorbance_{260nm}$ x dilution factor x 50 (or 40 for double-stranded RNA, or 30 for oligonucleotides) = μg/ml nucleic acid.

### A7.2.2 Guidelines for the Design of Primers for Polymerase Chain Reaction

When designing primer sets for PCR of the HPV 16 genome the following guidelines were followed (from Boehringer Mannheim):

- The primer length should be between 11 and 35 base pairs with 20 to 28 bases to be considered as the optimal length.
- Select primers with a random base distribution and a GC content of 50% or similar to that of the target.

- Avoid stretches of polypurines or pyrimidines (oligo(dT) and poly(dC) may be used as an exemption).
- Check the primers against each other for complementary sequences, i.e., avoiding 3' overlaps reduces the formation of "primer dimers".
- Avoid sequences with significant secondary structures, particularly at the 3' end.
- Add sequences not complementary to the template to the 5' end of the primer.
- the 3' ends should represent conservative amino acids and should not contain "wobble bases".
- Keep in mind that DNA polymerase does not extend mismatches between the primer and target.
- The distance between the primers is rather flexible, ranging up to 10kb.
- Preferably use purified primers although non-purified primers may also be used.
- The primers should be added in a concentration range of 0.1 to 1μmol/l with a ratio of 1:1.

## A7.2.3 Agarose Gel Electrophoresis

Agarose is a polymer, extracted from seaweed, and gel electrophoresis provides a qualitative means of obtaining results from PCR. During electrophoresis negatively charged DNA molecules migrate through the agarose under the influence of an electrical potential difference towards the anode, with the rate of migration through the gel dependent on a number of parameters including the size of the molecules. The fragments of DNA are visualised by staining with the dye ethidium bromide, which intercalates between DNA base pairs and fluoresces under ultraviolet (UV) transillumination. The size of bands on the gel is estimated by comparison of their position relative to those of an appropriate size marker.

### A7.2.3.1 Casting a Gel

1. For small PCR products (100 base pairs to 1000 base pairs), agarose concentrations of 2% to 3% were needed for clear resolution of bands. For a 2% gel, consisting of 2% Nu Sieve™ agarose (FMC Bioproducts), 1.2 grams of agarose was added to 60ml of x1 tris acetate electrophoresis (TAE) buffer. Sixty millilitres of melted agarose was required to produce a gel of approximately 3mm

in thickness for the 20 well Hoefer HE 99 Max submarine electrophoresis unit (20.5cm by 14.5cm gel). Alternatively, 30ml of melted agarose was required to produce a gel of approximately 3mm thickness for the eight well electrophoresis unit (10cm by 7cm).

2. A flask of TAE and powder agarose was covered with Gladwrap™, a hole poked in the top to allow steam to escape and the flask weighed.
3. The flask was microwaved on high for 1 minute to dissolve the agarose. The flask was reweighed and the appropriate volume of water added to restore the weight to its previous value.
4. The flask contents were cooled to about 55°C and the agarose poured into the assembled gel casting unit and allowed to set for approximately 30 minutes.
5. The well former was removed carefully, the gel lifted out of the casting tray and positioned in the electrophoresis unit.
6. The electrophoresis tank was filled with x1 TAE buffer so that the gel was just covered with about a millimetre of buffer.

### A7.2.4 Long Term Storage of Plasmid Constructs

A glycerol stock of each positive construct was prepared for safekeeping by combining 850µl of overnight bacterial culture with 150µl of sterile glycerol. This was vortexed and transferred to a cryovial for storage at -70°C.

Glycerinated cultures can be stored at -70°C for many years without loss of viability. Viable bacteria were recovered by scraping a sterile loop over the surface of a frozen stock, before returning the remainder of the suspension to the freezer.

### A7.2.5 Maintaining Cell Lines

HPV 18 positive HeLa cells (Pater and Pater, 1995) and HPV negative HT3 cells were maintained in cell culture flasks at 37°C with 5% $CO_2$. Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) with 7% fetal calf serum (FCS), penicillin (100 IU/ml) and gentamicin (40µg/ml). Cells were removed from the flask

once confluent using trypsin and either seeded into another flask, used for transfection experiments, or stored in liquid nitrogen.

## A7.2.6 Removing Cells from Flasks

1. The growth medium was decanted and the flask washed twice with phosphate-buffered saline (PBS) using sterile technique. Trypsin is not active at acid pH and works optimally at 37°C.
2. Approximately 2ml of trypsin was added to the flask and the flask incubated briefly at 37°C. The flask was then tapped to help unattach cells.
3. The cells were removed from flasks using a pasteur pipette, leaving some cells for further growth if required, and the remaining trypsin inactivated by adding media.

## A7.2.7 Storing Cells in Liquid Nitrogen

For each large flask ($75cm^2$) of confluent cells:

1. The growth medium was tipped off and the flask washed twice with warm PBS.
2. Approximately 2ml of trypsin was added to the flask and the flask incubated briefly at 37°C. The flask was then tapped to help remove the cells.
3. The cells were added to a 10ml tube, filled with PBS and spun for 5 minutes at 1,200 rpm (in an MSE large bench centrifuge).
4. The supernatant was tipped off and approximately 3.5ml of growth medium with 10% FCS added. The percentage of FCS was increased to 20% and 10% DMSO, (400µl) was added. The volume was made up to 4ml and the contents of the tube mixed well.
5. Two millilitres of cells were pipetted into freezer vials and placed at 4°C for 1 hr.
6. The vials were then placed at -20°C for 2 hrs. Then at -70°C for 1 to 2 hrs before storing long-term in liquid nitrogen.

### A7.2.8 Counting Cells

For cell counts a 1:2 dilution of the cell suspension was prepared using 0.1% Trypan Blue in PBS (100µl cell suspension + 200µl 0.1% Trypan Blue). The diluted specimen was allowed to just fill under the coverslip on both sides of a Neubauer counting chamber slide by capillary action. The slide was examined under x10 magnification and the living (unstained) cells counted in each of the 4 corner squares on each side of the counting chamber (Figure A7.2.1). Dead cells, which take up the stain, were not counted, as well as cells that landed on two sides of the square. The number of cells in the living suspension was determined using the calculation: n (average of the 4 squares) x 3 x $10^4$ = number of cells/ml.

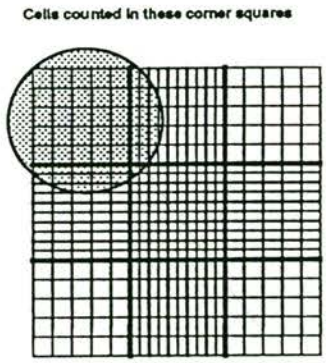



Figure A7.2.1: The squares counted in a Neubauer chamber in order to determine the number of cells in a suspension

# APPENDIX 7.3

## SIZE MARKERS FOR GEL ELECTROPHORESIS

The fragment sizes for each of the DNA size markers, pUC19 DNA/*Hpa*II (Biotech), and 1kb DNA Ladder (Life Technologies, Gibco BRL) used in this study, are given in Table A7.3.1 below.

Table A7.3.1: Fragments Sizes for DNA Markers Used in Gel Electrophoresis

| **pUC19 DNA/*Hpa*II (Biotech)** | | **1kb DNA Ladder (Life Technologies, Gibco BRL)** | |
|---|---|---|---|
| Fragment Number | Number of Base Pairs | Fragment Number | Number of Base Pairs |
| 1 | 501 | 1 | 12216 |
| 2 | 489 | 2 | 11198 |
| 3 | 404 | 3 | 10180 |
| 4 | 331 | 4 | 9162 |
| 5 | 242 | 5 | 8144 |
| 6 | 190 | 6 | 7126 |
| 7 | 147 | 7 | 6108 |
| 8 | 111 | 8 | 5090 |
| 9 | 110 | 9 | 4072 |
| 10 | 67 | 10 | 3054 |
| 11 | 34 | 11 | 2036 |
| 12 | 34 | 12 | 1636 |
| 13 | 26 | 13 | 1018 |
| | | 14 | 517 |
| | | 15 | 506 |
| | | 16 | 396 |
| | | 17 | 344 |
| | | 18 | 298 |

# APPENDIX 7.4

## SEQUENCING METHODS

DNA sequencing involves the enzymatic synthesis of labelled DNA, using fluorescently labelled dideoxynucleotides to terminate the elongating DNA strand (Sanger dideoxy method). Chain termination uses a DNA polymerase to extend the DNA stand until a chain terminating 2'3'-dideoxynucleotide (ddNTP) is incorporated. ddNTPs lack the 3-OH group necessary for elongation. This results in a nested set of terminated chains, that are separated according to size by a high resolution denaturing gel (adapted from Alphey, 1995).

### A7.4.1 Sequencing Using ABI PRISM® Big Dye™ Terminator Chemistry (PE Applied Biosystems)

Sequencing of purified PCR products and cloned DNA was carried out by three institutions, the Children's Medical Research Institute (CMRI), Sydney University Prince Alfred Molecular Analysis Centre (SUPAMAC) and the Australian Genome Research Facility (AGRF) at the University of Queensland. The method outlined below is that used by SUPAMAC, which was similar to that used by the other institutions. The procedure involved three basic steps; cycle sequencing, removal of excess dye-terminators, and sequence analysis on the Applied Biosystems Model 373 version 3.3 Sequencer.

### A7.4.2 Cycle Sequencing

Cycle sequencing was performed in a thermal sequencer using the ABI PRISM® Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit with *AmpliTaq*® DNA polymerase (Perkin Elmer) according to the manufacturer's instructions.

## A7.4.3 Preparation of the Reaction Mix

The following mix of reagents (Table A7.4.1) was prepared and mixed thoroughly. In general, the higher the DNA quantities used, the higher the resultant signal intensity.

Table A7.4.1: Reagent Mix for Cycle Sequencing Reaction

| **Reagent** | **Volume per Reaction** |
|---|---|
| Terminator Ready Reaction Mix | 4µl |
| Template:<br>dsDNA<br>PCR product<br>Primer | <br>200-500ng<br>30-90ng<br>3.2pmol<br>= 8µl |
| Total Volume | 12µl |

## A7.4.4 Cycling Conditions

Cycle sequencing was commenced immediately for 25 cycles in a GeneAmp PCR system 9700 thermal cycler as outlined in Table A7.4.2.

Table A7.4.2: Cycling Conditions

| |
|---|
| Rapid (1°C/1 sec) thermal ramp to 96°C |
| 96°C for 10 seconds |
| Rapid thermal ramp to 50°C |
| 50°C for 5 seconds |
| Rapid thermal ramp to 60°C |
| 60°C for 4 minutes |
| Rapid ramp to 4°C |

On completion of the extension reaction, excess dyedeoxy terminators were removed without delay as outlined below.

### A7.4.5 Purifying Extension Products to Remove Excess DyeDeoxy Terminators

Extension products were purified in Centri-Sep™ spin columns (Princeton Separations). The column was hydrated for two hours by adding 0.8ml of water and inverting the column several times to mix the water and gel material. The water was emptied from the column by gravity flow and then by spinning at 730 x g for two minutes. The column was inserted into a sample collection tube and the sample loaded in the centre of the column, making sure the sample did not touch the sides of the column and the pipette tip did not touch the gel surface. The column was then spun again for 2 minutes such that the sample was collected in the bottom of the tube. The sample was dried in a vacuum centrifuge for 10 to 15 minutes or until dry.

### A7.4.6 Preparing/Loading Samples and Analysis on the Applied Biosystems 377 DNA Sequencer

Each dry pelleted DNA sample was resuspended in a loading buffer containing a 5:1 ratio deionised formamide and 25mM EDTA pH 8.0, with 50mg/ml blue dextran. After vigorous agitating to dissolve the residue, the solution was centrifuged briefly to collect all the fluid at the bottom of the tube. Just prior to loading, the samples were heated at 95°C for two minutes to denature and then transferred immediately to ice. The samples were then loaded onto the Long Ranger™ acrylamide gel according to the manufacturer's instructions.

# APPENDIX 7.5

## SELECTED SEQUENCE ELECTROPHEROGRAMS

A typical sequence electropherogram for each of the primers used for DNA sequence analysis is presented on the following pages. The primers used and corresponding samples illustrated are given in Table A7.5.1 below.

Table A7.5.1: Primers and Corresponding Examples of Sequence Elelctropherograms

| Genomic Region | Primer | Patient Isolate | Figure Number |
|---|---|---|---|
| URR | URR-1 | M1 | A7.5.1 |
| | URR-2 | K1 | A7.5.2 |
| | URR-5 | K4 | A7.5.3 |
| | URR-4 | K1 | A7.5.4 |
| E6 | E6-1 | P2 | A7.5.5 |
| | E6-2 | M1 | A7.5.6 |
| E2 | E2-1 | M6 | A7.5.7 |
| | E2-2 | PRE262 | A7.5.8 |

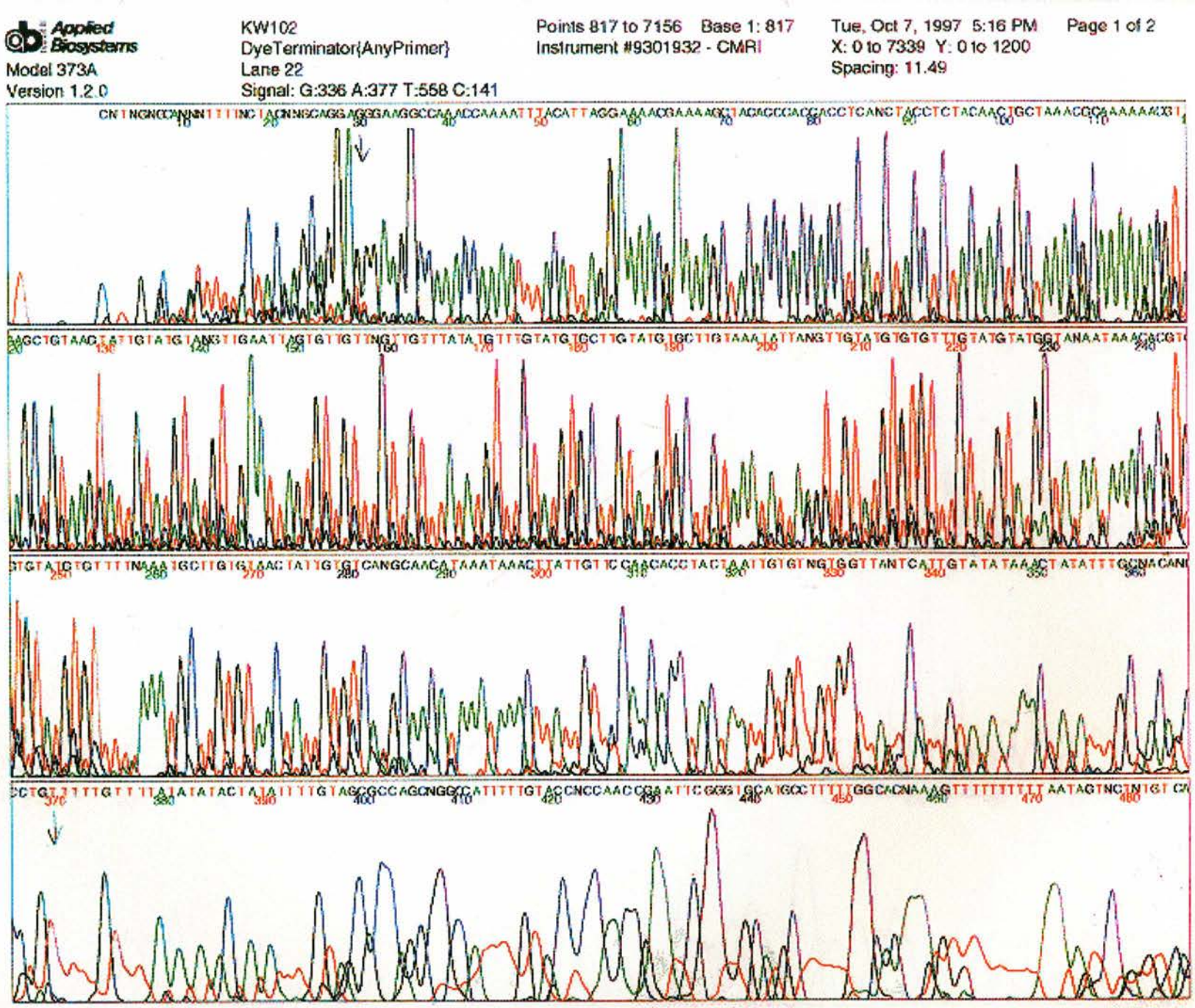



Figure A.7.5.1: Representative DNA sequence electropherogram for primer URR-1

Applied Biosystems
Model 373A
Version 1.2.0

KW032
DyeTerminator{AnyPrimer}
Lane 21
Signal: G:197 A:535 T:530 C:163

Points 749 to 7156 Base 1: 749
Instrument #9301932 - CMRI

Wed, Jul 16, 1997 5:25 PM Page 1 of 2
X: 0 to 6304 Y: 0 to 1200
Spacing: 12.67

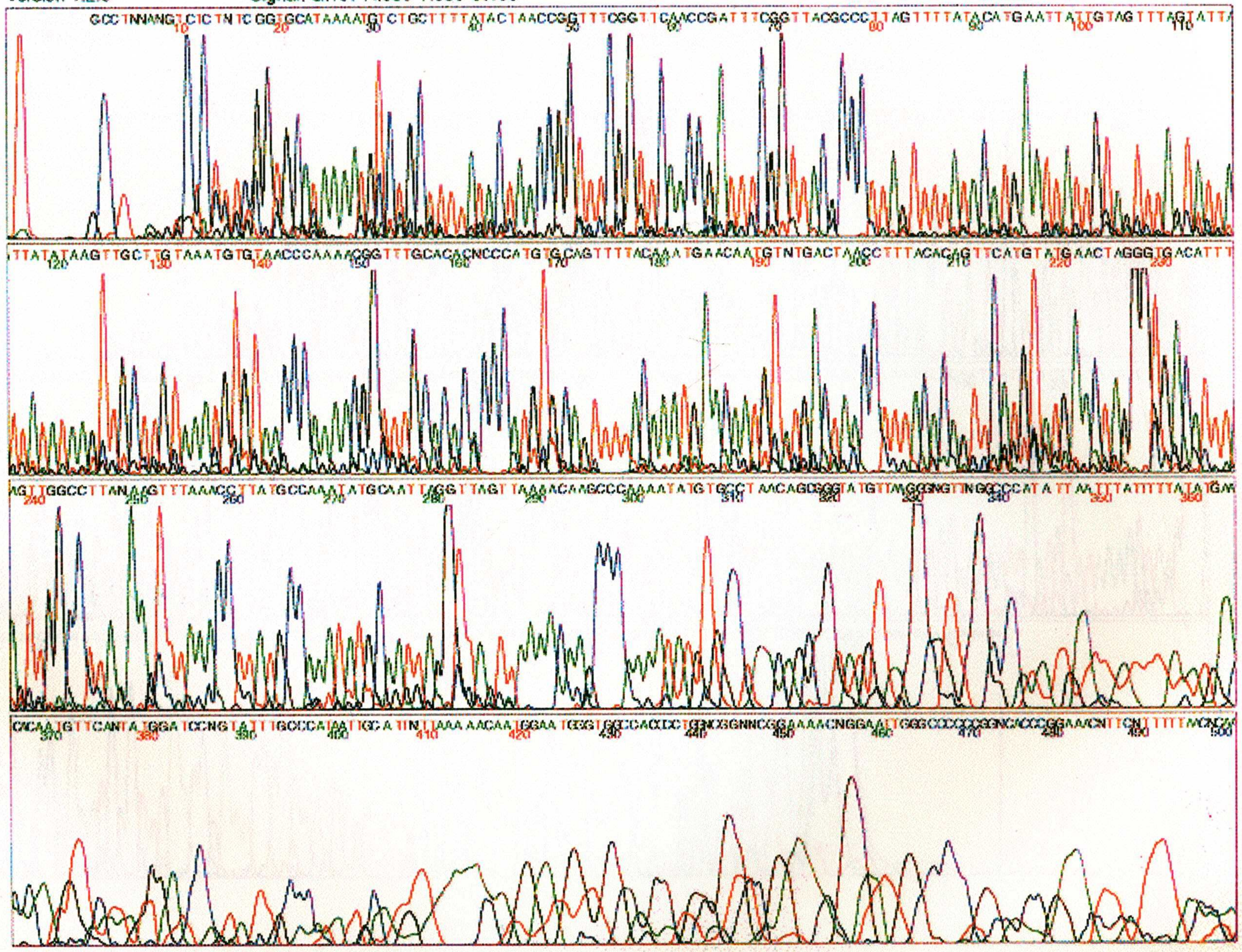


Figure A.7.5.2: Representative DNA sequence electropherogram for primer URR-2

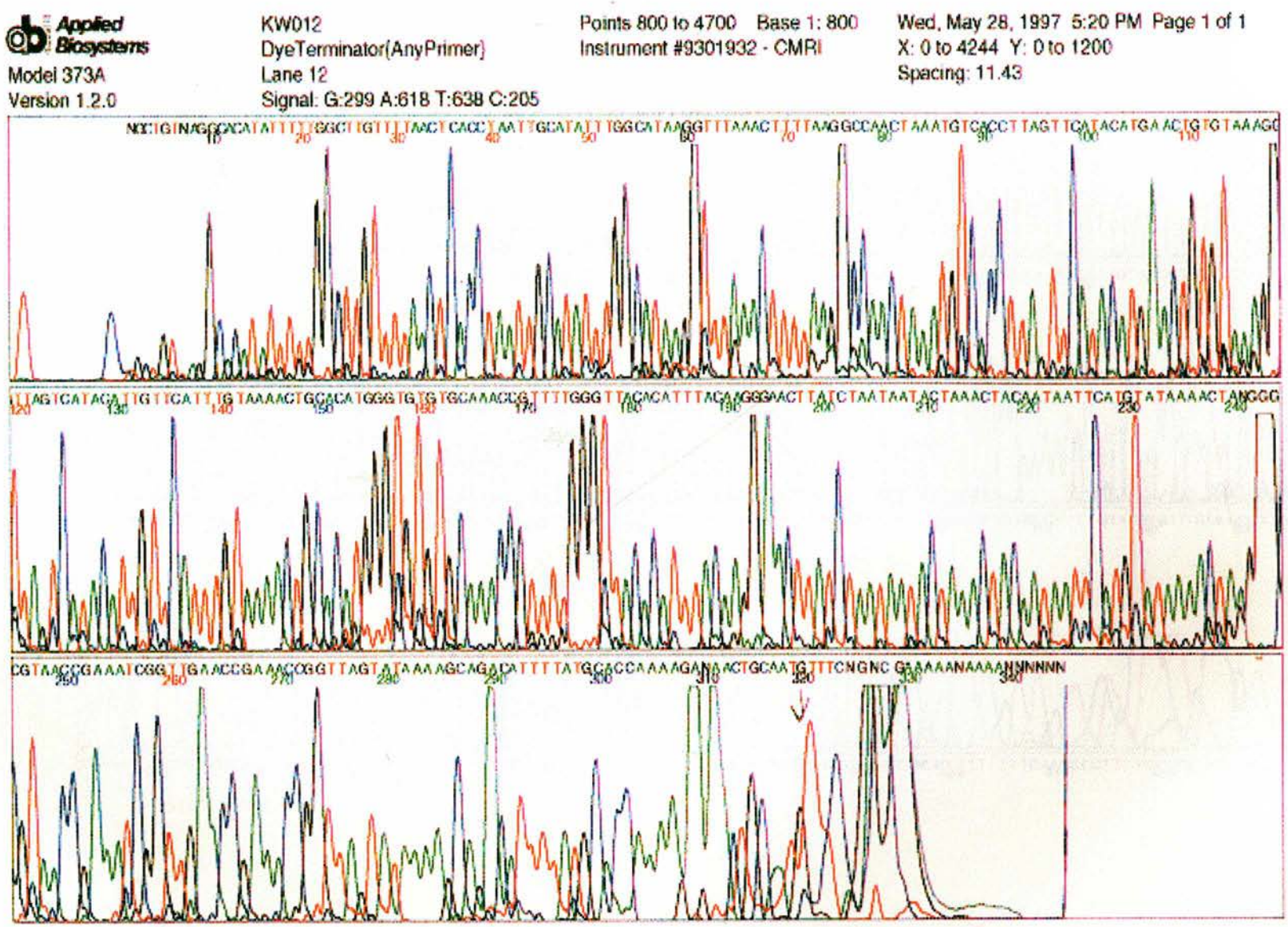



Figure A.7.5.3: Representative DNA sequence electropherogram for primer URR-5

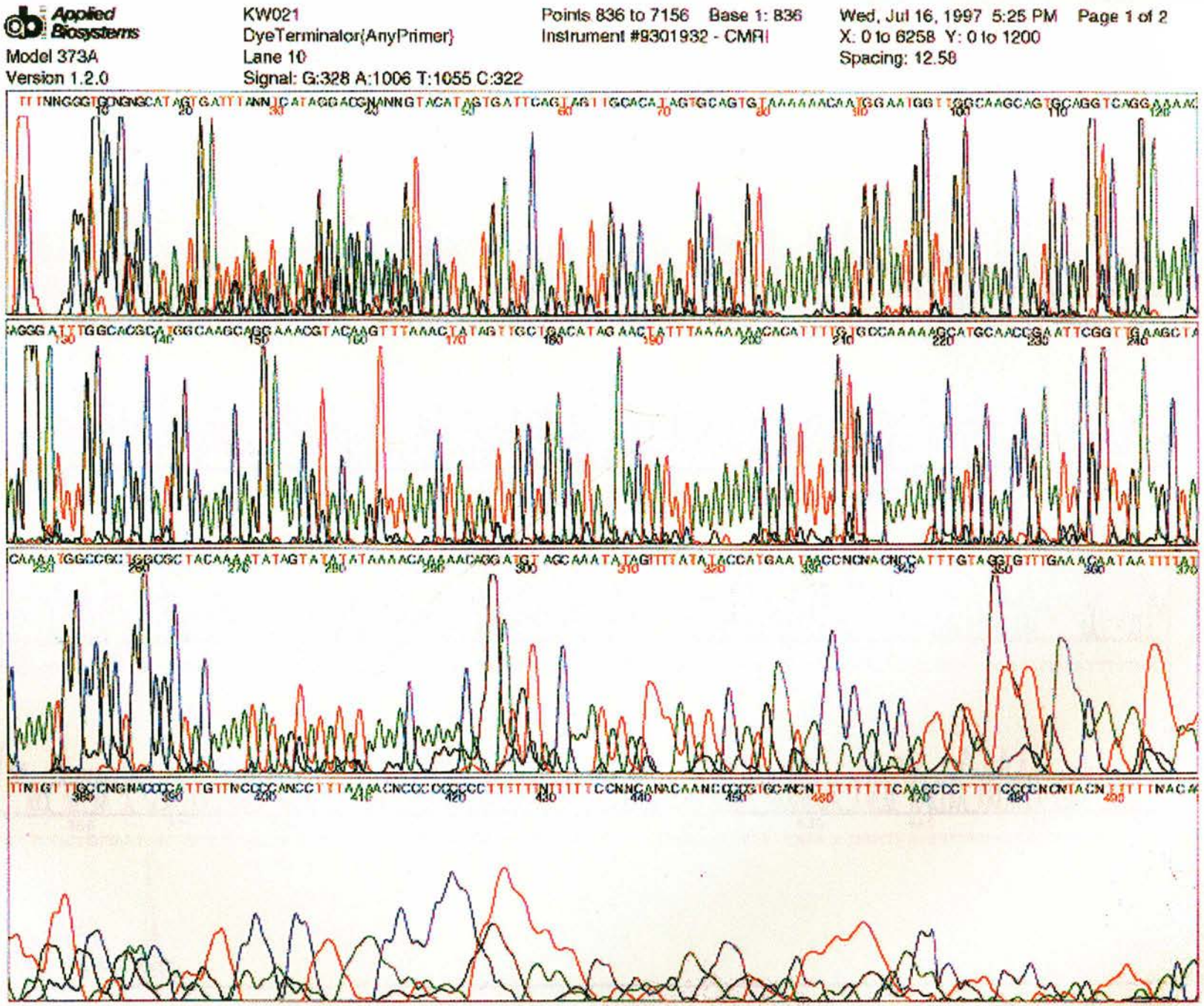



Figure A.7.5.4: Representative DNA sequence electropherogram for primer URR-4

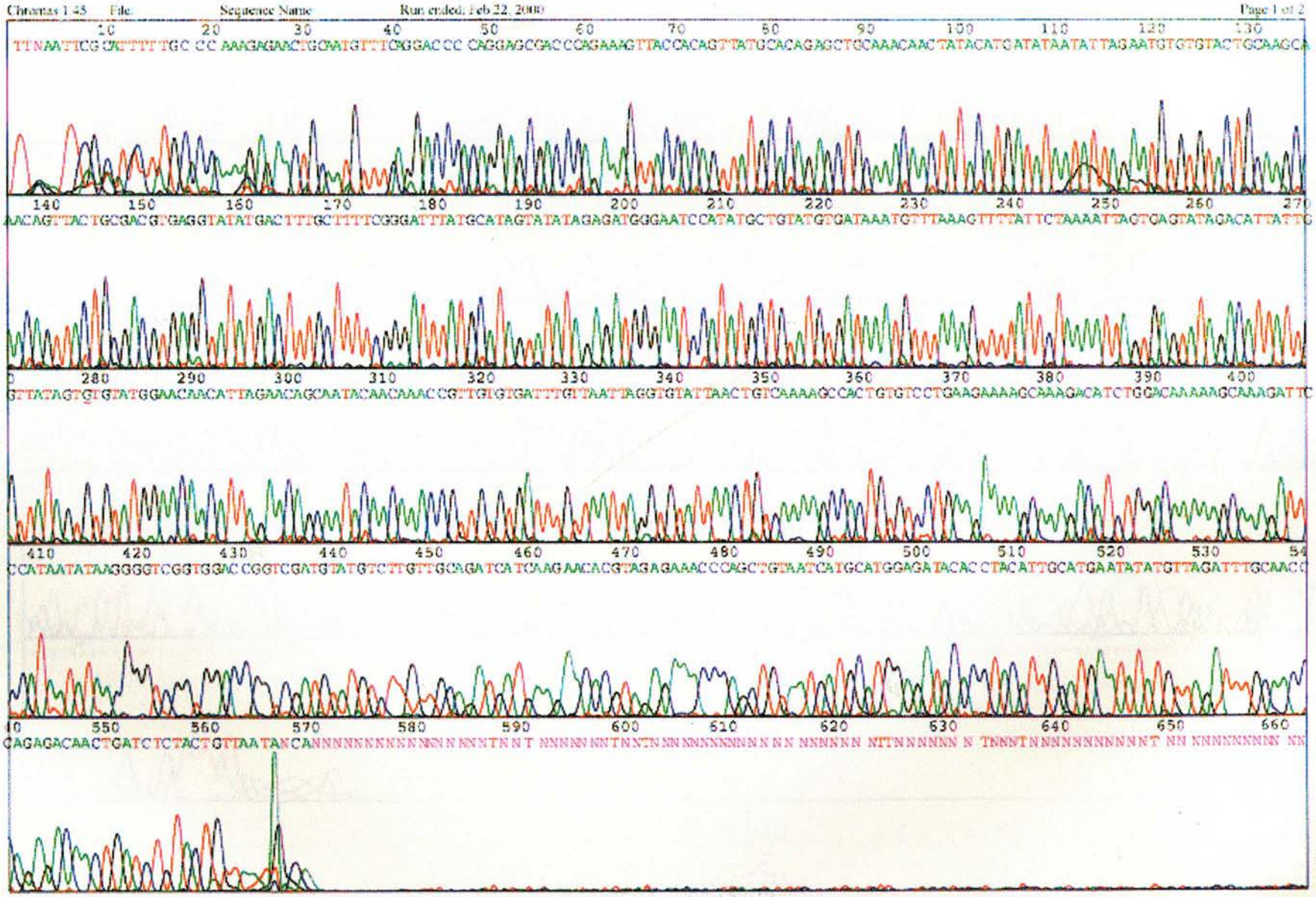



Figure A.7.5.5: Representative DNA sequence electropherogram for primer E6-1

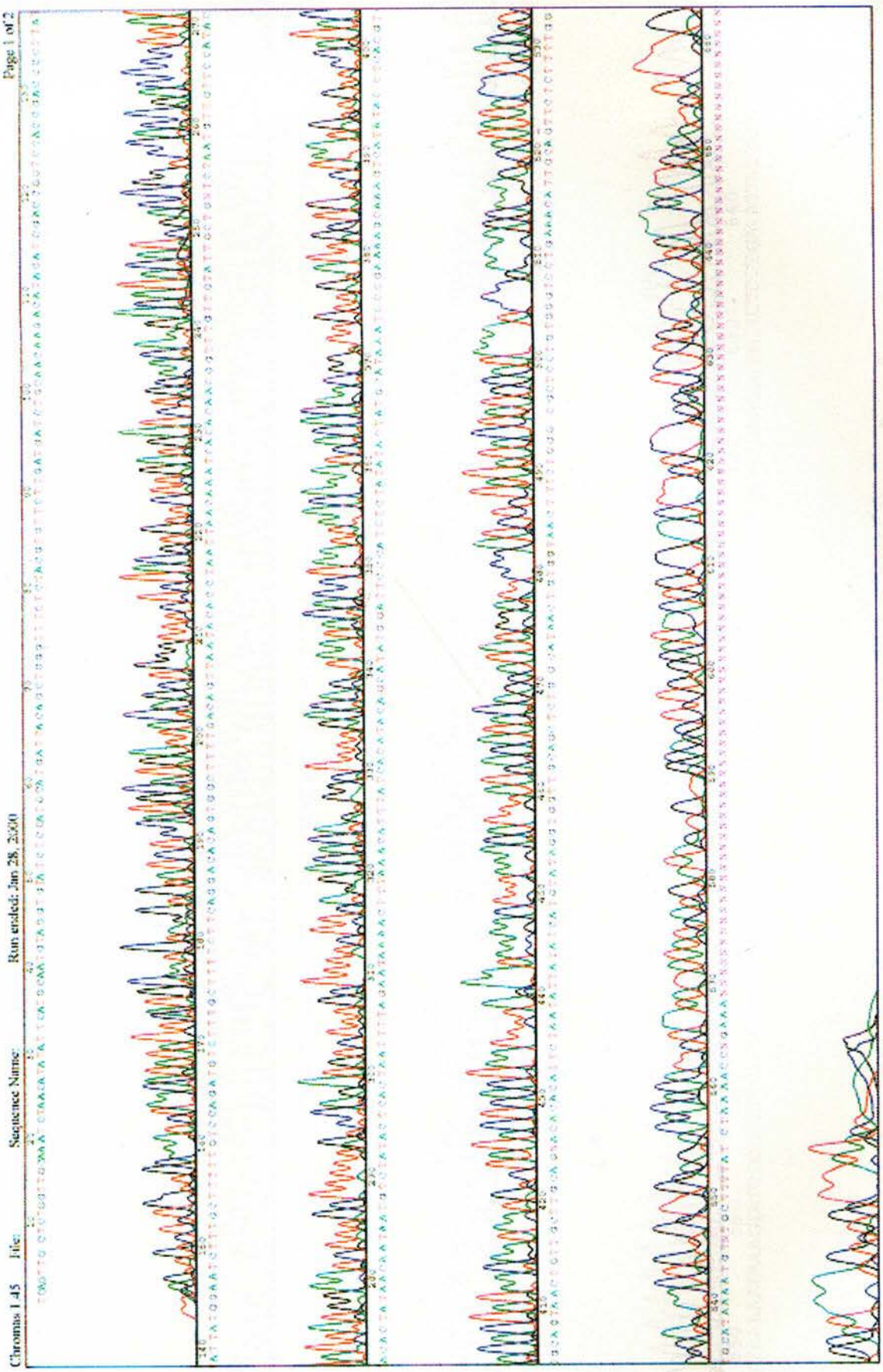



Figure A.7.5.6: Representative DNA sequence electropherogram for primer E6-2

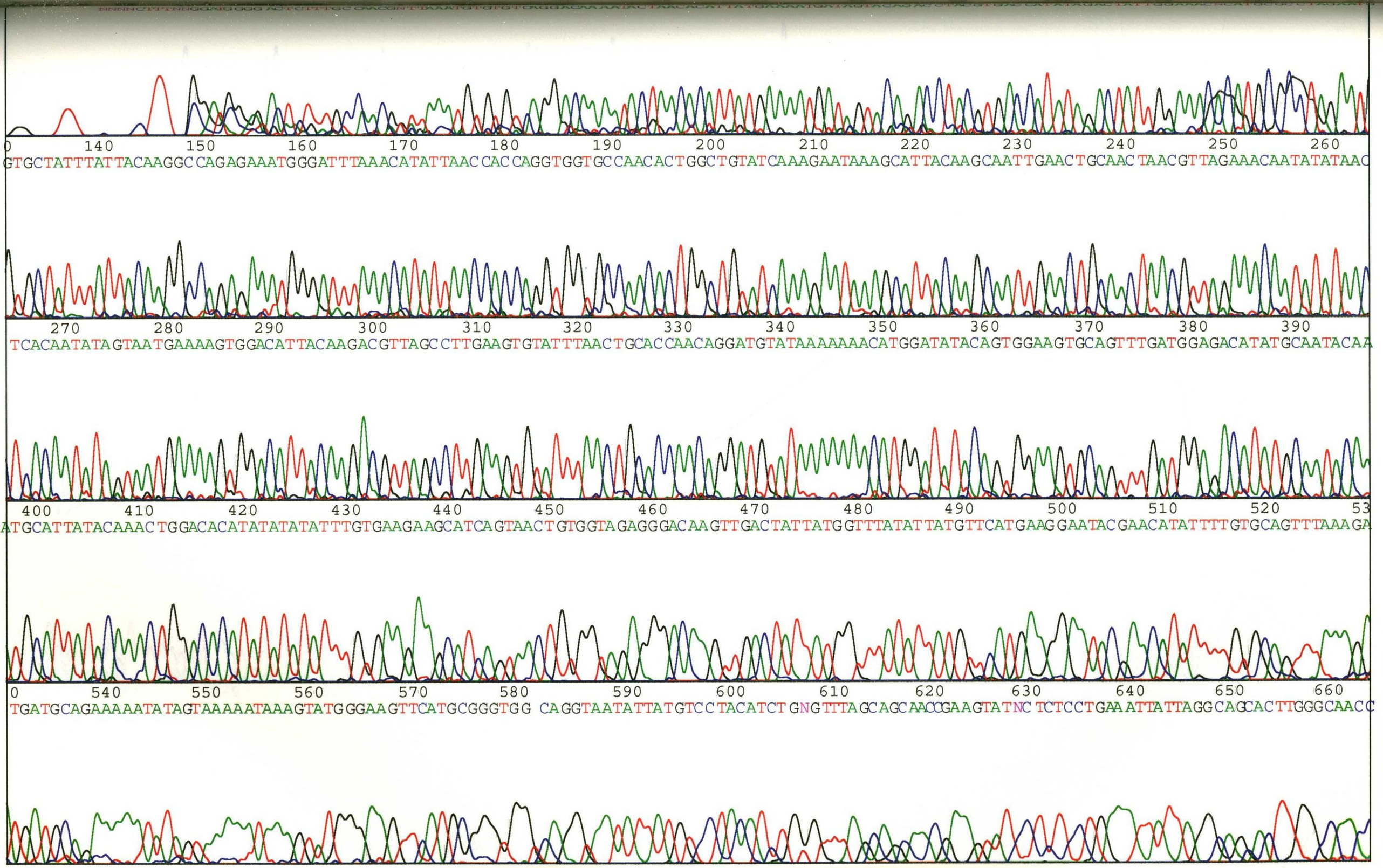



Figure A.7.5.7: Representative DNA sequence electropherogram for primer E2-1

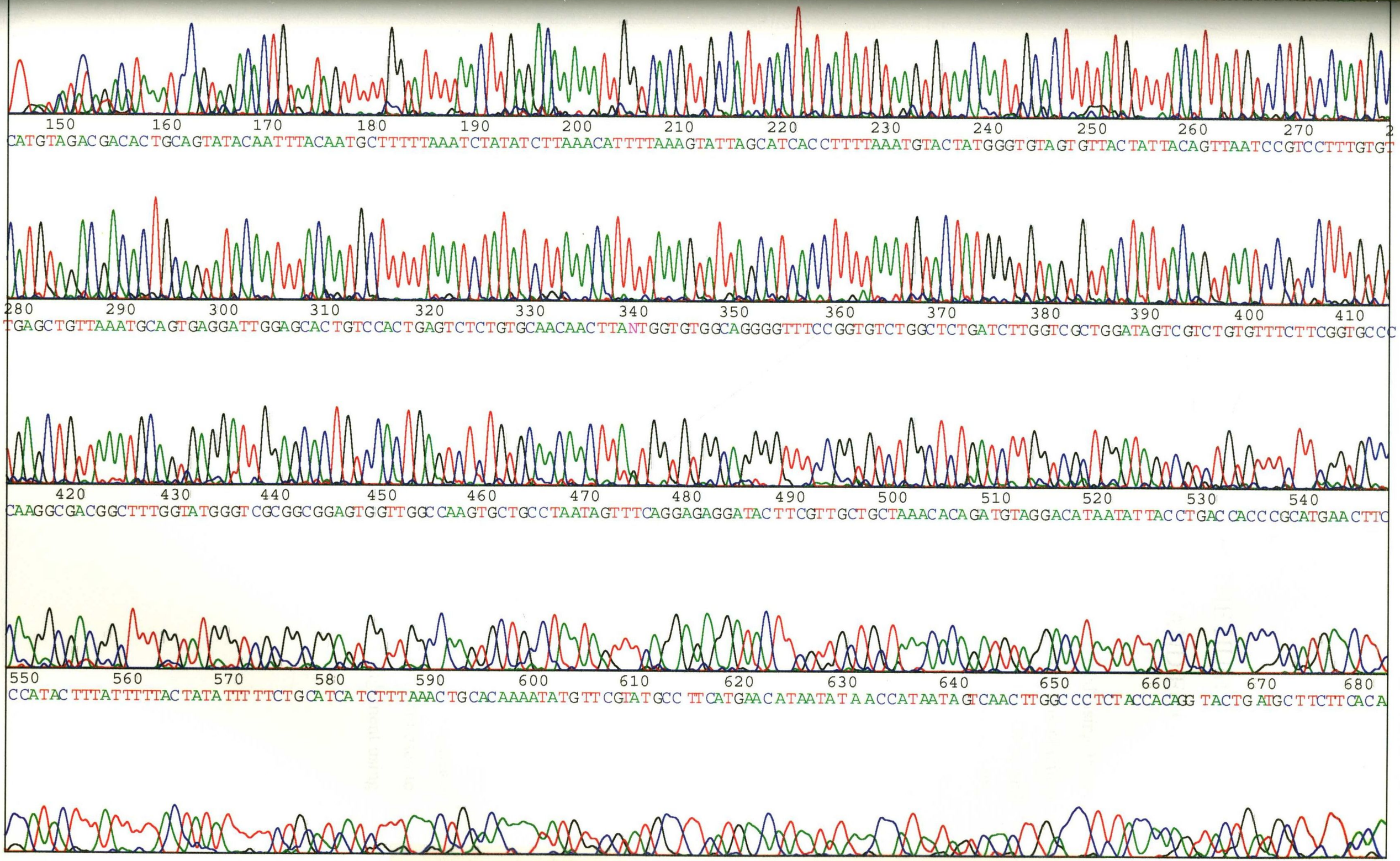



Figure A.7.5.8: Representative DNA sequence electropherogram for primer E2-2

# APPENDIX 7.6

# SEQUENCE ALIGNMENTS FOR THE HUMAN PAPILLOMAVIRUS TYPE 16 URR, E2/E4, AND E6 GENES

The following pages contain the DNA sequence alignments obtained using Pretty software from ANGIS, and the amino acid alignments of coding regions obtained using Translate and Eclustalw software also available from ANGIS. The DNA and amino acid sequences of the HPV 16 isolates are compared with the prototype HPV 16 sequence (highlighted in bold in each case). Each patient sample as well as the prototype is identified in brackets at the beginning of each row of sequence alignment. For nucleotide alignments, the consensus sequence is located at the bottom of each row of sequence, with a letter (A, T, G, or C) representing the consensus nucleotide and a hypen (-) representing a nucleotide change in at least one of the sequences at that position. For amino acid analysis, a star (*) at the bottom of the alignment represents the consensus sequence at that codon, and a space represents at least one amino acid change at that position.

## A7.6.1 URR

The sequence of all or the majority of the URR from 24 of the 34 isolates is presented here. This appendix enables differentiation between sequence data obtained during this study and that obtained during a previous study: the reader is referred to the sequence information in Figure 2.4 where all sequence data for the URR of the 34 isolates is presented for completion. Four separate sequence alignments are given for the cancer isolates, one for each of the URR primers used to obtain sequence information. This is followed by the URR sequence alignment using primer URR-3 for the noncancer isolates.

# CANCER ISOLATES

# NUCLEOTIDE ALIGNMENT FOR PRIMER URR-1 (nts 7010-7034)

```
                7060
  1pileup{1(M1)}  GAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(M2)}  .AAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(K5)}  .AAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(M4)}  .AAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(D3)}  GAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(D1)}  GAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(S1)}  AAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(K6)}  .AAGGCCAAA CCAAAATTTA CATAAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(M6)}  .AAGGCCAAA CCAAAATTTA CATAAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(J1)}  GAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(M5)}  GAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACGCCCACCA
  1pileup{1(C1)}  GAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
1pileup{1(prot)}  GAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(K1)}  .AAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(D2)}  .AAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(H3)}  .AAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(G1)}  .AAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(T1)}  GAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
  1pileup{1(P2)}  GAAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT ACACCCACCA
       Consensus  -AAGGCCAAA CCAAAATTTA CATTAGGAAA ACGAAAAGCT AC-CCCACCA

                7110
  1pileup{1(M1)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(M2)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(K5)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(M4)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(D3)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(D1)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(S1)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(K6)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(M6)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(J1)}  CCTTATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(M5)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(C1)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
1pileup{1(prot)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(K1)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(D2)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(H3)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(G1)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(T1)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
  1pileup{1(P2)}  CCTCATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA
       Consensus  CCT-ATCTAC CTCTACAACT GCTAAACGCA AAAAACGTAA GCTGTAAGTA


                7160
  1pileup{1(M1)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
  1pileup{1(M2)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
  1pileup{1(K5)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
  1pileup{1(M4)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
  1pileup{1(D3)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
  1pileup{1(D1)}  TTGTATGTAT GTTGACTCAG TGTTGTTTGT TGTTTATATG TCTGTATGTG
  1pileup{1(S1)}  TTGTATGTAT GTTGACTCAG TGTTGTTTGT TGTTTATATG TCTGTATGTG
  1pileup{1(K6)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTGTATATG TTTGTATGTG
  1pileup{1(M6)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTGTATATG TTTGTATGTG
  1pileup{1(J1)}  TTGTATGTAT GTCGAATTAG TGTTGTTTGT TGTGTATATG TTTGTATGTG
  1pileup{1(M5)}  TTGTATGTAT GTCGAATTAG TGTTGTTTGT TGTGTATATG TTTGTATGTG
  1pileup{1(C1)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
1pileup{1(prot)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTGTATATG TTTGTATGTG
  1pileup{1(K1)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
  1pileup{1(D2)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
  1pileup{1(H3)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
  1pileup{1(G1)}  TTGTATGTAT GTTGACTCAG TGTTGTTTGT TGTTTATATG TCTGTATGTG
  1pileup{1(T1)}  TTGTATGTAT GTTGAATTGG TGTTGTTTGT TGTTTATATG TTTGTATGTG
  1pileup{1(P2)}  TTGTATGTAT GTTGAATTAG TGTTGTTTGT TGTTTATATG TTTGTATGTG
       Consensus  TTGTATGTAT GT-GA-T--G TGTTGTTTGT TGT-TATATG T-TGTATGTG
```

```
                7210
lpileup{1(M1)}  CTTGTATGTG CTTGTAAATA TTATGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(M2)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(K5)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(M4)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(D3)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(D1)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(S1)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(K6)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(M6)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(J1)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(M5)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(C1)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(prot)} CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(K1)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(D2)}  CTTGTATGTG CTTGTAAATA TTACGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(H3)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(G1)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(T1)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
lpileup{1(P2)}  CTTGTATGTG CTTGTAAATA TTAAGTTGTA TGTGTGTTTG TATGTATGGT
     Consensus  CTTGTATGTG CTTGTAAATA TTA-GTTGTA TGTGTGTTTG TATGTATGGT

                7260
lpileup{1(M1)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(M2)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(K5)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(M4)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(D3)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(D1)}  ATAATAAACA TGTGTGTATG TGTTTTTCAA TGCTTGTGTA ACTATTGTGT
lpileup{1(S1)}  ATAATAAACA TGTGTGTATG TGTTTTTCAA TGCTTGTGTA ACTATTGTGT
lpileup{1(K6)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(M6)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(J1)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(M5)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(C1)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(prot)} ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(K1)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(D2)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(H3)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(G1)}  ATAATAAACA TGTGTGTATG TGTTTTTCAA TGCTTGTGTA ACTATTGTGT
lpileup{1(T1)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
lpileup{1(P2)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
     Consensus  ATAATAAACA -GTGTGTATG TGTTTTT-AA TGCTTGTGTA ACTATTGTGT


                7310
lpileup{1(M1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(M2)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(K5)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(M4)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(D3)}  CAGGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(D1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(S1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(K6)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(M6)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(J1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(M5)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(C1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(prot)} CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(K1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(D2)}  GATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(H3)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(G1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(T1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
lpileup{1(P2)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
     Consensus  -A-GCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
```

```
                7360
  lpileup{1(M1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACA....... ..........
  lpileup{1(M2)}  GTTATTCATT GTA....... .......... .......... ..........
  lpileup{1(K5)}  GTTATTCATT GTA....... .......... .......... ..........
  lpileup{1(M4)}  GTTATTCATT GTA....... .......... .......... ..........
  lpileup{1(D3)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  lpileup{1(D1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACA....... ..........
  lpileup{1(S1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATC..... ..........
  lpileup{1(K6)}  GTTATTCATT GTA....... .......... .......... ..........
  lpileup{1(M6)}  GTTATTCATT GTA....... .......... .......... ..........
  lpileup{1(J1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACA....... ..........
  lpileup{1(M5)}  GTTATTCATT GTATATAAAC TATATTTGCT ACA....... ..........
  lpileup{1(C1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACAACCTGTT TTTGTTTTAT
lpileup{1(prot)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  lpileup{1(K1)}  GTTATTCATT GTA....... .......... .......... ..........
  lpileup{1(D2)}  GTTATTCATT GTA....... .......... .......... ..........
  lpileup{1(H3)}  GTTATTCATT GTA....... .......... .......... ..........
  lpileup{1(G1)}  GTTATTCATT GTA...... .......... .......... ..........
  lpileup{1(T1)}  GTTATTCATT GTATAAAAAC TATATTTGCT ACA....... ..........
  lpileup{1(P2)}  GTTATTCATT GTA....... .......... .......... ..........
       Consensus  GTTATTCATT GTA------- ---------- ---------- ----------
```

# NUCLEOTIDE ALIGNMENT FOR PRIMER URR-4 (nts 7703-7685)

```
                7260
  4pileup{4(M2)}  .......... .....GTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(P2)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(G1)}  .......... .....GTATG TGTTTTTCAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(M1)}  .......... .......... .......... .......... ..........
  4pileup{4(D2)}  .......... .......ATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(H3)}  .......... .....GTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(J1)}  .......... .......... .......... .......... ..........
4pileup{4(prot)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(S2)}  .......... .......... .......... .......... ..........
  4pileup{4(K6)}  .......... .......... .......... .......... ..........
  4pileup{4(M6)}  .......... .......... .......... .......... ..........
  4pileup{4(C1)}  .......... .....GTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(B2)}  .......... .......... .......... TGCTTGTGTA ACTATTGTGT
  4pileup{4(H4)}  .......... .......... .......... .......... ..........
  4pileup{4(K1)}  .......... .......... .......... .......... ..........
  4pileup{4(M4)}  .......... .......... .......... .......... ..........
  4pileup{4(D3)}  .......... .......... .......... .......... ..........
  4pileup{4(P1)}  .......... .......... .......... .......... ..........
  4pileup{4(T1)}  .......... .......... .GTTTTTAAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(H2)}  .......... .......... .......... .......... ..........
  4pileup{4(S1)}  .......... .......... .......... .......... ..........
  4pileup{4(D1)}  .......... .......... .GTTTTTAAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(K5)}  ATAATAAACA CGTGTGTATG TGTTTTTAAA TGCTTGTGTA ACTATTGTGT
  4pileup{4(K2)}  .......... .......... .......... .......... ..........
  4pileup{4(M5)}  .......... .......... .......... .......... ..........
       Consensus  ---------- ---------- ---------- ---------- ----------

                7310
  4pileup{4(M2)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(P2)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(G1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(M1)}  .......... AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(D2)}  GATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(H3)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(J1)}  .......... .......... .......... .......... ..........
4pileup{4(prot)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(S2)}  .......... .......... .......... .......... ......TGTG
  4pileup{4(K6)}  .......... .......... .......... .......... ..........
  4pileup{4(M6)}  .......... .......... .......... .......... ..........
  4pileup{4(C1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(B2)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(H4)}  .....AACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(K1)}  .......... .......... .......... .......... ..........
  4pileup{4(M4)}  .......... .......... .......... .......... ..........
  4pileup{4(D3)}  .......... .......... .......... CACCTACTAA TTGTGTTGTG
  4pileup{4(P1)}  .......... .......... .......... CACCTACTAA TTGTGTTGTG
  4pileup{4(T1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(H2)}  .....AACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(S1)}  .......... .......... .......... .......... ..........
  4pileup{4(D1)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(K5)}  CATGCAACAT AAATAAACTT ATTGTTTCAA CACCTACTAA TTGTGTTGTG
  4pileup{4(K2)}  .......... .......... .......... .......... ..........
  4pileup{4(M5)}  .......... .......... .......... .......... ..........
       Consensus  ---------- ---------- ---------- ---------- ----------
```

```
                  7360
  4pileup{4(M2)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(P2)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(G1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TGTGTTTTAT
  4pileup{4(M1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(D2)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(H3)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(J1)}  .......... .......AAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
4pileup{4(prot)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(S2)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(K6)}  ....TTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(M6)}  ....TTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(C1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(B2)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(H4)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(K1)}  ....TTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(M4)}  .......... ..ATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(D3)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(P1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(T1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(H2)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(S1)}  ....TTCATT GTATATAAAC TATATTTGCT ACATTCTGTT TTTGTTTTAT
  4pileup{4(D1)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATTCTGTT TTTGTTTTAT
  4pileup{4(K5)}  GTTATTCATT GTATATAAAC TATATTTGCT ACATCCTGTT TTTGTTTTAT
  4pileup{4(K2)}  .......... .......AAC TATATTTGCA ACATCCTGTT TTTGTTTTAT
  4pileup{4(M5)}  .......... .......AAC TATATTGGCT ACATCCTGTT TTTGTTTTAT
      Consensus   ---------- -------AAC TATATT-G-- ACAT-CTGTT T-TGTTTTAT

                  7410
  4pileup{4(M2)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(P2)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(G1)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(M1)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(D2)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(H3)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(J1)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
4pileup{4(prot)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(S2)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(K6)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(M6)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(C1)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(B2)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(H4)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(K1)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(M4)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(D3)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(P1)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(T1)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(H2)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(S1)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(D1)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(K5)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(K2)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
  4pileup{4(M5)}  ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
      Consensus   ATATACTATA TTTTGTAGCG CCAGCGGCCA TTTTGTAGCT TCAACCGAAT
```

```
                7460
4pileup{4(M2)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(P2)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(G1)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(M1)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(D2)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(H3)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(J1)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(prot)} TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(S2)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(K6)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(M6)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(C1)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(B2)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(H4)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(K1)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(M4)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(D3)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTCTG
4pileup{4(P1)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(T1)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(H2)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(S1)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(D1)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(K5)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(K2)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
4pileup{4(M5)}  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG
     Consensus  TCGGTTGCAT GCTTTTTGGC ACAAAATGTG TTTTTTTAAA TAGTTCTATG

                7510
4pileup{4(M2)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(P2)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(G1)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(M1)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(D2)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(H3)}  TCAGCAACTA TAGTTTCAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(J1)}  TCAGCAACTA TGGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(prot)} TCAGCAACTA TGGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(S2)}  TCAGCAACTA TGGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(K6)}  TCAGCAACTA TGGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(M6)}  TCAGCAACTA TGGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(C1)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(B2)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(H4)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(K1)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(M4)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(D3)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(P1)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(T1)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(H2)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(S1)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(D1)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(K5)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(K2)}  TCAGCAACTA TAGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
4pileup{4(M5)}  TCAGCAACTA TGGTTTAAAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
     Consensus  TCAGCAACTA T-GTTT-AAC TTGTACGTTT CCTGCTTGCC ATGCGTGCCA
```

```
                7560
  4pileup{4(M2)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(P2)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(G1)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(M1)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(D2)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(H3)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(J1)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
4pileup{4(prot)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(S2)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(K6)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(M6)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(C1)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(B2)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(H4)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(K1)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(M4)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(D3)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(P1)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(T1)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(H2)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(S1)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(D1)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(K5)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(K2)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
  4pileup{4(M5)}  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT
       Consensus  AATCCCTGTT TTCCTGACCT GCACTGCTTG CCAACCATTC CATTGTTTTT

                7610
  4pileup{4(M2)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTAC.. ..........
  4pileup{4(P2)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(G1)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTA... ..........
  4pileup{4(M1)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTAC.. ..........
  4pileup{4(D2)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTAC.. ..........
  4pileup{4(H3)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGAACG. ..........
4pileup{4(J1)}    TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
4pileup{4(prot)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(S2)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(K6)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(M6)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTAC.. ..........
  4pileup{4(C1)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(B2)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(H4)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(K1)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTAC.. ..........
  4pileup{4(M4)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTA... ..........
  4pileup{4(D3)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(P1)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(T1)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(H2)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(S1)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(D1)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(K5)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGCGTCATAT
  4pileup{4(K2)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
  4pileup{4(M5)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
       Consensus  TACACTGCAC TATGTGCAAC TACTGAATCA CTATG----- ----------
```

```
                7660
  4pileup{4(M2)}  .......... .......... ...
  4pileup{4(P2)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(G1)}  .......... .......... ...
  4pileup{4(M1)}  .......... .......... ...
  4pileup{4(D2)}  .......... .......... ...
  4pileup{4(H3)}  .......... .......... ...
  4pileup{4(J1)}  AAAATAAATC ACTATGCGCC AAC
4pileup{4(prot)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(S2)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(K6)}  AAAATAAATC ACTATGCGCC ...
  4pileup{4(M6)}  .......... .......... ...
  4pileup{4(C1)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(B2)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(H4)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(K1)}  .......... .......... ...
  4pileup{4(M4)}  .......... .......... ...
  4pileup{4(D3)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(P1)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(T1)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(H2)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(S1)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(D1)}  AAAATAAATC ACTATGCGC. ...
  4pileup{4(K5)}  AAAATAAATC ACTATGCGCC AAC
  4pileup{4(K2)}  AAAATAAATC ACTATGAGCC AAC
  4pileup{4(M5)}  AAAATAAATC ACTAAGCGCC AAC
       Consensus  ---------- ---------- ---
```

# NUCLEOTIDE ALIGNMENT FOR PRIMER URR-5 (nts 7666-7684)

```
                    7711
  7661pileup{7661(K4)}  TTTTGGCTTG TTTTAACTCA CCTAATTGCA TATTTGGCAT AAGGTTTAAA
7661pileup{7661(prot)}  TTTTGGCTTG TTTTAACTAA CCTAATTGCA TATTTGGCAT AAGGTTTAAA
             Consensus  TTTTGGCTTG TTTTAACT-A CCTAATTGCA TATTTGGCAT AAGGTTTAAA

                    7761
  7661pileup{7661(K4)}  CTTTTAAGGC CAACTAAATG TCACCTTAGT TCATACATGA ACTGTGTAAA
7661pileup{7661(prot)}  CTTCTAAGGC CAACTAAATG TCACCCTAGT TCATACATGA ACTGTGTAAA
             Consensus  CTT-TAAGGC CAACTAAATG TCACC-TAGT TCATACATGA ACTGTGTAAA

                    7811
  7661pileup{7661(K4)}  GGTTAGTCAT ACATTGTTCA TTTGTAAAAC TGCACATGGG TGTGTGCAAA
7661pileup{7661(prot)}  GGTTAGTCAT ACATTGTTCA TTTGTAAAAC TGCACATGGG TGTGTGCAAA
             Consensus  GGTTAGTCAT ACATTGTTCA TTTGTAAAAC TGCACATGGG TGTGTGCAAA

                    7861
  7661pileup{7661(K4)}  CCGTTTTGGG TTACACATTT ACAAGGAACT TATCTAATAA TACTAAACTA
7661pileup{7661(prot)}  CCGTTTTGGG TTACACATTT ACAAGCAACT TATATAATAA TACTAAACTA
             Consensus  CCGTTTTGGG TTACACATTT ACAAG-AACT TAT-TAATAA TACTAAACTA

                    5
  7661pileup{7661(K4)}  CAATAATTCA TGTATAAAAC TAAGGGCGTA ACCGAAATCG GTTGAACCGA
7661pileup{7661(prot)}  CAATAATTCA TGTATAAAAC TAAGGGCGTA ACCGAAATCG GTTGAACCGA
             Consensus  CAATAATTCA TGTATAAAAC TAAGGGCGTA ACCGAAATCG GTTGAACCGA

                    55
  7661pileup{7661(K4)}  AACCGGTTAG TATAAAAGCA GACATTTTAT GCACCAAAAG AGA
7661pileup{7661(prot)}  AACCGGTTAG TATAAAAGCA GACATTTTAT GCACCAAAAG AGA
             Consensus  AACCGGTTAG TATAAAAGCA GACATTTTAT GCACCAAAAG AGA
```

# NUCLEOTIDE ALIGNMENT FOR PRIMER URR-2 (nts 123-98)

```
                7610
2pileup{2(D1)}  .......... .......... ....GAATCA CTATGTACAT TGTGTCATAT
2pileup{2(P2)}  TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
2pileup{2(M1)}  .......... .......... ....GAATCA CTATGTACAT TGTGTCATAT
2pileup{2(D3)}  .......... .......... ........CA CTATGTACAT TGTGTCATAT
2pileup{2(K1)}  .......... .......... .......... ......ACAT TGTGTCATAT
2pileup{2(D2)}  .......... .......... .......... ......ACAT TGTGTCATAT
2pileup{2(M4)}  .......... .......... ......ATCA CTATGTACAT TGTGTCATAT
2pileup{2(prot)} TACACTGCAC TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
2pileup{2(T1)}  .......... .......... ....GAATCA CTATGTACAT TGTGTCATAT
2pileup{2(M6)}  .......... TATGTGCAAC TACTGAATCA CTATGTACAT TGTGTCATAT
2pileup{2(H3)}  .......... .......... .......... ......ACAT TGTGTCATAT
2pileup{2(G1)}  .......... .......... ....GAATCA CTATGTACAT TGTGTCATAT
2pileup{2(M2}   .......... .......... .......... .......CAT TGTGTCATAT
2pileup{2(S1)}  .......... .......... .....AATCA CTATGTACAT TGTGTCATAT
2pileup{2(K5}   .......... .......... .......TCA CTATGTACAT TGTGTCATAT
2pileup{2(K6}   .......... .......... .......... .......... ...GTCATAT
Consensus       ---------- ---------- ---------- ---------- ----------

                7660
2pileup{2(D1)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(P2)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(M1)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(D3)}  AAAAAAAATC ACTATCCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(K1)}  AAAATAAATC ACTATGCCGC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(D2)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(M4)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(prot)} AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(T1)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(M6)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(H3)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(G1)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(M2}   AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(S1)}  AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(K5}   AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
2pileup{2(K6}   AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA
Consensus       AAAATAAATC ACTATGCGCC AACGCCTTAC ATACCGCTGT TAGGCACATA

                7710
2pileup{2(D1)}  TTTTTGGCTT GTTTTAACTA CCCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(P2)}  TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(M1)}  TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(D3)}  TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(K1)}  TTTTGGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(D2)}  TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(M4)}  TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(prot)} TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(T1)}  TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TATGGTTTAA
2pileup{2(M6)}  TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(H3)}  TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(G1)}  TTTTTGGCTT GTTTTAACTA CCCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(M2}   TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(S1)}  TTTTTGGCTT GTTTTAACTA CCCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(K5}   TTTTTGGGTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
2pileup{2(K6}   TTTTTGGCTT GTTTTAACTA ACCTAATTGC ATATTTGGCA TAAGGTTTAA
Consensus       TTTT-GG-TT GTTTTAACTA -CCTAATTGC ATATTTGGCA TA-GGTTTAA
```

```
                    7760
  2pileup{2(D1)}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(P2)}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(M1)}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(D3)}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(K1)}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(D2)}    ACTTTTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(M4)}    ACTTTTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
2pileup{2(prot)}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(T1)}    ACTTTTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(M6)}    ACTTTTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(H3)}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(G1)}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
   2pileup{2(M2}    ACTTTTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
  2pileup{2(S1)}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
   2pileup{2(K5}    ACTTCTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
   2pileup{2(K6}    ACTTTTAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA
       Consensus    ACTT-TAAGG CCAACTAAAT GTCACCCTAG TTCATACATG AACTGTGTAA

                    7810
  2pileup{2(D1)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTACACATGG GTGTGTGCAA
  2pileup{2(P2)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(M1)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(D3)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(K1)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(D2)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(M4)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
2pileup{2(prot)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(T1)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(M6)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(H3)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(G1)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTACACATGG GTGTGTGCAA
   2pileup{2(M2}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
  2pileup{2(S1)}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTACACATGG GTGTGTGCAA
   2pileup{2(K5}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
   2pileup{2(K6}    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CTGCACATGG GTGTGTGCAA
       Consensus    AGGTTAGTCA TACATTGTTC ATTTGTAAAA CT-CACATGG GTGTGTGCAA

                    7860
  2pileup{2(D1)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(P2)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(M1)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(D3)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(K1)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(D2)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(M4)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
2pileup{2(prot)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(T1)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(M6)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(H3)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(G1)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
   2pileup{2(M2}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
  2pileup{2(S1)}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
   2pileup{2(K5}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
   2pileup{2(K6}    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
       Consensus    ACCGTTTTGG GTTACACATT TACAAGCAAC TTATATAATA ATACTAAACT
```

```
                4
2pileup{2(D1)}  ACAATAATTC ATGTATAAAA .......... .......... ..........
2pileup{2(P2)}  ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(M1)}  ACAATAATTC ATGTATAAAA CTAAG..... .......... ..........
2pileup{2(D3)}  ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(K1)}  ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(D2)}  ACAATAATCC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(M4)}  ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(prot)} ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(T1)}  ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(M6)}  ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(H3)}  ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(G1)}  ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(M2}   ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(S1)}  ACAATAATTC ATGTATAAAA TTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(K5}   ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
2pileup{2(K6}   ACAATAATTC ATGTATAAAA CTAAGGGCGT AACCGAAATC GGTTGAACCG
    Consensus   ACAATAAT-C ATGTATAAAA ---------- ---------- ----------
```

```
                54
2pileup{2(D1)}  .......... .......... .......... ......
2pileup{2(P2)}  AAAC...... .......... .......... ......
2pileup{2(M1)}  .......... .......... .......... ......
2pileup{2(D3)}  AAACCGGTTA GTATAAAAGC AGACATTTTA TGCA..
2pileup{2(K1)}  AAACCGGTTA GTATAAAAGC AGACATTTTA TGCACC
2pileup{2(D2)}  AAACCGGTTA GTATAA.... .......... ......
2pileup{2(M4)}  AAACCGGTTA GTATAAAAGC AGACATTTT. ......
2pileup{2(prot)} AAACCGGTTA GTATAAAAGC AGACATTTTA TGCACC
2pileup{2(T1)}  AAACCGGTTA GTATAAAAGC AGACATTTTA TGCACC
2pileup{2(M6)}  AAACCGGTTA GTATAAAAGC AGACATTTTA TGCACC
2pileup{2(H3)}  AAACCGGTTA GTATAAAAGT AGACATTTTA TGCACC
2pileup{2(G1)}  AAACCGGTTA GTATAAAAGC AAACATTTTA TGCACC
2pileup{2(M2}   AAACCGGTTA GTATAAAAGC AGACATTTTA TGCACC
2pileup{2(S1)}  AAACCGGTTA GTATAAAAGC AGACATTTTA TGCACC
2pileup{2(K5}   AAACCGGTTA GTATAAAAGC AGACATTTTA TGCACC
2pileup{2(K6}   AAACCGGTTA GTATAAAAGC AGACATTTTA TGCACC
    Consensus   ---------- ---------- ---------- ------
```

# NONCANCER ISOLATES

# NUCLEOTIDE ALIGNMENT FOR PRIMER URR-3 (nts 7425-7434)

```
                    7457
pileup{PRE(PROT)}   AATTCGGTTG CATGCTTTTT GGCACAAAAT GTGTTTTTTT AAATAGTTCT
   pileup{PRE262}   AATTCGGTTG CATGCTTTTT GGCACAAAAT GTGTTTTTTT AAATAGTTCT
   pileup{PRE250}   AATTCGGTTG CATGCTTTTT GGCACAAAAT GTGTTTTTTT AAATAGTTCT
   pileup{PRE259}   AATTCGGTTG CATGCTTTTT GGCACAAAAT GTGTTTTTTT AAATAGTTCT
        Consensus   AATTCGGTTG CATGCTTTTT GGCACAAAAT GTGTTTTTTT AAATAGTTCT

                    7507
pileup{PRE(PROT)}   ATGTCAGCAA CTATGGTTTA AACTTGTACG TTTCCTGCTT GCCATGCGTG
   pileup{PRE262}   ATGTCAGCAA CTATAGTTTA AACTTGTACG TTTCCTGCTT GCCATGCGTG
   pileup{PRE250}   ATGTCAGCAA CTATAGTTTA AACTTGTACG TTTCCTGCTT GCCATGCGTG
   pileup{PRE259}   ATGTCAGCAA CTATAGTTTA AACTTGTACG TTTCCTGCTT GCCATGCGTG
        Consensus   ATGTCAGCAA CTAT-GTTTA AACTTGTACG TTTCCTGCTT GCCATGCGTG

                    7557
pileup{PRE(PROT)}   CCAAATCCCT GTTTTCCTGA CCTGCACTGC TTGCCAACCA TTCCATTGTT
   pileup{PRE262}   CCAAATCCCT GTTTTCCTGA CCTGCACTGC TTGCCAACCA TTCCATTGTT
   pileup{PRE250}   CCAAATCCCT GTTTTCCTGA CCTGCACTGC TTGCCAACCA TTCCATTGTT
   pileup{PRE259}   CCAAATCCCT GTTTTCCTGA CCTGCACTGC TTGCCAACCA TTCCATTGTT
        Consensus   CCAAATCCCT GTTTTCCTGA CCTGCACTGC TTGCCAACCA TTCCATTGTT

                    7607
pileup{PRE(PROT)}   TTTTACACTG CACTATGTGC AACTACTGAA TCACTATGTA CATTGTGTCA
   pileup{PRE262}   TTTTACACTG CACTATGTGC AACTACTGAA TCACTATGTA CATTGTGTCA
   pileup{PRE250}   TTTTACACTG CACTATGTGC AACTACTGAA TCACTATGTA CATTGTGTCA
   pileup{PRE259}   TTTTACACTG CACTATGTGC AACTACTGAA TCACTATGTA CATTGTGTCA
        Consensus   TTTTACACTG CACTATGTGC AACTACTGAA TCACTATGTA CATTGTGTCA

                    7657
pileup{PRE(PROT)}   TATAAAATAA ATCACTATGC GCCAACGCCT TACATACCGC TGTTAGGCAC
   pileup{PRE262}   TATAAAATAA ATCACTATGC GCCAACGCCT TACATACCGC TGTTAGGCAC
   pileup{PRE250}   TATAAAATAA ATCACTATGC GCCAACGCCT TACATACCGC TGTTAGGCAC
   pileup{PRE259}   TATAAAATAA ATCACTATGC GCCAACGCCT TACATACCGC TGTTAGGCAC
        Consensus   TATAAAATAA ATCACTATGC GCCAACGCCT TACATACCGC TGTTAGGCAC

                    7707
pileup{PRE(PROT)}   ATATTTTTGG CTTGTTTTAA CTAACCTAAT TGCATATTTG GCATAAGGTT
   pileup{PRE262}   ATATTTTTGG CTTGTTTTAA CTAACCTAAT TGCATATTTG GCATAAGGTT
   pileup{PRE250}   ATATTTTTGG CTTGTTTTAA CTAACCTAAT TGCATATTTG GCATAAGGTT
   pileup{PRE259}   ATATTTTTGG CTTGTTTTAA CTACCCTAAT TGCATATTTG GCATAAGGTT
        Consensus   ATATTTTTGG CTTGTTTTAA CTA-CCTAAT TGCATATTTG GCATAAGGTT

                    7757
pileup{PRE(PROT)}   TAAACTTCTA AGGCCAACTA AATGTCACCC TAGTTCATAC ATGAACTGTG
   pileup{PRE262}   TAAACTTCTA AGGCCAACTA AATGTCACCC TAGTTCATAC ATGAACTGTG
   pileup{PRE250}   TAAACTTCTA AGGCCAACTA AATGTCACCC TAGTTCATAC ATGAACTGTG
   pileup{PRE259}   TAAACTTCTA AGGCCAACTA AATGTCACCC TAGTTCATAC ATGAACTGTG
        Consensus   TAAACTTCTA AGGCCAACTA AATGTCACCC TAGTTCATAC ATGAACTGTG

                    7807
pileup{PRE(PROT)}   TAAA
   pileup{PRE262}   TAAA
   pileup{PRE250}   TAAA
   pileup{PRE259}   TAAA
        Consensus   TAAA
```

## A7.6.2 E2/E4 GENES AND PROTEINS

The E2 gene was sequenced for 13 selected cancer isolates and for two of the three noncancer isolates. The entire E2 sequence is presented for the cancer isolates whilst the first 293bp of the noncancer E2 gene sequences was not sequenced. The corresponding amino acid alignments for the E2 and E4 genes follow the nucleotide alignments.

## CANCER ISOLATES

## NUCLEOTIDE ALIGNMENT FOR PRIMER E2-1 (nts 2735-2753)

```
                2764
  E21pileup{E21C1}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21J1}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21H2}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21K1}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
E21pileup{E21prot}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21T1}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21K6}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21G1}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21H3}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21M6}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21S1}  .........A CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21M2}  TCTTTGCCAA CGTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21K2}  .......... ..TTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
  E21pileup{E21K4}  .........A CCTTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG
         Consensus  ---------- --TTTAAATG TGTGTCAGGA CAAAATACTA ACACATTATG

                2814
  E21pileup{E21C1}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21J1}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21H2}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21K1}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
E21pileup{E21prot}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21T1}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21K6}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21G1}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21H3}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21M6}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21S1}  AAAATGATAG TACAAACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21M2}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21K2}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACACATG
  E21pileup{E21K4}  AAAATGATAG TACAGACCTA CGTGACCATA TAGACTATTG GAAACAAATG
         Consensus  AAAATGATAG TACA-ACCTA CGTGACCATA TAGACTATTG GAAACA-ATG

                2864
  E21pileup{E21C1}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21J1}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21H2}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21K1}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
E21pileup{E21prot}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21T1}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21K6}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21G1}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21H3}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21M6}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21S1}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21M2}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21K2}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
  E21pileup{E21K4}  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
         Consensus  CGCCTAGAAT GTGCTATTTA TTACAAGGCC AGAGAAATGG GATTTAAACA
```

```
                    2914
  E21pileup{E21C1}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21J1}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21H2}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21K1}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
E21pileup{E21prot}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21T1}  TATTAACCAC CAGGTGGTGC CAACGCTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21K6}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21G1}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21H3}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21M6}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21S1}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21M2}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21K2}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
  E21pileup{E21K4}  TATTAACCAC CAGGTGGTGC CAACACTGGC TGTATCAAAG AATAAAGCAT
        Consensus   TATTAACCAC CAGGTGGTGC CAAC-CTGGC TGTATCAAAG AATAAAGCAT

                    2964
  E21pileup{E21C1}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21J1}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21H2}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21K1}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
E21pileup{E21prot}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21T1}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21K6}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21G1}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21H3}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21M6}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21S1}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21M2}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21K2}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
  E21pileup{E21K4}  TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA
        Consensus   TACAAGCAAT TGAACTGCAA CTAACGTTAG AAACAATATA TAACTCACAA

                    3014
  E21pileup{E21C1}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21J1}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21H2}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21K1}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
E21pileup{E21prot}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21T1}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21K6}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21G1}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21H3}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21M6}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21S1}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21M2}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21K2}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
  E21pileup{E21K4}  TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT
        Consensus   TATAGTAATG AAAAGTGGAC ATTACAAGAC GTTAGCCTTG AAGTGTATTT

                    3064
  E21pileup{E21C1}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21J1}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21H2}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21K1}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
E21pileup{E21prot}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21T1}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21K6}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21G1}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21H3}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21M6}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21S1}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21M2}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21K2}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
  E21pileup{E21K4}  AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
        Consensus   AACTGCACCA ACAGGATGTA TAAAAAAACA TGGATATACA GTGGAAGTGC
```

```
                    3114
  E21pileup{E21C1}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21J1}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21H2}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21K1}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
E21pileup{E21prot}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21T1}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21K6}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21G1}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGAAACAT
  E21pileup{E21H3}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21M6}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21S1}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGAAACAT
  E21pileup{E21M2}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21K2}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGACACAT
  E21pileup{E21K4}  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGAAACAT
         Consensus  AGTTTGATGG AGACATATGC AATACAATGC ATTATACAAA CTGGA-ACAT

                    3164
  E21pileup{E21C1}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21J1}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21H2}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21K1}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
E21pileup{E21prot}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21T1}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21K6}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GACAAGTTGA
  E21pileup{E21G1}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21H3}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21M6}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GACAAGTTGA
  E21pileup{E21S1}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21M2}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21K2}  ATATATATTT GTGAAGAAGC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
  E21pileup{E21K4}  ATATATATTT GTGAAGAAAC ATCAGTAACT GTGGTAGAGG GTCAAGTTGA
         Consensus  ATATATATTT GTGAAGAA-C ATCAGTAACT GTGGTAGAGG G-CAAGTTGA

                    3214
  E21pileup{E21C1}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
  E21pileup{E21J1}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTT.....
  E21pileup{E21H2}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
  E21pileup{E21K1}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
E21pileup{E21prot}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
  E21pileup{E21T1}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
  E21pileup{E21K6}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
  E21pileup{E21G1}  CTATTATGGT TTATATTATG TTCATGAAGG AATACAAACA TATTTTGTGC
  E21pileup{E21H3}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
  E21pileup{E21M6}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
  E21pileup{E21S1}  CTATTATGGT TTATATTATG TTCATGAAGG AATACAAACA TATTTTGTGC
  E21pileup{E21M2}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
  E21pileup{E21K2}  CTATTATGGT TTATATTATG TTCATGAAGG AATACGAACA TATTTTGTGC
  E21pileup{E21K4}  CTATTATGGT ATATATTATG TTCATGAAGG AATACAAACA TATTTTGTGC
         Consensus  CTATTATGGT -TATATTATG TTCATGAAGG AATAC-AACA TATTT-----

                    3264
  E21pileup{E21C1}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21J1}  .......... .......... .......... .......... ..........
  E21pileup{E21H2}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21K1}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
E21pileup{E21prot}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21T1}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21K6}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21G1}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21H3}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21M6}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21S1}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21M2}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21K2}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
  E21pileup{E21K4}  AGTTTAAAGA TGATGCAGAA AAATATAGTA AAAATAAAGT ATGGGAAGTT
         Consensus  ---------- ---------- ---------- ---------- ----------
```

```
                    3314
  E21pileup{E21C1}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21J1}  .......... .......... .......... .......... ..........
  E21pileup{E21H2}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21K1}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
E21pileup{E21prot}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21T1}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21K6}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21G1}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT T.........
  E21pileup{E21H3}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21M6}  CATGCGGGTG GTTAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21S1}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21M2}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21K2}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCAA
  E21pileup{E21K4}  CATGCGGGTG GTCAGGTAAT ATTATGTCCT ACATCTGTGT TTAGCAGCGA
        Consensus   ---------- ---------- ---------- ---------- ----------

                    3364
  E21pileup{E21C1}  CGAAGTATCC TCTCCTGAAA TTATTAGGCA GCACTTGGCC AACCACTCCG
  E21pileup{E21J1}  .......... .......... .......... .......... ..........
  E21pileup{E21H2}  CGAAGTATCC TCTCCTGAAA TTATTAGGCA GCACTTGGCC AACCACTCCG
  E21pileup{E21K1}  CGAAGTATCC TCTCCTGAAA TTATTAGGCA GCACTTGGCC AACCACTCCG
E21pileup{E21prot}  CGAAGTATCC TCTCCTGAAA TTATTAGGCA GCACTTGGCC AACCACCCCG
  E21pileup{E21T1}  CGAAGTATCC TCTCCTGAAA CTATTAGGCA GCACTTGGCC AACCACTCCG
  E21pileup{E21K6}  CGAAGTAT.. .......... .......... .......... ..........
  E21pileup{E21G1}  .......... .......... .......... .......... ..........
  E21pileup{E21H3}  CGAAGTATCC TCTCCTGAAA TTATTAGGCA GCACTTGGCC AACCACTCCG
  E21pileup{E21M6}  CGAAGTAT.. .......... .......... .......... ..........
  E21pileup{E21S1}  CGAAGTATCC TCTCCTGAAA CTATTAGGCA GCACTTGGCC AACCACTCCG
  E21pileup{E21M2}  CGAAGTATCC TCTCCTGAAA TTATTAGGCA GCACTTGGCC AACCACTCCG
  E21pileup{E21K2}  CGAAGTATCC TCTCCTGAAA TTATTAGGCA GCACTTGGCC AACCACTCCG
  E21pileup{E21K4}  CGAAGTATCC TCTGCTGAAA TTATTAGGCA GCACTTGGCC AACCACTCCG
        Consensus   ---------- ---------- ---------- ---------- ----------

                    3414
  E21pileup{E21C1}  CCGCGACCCA TACCAAA... ..........
  E21pileup{E21J1}  .......... .......... ..........
  E21pileup{E21H2}  CCGCGACCCA TACCAAA... ..........
  E21pileup{E21K1}  CCGCGACCCA TACCAAA... ..........
E21pileup{E21prot}  CCGCGACCCA TACCAAAGCC GTCGCCTTGG
  E21pileup{E21T1}  CCGCGACCCA TACCAAA... ..........
  E21pileup{E21K6}  .......... .......... ..........
  E21pileup{E21G1}  .......... .......... ..........
  E21pileup{E21H3}  CCGCGACCTA TACCAAAGCC ..........
  E21pileup{E21M6}  .......... .......... ..........
  E21pileup{E21S1}  CCGCGACCCA TACCAAA... ..........
  E21pileup{E21M2}  CCGCGACCCA TACCAAAGCC GTCGCCTTGG
  E21pileup{E21K2}  CCGCGACCCA TACCAAAGCC GTCGCCTTGG
  E21pileup{E21K4}  CCGCGACCCA TACCAAAGCC GTCGCCTTGG
        Consensus   ---------- ---------- ----------
```

# AMINO ACID ALIGNMENT FOR PRIMER E2-1

```
                  4
                  .         .         .         .         .         .
pepE21C1          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21G1          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21H2          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21H3          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21J1          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21K1          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21K2          ----LNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21K4          ---ALNVCQDKILTHYENDSTDLRDHIDYWKQMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21K6          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21M2          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21M6          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21S1          ---RLNVCQDKILTHYENDSTNLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21T1          LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
pepE21prot        LCQRLNVCQDKILTHYENDSTDLRDHIDYWKHMRLECAIYYKAREMGFKHINHQVVPTLA
                  ***************** ********* ****************************
```

```
                64
pepE21C1        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21G1        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21H2        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21H3        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21J1        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21K1        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21K2        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21K4        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21K6        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21M2        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21M6        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21S1        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21T1        VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
pepE21prot      VSKNKALQAIELQLTLETIYNSQYSNEKWTLQDVSLEVYLTAPTGCIKKHGYTVEVQFDG
                ************************************************************

                124
pepE21C1        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepE21G1        DICNTMHYTNWKHIYICEEASVTVVEGQVDYYGLYYVHEGIQTYFVQFKDDAEKYSKNKV
pepE21H2        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepE21H3        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepE21J1        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTY----------------
pepE21K1        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepE21K2        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepE21K4        DICNTMHYTNWTHIYICEETSVTVVEGQVDYYGIYYVHEGIQTYFVQFKDDAEKYSKNKV
pepE21K6        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepE21M2        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepE21M6        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepE21S1        DICNTMHYTNWKHIYICEEASVTVVEGQVDYYGLYYVHEGIQTYFVQFKDDAEKYSKNKV
pepE21T1        DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepE21prot      DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
                *********** ******* ************* *******  *

                184
pepE21C1        WEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTK----
pepE21G1        WEVHAGGQVILCPTSV------------------------------
pepE21H2        WEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTK----
pepE21H3        WEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATYTKA---
pepE21J1        ---------------------------------------------
pepE21K1        WEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTK----
pepE21K2        WEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHSAATHTKAVAL
pepE21K4        WEVHAGGQVILCPTSVFSSDEVSSAEIIRQHLANHSAATHTKAVAL
pepE21K6        WEVHAGGQVILCPTSVFSSNEV-----------------------
pepE21M2        WEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHSAATHTKAVAL
pepE21M6        WEVHAGG-VILCPTSVFSSNEV-----------------------
pepE21S1        WEVHAGGQVILCPTSVFSSNEVSSPETIRQHLANHPAANHTK----
pepE21T1        WEVHAGGQVILCPTSVFSSNEVSSPETIRQHLANHSAATHTK----
pepE21prot      WEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTKAVAL
```

# NUCLEOTIDE ALIGNMENT FOR PRIMER E2-2 (nts 3873-3854)

```
                    3122
  E22pileup{E22J1}  .......... .......... .......... .......... ..........
  E22pileup{E22K6}  .......... .......... .......... .......... ..........
  E22pileup{E22M6}  .......... .......... .......... .......... ..........
  E22pileup{E22K2}  ....ACATAT GCAATACAAT GCATTATACA AACTGGACAC ATATATATAT
  E22pileup{E22M2}  GGAGACATAT GCAATACAAT GCATTATACA AACTGGACAC ATATATATAT
  E22pileup{E22C1}  .......... .......... .......... .......... ..........
  E22pileup{E22K1}  .......... .......... .......... .......... ..........
E22pileup{E22prot}  GGAGACATAT GCAATACAAT GCATTATACA AACTGGACAC ATATATATAT
  E22pileup{E22H3}  .......... .......... .......... ........AC ATATATATAT
  E22pileup{E22H2}  .......... .......... .......... .......... ..........
  E22pileup{E22T1}  .......... .......... .......... .......... ..........
  E22pileup{E22G1}  .......... .......... .......... .......... ..........
  E22pileup{E22S1}  .......... .......... .......... .......... .........T
  E22pileup{E22K4}  ....ACATAT GCAATACAAT GCATTATACA AACTGGAAAC ATATATATAT
        Consensus   ---------- ---------- ---------- ---------- ----------


                    3172
  E22pileup{E22J1}  .......... .......... .......... .......... .........G
  E22pileup{E22K6}  .......... .......... .......... .......... .........G
  E22pileup{E22M6}  .......... .......... .......... .......... .........G
  E22pileup{E22K2}  TTGTGAAGAA GCATCAGTAA CTGTGGTAGA GGGTCAAGTT GACTATTATG
  E22pileup{E22M2}  TTGTGAAGAA GCATCAGTAA CTGTGGTAGA GGGTCAAGTT GACTATTATG
  E22pileup{E22C1}  .......... .......... .......... ....CAAGTT GACTATTATG
  E22pileup{E22K1}  .......... .......... .......... ....CAAGTT GACTATTATG
E22pileup{E22prot}  TTGTGAAGAA GCATCAGTAA CTGTGGTAGA GGGTCAAGTT GACTATTATG
  E22pileup{E22H3}  TTGTGAAGAA GCATCAGTAA CTGTGGTAGA GGGTCAAGTT GACTATTATG
  E22pileup{E22H2}  .......... .......... .......... .......... ..........
  E22pileup{E22T1}  .......... .......... .......... .......... .........G
  E22pileup{E22G1}  .......... .......... .......... .......... .........G
  E22pileup{E22S1}  TTGTGAAGAA GCATCAGTAA CTGTGGTAGA GGGTCAAGTT GACTATTATG
  E22pileup{E22K4}  TTGTGAAGAA ACATCAGTAA CTGTGGTAGA GGGTCAAGTT GACTATTATG
        Consensus   ---------- ---------- ---------- ---------- ----------


                    3222
  E22pileup{E22J1}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTTT GCAGTTTAAA
  E22pileup{E22K6}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22M6}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22K2}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22M2}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22C1}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22K1}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTGT GCAGTTTAAA
E22pileup{E22prot}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22H3}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22H2}  .......... .......... .......... .......... ..........
  E22pileup{E22T1}  GTTTATATTA TGTTCATGAA GGAATACGAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22G1}  GTTTATATTA TGTTCATGAA GGAATACAAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22S1}  GTTTATATTA TGTTCATGAA GGAATACAAA CATATTTTGT GCAGTTTAAA
  E22pileup{E22K4}  GTATATATTA TGTTCATGAA GGAATACAAA CATATTTTGT GCAGTTTAAA
        Consensus   ---------- ---------- ---------- ---------- ----------


                    3272
  E22pileup{E22J1}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22K6}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22M6}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22K2}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22M2}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22C1}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22K1}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
E22pileup{E22prot}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22H3}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22H2}  .......... .......... .......AAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22T1}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22G1}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22S1}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
  E22pileup{E22K4}  GATGATGCAG AAAAATATAG TAAAAATAAA GTATGGGAAG TTCATGCGGG
        Consensus   ---------- ---------- -------AAA GTATGGGAAG TTCATGCGGG
```

```
                    3322
  E22pileup{E22J1}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22K6}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22M6}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22K2}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22M2}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22C1}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22K1}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
E22pileup{E22prot}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22H3}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22H2}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22T1}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22G1}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22S1}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC AACGAAGTAT
  E22pileup{E22K4}  TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC GACGAAGTAT
          Consensus TGGTCAGGTA ATATTATGTC CTACATCTGT GTTTAGCAGC -ACGAAGTAT

                    3372
  E22pileup{E22J1}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACCC CGCCGCGACC
  E22pileup{E22K6}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACCC CGCCGCGACC
  E22pileup{E22M6}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACCC CGCCGCGACC
  E22pileup{E22K2}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACTC CGCCGCGACC
  E22pileup{E22M2}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACTC CGCCGCGACC
  E22pileup{E22C1}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACTC CGCCGCGACC
  E22pileup{E22K1}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACTC CGCCGCGACC
E22pileup{E22prot}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACCC CGCCGCGACC
  E22pileup{E22H3}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACTC CGCCGCGACC
  E22pileup{E22H2}  CCTCTCCTGA AATTATTAGG CAGCACTTGG CCAACCACTC CGCCGCGACC
  E22pileup{E22T1}  CCTCTCCTGA AACTATTAGG CAGCACTTGG CCAACCACTC CGCCGCGACC
  E22pileup{E22G1}  CCTCTCCTGA AACTATTAGg CAGCACTTGG CCAACCACTC CGCCGCGACC
  E22pileup{E22S1}  CCTCTCCTGA AACTATTAGG CAGCACTTGG CCAACCACTC CGCCGCGACC
  E22pileup{E22K4}  CCTCTGCTGA AATTATTAGG CAGCACTTGG CCAACCACTC CGCCGCGACC
          Consensus CCTCT-CTGA AA-TATTAGG CAGCACTTGG CCAACCAC-C CGCCGCGACC

                    3422
  E22pileup{E22J1}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22K6}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22M6}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22K2}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22M2}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22C1}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22K1}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
E22pileup{E22prot}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22H3}  TATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22H2}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22T1}  CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAGA CGACTATCCA
  E22pileup{E22G1}  CATACCAAAG CCGTCGCCTT GGGCACCAAA GAAACACAGA CGACTATCCA
  E22pileup{E22S1}  CATACCAAAG CCGTCGCCTT GGGCACCAAA GAAACACAGA CGACTATCCA
  E22pileup{E22K4}  CATACCAAAG CCGTCGCCTT GGGCACCAAA GAAACACAGA CGACTATCCA
          Consensus -ATACCAAAG CCGTCGCCTT GGGCACC-AA GAAACACAGA CGACTATCCA

                    3472
  E22pileup{E22J1}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22K6}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22M6}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22K2}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22M2}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22C1}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22K1}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
E22pileup{E22prot}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22H3}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22H2}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22T1}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22G1}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22S1}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCACTAAGT
  E22pileup{E22K4}  GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCAACAAGT
          Consensus GCGACCAAGA TCAGAGCCAG ACACCGGAAA CCCCTGCCAC ACCA--AAGT
```

```
                    3522
  E22pileup{E22J1}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22K6}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22M6}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22K2}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22M2}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22C1}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22K1}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
E22pileup{E22prot}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22H3}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22H2}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22T1}  TGTTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22G1}  TGCTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATGTAAC
  E22pileup{E22S1}  TGCTGCACAG AGACTCAGTG GACAGTGCTC CAATCCTCAC TGCATTTAAC
  E22pileup{E22K4}  TGTTGCACAG AGACTCCGTG GACAGTGCTC CAATCCTCAC TGCAGTTAAC
        Consensus   TG-TGCACAG AGACTC-GTG GACAGTGCTC CAATCCTCAC TGCA--TAAC

                    3572
  E22pileup{E22J1}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22K6}  AGCGCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22M6}  AGCGCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22K2}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22M2}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22C1}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22K1}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
E22pileup{E22prot}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22H3}  AGCTCACACA AAGGACGGAT TAACTGTAAT AATAACACTA CACCCATAGT
  E22pileup{E22H2}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22T1}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22G1}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22S1}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
  E22pileup{E22K4}  AGCTCACACA AAGGACGGAT TAACTGTAAT AGTAACACTA CACCCATAGT
        Consensus   AGC-CACACA AAGGACGGAT TAACTGTAAT A-TAACACTA CACCCATAGT

                    3622
  E22pileup{E22J1}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22K6}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22M6}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TACAGATTTA
  E22pileup{E22K2}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22M2}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22C1}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22K1}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
E22pileup{E22prot}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22H3}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22H2}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22T1}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22G1}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22S1}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
  E22pileup{E22K4}  ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TATAGATTTA
        Consensus   ACATTTAAAA GGTGATGCTA ATACTTTAAA ATGTTTAAGA TA-AGATTTA

                    3672
  E22pileup{E22J1}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22K6}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22M6}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22K2}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22M2}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22C1}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22K1}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
E22pileup{E22prot}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22H3}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22H2}  AAAAGCATTG TACATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22T1}  AAAAGCATTG TAAATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22G1}  AAAAGCATTG TAAATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22S1}  AAAAGCATTG TAAATTGTAT ACTGCAGTGT CGTCTACATG GCATTGGACA
  E22pileup{E22K4}  AAAAGCATTG TAAATTGTAT ACAGCAGTGT CGTCCACATG GCATTGGACA
        Consensus   AAAAGCATTG TA-ATTGTAT AC-GCAGTGT CGTC-ACATG GCATTGGACA
```

```
                    3722
  E22pileup{E22J1}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22K6}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22M6}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22K2}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22M2}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22C1}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22K1}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
E22pileup{E22prot}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22H3}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22H2}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22T1}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22G1}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22S1}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
  E22pileup{E22K4}  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG
         Consensus  GGACATAATG TAAAACATAA AAGTGCAATT GTTACACTTA CATATGATAG

                    3772
  E22pileup{E22J1}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22K6}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22M6}  TGAATGGCAA CGTGACCAAT TTTTGTTTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22K2}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22M2}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22C1}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22K1}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
E22pileup{E22prot}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22H3}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22H2}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22T1}  TGAATGGCAA CGTGACCAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22G1}  TGAATGGCAA CGTGAACAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22S1}  TGAATGGCAA CGTGAACAAT TTTTGTCTCA AGTTAAAATA CCAAAAACTA
  E22pileup{E22K4}  TGAATGTCAA CGTGAACAAT TTTTGTCTCA AGTGAAAATA CCAAAAACTA
         Consensus  TGAATG-CAA CGTGA-CAAT TTTTGT-TCA AGT-AAAATA CCAAAAACTA

                    3822
  E22pileup{E22J1}  TTACAGTGTC TACTGGATTT ATGTCTAT
  E22pileup{E22K6}  TTACAGTGTC TACTGGATTT ATGTCTAT
  E22pileup{E22M6}  TTACAGTGTC TACTGGATTT ATGTCTAT
  E22pileup{E22K2}  TTACAGTGTC TACTGGATTT A.......
  E22pileup{E22M2}  TTACAGTGTC TACTGGATTT A.......
  E22pileup{E22C1}  TTACAGTGTC TACTGGATTT A.......
  E22pileup{E22K1}  TTACAGTGTC TACTGGATTT A.......
E22pileup{E22prot}  TTACAGTGTC TACTGGATTT ATGTCTAT
  E22pileup{E22H3}  TTACAGTGTC TACTGGATTT ATGTCTA.
  E22pileup{E22H2}  TTACAGTGTC .......... ........
  E22pileup{E22T1}  TTACAGTGTC TACTGGATTT ATGTCTAT
  E22pileup{E22G1}  TTACAGTGTC TACTGGATTT ATGTCTAT
  E22pileup{E22S1}  TTACAGTGTC TACTGGATTT ........
  E22pileup{E22K4}  TTACAGTGTC TACTGGATTT A.......
         Consensus  TTACAGTGTC ---------- --------
```

# AMINO ACID ALIGNMENT FOR PRIMER E2-2

```
                  123
                        .         .         .         .         .         .
pepE22C1          ----------------------------QVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNK
pepE22G1          -------------------------------GLYYVHEGIQTYFVQFKDDAEKYSKNK
pepE22H2          ------------------------------------------------------------K
pepE22H3          -------------HIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNK
pepE22J1          --------------------------------GLYYVHEGIRTYFLQFKDDAEKYSKNK
pepE22K1          ----------------------------QVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNK
pepE22K2          --ICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNK
pepE22K4          --ICNTMHYTNWKHIYICEETSVTVVEGQVDYYGIYYVHEGIQTYFVQFKDDAEKYSKNK
pepE22K6          -------------------------------GLYYVHEGIRTYFVQFKDDAEKYSKNK
pepE22M2          GDICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNK
pepE22M6          --------------------------------GLYYVHEGIRTYFVQFKDDAEKYSKNK
pepE22S1          -----------------CEEASVTVVEGQVDYYGLYYVHEGIQTYFVQFKDDAEKYSKNK
pepE22T1          --------------------------------GLYYVHEGIRTYFVQFKDDAEKYSKNK
pepE22prot        GDICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNK
                                                                             *
```

```
                183
pepE22C1        VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHSAATHTKAVALGTEETQTTIQRPR
pepE22G1        VWEVHAGGQVILCPTSVFSSNEVSSPETIRQHLANHSAATHTKAVALGTKETQTTIQRPR
pepE22H2        VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHSAATHTKAVALGTEETQTTIQRPR
pepE22H3        VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHSAATYTKAVALGTEETQTTIQRPR
pepE22J1        VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTKAVALGTEETQTTIQRPR
pepE22K1        VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHSAATHTKAVALGTEETQTTIQRPR
pepE22K2        VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHSAATHTKAVALGTEETQTTIQRPR
pepE22K4        VWEVHAGGQVILCPTSVFSSDEVSSAEIIRQHLANHSAATHTKAVALGTKETQTTIQRPR
pepE22K6        VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTKAVALGTEETQTTIQRPR
pepE22M2        VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHSAATHTKAVALGTEETQTTIQRPR
pepE22M6        VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTKAVALGTEETQTTIQRPR
pepE22S1        VWEVHAGGQVILCPTSVFSSNEVSSPETIRQHLANHSAATHTKAVALGTKETQTTIQRPR
pepE22T1        VWEVHAGGQVILCPTSVFSSNEVSSPETIRQHLANHSAATHTKAVALGTEETQTTIQRPR
pepE22prot      VWEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTKAVALGTEETQTTIQRPR
                ********************* **** * ******** *** ******** **********

                243
pepE22C1        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22G1        SEPDTGNPCHTTKLLHRDSVDSAPILTACNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22H2        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22H3        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNNNTTPIVHLKGDANTLKCLR
pepE22J1        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22K1        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22K2        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22K4        SEPDTGNPCHTNKLLHRDSVDSAPILTAVNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22K6        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSAHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22M2        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22M6        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSAHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22S1        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22T1        SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
pepE22prot      SEPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLR
                *********** ********** ** ******** ******************

                303
pepE22C1        YRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGF
pepE22G1        YRFKKHCKLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQREQFLSQVKIPKTITVSTGF
pepE22H2        YRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITV----
pepE22H3        YRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGF
pepE22J1        YRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGF
pepE22K1        YRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGF
pepE22K2        YRFKKHCTLYTAVSSTWHWTGHNVKHKSAXVTLTYDSEWQRDQFLSQVKIPKTITVSTGF
pepE22K4        YRFKKHCKLYTAVSSTWHWTGHNVKHKSAIVTLTYDSECQREQFLSQVKIPKTITVSTGF
pepE22K6        YRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGF
pepE22M2        YRFKKHCTLYTAVSSTWHWTGHNVKHKSXIVTLTYDSEWQRDQFLSQVKIPKTITVSTGF
pepE22M6        YRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLFQVKIPKTITVSTGF
pepE22S1        YRFKKHCKLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQREQFLSQVKIPKTITVSTDL
pepE22T1        YRFKKHCKLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTXF
pepE22prot      YRFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGF
                ******* ******************************** ** *** **********
```

# AMINO ACID ALIGNMENT FOR E4 (PRIMER E2-2)

```
                1       .       .       .       .       .
pepE4C1         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4G1         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4H2         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4H3         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4J1         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4K1         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4K2         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4K4         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPTSCCTET
pepE4K6         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4M2         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4M6         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4S1         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4T1         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
pepE4prot       AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
                ************************************************** ******

                61
pepE4C1         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4G1         QWTVLQSSLHVTAHTKDGLTVIVTLHP
pepE4H2         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4H3         QWTVLQSSLHLTAHTKDGLTVIITLHP
pepE4J1         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4K1         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4K2         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4K4         PWTVLQSSLQLTAHTKDGLTVIVTLHP
pepE4K6         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4M2         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4M6         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4S1         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4T1         QWTVLQSSLHLTAHTKDGLTVIVTLHP
pepE4prot       QWTVLQSSLHLTAHTKDGLTVIVTLHP
                 ********  *********** ****
```

# NONCANCER ISOLATES

## NUCLEOTIDE ALIGNMENT FOR PRIMER E2-2 (nts 3873-3854)

```
                    3125
  PREpileup{PRE250}  GACATATGCA ATACAATGCA TTATACAAAC TGGACACATA TATATATTTG
**PREpileup{PREPROTO}** **GACATATGCA ATACAATGCA TTATACAAAC TGGACACATA TATATATTTG**
  PREpileup{PRE262}  .......... .......... .......... ......CATA TATATATTTG
          Consensus  ---------- ---------- ---------- ------CATA TATATATTTG

                    3175
  PREpileup{PRE250}  TGAAGAAGCA TCAGTAACTG TGGTAGAGGG TCAAGTTGAC TATTATGGTT
**PREpileup{PREPROTO}** **TGAAGAAGCA TCAGTAACTG TGGTAGAGGG TCAAGTTGAC TATTATGGTT**
  PREpileup{PRE262}  TGAAGAAGCA TCAGTAACTG TGGTAGAGGG TCAAGTTGAC TATTATGGTT
          Consensus  TGAAGAAGCA TCAGTAACTG TGGTAGAGGG TCAAGTTGAC TATTATGGTT

                    3225
  PREpileup{PRE250}  TATATTATGT TCATGAAGGA ATACGAACAT ATTTTGTGCA GTTTAAAGAT
**PREpileup{PREPROTO}** **TATATTATGT TCATGAAGGA ATACGAACAT ATTTTGTGCA GTTTAAAGAT**
  PREpileup{PRE262}  TATATTATGT TCATGAAGGC ATACGAACAT ATTTTGTGCA GTTTAAAGAT
          Consensus  TATATTATGT TCATGAAGG- ATACGAACAT ATTTTGTGCA GTTTAAAGAT

                    3275
  PREpileup{PRE250}  GATGCAGAAA AATATAGTAA AAATAAAGTA TGGGAAGTTC ATGCGGGTGG
**PREpileup{PREPROTO}** **GATGCAGAAA AATATAGTAA AAATAAAGTA TGGGAAGTTC ATGCGGGTGG**
  PREpileup{PRE262}  GATGCAGAAA AATATAGTAA AAATAAAGTA TGGGAAGTTC ATGCGGGTGG
          Consensus  GATGCAGAAA AATATAGTAA AAATAAAGTA TGGGAAGTTC ATGCGGGTGG

                    3325
  PREpileup{PRE250}  TCAGGTAATA TTATGTCCTA CATCTGTGTT TAGCAGCAAC GAAGTATCCT
**PREpileup{PREPROTO}** **TCAGGTAATA TTATGTCCTA CATCTGTGTT TAGCAGCAAC GAAGTATCCT**
  PREpileup{PRE262}  TCAGGTAATA TTATGTCCTA CATCTGTGTT TAGCAGCAAC GAAGTATCCT
          Consensus  TCAGGTAATA TTATGTCCTA CATCTGTGTT TAGCAGCAAC GAAGTATCCT

                    3375
  PREpileup{PRE250}  CTCCTGAAAT TATTAGGCAG CACTTGGCCA ACCACTCCGC CGCGACCCAT
**PREpileup{PREPROTO}** **CTCCTGAAAT TATTAGGCAG CACTTGGCCA ACCACCCCGC CGCGACCCAT**
  PREpileup{PRE262}  CTCCTGAAAC TATTAGGCAG CACTTGGCCA ACCACTCCGC CGCGACCCAT
          Consensus  CTCCTGAAA- TATTAGGCAG CACTTGGCCA ACCAC-CCGC CGCGACCCAT

                    3425
  PREpileup{PRE250}  ACCAAAGCCG TCGCCTTGGG CACCGAAGAA ACACAGACGA CTATCCAGCG
**PREpileup{PREPROTO}** **ACCAAAGCCG TCGCCTTGGG CACCGAAGAA ACACAGACGA CTATCCAGCG**
  PREpileup{PRE262}  ACCAAAGCCG TCGCCTTGGG CACCGAAGAA ACACAGACGA CTATCCAGCG
          Consensus  ACCAAAGCCG TCGCCTTGGG CACCGAAGAA ACACAGACGA CTATCCAGCG

                    3475
  PREpileup{PRE250}  ACCAAGATCA GAGCCAGACA CCGGAAACCC CTGCCACACC ACTAAGTTGT
**PREpileup{PREPROTO}** **ACCAAGATCA GAGCCAGACA CCGGAAACCC CTGCCACACC ACTAAGTTGT**
  PREpileup{PRE262}  ACCAAGATCA GAGCCAGACA CCGGAAACCC CTGCCACACC ACTAAGTTGT
          Consensus  ACCAAGATCA GAGCCAGACA CCGGAAACCC CTGCCACACC ACTAAGTTGT

                    3525
  PREpileup{PRE250}  TGCACAGAGA CTCAGTGGAC AGTGCTCCAA TCCTCACTGC ATTTAACAGC
**PREpileup{PREPROTO}** **TGCACAGAGA CTCAGTGGAC AGTGCTCCAA TCCTCACTGC ATTTAACAGC**
  PREpileup{PRE262}  TGCACAGAGA CTCAGTGGAC AGTGCTCCAA TCCTCACTGC ATTTAACAGC
          Consensus  TGCACAGAGA CTCAGTGGAC AGTGCTCCAA TCCTCACTGC ATTTAACAGC

                    3575
  PREpileup{PRE250}  TCACACAAAG GACGGATTAA CTGTAATAGT AACACTACAC CCATAGTACA
**PREpileup{PREPROTO}** **TCACACAAAG GACGGATTAA CTGTAATAGT AACACTACAC CCATAGTACA**
  PREpileup{PRE262}  TCACACAAAG GACGGATTAA CTGTAATAGT AACACTACAC CCATAGTACA
          Consensus  TCACACAAAG GACGGATTAA CTGTAATAGT AACACTACAC CCATAGTACA

                    3625
  PREpileup{PRE250}  TTTAAAAGGT GATGCTAATA CTTTAAAATG TTTAAGATAT AGATTTAAAA
**PREpileup{PREPROTO}** **TTTAAAAGGT GATGCTAATA CTTTAAAATG TTTAAGATAT AGATTTAAAA**
  PREpileup{PRE262}  TTTAAAAGGT GATGCTAATA CTTTAAAATG TTTAAGATAT AGATTTAAAA
          Consensus  TTTAAAAGGT GATGCTAATA CTTTAAAATG TTTAAGATAT AGATTTAAAA

                    3675
  PREpileup{PRE250}  AGCATTGTAC ATTGTATACT GCAGTGTCGT CTACATGGCA TTGGACAGGA
**PREpileup{PREPROTO}** **AGCATTGTAC ATTGTATACT GCAGTGTCGT CTACATGGCA TTGGACAGGA**
  PREpileup{PRE262}  AGCATTGTAA ATTGTATACT GCAGTGTCGT CTACATGGCA TTGGACAGGA
          Consensus  AGCATTGTA- ATTGTATACT GCAGTGTCGT CTACATGGCA TTGGACAGGA
```

```
                    3725
  PREpileup{PRE250}  CATAATGTAA AACATAAAAG TGCAATTGTT ACACTTACAT ATGATAGTGA
PREpileup{PREPROTO}  CATAATGTAA AACATAAAAG TGCAATTGTT ACACTTACAT ATGATAGTGA
  PREpileup{PRE262}  CATAATGTAA AACATAAAAG TGCAATTGTT ACACTTACAT ATGATAGTGA
          Consensus  CATAATGTAA AACATAAAAG TGCAATTGTT ACACTTACAT ATGATAGTGA

                    3775
  PREpileup{PRE250}  ATGGCAACGT GACCAATTTT TGTCTCAAGT TAAAATACCA AAAACTATTA
PREpileup{PREPROTO}  ATGGCAACGT GACCAATTTT TGTCTCAAGT TAAAATACCA AAAACTATTA
  PREpileup{PRE262}  ATGGCAACGT GACCAATTTT TGTCTCAAGT TAAAATACCA AAAACTATTA
          Consensus  ATGGCAACGT GACCAATTTT TGTCTCAAGT TAAAATACCA AAAACTATTA

                    3825
  PREpileup{PRE250}  CAGTGTCTAC TGGATTTATG TCTA
PREpileup{PREPROTO}  CAGTGTCTAC TGGATTTATG TCTA
  PREpileup{PRE262}  CAGTGTCTAC TGGATTTATG TCTA
          Consensus  CAGTGTCTAC TGGATTTATG TCTA
```

# AMINO ACID ALIGNMENT FOR PRIMER E2-2

```
                    124
                     .         .         .         .         .         .
pepPRE250           DICNXMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepPRE262           ------------HIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
pepPREPROTO         DICNTMHYTNWTHIYICEEASVTVVEGQVDYYGLYYVHEGIRTYFVQFKDDAEKYSKNKV
                           **************************************************

                    184
pepPRE250           WEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHSAATHTKAVALGTEETQTTIQRPRS
pepPRE262           WEVHAGGQVILCPTSVFSSNEVSSPETIRQHLANHSAATHTKAVALGTEETQTTIQRPRS
pepPREPROTO         WEVHAGGQVILCPTSVFSSNEVSSPEIIRQHLANHPAATHTKAVALGTEETQTTIQRPRS
                    ************************** ******** ************************

                    244
pepPRE250           EPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLRY
pepPRE262           EPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLRY
pepPREPROTO         EPDTGNPCHTTKLLHRDSVDSAPILTAFNSSHKGRINCNSNTTPIVHLKGDANTLKCLRY
                    ************************************************************

                    304
pepPRE250           RFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGFS
pepPRE262           RFKKHCKLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGFM
pepPREPROTO         RFKKHCTLYTAVSSTWHWTGHNVKHKSAIVTLTYDSEWQRDQFLSQVKIPKTITVSTGFM
                    ****** ***************************************************
```

# AMINO ACID ALIGNMENT FOR E4

```
                    1
                     .         .         .         .         .         .
E4pepPRE250         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
E4pepPRE262         AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
E4pepPREPROT        AATKYPLLKLLGSTWPTTPPRPIPKPSPWAPKKHRRLSSDQDQSQTPETPATPLSCCTET
                    ************************************************************

                    61
E4pepPRE250         QWTVLQSSLHLTAHTKDGLTVIVTLHP
E4pepPRE262         QWTVLQSSLHLTAHTKDGLTVIVTLHP
E4pepPREPROT        QWTVLQSSLHLTAHTKDGLTVIVTLHP
                    ***************************
```

## A7.6.3 E6 GENE AND PROTEIN

The sequence of the E6 gene was obtained for 27 of the original 34 cancer isolates using primer E6-2. The corresponding amino acid alignment follows the nucleotide alignment.

## NUCLEOTIDE ALIGNMENT FOR PRIMER E6-2 (nts 640-625)

```
                    73
  E6pileup{E62E6K4} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6P2} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6K2} .......... .......... .......... .......... ..........
  E6pileup{E62E6K1} .......... .......... .......... .......... ..........
  E6pileup{E62E6M4} .......... .......... .......... .......... ..........
  E6pileup{E62E6M2} .......... .......... .......... .......... ..........
  E6pileup{E62E6M5} .......... .......... .......... .......... ..........
  E6pileup{E62E6D3} .......... .......... .......... .......... ..........
  E6pileup{E62E6P1} .......... .......... .......... .......... ..........
  E6pileup{E62E6M6} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6S2} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6O2} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
E6pileup{E62E6PROT} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6K5} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6M1} .......... .......... .......... .......... ..........
  E6pileup{E62E6T1} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6J1} CAgACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6K3} .......... .......... .......... .......... ..........
  E6pileup{E62E6S1} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6B2} .......... .......... .......... .......... ..........
  E6pileup{E62E6E1} .......... .......... .......... .......... ..........
  E6pileup{E62E6H4} .......... .......... .......... .......... ..........
  E6pileup{E62E6H2} CAGACATTTT ATGCACCAAA AGAGAACTGC AATGTTTCAG GACCCACAGG
  E6pileup{E62E6R1} .......... .......... .......... .......... ..........
  E6pileup{E62E6K6} .......... .......... .......... .......... ..........
  E6pileup{E62E6G1} .......... .......... .......... .......... ..........
  E6pileup{E62E6H3} .......... .......... .......... .......... ..........
  E6pileup{E62E6C1} .......... .......... .......... .......... ..........
          Consensus ---------- ---------- ---------- ---------- ----------

                    123
  E6pileup{E62E6K4} AGCGACCCAG AAAGTTACCA CATTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6P2} AGCGACCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6K2} .......... .......... .......... .......... ..CAACTATC
  E6pileup{E62E6K1} .......... .......... .......... .......... ..CAACTATC
  E6pileup{E62E6M4} .......... .......... .......... .......... ..CAACTATC
  E6pileup{E62E6M2} .......... .......... .......... .......... ..CAACTATC
  E6pileup{E62E6M5} .......... .......... .......... .......... ..........
  E6pileup{E62E6D3} .......... .......... .......... .......... ..CAACTATA
  E6pileup{E62E6P1} .......... .......... .......... .......... ..CAACTATA
  E6pileup{E62E6M6} AGCGACCCAG AAAGTTACCC CAGTTATGCA CAGAGCTGCA ACCAACTATA
  E6pileup{E62E6S2} AGCGACCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6O2} AGCGACCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
E6pileup{E62E6PROT} AGCGACCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6K5} AGCGACCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6M1} .......... .......... .......... .......... ..CACCTATA
  E6pileup{E62E6T1} AGCGACCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6J1} AGCGACCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6K3} .....CCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6S1} AGCGACCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6B2} .......... .......... .......... .......... ..CAACTATA
  E6pileup{E62E6E1} .......... .......... .......... .......... ..CAACTATA
  E6pileup{E62E6H4} .......... .......... CAATTATGCA CAGAGCTGCA AACAACTATC
  E6pileup{E62E6H2} AGCGACCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6R1} .......... .......... .......... .......... ..CAACTATA
  E6pileup{E62E6K6} .......... .......... .......... .......... ..CAACTATA
  E6pileup{E62E6G1} .......... .......... .......... .......... ..........
  E6pileup{E62E6H3} .....CCCAG AAAGTTACCA CAGTTATGCA CAGAGCTGCA AACAACTATA
  E6pileup{E62E6C1} .......... .......... .......... .......... ..CAACTATA
          Consensus ---------- ---------- ---------- ---------- ----------
```

```
                    173
  E6pileup{E62E6K4}  CATGATATAA GATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6P2}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6K2}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6K1}  CATAATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6M4}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6M2}  CATAATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6M5}  .......... .......... .......... .......... ..........
  E6pileup{E62E6D3}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6P1}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6M6}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6S2}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6O2}  CATGAGATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
E6pileup{E62E6PROT}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6K5}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6M1}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6T1}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6J1}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6K3}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6S1}  CATGAGATAA TATTAGAATG TGTGCACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6B2}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TAATGCGACG
  E6pileup{E62E6E1}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6H4}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TAATGCGACG
  E6pileup{E62E6H2}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6R1}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TAATGCGACG
  E6pileup{E62E6K6}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TAATGCGACG
  E6pileup{E62E6G1}  ..TGATATTA TATTAGAATG GGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6H3}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
  E6pileup{E62E6C1}  CATGATATAA TATTAGAATG TGTGTACTGC AAGCAACAGT TACTGCGACG
          Consensus  ---------- ---------- ---------- ---------- ----------

                    223
  E6pileup{E62E6K4}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6P2}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6K2}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6K1}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6M4}  TGAGGTATAT GGCTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6M2}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6M5}  .......... .......... ....GGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6D3}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6P1}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6M6}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6S2}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6O2}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
E6pileup{E62E6PROT}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6K5}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6M1}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6T1}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6J1}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6K3}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCGTAGTA TATAGAGATG
  E6pileup{E62E6S1}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6B2}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6E1}  TGAGGTATAT GACCTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6H4}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6H2}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6R1}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6K6}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6G1}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6H3}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
  E6pileup{E62E6C1}  TGAGGTATAT GACTTTGCTT TTCGGGATTT ATGCATAGTA TATAGAGATG
          Consensus  ---------- ---------- ----GGATTT ATGC-TAGTA TATAGAGATG
```

```
                    273
  E6pileup{E62E6K4}  GGAATCCATA TGCAGTGTGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6P2}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6K2}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6K1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6M4}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6M2}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6M5}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6D3}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6P1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6M6}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6S2}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6O2}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
E6pileup{E62E6PROT}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6K5}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6M1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6T1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6J1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6K3}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6S1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6B2}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6E1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6H4}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6H2}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6R1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6K6}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6G1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6H3}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
  E6pileup{E62E6C1}  GGAATCCATA TGCTGTATGT GATAAATGTT TAAAGTTTTA TTCTAAAATT
          Consensus  GGAATCCATA TGC-GT-TGT GATAAATGTT TAAAGTTTTA TTCTAAAATT

                    323
  E6pileup{E62E6K4}  AGTGAGTATA GATATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6P2}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6K2}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6K1}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6M4}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6M2}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6M5}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6D3}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6P1}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6M6}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6S2}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6O2}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
E6pileup{E62E6PROT}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6K5}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6M1}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6T1}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6J1}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6K3}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6S1}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6B2}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6E1}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6H4}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6H2}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6R1}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6K6}  AGTGAGTATA GGCATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6G1}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAA CATTAGAACA
  E6pileup{E62E6H3}  AGTGAGTATA GACATTATTG TTATAGTTTG TATGGAACAG CATTAGAACA
  E6pileup{E62E6C1}  AGTGAGTATA GACATTATTG TTATAGTGTG TATGGAACAA CATTAGAACA
          Consensus  AGTGAGTATA G--ATTATTG TTATAGT-TG TATGGAACA- CATTAGAACA
```

```
                    373
  E6pileup{E62E6K4} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6P2} GCAATACAAC AATCCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6K2} GCAATACAAC AATCCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6K1} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6M4} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6M2} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6M5} GCAATTCAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6D3} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6P1} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6M6} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6S2} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6O2} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
E6pileup{E62E6PROT} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6K5} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6M1} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6T1} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6J1} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6K3} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6S1} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6B2} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6E1} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6H4} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6H2} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6R1} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6K6} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6G1} GCAATACAAC AAGCCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6H3} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
  E6pileup{E62E6C1} GCAATACAAC AAACCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC
          Consensus GCAAT-CAAC AA-CCGTTGT GTGATTTGTT AATTAGGTGT ATTAACTGTC

                    423
  E6pileup{E62E6K4} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6P2} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6K2} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6K1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6M4} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6M2} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6M5} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6D3} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6P1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6M6} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6S2} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6O2} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
E6pileup{E62E6PROT} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6K5} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6M1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6T1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6J1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6K3} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6S1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6B2} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6E1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6H4} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6H2} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6R1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6K6} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6G1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6H3} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
  E6pileup{E62E6C1} AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
          Consensus AAAAGCCACT GTGTCCTGAA GAAAAGCAAA GACATCTGGA CAAAAAGCAA
```

```
                    473
E6pileup{E62E6K4}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGCCGATGTA TGTCTTGTTG
E6pileup{E62E6P2}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6K2}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6K1}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6M4}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6M2}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6M5}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6D3}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6P1}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6M6}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6S2}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6O2}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6PROT} AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6K5}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGCCGATGTA TGTCTTGTTG
E6pileup{E62E6M1}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6T1}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGCCGATGTA TGTCTTGTTG
E6pileup{E62E6J1}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6K3}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6S1}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6B2}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6E1}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6H4}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6H2}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6R1}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6K6}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGCCGATGTA TGTCTTGTTG
E6pileup{E62E6G1}   AGATT..... .......... .......... .......... ..........
E6pileup{E62E6H3}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
E6pileup{E62E6C1}   AGATTCCATA ATATAAGGGG TCGGTGGACC GGTCGATGTA TGTCTTGTTG
        Consensus   AGATT----- ---------- ---------- ---------- ----------

                    523
E6pileup{E62E6K4}   CAGATCATCG AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6P2}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6K2}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6K1}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6M4}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6M2}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6M5}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6D3}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6P1}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6M6}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6S2}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6O2}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6PROT} CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6K5}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6M1}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6T1}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6J1}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6K3}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6S1}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6B2}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6E1}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6H4}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6H2}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6R1}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6K6}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6G1}   .......... .......... .......... .......... ..........
E6pileup{E62E6H3}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
E6pileup{E62E6C1}   CAGATCATCA AGAACACGTA GAGAAACCCA GCTGTAATCA TGCATGGAGA
        Consensus   ---------- ---------- ---------- ---------- ----------
```

```
                    573
   E6pileup{E62E6K4}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6P2}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6K2}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCACCCA GAG..
   E6pileup{E62E6K1}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6M4}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6M2}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6M5}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAGA.
   E6pileup{E62E6D3}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6P1}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCACCCA GAG..
   E6pileup{E62E6M6}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6S2}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6O2}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
E6pileup{E62E6PROT}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAGAC
   E6pileup{E62E6K5}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6M1}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6T1}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6J1}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6K3}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAA... .....
   E6pileup{E62E6S1}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA .....
   E6pileup{E62E6B2}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6E1}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6H4}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6H2}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAGAC
   E6pileup{E62E6R1}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6K6}  TACACCGACA TTGCATGGAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6G1}  .......... .......... .......... .......... .....
   E6pileup{E62E6H3}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
   E6pileup{E62E6C1}  TACACCTACA TTGCATGAAT ATATGTTAGA TTTGCAACCA GAG..
           Consensus  ---------- ---------- ---------- ---------- -----
```

# AMINO ACID ALIGNMENT FOR PRIMER E6-2

```
                    1
                       .         .         .         .         .         .
pepE6B2             ----------------------TIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6C1             ----------------------TIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6D3             ----------------------TIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6E1             ----------------------TIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6G1             ----------------------DIILECVYCKQQLLRXEVYDFAFGDLCIVYRDGNPY
pepE6H2             MXQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6H3             --------PRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6H4             -------------QLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGTPY
pepE6J1             MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6K1             ----------------------TIHNIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6K2             ----------------------TIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6K3             --------PRKLPQLCTELQTLIHDIILECVYCXQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6K4             MFQDPQERPRKLPHLCTELQTTIHDIRLECVYCKQQLLRREVYDFAFRDLCVVYRDGNPY
pepE6K5             MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6K6             ----------------------TIHDIILECVYCKQQLLRREVYDFAVRDLCIVYRDGTPY
pepE6M1             ----------------------SIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6M2             ----------------------TIHNIILECVYCKQQLLRREVYDFAFRDLCIVYRDGTPY
pepE6M4             ----------------------TIHDIILECVYCKQQLLRREVYGFAFRDLCIVYRDGNPY
pepE6M5             ------------------------------------------------DLCIVYRDGTPY
pepE6M6             MFQDPQERPRKLPQLCTELQPTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6O2             MFQDPQERPRKLPQLCTELQTTIHEIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6P1             ----------------------TIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6P2             MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6R1             ----------------------TIHDIILECVYCXQQLLRREVYDFAVRDLCIVYRDGNPY
pepE6S1             MFQDPQERPRKLPQLCTELQTTIHEIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6S2             MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6T1             MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
pepE6prot           MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVYRDGNPY
                                                                *** ********
```

```
                61
pepE6B2         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6C1         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6D3         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6E1         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6G1         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6H2         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6H3         AVCDKCLKFYSKISEYRHYCYSLYGTALEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6H4         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6J1         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6K1         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6K2         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6K3         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6K4         AVCDKCLKFYSKISEYRYYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6K5         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6K6         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6M1         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6M2         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6M4         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6M5         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQFNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6M6         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6O2         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6P1         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6P2         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6R1         AVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6S1         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6S2         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6T1         AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
pepE6prot       AVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD
                ********************************************************************

                121                                 1 (E7)
pepE6B2         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6C1         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6D3         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6E1         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6G1         KKQR-----------------------------------------------
pepE6H2         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQR
pepE6H3         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQR
pepE6H4         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6J1         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6K1         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6K2         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6K3         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-IC---
pepE6K4         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6K5         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6K6         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6M1         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6M2         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6M4         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6M5         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQR
pepE6M6         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6O2         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6P1         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6P2         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNX-
pepE6R1         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6S1         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQR
pepE6S2         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6T1         KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQ-
pepE6prot       KKQRFHNIRGRWTGRCMSCCRSSRTRRETQL-SCMEIHLHCMNIC-ICNQR
                ****
```

# CHAPTER 8

# REFERENCES

# 8.1 REFERENCES


Adam, E., Kaufman, R.H., Berkova, Z., Icenogle, J., and Reeves, W.C. Is human papillomavirus testing an effective triage method for detection of high-grade (grade 2 or 3) cervical intraepithelial neoplasia. American Journal of Obstetrics and Gynecology, 178(6): 1235-1244, 1998.

Adimora, A.A., and Quinlivan, E.B. Human papillomavirus infection. Recent findings on progression to cervical cancer. Postgraduate Medicine, 98(3): 109-112, 1995.

Ai, W., Toussaint, E., and Roman, A. CCAAT displacement protein binds to and negatively regulates human papillomavirus type 6 E6, E7, and E1 promoters. Journal of Virology, 73(5): 4220-4229, 1999.

Alexander, M., Salgaller, M.L., Celis, E., Sette, A., Barnes, W.A., Rosenberg, S.A., and Steller, M.A. Generation of tumor-specific cytolytic T lymphocytes from peripheral blood of cervical cancer patients by in vitro stimulation with a synthetic human papillomavirus type 16 E7 epitope. American Journal of Obstetrics & Gynecology, 175(6): 1586-1593, 1996.

Almeida, J.D., Oriel, J.D., and Stannard, L.M. Characterisation of the virus found in human genital warts. Microbios, 3: 225-232, 1969.

Alphey, A. DNA sequencing, Chapter 7. *In:* Glover, D.M. and Homes, B.H. (eds.), DNA Cloning I – A Practical Approach. Core Techniques – Second Edition. pp 225-260, IRL Press at Oxford University Press, Oxford, UK, 1995.

Alphey, A and Parry, H.D. Making Nucleic Acid Probes, Chapter 4. *In:* Glover, D.M. and Homes, B.H (eds.), DNA Cloning I – A Practical Approach. Core Techniques – Second Edition. pp 121-141, IRL Press at Oxford University Press, Oxford, UK, 1995.

Alvarez-Salas, L.M., Wilczynski, S.P., Burger, R.A., Monk, B.J., and Dipaolo, J.A. Polymorphism of the HPV-16 E6 gene of cervical carcinoma. International Journal of Oncology, 7: 261-266, 1995.

Alvarez-Salas, L.M., Arpawong, T.E., and Dipaolo, J.A. Growth inhibition of cervical tumour cells by antisense oligodeoxynucleotides directed to the human papillomavirus type 16 E6 gene. Antisense and Nucleic Acid Drug Development, 9: 441-450, 1999.

Andersson, S., Alemi, M., Rylander, E., Strand, A., Larsson, B., Sallstrom, J., and Wilander, E. Uneven distribution of HPV 16 E6 prototype and variant (L83V) oncoprotein in cervical neoplastic lesions. British Journal of Cancer, 83(3): 307-310, 2000.

Antson, A.A., Burns, J.E., Moroz, O.V., Scott, D.J., Sanders, C.M., Bronstein, I.B., Dodson, G.G., Wilson, K.S., and Maitland, N.J. Structure of the intact transactivation domain of the human papillomavirus E2 protein. Nature, 403: 805-809, 2000.

Anwar, K., Nakakuki, K., Shiraishi, T., Naiki, H., Yatani, R., and Inuzuka, M. Presence of *ras* oncogene mutations and human papillomavirus DNA in human prostate cancer. Cancer Research, 52: 5991-5996, 1992.

Aoyama, C., Pters, J., Senadheera, S., Liu, P., and Shimada, H. Uterine cervical dysplasia and cancer: Identification of c-*myc* status by quantitative polymerase chain reaction. Diagnostic Molecular Pathology, 7(6): 324-330, 1998.

Apple, R.J., Erlich, H.A., Kiltz, W., Manos, M.M., Becker, T.M., and Wheeler, C.M. HLA DR-DQ associations with cervical carcinoma show papillomavirus-type specificity. Nature Genetics, 6: 157-162, 1994.

Apple, R.J., Becker, T.M., Wheeler, C.M., and Erlich, H.A. Comparison of human leukocyte antigen DR-DQ disease associations found with cervical dysplasia and invasive cervical carcinoma. Journal of the National Cancer Institute, 87: 427-436, 1995.

Appleby, J. Management of the abnormal Papanicolaou smear. Medical Clinics of North America. 79: 345-360, 1995.

Apt, D., Chong, T., Liu, Y., and Bernard, H.U. Nuclear factor I and epithelial cell-specific transcription of human papillomavirus type 16. Journal of Virology, 67(8): 4455-4463, 1993.

Apt, D., Liu, Y., and Bernard, H.U. Cloning and functional analysis of spliced isoforms of human nuclear factor I-X: Interference with transcriptional activation by NFI/CTF in a cell-type specific manner. Nucleic Acids Research, 22(19): 3825-3833, 1994.

Apt, D., Watts, R.M., Suske, G., and Bernard, H.U. High Sp1/Sp3 ratios in epithelial cells during epithelial differentiation and cellular transformation correlate with the activation of the HPV-16 promoter. Virology, 224: 281-291, 1996.

Armstrong, BK., Munoz, N., and Bosch, FX. The epidemiology of cancer of the cervix. *In:* Coppleson M, (ed.), Gynecologic Oncology, pp 11-30, Edinburgh, 1992.

Auborn, K.J., Woodworth, C., DiPaolo, J.A., and Bradlow, H.L. The interaction between HPV infection and estrogen metabolism in cervical carcinogenesis. International Journal of Cancer, 49: 867-869, 1991.

Australasian Institute of Health and Welfare and Australasian Society of Cancer Registries – Cancer in Australia 1991-1994 (with projections to 1998). Canberra: Australian Institute of Health and Welfare (Cancer series No. 7).

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. Short Protocols in Molecular Biology. Second Edition. Harvard Medical School. John Wiley and Sons, New York, USA, 1992.

Awady, M.K., Kaplan, J.B., O'Brien, S.J., and Burk, R.D. Molecular analysis of integrated human papillomavirus 16 sequences in the cervical cancer cell line SiHa. Virology, 159: 389-398, 1987.

Azoury-Ziadeh, R., Herd, K., Fernando, G.J.P., Frazer, I.H., and Tindle, R.W. T-helper epitopes identified within the E6-transforming protein of cervical cancer-associated human papillomavirus type 16. Viral Immunology, 12(4): 297-312, 1999.

Bain, R.W., and Crocker, D.W. Rapid onset of cervical cancer in an upper socioeconomic group. American Journal of Obstetrics and Gynecology, 146: 366-371, 1983.

Baker, C.C., Phelps, W.C., Lindgren, V., Braun, M.J., Gonda, M.A., and Howley, P.M. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. Journal of Virology, 61(4): 962-971, 1987.

Baker, T.M., Newcomb, W.W., Olsen, N.H., Cowsert, L.M., Olsen, C., and Brown, J. Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. Biophysical Journal, 60: 1445-1456, 1991.

Barber, H.R.K. Incidence, prevalence, and median survival rates of gynecologic cancer. *In:* Van Nagell, J.R., and Barber, H.R.K. (eds.), Modern Concepts of Gynecologic Oncology. pp 1-19, John Wright-PSG, Boston, USA, 1982.

Barbosa, M.S., Lowy, D., and Schiller, J. Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. Journal of Virology, 63(3): 1404-1407, 1989.

Barbosa, M., and Schlegel, R. The E6 and E7 genes of HPV 18 are sufficient for inducing two-stage *in vitro* transformation of human keratinocytes. Oncogene, 4: 1529-1532, 1989.

Barbosa, M.S., Vass, W.C., Lowy, D.R., and Schiller, J.T. *In vitro* biological activities of the E6 and E7 genes vary among human papillomaviruses of different oncogenic potential. Journal of Virology, 65: 292-298, 1991.

Barbosa, M.S. The oncogenic role of human papillomavirus proteins. Critical Reviews in Oncogenesis, 7(1 and 2): 1-18, 1996.

Barnes, W., Delgado, G., Kurman, R.J., Petrilli, E.S., Smith, D.M., Ahmed, S., Lorincz, A.T., Temple, G.F., Jensen, A.B., and Lancaster, W.D. Possible prognostic significance of human papillomavirus type in cervical cancer. Gynecologic Oncology, 29: 267-273, 1988.

Bartholomew, J.S., Stacey, S.N., Coles, B., Burt, D.J., Arrand, J.R., and Stern, P.L. Identification of a naturally processed HLA A0201-restricted viral peptide from cells expressing human papillomavirus type 16 E6 oncoprotein. European Journal of Immunology, 24: 3175-3179, 1994.

Barrasso, R., and Guillemotonia, A. Cervix and vagina: Diagnosis, Chapter 6. *In:* Gross, G.E., Barrasso, R. (eds), Human Papilloma Virus Infection. A Clinical Atlas. pp 147-276, Ullstein Mosby, Berlin, Germany, 1997.

Barrasso, R., Zanardi, C., Huynh, B., and Ferenczy, A. Cervix and vagina: Treatment, Chapter 7. *In:* Gross, G.E., Barrasso, R. (eds), Human papilloma virus infection. A clinical atlas. pp 277-287, Ullstein Mosby, Berlin, Germany, 1997.

Bauer, H.M., Ting, Y., Greer, C.E., Chambers, J.C., Tashiro, C.J., Chimera, J., Reingold, A., and Manos, M.M. Genital human papillomavirus infection in female university students as determined by a PCR-based method. Journal of the American Medical Association, 265(4): 472-477, 1991.

Bauknecht, T., Angel, P., Royer, H.D., and zur Hausen, H. Identification of a negative regulatory domain in the human papillomavirus type 18 promoter: Interaction with the transcriptional repressor YY1. The EMBO Journal, 11(12): 4607-4617, 1992.

Bauknecht, T., Jundt, F., Herr, I., Oehler, T., Delius, H., Shi, Y., Angel, P., and zur Hausen, H. A switch region determines the cell type-specific positive or negative action of YY1 on the activity of the human papillomavirus type 18 promoter. Journal of Virology, 69(1): 1-12, 1995.

Bauknecht, T., See, R.H., and Shi, Y. A novel C/EBP β-YY1 complex controls the cell-type-specific activity of the human papillomavirus type 18 upstream regulatory region. Journal of Virology, 70(11): 7695-7705, 1996.

Bavin, P.J., Walker, P.G., and Emery, V.C. Sequence microheterogeneity in the long control region of clinical isolates of human papillomavirus type 16. Journal of Medical Virology, 39: 267-272, 1993.

Beaudenon, S., Kremsdorf, D., Croissant, O., Jablonska, S., Wain-Hobson, S., and Orth, G. A novel type of human papillomavirus associated with genital neoplasias. Nature, 321(6067): 246-249, 1986.

Bedell, M.A., Jones, K.H., Grossman, S.R., and Laimins, L.A. Identification of human papillomavirus type 18 transforming genes in immortalized and primary cells. Journal of Virology, 63: 1247-1255, 1989.

Bedell, M.A., Hudson, J.B., Golub, T.R., Turyk, M.E., Hosken, M., Wilbanks, G.D., and Laimins, L.A. Amplification of human papillomavirus genomes *in vitro* is dependent on epithelial differentiation. Journal of Virology, 65: 2254-2260, 1991.

Benoit, A.G., Krepart, G.V., and Lotocki, R.J. Results of prior cytologic screening in patients with a diagnosis of stage I carcinoma of the cervix. American Journal of Obstetrics and Gynecology, 148: 690-694, 1984.

Berkhout, R.J.M., Tieben, L.M., Smites, H.L., Bavinck, JNB., Vermeer, B.J., and Terschegget, J. Nested PCR approach for detection and typing of epidermodysplasia verruciformis-associated human papillomavirus types in cutaneous cancers from renal transplant recipients. Journal of Clinical Microbiology, 33: 690-695, 1995.

Bernard, H.U., Seedorf, K., and Klock, G. The upstream regulatory region of the human papilloma virus-16 contains an E2 protein-independent enhancer which is specific for cervical carcinoma cells and regulated by glucocortioid hormones. The EMBO Journal, 6(12): 3735-3743, 1987.

Bernard, H.U., and Apt, D. Transcriptional control and cell type specificity of HPV gene expression. Achives of Dermatology, 130: 210-215, 1994.

Bernard, H.U., Chan, S.Y., Manos, M.M., Ong, C.K., Villa, L.L., Delius, H., Peyton, C.L., Bauer, H.M., and Wheeler, C.M. Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms. The Journal of Infectious Diseases, 170: 1077-1085, 1994.

Berumen, J., Casas, L., Segura, E., Amezcua, J.L., and Garcia-Carranca, A. Genome amplification of human papillomavirus type 16 and 18 in cervical carcinomas is related to the retention of the E1/E2 genes. International Journal of Cancer, 56: 640-645, 1994.

Beutner, K.R., and Ferenczy, A. Therapeutic approaches to genital warts. American Journal of Medicine, 102: 28-37, 1997.

Bible, J.M., Mant, C., Best, J.M., Kell, B., Starkey, W.G., Raju, K.S., Seed, P., Biswas, C., Muir, P., Banatvala, J.E., and Cason, J. Cervical lesions are associated with human papillomavirus type 16 intratypic variants that have high transcriptional activity and increased usage of common mammalian codons. Journal of General Virology, 81: 1517-1527, 2000.

Blohme, I., and Brynger, H. Malignant disease in renal transplant patients. Transplantation, 39: 23-25, 1985.

Böhm, S., Wilczynski, S.P., Pfister, H., and Iftner, T. The predominant mRNA class in HPV 16-infected genital neoplasias does not encode the E6 or the E7 protein. International Journal of Cancer, 55: 791-798, 1993.

Bontkes, H.J., van Duin, M., de Gruijl, T.D., Duggan-Keen, M.F., Walboomers, J.M.M., Stukart, M.J., Verheijen, R.H.M., Helmerhorst, T.J.M., Meijer, C.J.L.M., Scheper, R.J., Stevens, A., Dyer, P.A., Sinnott, P., and Stern, P.L. HPV 16 infection and progression of cervical intra-epithelial neoplasia: Analysis of HLA polymorphism and HPV 16 E6 sequence variants. International Journal of Cancer, 78: 166-171, 1998.

Bontkes, H.J., de Gruijl, T.D., Walboomers, J.M., Schiller, J.T., Dillner, J., Helmerhorst, T.J., Verheijen, R.H., Scheper, R.J., and Meijer, C.J. Immune responses against human papillomavirus (HPV) type 16 virus-like particles in a cohort study of women with cervical intraepithelial neoplasia. II. Systemic but not local IgA responses correlate with clearance of HPV-16. Journal of General Virology, 80 ( Pt 2): 409-417, 1999a.

Bontkes, H., de Gruijl, T.D., Bigl, A., Verheijen, R.H.M., Meijer, C.J.L.M., Scheper, R.J., Stern, P.L., Burns, J.E., Maitland, N.J., and Walboomers, J.M.M. Human papillomavirus type 16 E2-specific T-helper lymphocyte responses in patients with cervical intraepithelial neoplasia. Journal of General Virology, 80: 2453-2459, 1999b.

Bosch, F.X., Munoz, N., de Sanjose, S., Izarzugaza, I., Gili, M., Viladiu, P., Tormo, M.J., Moreo, P., Ascunce, N., and Gonzalez, L.C. *et al.* Risk factors for cervical cancer in Colombia and Spain. International Journal of Cancer, 52(5): 750-758, 1992.

Bosch, F.X., Manos, M.M., Muñoz, N., Sherman, M., Jansen, A.M., Peto, J., Schiffman, M.H., Moreno, V., Kurman, R., Shah, K.V., and The International Biological Study on Cervical Cancer (IBSCC) Study Group. Prevalence of human papillomavirus in cervical cancer: A worldwide perspective. Journal of the National Cancer Institute, 87(11): 796-802, 1995.

Bosch, F.X., Castellsagué, X., Muñoz, N., de Sanjosé, S., Ghaffarl, A.M., Gonzálcz, L.C Gili M. Izarzugaza, I., Viladiu, P., Navarro, C., Vergara, A., Ascunce, N., Guerrero, E., and Shah, K.V. Male

sexual behaviour and human papillomavirus DNA: Key risk factors for cervical cancer in Spain. Journal of the National Cancer Institute, 88: 1060-1067, 1996.

Bosch, F.X., Muñoz, N., and de Sanjosé, S. Human papillomavirus and other risk factors for cervical cancer. Biomedicine and Pharmacotherapy, 51: 268-275, 1997.

Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W., and zur Hausen, H. A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. The EMBO Journal, 3: 1151-1157, 1984.

Bouvard, V., Storey, A., Pim, D., and Banks, L. Characterization of the human papillomavirus E2 protein: Evidence of trans-activation and trans-repression in cervical keratinocytes. The EMBO Journal, 13: 5451-5459, 1994a.

Bouvard, V., Matlashewski, G., Gu, Z.M., Storey, A., and Banks, L. The human papillomavirus type 16 E5 gene cooperates with the E7 gene to stimulate proliferation of primary cells and increases viral gene expression. Virology, 203: 73-80, 1994b.

Brady, C.S., Duggan-Keen, M.F., Davidson, J.A., Varley, J.M., and Stern, P.L. Human papillomavirus type 16 E6 variants in cervical carcinoma: Relationship to host genetic factors and clinical parameters. Journal of General Virology, 80: 3233-3240, 1999.

Brand, E., Berek, J.S., and Hacker, N.F. Controversies in the management of cervical adenocarcinomas. Obstetrics and Gynecology, 71: 261-269, 1988.

Breitburd, F., Croissant, O, and Orth, G. Expression of human papillomavirus type 1 E4 gene products in warts. *In:* Steinberg, B.M., Brandsma, J.L., and Taichman, L.B. (eds.), Papillomaviruses: Cancer Cells 5. pp 115-122, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA, 1987.

Brinton, L.A., Huggins, G.R., Lehman, H.F., Mallin, K., Savitz, D.A., Trapido, E., Rosenthal, J., and Hoover, R. Long-term use of oral contraceptives and risk of invasive cervical cancer. International Journal of Cancer, 38(3): 399-444, 1986.

Buckley, C.H., and Fox, H. Pathology of clinical invasive carcinoma of cervix. *In:* Coppleson, M. (ed.), Gynaecologic Oncology, Second Edition. pp 649-662, Churchill Livingstone, Edinburgh, UK, 1992.

Burger, R.A., Monk, B.J., Kurosaki, T., Anton-Culver, H., Vasilev, S.A., Berman, M.L., and Wilczynski, S.P. Human Papillomavirus Type 18: Association with poor prognosis in early stage cervical cancer. Journal of the National Cancer Institute, 88: 1361-1368, 1996.

Burghardt, E., Webb, M.J., Monaghan, J.M., Kinderman, D (eds.), Surgical Gynecologic Oncology. Thieme Medical Publishers, New York, USA, 1993.

Burnet, F.M. The concept of immunological surveillance. Progress in Experimental Tumor Research, 13: 1-27, 1970.

Cannistra, S.A., and Niloff, J.M. Cancer of the uterine cervix. The New England Journal of Medicine, 334(16): 1030-1038, 1996.

Carson, L.F., Twigg, L.B., Fukushima, M., Ostrow, R.S., Faras, A.J., and Okagaki, T. Human genital papilloma infections: An evaluation of immunologic competence in the genital neoplasia-papilloma syndrome. American Journal of Obstetrics and Gynecology, 155: 784-789, 1986.

Casas, L., Galvan, S.C., Ordoñez, R.M., Lopez, N., Guido, M., and Berumen, J. Asian-American variants of human papillomavirus type 16 have extensive mutations in the E2 gene and are highly amplified in cervical carcinomas. International Journal of Cancer, 83: 449-455, 1999.

Chan, S.Y., Ho, L., Ong, C.K., Chow, V., Drescher, B., Dürst, M., ter Meulen, J., Villa, L., Luande, J., Mgaya, H.N., and Bernard, H.U. Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind. Journal of Virology, 66(4): 2057-2066, 1992a.

Chan, S.Y., Bernard, H.U., Ong, C.K., Chan, S.P., Hofmann, B., and Delius, H. Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and variants: A showcase for the molecular evolution of DNA viruses. Journal of Virology, 66(10): 5714-5725, 1992b.

Chan, S.Y., Delius, H., Halpern, A.L., and Bernard, H.U. Analysis of genomic sequences of 95 papillomavirus types: Uniting typing, phylogeny, and taxonomy. Journal of Virology, 69(5): 3074-3083, 1995.

Chan, W.K., Klock, G., and Bernard, H.U. Progesterone and glucocorticoid response elements occur in the long control regions of several human papillomaviruses involved in anogenital neoplasia. Journal of Virology, 63(8): 3261-3269, 1989.

Chan, W.K., Chong, T., Bernard, H.U., and Klock, G. Transcription of the transforming genes of the oncogenic human papillomavirus-16 is stimulated by tumor promoters through AP1 binding sites. Nucleic Acids Research, 18(4): 763-769, 1990.

Chellappan, S., Kraus, V., Kroger, B., Münger, K., Howley, P.M., Phelps, W., and Nevins, J.R. Adenovirus E1A, simian virus 40 tumour antigen, human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. Proceedings of the National Academy of Sciences of the United States of America, 89: 4549-4553, 1992.

Chen, C.M., Shyu, M.P., Au, L.C., Chu, H.W., Cheng, W.T.K., and Choo, K.B. Analysis of deletion of the integrated human papillomavirus 16 sequence in cervical cancer: A rapid multiplex polymerase chain reaction approach. Journal of Medical Virology, 44: 206-211, 1994b.

Chen, J.J., Reid, C.E., Band, V., and Androphy, E.J. Interaction of papillomavirus E6 oncoproteins with a putative calcium-binding protein. Science, 269: 529-531, 1995.

Chen, L., Miruno, M.T., Singhal, M.C., Hu, S, L., Galloway, D.A., Hellstrom, I and Hellstrom, K.E. Induction of cytotoxic T lymphocytes specific for syngeneic tumor expressing the E6 oncoprotein of human papillomaqvirus type 16. Journal of Immunology, 148: 2618-2627, 1992.

Chen, T.M., Chen, C.A., Wu, C.C., Huang, S.C., Chang, C.F., and Hsieh, C.Y. The genotypes and prognostic significance of human papillomaviruses in cervical cancer. International Journal of Cancer, 57: 181-184, 1994a.

Chen, Z., Storthz, K.A., and Shillitoe, E.J. Mutations in the long control regions of human papillomavirus DNA in oral cancer cells, and their functional consequences. Cancer Research, 57: 1614-1619, 1997.

Chesters, P.M., and McCance, D.J. Human papillomavirus type 6 and 16 in cooperation with Ha-ras transform secondary rat embryo fibroblasts. Journal of General Virology, 70(2): 353-365, 1989.

Chong, T., Chan, W.K., and Bernard, H.U. Transcriptional activation of human papillomavirus 16 by nuclear factor I, AP1, steroid receptors and a possibly novel transcription factor, PVF: A model for the composition of genital papillomavirus enhancers. Nucleic Acids Research, 18(3): 465-470, 1990.

Chong, T., Apt, D., Gloss, B., Isa, M., and Bernard, H.U. The enhancer of human papillomavirus type 16: Binding sites for the ubiquitous transcription factors Oct-1, NFA, TEF-2, NF1, and AP-1 participate in epithelial cell-specific transcription. Journal of Virology, 65(11): 5933-5943, 1991.

Choo, K.B., Pan, C.C., and Han, S.H. Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: Preferential deletion of the E2 gene and variable retention of the long control region and the E6/E7 open reading frames. Virology, 162: 259-261, 1987.

Choo, K.B., Lee, H.H., Pan, C.C., Wu, S.M., Liew, L.N., Cheung, W.F., and Han, S.H. Sequence duplication and internal deletion in the integrated human papillomavirus type 16 genome cloned from a cervical carcinoma. Journal of Virology, 62: 1659-1666, 1988.

Chow, L.T, and Broker, T.R. Papillomavirus DNA replication. Intervirology, 37(3-4): 150-158, 1994.

Ciccolini, F., Di Passquale, G., Carlotti, F., Crawford, L., and Tommasino, M. Functional studies of E7 proteins from different HPV types. Oncogene, 9: 2633-2638, 1994.

Cirisano, F.D. Management of pre-invasive disease of the cervix. Seminars in Surgical Oncology, 16: 222-227, 1999.

Ciuffo, G. Innesto positivo con filtrado di verrucae volgara. G. Ital. Mal Venereol, 48: 12-17, 1907.

Clertant, P., and Seif, I. A common function for polyomavirus large T and papillomavirus E1 proteins? Nature, 311: 276-279, 1984.

Clewley, J.P. Introduction to PCR for viral diagnosis. *In:* Clewley, J.P. (ed.), The polymerase chain reaction for human viral diagnosis. pp 1-12, CRC Press, Inc. Florida, USA, 1995.

Cohen, S.N, Chang, A. C.Y., and Hsu, L. Nonchromosomal antibiotic resistance in bacterial genetic transformation of *Escherichia coli* by R-Factor DNA. Proceedings of the National Academy of Sciences of the United States of America, 69(8): 2110-2114, 1972.

Cole, S.T., and Danos, O. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. Journal of Molecular Biology, 193: 599-608, 1987.

Coleman, N., Birley, H.D., Renton, A.M., Hanna, N.F., Ryait, B.K., Byrne, M., Taylor-Robinson, D., and Stanley, M.A. Immunological events in regressing genital warts. American Journal of Clinical Pathology, 102: 768-774, 1994.

Conrad, M., Bubb, V.J., and Schlegel, R. The HPV-6 and HPV-16 E5 proteins are membrane-associated proteins which associate with the 16 kD pore-forming protein. Journal of Virology, 67: 6170-6178, 1993.

Conrad Stöppler, M., Ching, K., Stöppler, H., Clancey, K., Schlegel, R., and Icenogle, J. Natural variants of the human papillomavirus type 16 E6 protein differ in their abilities to alter keratinocyte differentiation and to induce p53 degradation. Journal of Virology, 70(10): 6987-6993, 1996.

Corden, S.A., Sant-Cassia, L.J., Easton, A.J., and Morris, A.G. The integration of HPV-18 DNA in cervical carcinoma. Journal of Clinical Pathology and Molecular Pathology, 52: 275-282, 1999.

Cossart, Y.E., Thompson, C., and Rose, B. Virology. In: Mindel, A (ed.). Genital Warts – HPV Infection. pp 1-34, Edward Arnold, A member of the Hodder Headline Group, London, UK, 1995.

Cotten, M., Schaffner, G., and Birnstiel, M.L. Ribozyme, antisense RNA, and antisense DNA inhibition of U7 small nuclear ribonucleoprotein-mediated histone pre-mRNA processing *in vitro*. Molecular and Cellular Biology, 9: 4479-4487, 1989.

Coutlée, F., Mayrand, M.H., Provencher, D., and Franco, E. The future of HPV testing in clinical laboratories and applied virology research. Clinical and Diagnostic Virology, 8: 123-141, 1997.

Couturier, J., Sastre-garau, X., Schneider-Maunoury, S., Labib, A., and Orth, G. Integration of papillomavirus DNA near c-myc genes in genital carcinomas and its consequences for proto-oncogene expression. Journal of Virology, 65: 4534-4538, 1991.

Cowsert, M., Lake, P., and Jenson, A.B. Topographical and conformational epitopes of bovine papillomavirus type 1 defined by monoclonal antibodies. Journal of the National Cancer Institute, 79: 1053-1057, 1987.

Cox, J.T. Clinical role of HPV testing, Obstetrics and Gynecology Clinics of North America, 23: 811-851, 1996.

Cramer, D.W., and Cutler, S.J. Incidence and histopathology of malignancies of the female genital organs in the United States. American Journal of Obstetrics and Gynecology, 118: 443-460, 1974.

Crawford, L.V., and Crawford, E.M. A comparative study of polyoma and papilloma viruses. Virology, 21: 258-263, 1963.

Crawford, L.V. A study of human papillomavirus DNA. Journal of Molecular Biology, 13: 362, 1965.

Crawford, L., and Tommasino, M. Oncogenes and antioncogenes in the development of HPV associated tumors. Clinics in Dermatology, 15: 207-215, 1997.

Cripe, T.P., Alderborn, A., Anderson, R.D., Parkkinen, S., Bergmna, P., Haugen, T.H., Pettersson, U., and Turek, L.P. Transcriptional activation of the human papillomavirus-16 P97 promoter by an 88-nucleotide enhancer containing distinct cell-dependent and AP-1-responsive modules. The New Biologist, 2(5): 450-463, 1990.

Critchlow, C.W., Surawicz, C.M., Holmes, K.K., Kuypers, J., Daling, J.R., Hawes, S.E., Goldbaum, G.M., Sayer, J., Hurt, C., Dunphy, C., and Kiviat, N.B. Prospective study of high grade anal squamous intraepithelial neoplasia in a cohort of homosexual men: influence of HIV infection, immunosupression and human papillomavirus infection. AIDS, 9: 1255-1262, 1995.

Croissant, O., Breitburd, F., and Orth, G. Specificity of cytopathic effect of cutaneous papillomaviruses. Clinics in Dermatology, 3: 43-55, 1985.

Crook, T., Storey, A., Almond, N., Osborn, K., and Crawford, L. Human papillomavirus type 16 cooperates with activated ras and fos oncogenes in the hormone-dependent transformation of primary mouse cells. Proceedings of the National Academy of Sciences of the United States of America, 85(23): 8820-8824, 1988.

Crook, T., Morgenstein, J., Crawford, L., and Banks, L. Continued expression of HPV 16 E7 protein is required for maintenance of the transformed phenotype of cells transformed with HPV 16 plus EJ-ras. The EMBO Journal, 8: 513-519, 1989.

Crook, T., Greenfiled, I., Howard, J,. and Stanley, M. Alterations in growth properties of human papillomavirus type 16 immortalised human cervical keratinocyte cell line correlates with amplification and overexpression of c-myc oncogene. Oncogene, 5: 619-622, 1990.

Crook, T., Wrede, D., and Vousden, K.H. p53 point mutation in HPV negative human cervical carcinoma cell lines. Oncogene, 6(5): 873-875, 1991a.

Crook, T., Tidy, J.A., and Vousden, K.H. Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation. Cell, 67: 547-556, 1991b.

Cullen, A.P., Reid, R., Campion, M. And Lörincz, A.T. Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm. Journal of Virology, 65(2): 606-612, 1991.

Cuthill, S., Sibbet, G,J., and Campo, M.S. Characterisation of a nuclear factor, papilloma enhancer binding factor-1, that binds the long control region of human papillomavirus type-16 and contributes to enhancer activity. Molecular Carcinogenesis, 8(2): 96-104, 1993.

Daniel, B., Mukherjee, G., Seshadri, L., Vallikad, E., and Krishna, S. Changes in the physical state and expression of human papillomavirus type 16 in the progression of cervical intraepithelial neoplasia lesions analysed by PCR. Journal of General Virology, 76: 2589-2593, 1995.

Das, B.C., Sharma, J.K., Gopalakrishna, V., and Luthra, U.K. Analysis by polymerase chain reaction of the physical state of human papillomavirus type 16 DNA in cervical preneoplastic and neoplastic lesions. Journal of General Virology, 73: 2327-2336, 1992a.

Das, B.C., Sharma, J.K., Gopalakrishna, V, Luthra, U.K. A high frequency of human papillomavirus DNA sequences in cervical carcinomas of Indian women as revealed by Southern blot hybridisation and polymerase chain reaction. Journal of Medical Virology, 36: 239-245, 1992b.

Davies, R., Hicks, R., Crook, T Morris J., and Vousden, K. Human papillomavirus type 16 E7 associates with a histone H1 kinase and with p107 through sequences necessary for transformation. Journal of Virology, 67: 2521-2528, 1993.

D'Costa, J., Saranath, D., Dedhia, P., Sanghvi, V., and Mehta, A.R. Detection of HPV-16 genome in human oral cancers and potentially malignant lesions from India. Oral Oncology, 34: 413-420, 1998.

de Sanjose, S., Bosch, F.X., Munoz, N., Tafur, L., Gili, M., Izarzugaza, I., Izquierdo, A., Navarro, C., Vergara, A., Munoz, M.T., Ascunce, N., and Shah, K.V. Socioeconomic differences in cervical cancer: two case-control studies in Colombia and Spain. American Journal of Public Health, 86(11): 1532-1538, 1996.

de Villiers, E.M. Laboratory techniques in the investigation of human papillomavirus infection. Genitourinary Medicine. 68: 50-54, 1992.

de Villiers, E.M. Human pathogenic papillomavirus types: An update. In: Human Papillomaviruses. pp 1-12, Springer Verlag, Berlin, Germany, 1994.

de Villiers, E.M. Papillomaviruses and HPV typing. Clinics in Dermatology, 15: 199-206, 1997.

de Wet, J.R., Wood, K.V., de Luca, M., Helinski, D.R., and Subramani, S. Firefly luciferase gene: Structure and expression in mammalian cells. Molecular and Cellular Biology, 7(2): 725-737, 1987.

Della Torre, G., Pilotti, S., de Palo, G., and Rilke, F. Viral particles in cervical condylomatous lesions. Tumorigenesis, 64(5): 549-553, 1978.

Demeret, C., Yaniv, M., and Thierry, F. The E2 transcriptional repressor can compensate for Sp1 activation of the human papillomavirus type 18 early promoter. Journal of Virology, 68(11): 7075-7082, 1994.

Dent, C.L., and Latchman, D.S. The DNA mobility shift assay. *In:* Latchman, D.S. (ed.), Transcription Factors. A Practical Approach. pp 1-26, IRL Press at Oxford University, Oxford, 1993.

Desaintes, C., Hallez, S., van Alphen, P., and Burny, A. Transcriptional activation of several homologous promoters by the E6 protein of human papillomavirus type 16. Journal of Virology, 66: 325-333, 1992.

Desaintes, C., Demeret, C., Goyat, S., Yaniv, M., and Thierry, F. Expression of the papillomavirus E2 protein in HeLa cells leads to apoptosis. The EMBO Journal, 16(3): 504-514, 1997.

Dillner, J., Wiklund, F., Lenner, P., Eklund, C., Fredriksson-Shanazarian, V., Schiller, J.T., Hibma, M., Hallmans, G., and Stendahl, U. Antibodies against linear and conformational epitopes of human papillomavirus type 16 that independently associate with incident cervical cancer. International Journal of Cancer, 60: 377-382, 1995.

DiLorenzo, T.P., Tamsen, A., Abramson, A.L., and Steinberg, B.M. Human papillomavirus type 6a DNA in the lung carcinoma of a patient with recurrent laryngeal papillomatosis is characterized by a partial duplication. Journal of General Virology, 73: 423-428, 1992.

DiPaolo, J.A., Woodworth, C.D., Popescu, N.C., Notario, V., and Doniger, J. Induction of human cervical squamous cell carcinoma by sequential transfection with human papillomavirus 16 DNA and viral Harvey ras. Oncogene, 4(4): 395-399, 1989.

DiPaolo, J.A., Woodworth, C.D., Popescu, N.C., Koval, D.L., Lopez, J.V., and Doniger, J. HSV-2-induced tumorigenicity in HPV 16-immortalized human genital keratinocytes. Virology, 177: 777-779, 1990.

DiPaolo, J.A., Popescu, N.C., Woodworth, C.D., and Zimonjic, D.B. Papillomaviruses and potential copathogens. Toxicology Letters, 88: 1-7, 1996.

Dolei, A., Curreli, S., Marongiu, P., Pierangeli, A., Gomes, E., Bucci, M., Serra, C., and Degener, A.M. Human immunodeficiency virus infection *in vitro* activates naturally integrated human papillomavirus type 18 and induces synthesis of the L1 capsid. Journal of General Virology, 80: 2937-2944, 1999.

Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgomery, C.A. Jr., Butel, J.S., and Bradley, A. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature, 356: 215-221, 1992.

Dong, X.P., Stubenrauch, F., Beyer-Finkler, E., and Pfister, H. Prevalence of deletions of YY1-binding sites in episomal HPV 16 DNA from cervical cancers. International Journal of Cancer, 58: 803-808, 1994.

Dong, X.P., and Pfister, H. Overlapping YY1 - and aberrant SP1-binding sites proximal to the early promoter of human papillomavirus type 16. Journal of General Virology, 80: 2097-2101, 1999.

Doorbar, J., Campbell, D., Grand, R.J., and Gallimore, P.H. Identification of the human papilloma virus-1a E4 gene products. The EMBO Journal, 5(2): 355-362, 1986.

Doorbar, J., Ely, S., Sterling, J., McLean, C., and Crawford, L. Specific interaction between HPV-16 E1-E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. Nature, 352: 824-827, 1991.

Doorbar, J. The E4 proteins and their role in the viral life cycle. *In:* Lacey, C. (ed.), Papillomavirus Reviews: Current Research on Papillomaviruses. pp 31-38, Leeds University Press, Leeds, UK, 1996.

Doorbar, J and Myers, G. The E4 Protein. *In:* Myers, G., Halpern, A., Baker, C., McBride, A., Wheeler, C., and Doorbar, J. Human Papillomaviruses 1996. pp III-58-80, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, 1996.

Dowhanick, J.J., McBride, A.A., and Howley, P.M. Supression of cellular proliferation by the human papillomavirus E2 protein. Journal of Virology, 69: 7791-7799, 1995.

Dulic, V., Kaufmann, W., Wison, S., Tisty, T., Lees, E., Harper, J., Elledge, S., and Reed, S. p53-dependent inhibition of cyclin-dependent kinase activities in human fibroblasts during radiation-induced G1 arrest. Cell, 76: 1013-1023, 1994.

Dunn, A.E., and Ogilvie, M.M. Intranuclear virus particles in human genital wart tissue: Observation on the ultrastructure of epidermal layer. Journal of Ultrastructural Research, 22: 282-295, 1968.

Dupuy, C., Buzoni-Gatel, D., Touze, A., Le Cann, P., Bout, D., and Coursaget, P. Cell mediated immunity induced in mice by HPV 16 L1 virus-like particles. Microbial Pathogenesis, 22: 219-225, 1997.

Dürst, M., Gissmann, L., Ikenberg, H., and zur Hausen, H. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proceedings of the National Academy of Sciences of the United States of America, 80: 3812-3815, 1983.

Dürst, M., Kleinheinz, A., Hotz, M., and Gissman, L. The physical state of human papillomavirus type 16 DNA in benign and malignant genital tumours. Journal of General Virology, 66 (7): 1515-1522, 1985.

Dürst, M., Dzarlieva-Petrusevska, R.T., Boukamp, P., Fusenig, N.E., and Gissmann, L. Molecular and cytogenetic analysis of immortalised human primary keratinocytes obtained after transfection with human papillomavirus type 16 DNA. Oncogene, 1(3): 251-256, 1987.

Dürst, M., Gallahan, D., Jay, G., and Rhim, J.S. Glucocorticoid-enhanced neoplastic transformation of human keratinocytes by human papillomavirus type 16 and an activated *ras* oncogene. Virology, 173: 767-771, 1989.

Dürst, M., Glitz, D., Schnieder, Z., and zur Hausen, H. Human papillomavirus type 16 (HPV 16) gene expression and DNA replication in cervical neoplasia: Analysis by *in situ* hybridisation. Virology, 189: 132-140, 1992.

Dussurget, O., and Roulland-Dussoix, D. Rapid, sensitive PCR-based detection of mycoplasmas in simulated samples of animal sera. Applied and Environmental Microbiology, 60(3): 953-959, 1994.

Dynan, W.S., and Tjian, R. Isolation of transcription factors that discriminate between different promoters recognised by RNA polymerase II. Cell, 32: 669-680, 1983.

Dyson, N., Howley, P.M., Münger, K., and Harlow, E. The human papillomavirus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science, 243(4893): 934-936, 1989.

Dyson, N., Guida, P., Münger, K., and Harlow, E. Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins. Journal of Virology, 66: 6893- 6902, 1992.

Eckert, R.L., Crish, J.F., Banks, E.B., and Welter, J.F. The epidermis: Genes on – genes off. The Journal of Investigative Dermatology, 109(4): 501-509, 1997.

Edmonds, C., and Vousden, K.H. A point mutational analysis of human papillomavirus type 16 E7 protein. Journal of Virology, 63: 2650-2656, 1989.

Edwards, R.P., Kuykendall, K., Crowley-Nowick, P., partrigde, E.E., Shingleton, H.M., and Mestcky, J. T lymphocytes infiltrating advanced grades of cervical neoplasia: CD8-positive cells are recruited to invasion. Cancer, 76: 1411-1415, 1995.

Eide, T.J. Cancer of the uterine cervix in Norway by histologic type, 1970-84. Journal of the National Cancer Institute, 79: 199-205, 1987.

Ellis, J.R.M., Keatong, P.J., Baird, J., Hounsell, E.F., Renouf, D.V., Rowe, M., Hopkins, D., Duggan-Keen, M.F., Bartholomew, J.S., Young, L.S., and Stern, P.L. The association of an HPV 16 oncogene

variant with HLA-B7 has implications for vaccine design in cervical cancer. Nature Medicine, 1(5): 464-470, 1995.

Ellis, J.R.M., Etherington, I., Galloway, D., Luesley, D., and Young, L.S. Antibody responses to HPV16 virus-like particles in women with cervical intraepithelial neoplasia infected with a variant HPV16. The Lancet, 348: 1069-1070, 1997.

Emini, E.A., Hughes, J.V., Perlow, D.S., and Boger, J. Induction of Hepatitis A virus-neutralising antibody by a virus-specific synthetic peptide. Journal of Virology, 55(3): 836-839, 1985.

Enomoto, T., Inoue, M., Perantoni, A.O., Terakawa, N., Tanizawa, O., and Rice, JM. K-ras activation in neoplasms of the human female reproductive tract. Cancer Research, 50(19): 6139-6145, 1990.

Enomoto, T., Inoue, M., Perantoni, A.O., Buzard, G.S., Miki, H., Tanizawa, O., and Rice, J.M. K-ras activation in premalignant and malignant epithelial lesions of the human uterus. Cancer Research, 51(19): 5308-5314, 1991.

Eriksson, A., Herron, J.R., Yamada, T., and Wheeler, C.M. Human papillomavirus type 16 variant lineages characterized by nucleotide sequence analysis of the E5 coding segment and E2 hinge region. Journal of General Virology, 80: 595-600, 1999.

Etherington, I., Ellis, J., Luesley, D., Moffitt, D., and Young, L. Histologic and immunologic associations of an HPV 16 variant in LoSIL smears. Gynecological Oncology, 72: 56-59, 1999.

Evander M, Frazer IH, Payne E, Qi YM, Hengst K and McMillan NA. Identification of the alpa6 integrin as a candidate receptor for papillomaviruses. Journal of Virology, 71(3): 2449-56, 1997.

Fang, B.S., Guedes, A.C., Muñoz, L.C., and Villa, L.L. Human papillomavirus type 16 variants isolated from vulvar bowenoid papulosis. Journal of Medical Virology, 41: 49-54, 1993.

Favre, M., Ramoz, N., and Orth, G. Human papillomaviruses: General features. Clinics in Dermatology, 15: 181-198, 1997a.

Favre, M., Breitburd, F., Croissant, O., and Orth, G. Chromatin-like structures obtained after alkaline disruption of bovine and human papillomaviruses. Journal of Virology, 15: 1239-1247, 1997b.

Feltkamp, M.C., Smits, H.L., Vierboom, M.P., Minnaar, R.P., de Jongh, B.M., Drijfhout, J.W., ter Schegget, J., Melief, C.J., and Kast, W.M. Vaccination with cytotoxic T lymphocyte epitope-containing peptide protects against a tumor induced by human papillomavirus type 16-transformed cells. European Journal of Immunology, 23(9): 2242-2249, 1993.

Figge, D.C., Bennington, J.L., and Schweid, A.I. Cervical cancer after initial negative and atypical vaginal cytology. American Journal of Obstetrics and Gynecology, 108: 422-428, 1970.

Flores, E.R., and Lambert, P.F. Evidence for a switch in the mode of human papillomavirus type 16 DNA replication during the viral life cycle. Journal of Virology, 71: 7167-7179, 1997.

Foster, S.A., Demers, G.W., Etscheid, B.G., and Galloway, D.A. The ability of human papillomavirus E6 proteins to target p53 for degradation *in vivo* correlates with their ability to abrogate actinomycin D-induced growth arrest. Journal of Virology, 68: 5698-5705, 1994.

Franco, E.L.F. Epidemiology of anogential warts and cancer. *In:* Lörincz, A.T., Reid, R, (eds.), Human Papillomavirus I. Second edition. pp 597-623, Saunders, Philadelphia, USA, 1996.

Franco, E.L., Villa, L.L., Richardson, H., Rohan, T., and Ferenczy, A. Epidemiology of cervical human papillomavirus infection. *In:* Franco, E.L., Monsonego, J (eds.), New Developments in Cervical Cancer Screening and Prevention. pp 14-22, Blackwell Science, Oxford, UK, 1997.

Franco, E.L., Villa, L.L., Sobrinho, J.P., Prado, J.M., Rousseau, M.C., Désy, M., and Rohan, T.E. Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. The Journal of Infectious Diseases, 180: 1415-1423, 1999.

Frattini, M.G., Lim, H.B., and Laimins, L.A. *In vitro* synthesis of oncogenic human papillomavirus requires episomal genomes for differentiation-dependent late expression. Proceedings of the National Academy of Sciences of the United States of America, 93: 3062-3067, 1996.

Fuchs, P.G. and Pfister, H. Transcription of papillomavirus genomes. Intervirology, 37: 159-167, 1994.

Fujinaga, Y., Okazawa, K., Nishikawa, A., Yamakawa, Y., Fukushima, M., Kato, I., and Fujinaga, K. Sequence variation of human papillomavirus type 16 E7 in preinvasive and invasive neoplasias. Virus Genes, 9(1): 85-92, 1994.

Funk, J.O., Waga, S., Harry, J., Espling, E., Stillman, B., and Galloway, D.A. Inhibition of CDK activity and PCNA-dependent DNA replication by p21 is blocked by interaction with the HPV-16 E7 oncoprotein. Genes in Development, 11: 2090-2100, 1997.

Galloway, D.A. Human papillomavirus vaccines: A warty problem. Infectious Agents of Disease, 3: 187-193, 1994.

Galvin, K.M., and Shi, Y. Multiple mechanisms of transcriptional repression by YY1. Molecular and Cellular Biology, 17(7): 3723-3732, 1997.

Gao, L., Chain, B., Sinclair, C., Crawford, L., Zhou, J., Morris, J,. Zhu, X., and Strass, H. Immune responses to human papillomavirus type 16 E6 gene in a live vaccinia vector. Journal of General Virology, 75: 157-164, 1994.

Garrido-Guerrero, E., Carrillo, E., Guido, M., Zamorano, R., Garcia-Carranca, A., and Gariglio, P. Different arrangement of human papillomavirus E2 binding sites distinguishes cutaneous types from those associated with mucosal lesions. Achives of Medical Research, 27(3): 389-394, 1996.

Giles, R.H., Peters, D.J., and Breuning, M.H. Conjunction dysfunction: CDP/p300 in human disease. Trends in Genetics, 14: 178-183, 1998.

Giri, I., and Yaniv, M. Structural and mutational analysis of E2 trans-activating proteins of papillomaviruses reveals three distinct functional domains. The EMBO Journal, 7(9): 2823-2829, 1988.

Gloss, B., Bernard, H.U., Seedorf, K., and Klock, G. The upstream regulatory region of the human papilloma virus-16 contains an E2 protein-independent enhancer which is specific for cervical carcinoma cells and regulated by glucocorticoid hormones. The EMBO Journal, 6(12): 3735-3743, 1987.

Gloss, B., Chong, T., and Bernard, H.U. Numerous nuclear proteins bind the long control region of human papillomavirus type 16: A subset of 6 of 23 DNase I-protected segments coincides with the location of the cell-type-specific enhancer. Journal of Virology, 63(3): 1142-1152, 1989a.

Gloss, B., Yeo-Gloss, M., Meisterenst, M., Rogge, L., Winnacker, E.L., and Bernard, H.U. Clusters of Nuclear Factor I binding sites identify enhancers of several papillomaviruses but alone are not sufficient for enhancer function. Nucleic Acids Research, 17(9): 3519-3933, 1989b.

Gloss, B., and Bernard, H.U. The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant Sp1 distal element. Journal of Virology, 64(11): 5577-5584, 1990.

Graham, D.A., and Herrington, C.S. HPV-16 E2 gene disruption and sequence variation in CIN 3 lesions and invasive squamous cell carcinomas of the cervix: Relation to numerical chromosome abnormalities. Molecular Pathology, 53(4): 201-206, 2000.

Grassmann, K., Wilczynski, S.P., Cook, N., Rapp, B., and Iftner, T. HPV 6 variants from malignant tumors with sequence alterations in the regulatory region do not reveal differences in the activities of the oncogene promoters but do contain amino acid exchanges in the E6 and E7 proteins. Virology, 223: 185-197, 1996a.

Grassmann, K., Krätzer, F., Petry, K.U., and Iftner, T. Functional characterization of naturally occurring mutants of human papillomavirus type 16 with large deletions in the non-coding region. International Journal of Cancer, 68: 265-269, 1996b.

Grendys, E.C., Barnes, W.A., Weitzel, J., Sparkowski, J., and Schlegel, R. Identification of H, K, and N-*ras* point mutations in stage IB cervical carcinoma. Gynecologic Oncology, 65: 343-347, 1997.

Gross, G. Interferons in genital human papillomavirus disease. *In:* Gross, G., Jablonska, S., Pfister, H., and Stegner, H.E. (eds.), Genital Papillomavirus Infections- Modern Diagnosis and Treatment. pp 393-412, Springer-Verlag, Berlin, Germany, 1990.

Gross, G. Treatment of human papillomavirus infection. *In:* Mindel, A. (ed.), Genital Warts - Human Papillomavirus Infection. pp 198-236, Edward Arnold- A member of the Hodder Headline Group, London, UK, 1995.

Gross, G.E., and Barrasso, R. General principles of treatment. Chapter 3. *In:* Gross, G.E., Barrasso, R. (eds), Human papilloma virus infection. A clinical atlas. pp 48-62, Ullstein Mosby, Berlin, Germany, 1997.

Grossman, S.R., and Laimins, L.A. E6 protein of human papillomavirus type 18 binds zinc. Oncogene, 4: 1089-1093, 1989.

Gross-Mesilaty, S., Reinstein, E., Bercovich, B., Tobias, K., Schwartz, A., Kahana, C., and Ciechanover, A. Basal and human papillomavirus E6 oncoprotein-induced degradation of Myc proteins by the ubiquitin pathway. Proceedings of the National Academy of Sciences of the United States of America. 95: 8058-8063, 1998.

Gründemann, D., and Schömig, E. Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. BioTechniques, 21: 898-903, 1996.

Hakama, M., Magnus, K., Pettersson, F et al. Effect of organized screening on the risk of cervical cancer in the Nordic countries. *In:* Miller, A.B., Chamberlain, J., Day, N.E et al. (eds.), Cancer Screening. pp 153-162, UICC Projects on evaluation of screening for cancer, Cambridge, UK, 1991.

Halbert, C.L., Demers, G.W., and Galloway, D.A. The E7 gene of human papillomavirus type 16 is sufficient for immortalization of human epithelial cells. Journal of Virology, 65: 473-478, 1991.

Halbert, C.L., Demers, G.W., and Galloway, D.A. The E6 and E7 genes of human papillomavirus type 6 have weak immortalizing activity in human epithelial cells. Journal of Virology, 66(4): 2125-2134, 1992.

Hall, B.D., and Spiegelman, S. Sequence complementarity of T2-DNA and T2-specific RNA. Proceedings of the National Academy of Sciences of the United States of America, 47: 137-146, 1961.

Hall, W.S., Goto-Mandeville, R., Shih, H.A., Shank, P.R., and Braun, L. Molecular analysis of episomal human papillomavirus type 16 DNA in a cervical carcinoma cell line. Virus Research, 51: 183-195, 1997.

Halpern, A.L., and McCance, D.J. Papillomavirus E5 Proteins. *In:* Myers, G., Halpern, A., Baker, C., McBride, A., Wheeler, C., and Doorbar, J., (eds.), Human Papillomaviruses. pp III-81-111, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, 1996.

Ham, J., Dostatni, N., Gauthier, J.M., and Yaniv, M. The papillomavirus E2 protein: A factor with many talents. TIBS, 16: 440-444, 1991a.

Ham, J., Dostatni, N., Arnos, F., and Yaniv, M. Several different upstream promoter elements can potentiate transactivation by the BPV-1 E2 protein. The EMBO Journal, 10: 2931-2940, 1991b.

Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology, 166: 557-580, 1983.

Hanahan, D., Jessee, J., and Bloom, F.R. Techniques for transformation of *E.coli*, Chapter 1. *In:* Glover, D.M. and Homes, B.H. (eds.), DNA Cloning I – A Practical Approach. Core Techniques – Second Edition. pp 1-36, IRL Press at Oxford University Press, Oxford, UK, 1995.

Harris, R.W., Brinton, L.A., Cowdell, R.H., Skegg, D.C., Smith, P.G., Vessey, M.P., and Doll, R. Characteristics of women with dysplasia or carcinoma *in situ* of the cervix uteri. British Journal of Cancer, 42(3): 359-369, 1980.

Hecht, J.L., Kadish, A.S., Jiang, G., and Burk, R.D. Genetic characterization of the human papillomavirus (HPV) 18 E2 in clinical specimens suggests the presence of a subtype with decreased oncogenic potential. International Journal of Cancer, 60: 369-376, 1995.

Heck, D.V., Yee, C.L., Howley, P.M., and Münger, K. Efficiency of binding the retinoblastoma protein correlates with the transforming capacity of the E7 oncoproteins of the human papillomaviruses. Proceedings of the National Academy of Sciences of the United States of America, 89(10): 4442-4446, 1992.

Helland, A., Borresen, A-L., Kristensen, G., and Ronningen, K.S. DQA1 and DQB1 genes in pateints with squamous cell carcinoma of the cervix: Relationship to human papillomavirus infection and prognosis. Cancer Epidemiology, Biomarkers and Prevention, 3: 479-486, 1994.

Hermonat, PL. The adeno-associated virus Rep78 gene inhibits cellular transformation induced by bovine papillomavirus. Virology, 172(1): 253-261, 1989.

Hermonat, P.L. Adeno-associated virus inhibits human papillomavirus type 16: A viral interaction implicated in cervical cancer. Cancer Research, 54(8): 2278-2281, 1994.

Hermonat, P.L., Plott, R.T., Santin, A.D., Parham, G.P., and Flick, J.T. Adeno-associated virus Rep78 inhibits oncogenic transformation of primary human keratinocytes by a human papillomavirus type 16-ras chimeric. Gynecologic Oncology, 66(3): 487-494, 1997.

Herrington, C.S., and Wells, M. Can HPV typing predict the behaviour of cervical epithelial neoplasia? Histopathology, 31: 301-303, 1997.

Herrington, C.S. Do HPV-negative cervical carcinomas exist? - Revisited. Journal of Pathology, 189: 1-3, 1999.

Heselmeyer, K., Schröck, E., Du Manoir, S., Blegen, H., Shah, K., Steinbeck, R., Auer, G., and Ried, T. Gain of chromosome 3q defines the transition from severe dysplasia to invasive carcinoma of the uterine cervix. Proceedings of the National Academy of Sciences of the United States of America, 93: 479-484, 1996.

Higgens, G.D., Uzelin, D.M., Phillips, G.E., and Burrell, C.J. Presence and distribution of papillomavirus sense and anti-sense RNA transcripts in genital cancers. Journal of General Virology, 72: 885-895, 1991.

Hilders, C.G., Houbiers, J.G., van Ravenswaay Claasen, H.H., Veldhuizen, R.W., and Fleuren, G.J. Association between HLA-expression and infiltration of immune cells in cervical carcinoma. Laboratory Investigations, 69: 651-659, 1993.

Hildesheim, A., Mann, V., Brinton, L., Szklo, M., Reeves, W., and Rawls, W. Herpes simplex virus type 2: A possible interaction with human papillomavirus types 16/18 in the development of invasive cervical cancer. International Journal of Cancer, 49: 335-340, 1991.

Hildesheim, A., Schiffman, M., Scott, D.R., Marti, D., Kissner, T., Sherman, M.E., Glass, A.G., Manos, M.M., Lorincz, A.T., Kurman, R.J., Buckland, J., Rush, B.B., and Carrington, M. Human leukocyte antigen class I/II alleles and development of human papillomavirus-related cervical neoplasia: Results from a case-control study conducted in the United States. Cancer Epidemiology, Biomarkers and Prevention, 7: E1035-1041, 1998.

Hildesheim, A., Hadjimichael, O., Schwartz, P., Wheeler, C., Barnes, W., Lowell, D.M., Willet, J., and Schiffman, M. Risk factors for rapid-onset cervical cancer. American Journal of Obstetrics and Gynecology, 180(3): 571-577, 1999.

Hills, E., and Laverty, C. Electronmicroscopic detection of papillomavirus particles in selected koilocytotic cells in a routine cervical smear. Acta Cytologica, 23: 53-56, 1979.

Ho, L., Chan, S.Y., Chow, V., Ching, T., Tay, S.K., Villa, L.L., and Bernard, H.U. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree. Journal of Clinical Microbiology, 29(9): 1765-1772, 1991.

Ho, L., Chan, S.Y., Burk, R.D., Das, B.C., Fujinaga, K., Icenogle, J.P., Kahn, T., Kiviat, N., Lancaster, W., Mavromara-Nazos, P., Labropoulou, V., Mitrani-Rosenbaum, S., Norrild, B., Pillai, M.R., Stoerker, J., Syrjaenen, K., Syrjaenen, S., Tay, S.K., Villa, L.L., Wheeler, C.M., Williamson, A.L., and Bernard, H.U. The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. Journal of Virology, 67(11): 6413-6423, 1993a.

Ho, L., Tay, S.K., Chan, S.Y., and Bernard, H.U. Sequence variants of human papillomavirus type 16 from couples suggest sexual transmission with low infectivity and polyclonality in genital neoplasia. The Journal of Infectious Diseases, 168: 803-809, 1993b.

Ho YF, Bierman R, Beardsley L, Chang CJ and Burk RD Natural history of cervicovaginal papillomavirus infection in young women. The New England Journal of Medicine, 338(7): 423-428, 1998.

Hopkins, M.P., and Morley, G.W. A comparison of adenocarcinoma and squamous cell carcinoma of the cervix. Obstetrics and Gynecology, 77: 912-917, 1991.

Hoppe-Seyler, F., and Butz, K. Cellular control of human papillomavirus oncogene transcription. Molecular Carcinogenesis. 10: 134-141, 1994.

Hou, S.Y., Wu, S.Y., Zhou, T., Thomas, M.C., and Chiang, C.M. Alleviation of human papillomavirus E2-mediated transcriptional repression via formation of a TATA binding protein (or TFIID)-TFIIB-RNA polymerase II-TFIIF preinitiation complex. Molecular and Cellular Biology, 20(1): 113-125, 2000.

Howley, P.M. *Papillomaviridae*: The viruses and their replication. *In:* Fields, B.N., Knipe, D.M., Howley, P.M. (eds.), Virology, volume 2. pp 2045-2076, Lippincott-Raven, Philadelphia, USA, 1996.

Howley-Nelson, P., Vousden, K.H., Hubbert, N.L., Lowy, D.R., and Schiller, J.T. HPV 16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. The EMBO Journal, 8: 3905-3910, 1989.

Hu, X., Guo, Z., Tianyun, O., Pontén, F., Wilander, E., Andersson, S., and Pontén, J. HPV typing and HPV 16 E6-sequence variations in synchronous lesions of cervical squamous-cell carcinoma from Swedish patients. International Journal of Cancer, 83: 34-37, 1999.

Hudson, J.B., Bedell, M.A., McCance, D.J., and Laiminis, L.A. Immortalization and altered differentiation of human keratinocytes *in vitro* by the E6 and E7 open reading frames of human papillomavirus type 18. Journal of Virology, 64(2): 519-526, 1990.

Icenogle, J.P., Clancey, K.A., and Lin, S.Y. Sequence variation in the capsid protein genes of human papillomavirus type 16 and 31. Virology, 214(2): 664-669, 1995.

Imai, T., Matsuda, K., Shimojima, T., Hashimoto, T., Matsuhiro, Y., Kitamoto, T., Sugita, A., Suzuki, K., Matsumoto, H., Masushige, S., Nogi, Y., Muramatsu, M., Handa, H., and Kato, S. ERC-55, a binding protein for the papilloma virus E6 oncoprotein, specifically interacts with vitamin D receptor among nuclear receptors. Biochemical and Biophysical Research and Communication, 233: 765-769, 1997.

Ip, E.W., Collins, R.J., Srivastava, G., Cheung, A.N., and Wong, L.C. Human papillomavirus and its prognostic significance in invasive carcinoma of the cervix in young patients. International Journal of Gynecological Pathology, 11(4):258-265, 1992.

Ishiji, T., Lace, M.J., Parkkinen, S., Anderson, R.D., Haugen, T.H., Cripe, T.P., Xiao, J.H., Davidson, I., Chambon, P., and Turek, L.P. Transcriptional enhancer factor (TEF)-1 and its cell-specific co-activator activate human papillomavirus-16 E6 and E7 oncogene transcription in keratinocytes and cervical carcinoma cells. The EMBO Journal, 11(6): 2271-2281, 1992.

Jameson, B.A., and Wolf, H. The antigenic index: A novel algorithm for predicting antigenic determinants. Computer Applications in the Biosciences. 4(1): 181-186, 1988.

Janerich, D.T., Hadjimichael, O., Schwartz, P.E., Lowell, D.M., Meigs, J.W., Merino, M.J., Flannery, J.T., and Polednak, A.P. The screening histories of women with invasive cervical cancer, Connecticut. American Journal of Public Health, 85: 791-794, 1995.

Janin, J., and Wodak, S. Conformation of amino acid side-chains in proteins. Journal of Molecular Biology, 125(3): 357-386, 1978.

Jelfs P. Cervical Cancer in Australia, Australian Institute of Health and Welfare: Cancer Series No. 3, AIHW, Canberra, Australia, 1995.

Jenison, S.A., Yu, X.P., Valentine, J.M., Koutsky, L.A., Christiansen, A.E., Beckmann, A.M., and Galloway, D.A. Evidence of prevalent genital-type human papillomavirus infections in adults and children. Journal of Infectious Diseases, 162: 60-69, 1990.

Jeon, S., Allen-Hoffman, B.L., and Lambert, P.F. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. Journal of Virology, 69: 2989-2997, 1995.

Jeon, S., and Lambert, P.F. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: Implications for cervical carcinogenesis. Proceedings of the National Academy of Sciences of the United States of America, 92: 1654-1658, 1995.

Jewers, R.J., Jildebrandt, P., Ludlow, J.W., Kell, B., and McCance, D.J. Regions of papillomavirus type 16 E7 oncoprotein required for immortalization of human keratinocytes. Journal of Virology, 66: 1329-1335, 1992.

Jones, D.L., Alani, R.M., and Münger, K. The human papillomavirus E7 oncoprotein can uncouple cellular differentiation and proliferation in human keratinocytes by abrogating $p21^{Cip1}$-mediated inhibition of cdk2. Genes and Development, 11: 2101-2111, 1997.

Jones, K.A., Kadonga, J.T., Rosenfeld, P.J., Keey, T.J., Tjian, R. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell, 48: 79-89, 1987.

Josefsson, A.M., Magnusson, P.K.E., Ylitalo, N., Sorensen, P., Qwarforth-Tubbin, P., Andersen, P.K., Melbye, M., Adami, H.O., and Gyllensten, U.B. Viral load of human papilloma virus 16 as a

determinant for development of cervical carcinoma in situ: A nested case-control study. The Lancet, 355(9222): 2189-2193, 2000.

Kadish, A.S., Ho, G.Y.F., Burk, R.D., Wang, Y., Romney, S.L., Ledwidge, R., and Angeletti, R.H. Lymphoproliferative responses to human papillomavirus (HPV) type 16 proteins E6 and E7: Outcome of HPV infection and associated neoplasia. Journal of the National Cancer Institute, 89(17): 1258-1293, 1997.

Kalantari, M., Karlsen, F., Kristensen, G., Holm, R., Hagmar, B., and Johansson, B. Disruption of the E1 and E2 reading frames of HPV 16 in cervical carcinoma is associated with poor prognosis. International Journal of Gynecological Pathology, 17; 146-153, 1998.

Kammer, C., Warthorst, U., Torrez-Martinez, N., Wheeler, C.M., and Pfister, H. Sequence analysis of the long control region of human papillomavirus type 16 variants and functional consequences for P97 promoter activity. Journal of General Virology, 81: 1975-1981, 2000.

Kanda, T., Watanabe, S., and Yoshiike, K. Immortalization of primary rat cells by human papillomavirus type 16 subgenomic DNA fragments controlled by the SV40 promoter. Virology, 165, 321-325, 1988.

Kanda, T., Watanabe, S., Zanma, S., Sato, H., Furuno, A., and Yoshiike, K. Human papillomavirus type 16 E6 proteins with glycine substitution for cysteine in the metal-binding motif. Virology, 185: 536-543, 1991.

Kast, W.M., Brandt, R.M.P., Sidney, J., Drijfhout, J.W., Kubo, R.T., Grey, H.M., Melief, C.J.M., and Sette, A. Role of HLA-A motifs in identification of potential CTL epitopes in human papillomavirus type 16 E6 and E7 proteins. Journal of Immunology, 152: 3904-3912, 1994.

Kaufman, R.H. Atypical squamous cells of undetermined significance and low-grade squamous intraepithelial lesions: Diagnostic criteria and management. American Journal of Obstetrics and Gynecology, 175: 1120-1128, 1996.

Kaur, P., and McDougall, J.K. Characterization of primary human keratinocytes transformed by human papilloamvirus type 18. Journal of Virology, 62: 1917-1924, 1988.

Kennedy, I.M., Haddow, J.K., and Barklie Clements, J. A negative element in the human papillomavirus type 16 genome acts at the level of late mRNA stability. Journal of Virology, 65(4): 2093-2097, 1991.

Kersemaekers, A.M.F., Fleuren, G.J., Kenter, G.G., van den Broek, L.J.C.M., Uljee, S.M., Hermans, J., and van de Vijver, M.J. Oncogene alterations in carcinomas of the uterine cervix: Overexpression of the epidermal growth factor receptor is associated with poor prognosis. Clinical Cancer Research, 5: 577-586, 1999.

Khare, S., Kumar, K.U., Tang, S.C., Pater, M.M., and Pater, A. Up-regulation of hormone response of human papillomavirus type 16 expression and increased DNA-protein binding by consensus mutations of viral glucocorticoid response elements. Journal of Medical Virology, 50: 254-262, 1996.

Kim J., and Shapiro, D.J. In simple synthetic promoters YY1-induced DNA bending is important in transcription activation and repression. Nucleic Acids Research, 24(21): 4341-4348, 1996.

Kim, J.W., Kim, Y.T., Kim, D.K., Song, C.H., and Lee, J.W. Expression of epidermal growth factor receptor in carcinoma of the cervix. Gynecologic Oncology, 60: 283-287, 1996.

Kinoshita, M., Katsuragi, K., Shin, S., and Aono, T. Amplification and point-mutation alterations of oncogenes in patients with cervical cancer-associated human papillomavirus. Biomedical Research, 14(6): 403-410, 1993.

Kirnbauer, R., Booy, F., Cheng, N., Lowy, D.R., and Schiller, J.T. Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. Proceedings of the National Academy of Sciences of the United States of America, 89: 12180-12184, 1992.

Kirnbauer, R., Taub, J., Greenstone, H., Roden, R., Durst, M., Gissmann, L., Lowy, D.R., and Schiller, J.T. Efficient self-assembly of human papillomavirus type-16 L1 and L1-L2 into virus-like particles. Journal of Virology, 67: 6929-6936, 1993.

Kirnbauer, R., Hubbert, N.L., Wheeler, C.M., Becker, T.M., Lowy, D.R., and Schiller, J.T. A virus-like particle enzyme-linked immunosorbent assay detects serum antibodies in a majority of women infected with human papillomavirus type 16. Journal of the National Cancer Institute, 86(7): 494-499, 1994.

Kiviat, N. Natural history of cervical neoplasia: Overview and update. American Journal of Obstetrics and Gynecology, 175(4 Part 2): 1099-1104, 1996.

Kjaer, S.K., van sen Brule, A.J.C., Bock, J.E., Poll, P.A., Engholm, G., Sherman, M.E., Walboomers, J.M.M., and Meijer, C.J.L.M. Human papillomavirus – The most significant risk determinant of cervical intraepithelial neoplasia. International Journal of Cancer, 65: 601-606, 1996.

Kjellberg, L., Hallmans, G., Ahren, A.M., Johansson, R., Bergman, F., Wadell, G., Angstrom, T., and Dillner, J. Smoking, diet, pregnancy and oral contraceptive use as risk factors for cervical intra-epithelial neoplasia in relation to human papillomavirus infection. British Journal of Cancer, 82(7): 1332-1338, 2000.

Kjellgren, O. Clinical invasive carcinoma of cervix: Place of radiotherapy as primary treatment. *In:* Coppleson, M. (ed.) Gynaecologic Oncology Second Edition. pp 673-695, Churchill Livingstone, Edinburgh, UK, 1992.

Klaes, R., Woerner, S.M., Ridder, R., Wentzensen, N., Deurst, M., Schneider, A., Lotz, B., Melsheimer, P., and von Knebel Doeberitz, M. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Research, 59: 6132-6136, 1999.

Kleiner, E., Dietrich, W., and Pfister, H. Differential regulation of papilloma virus early gene expression in transformed fibroblasts and carcinoma cell lines. The EMBO Journal, 5: 1945-1950, 1986.

Klingelhutz, A.J., Foster, S.A., and McDougall, J.K. Telomerase activation by the E6 gene product of human papillomavirus type 16. Nature, 380: 79-82, 1996.

Kónya, J., Eklund, C., Geijersstam, V.,Yuan, F., Stuber, G., and Dillner, J. Identification of a cytotoxic T-lymphocyte epitope in the human papillomavirus type 16 E2 protein. Journal of General Virology, 78: 2615-2620, 1997.

Kook, J.K., Kim, J.H., and Min, B.M. Activity of human papillomavirus type 16 P97 promoter in immortal and tumorigenic human oral keratinocytes. International Journal of Oncology, 13: 765-771, 1998.

Koss, L.G., and Durfee, G.R. Unusual pattern of squamous epithelium of the uterine cervix: Cytologic and pathologic study of koilocytotic atypica. Annual NY Academy of Science, 63: 1245-1261, 1956.

Koss, L.G. The Papanicolaou test for cervical cancer detection: A triumph and a tradegy. Journal of the American Medical Association, 261: 737-743, 1989.

Koss, L.G. Cervical (Pap) smear: new directions. Cancer, 71(Suppl.): 1406-1412, 1993.

Koutsky, L.A., Galloway, D.A., and Holmes, K.K. Epidemiology of genital papillomavirus infection. Epidemiology Reviews, 10: 122-163, 1988.

Koutsky, L.A., Holmes, K.K., Critchlow, C.W., Stevens, C.W., Paavonen, J., Beckmann, A.M., DeRouen, T.A., Galloway, D.A., Vernon, D., and Kiviat, N.B. A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection. New England Journal of Medicine, 327: 1272-1278, 1992.

Koutsky, L. Epidemiology of genital human papillomavirus infection. American Journal of Medicine, 102: 3-8, 1997.

Kovelman, R., Bilter, G.K., Glezer, E., Tsou, A.Y., and Barbosa, M.S. Enhanced transcriptional activation by E2 proteins from the oncogenic human papillomaviruses. Journal of Virology, 70: 7549-7560, 1996.

Kovelman, R., Bilter, G.K., Roman, A., Brown, D.R., and Barbosa, M.S. Human papillomavirus type 6: Classification of clinical isolates and functional analysis of E2 proteins. Journal of General Virology, 80: 2445-2451, 1999.

Kozuka, T., Aoki, Y., Nakagawa, K., Ohtomo, K., Yoshikawa, H., Matsumoto, K., Yoshiike, K., and Kanda, T. Enhancer-promoter activity of human papillomavirus type 16 long control regions isolated from cell lines SiHa and CaSki and cervical cancer biopsies. Japanese Journal of Cancer Research, 91(3): 271-279, 2000.

Krajinovic, M., Lazic, J., Stanimirovic, B., Diklic, V., and Savic, A. The E2 region of HPV 16 in relation to different types of cervical lesions. Journal of Medical Virology, 41: 1-5, 1993.

Kristensen, G.B., Holm, R., Abeler, V.M., and Trope, C.G. Evaluation of the prognostic significance of cathepsin D, epidermal growth factor receptor, and c-erbB-2 in early squamous cell carcinoma. Cancer, 78: 433-440, 1996.

Kulke, R., Gross, G.E., and Pfister, H. Duplication of enhancer sequences in human papillomavirus 6 from condylomas of the mamilla. Virology, 173: 284-290, 1989.

Kurvinen, K., Yliskoski, M., Saarikoski, S., Syrjänen, K., and Syrjänen, S. Variants of the long control region of human papillomavirus type 16. European Journal of Cancer, 36: 1402-1410, 2000.

Kyo, S., Inoue, M., Nishio, Y., Nakanishi, K., Akira, S., Inoue, H., Yutsudo, M., Tanizawa, O., and Hakura, A. NF-IL6 represses early gene expression of human papillomavirus type 16 through binding to the noncoding region. Journal of Virology, 67(2): 1058-1066, 1993.

Kyo, S., Inoue, M., Hayasaka, N., Inoue, T., Yutsudo, M., Tanizawa, O., and Hakura, A. Regulation of early gene expression of human papillomavirus type 16 by inflammatory cytokines. Virology, 200: 130-139, 1994.

Kyo, S., Klumpp, D.J., Inoue, M., Kanaya, T., and Laimins, L.A. Expression of AP1 during cellular differentiation determines human papillomavirus E6/E7 expression in stratified epithelial cells. Journal of General Virology, 78: 401-411, 1997.

Kyte, J., and Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology, 157(1): 105-132, 1982.

LaGuardia, K.D., McGuinness, K., and Hunter, D. Cervical disease among HIV-infected women by immune status. VII International Conference on AIDS, Florence, Italy, Abstract MC 97, 1991.

LaGuardia, K.D. Other sexually transmitted disease: Cervical intraepithelial neoplasia. *In:* Johnson, M.A., and Johnstone, F.D. (eds.), HIV Infection in Women. pp 247-261, Churchill Livingstone, Edinburgh, UK, 1993.

Lambert, P.F. Papillomavirus DNA replication. Journal of Virology, 65(7): 3417-3420, 1991.

Lanham, S., Campbell, I., Watt, P., and Gornall, R. p53 polymorphism and risk of cervical cancer. The Lancet, 352: 1631, 1998.

Latchman, D. Gene regulation and cancer. *In:* Latchman, D. (ed.), Gene Regulation. A Eukaryotic Perspective. pp 231-256, Unwin Hyman, London, UK, 1990.

Latchman, D.S. Transcriptional regulation of viral gene expression. Reviews in Medical Virology, 3: 115-122, 1993.

Latchman, D.S. Regulation of DNA virus transcription by cellular POU family transcription factors. Reviews in Medical Virology, 9: 31-38, 1999.

Laverty, C.R., Booth, N., Hills, E., Cossart, Y., and Wills, E.J. Noncondylomatous wart virus infection of the postmenopausal cervix. Pathology, 10(4): 373-378, 1978.

Le, J.Y., and Defendi, V. A viral-cellular junction fragment from a human papilloamvirus type 16-positive tumour is competent in transformation of NIH 3T3 cells. Journal of Virology, 62: 4420-4426, 1988.

Le Buanec, H., Lachgar, A., Anna, R.D., Zagury, J.F., Bizzini, B., Bernard, J., Ittele, D., Hallez, S., Giannouli, C., Burny, A., and Zagury, D. Induction of cellular immunosupression by the human papillomavirus type 16 E7 oncogenic protein. Biomedicine and pharmacotherapy, 53: 323-328, 1999.

Lechner, M.S., and Laimins, L.A. Inhibition of p53 binding by human papillomaviurs E6 proteins. Journal of Virology, 68(7): 4262-4273, 1994.

Lee, D., Lee, B., Kim, J., Kim, D.W., and Choe, J. cAMP response element-binding protein-binding protein binds to human papillomavirus E2 protein and activates E2-dependent transcription. The Journal of Biological Chemistry, 275(10): 7045-7051, 2000.

Leechanachai, P., Banks, L., Moreau, F., and Matlashewski, G. The E5 gene from human papillomavirus type 16 is an oncogene which enhances growth factor-mediated signal transduction to the nucleus. Oncogene, 7: 19-25, 1992.

Lehn, H., Villa, L.L., Marziona, F., Hilgarth, M., Hillemans, H.G., and Sauer, G. Physical state and biological activity of human papillomavirus genomes in precancerous lesions of the female genital tract. Journal of General Virology, 69: 187-196, 1988.

Lehninger, A.L., Nelson, D.L., and Cox, M.M. Recombinant DNA Technology. *In:* Principles of Biochemistry Second Edition. pp 984-1013, Worth Publishers, New York, USA, 1993.

Liaw, K.L., Glass, A.G., Manos, M.M., Greer, C.E., Scott, D.R., Sherman, M., Burk, R.D., Kurman, R.J., Wacholder, S., Rush, B.B., Cadell, D.M., Lawler, P., Tabor, D., and Schiffman, M. Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions. Journal of the National Cancer Institute, 91(11): 954-960, 1999.

Lis, J.T. Size fractionation of DNA fragments by polyethylene glycol induced precipitation. Methods in Enzymology, 65: 347-353, 1980.

Liu, Z., Ghai, J., Ostrow, R.S., and Faras, A.J. The expression levels of the human papillomavirus type 16 E7 correlate with its transforming potential. Virology, 207: 260-270, 1995.

Lombard, I., Vincent-Salomon, A., Validire, P., Zafrani, B., de la Rochefordière, A., Clough, K., Favre, M., Pouillart, P., and Sastre-Garau, X. Human papillomavirus genotype as a major determinant of the course of cervical cancer. Journal of Clinical Oncology, 16(8): 2613-2619, 1998.

Londesborough, P., Ho, L., Terry, G., Cuzick, J., Wheeler, C.M., and Singer, A. Human papillomavirus genotype as a predictor of persistence and development of high-grade lesions in women with minor cervical abnormalities. International Journal of Cancer, 69: 364-368, 1996.

Lörincz, A.T., Lancaster, W.D., and Temple, G.F. Cloning and characterization of the DNA of a new human papillomavirus from a woman with dysplasia of the uterine cervix. Journal of Virology, 58(1): 225-259, 1986.

Lörincz, A.T., Reid, R., Jenson, A.B., Greenberg, M.D., Lancaster, W., and Kurman, R.J. Human papillomavirus infection of the cervix: Relative risk associations of 15 common anogenital types. Obstetrics and Gynecology, 79: 328-337, 1992.

Lörincz, A.T. Human Papillomaviruses. *In:* Lennette, E.H., and Smith, T.F. (eds.), Laboratory Diagnosis of Viral Infections. Third Edition. pp 635-663, Marcel Dekker, Inc. New York, USA, 1999.

Lowy, D.R., Kirnbauer, R., and Schiller, J.T. Genital human papillomavirus infection. Proceedings of the National Academy of Sciences of the United States of America, 91: 2436-2440, 1994.

Lowy, D.R. HPV infection: future prospects. Clinics in Dermatology, 15(2): 299-302, 1997.

Lungu, O., Sun, X.W., Felix, J., Richart, R.M., Silverstein, S., and Wright, T.C. Relationship of human papillomavirus type to grade of cervical intraepithelial neoplasia. Journal of the American Medical Association, 267: 2493-2496, 1992.

Luostarinen, T., Geijersstam, V.A.F., Bjorge, T., Eklund, C., Hakama, M., Hakulinen, T., Jellum, E., Koskela, P., Paavonen, J., Pukkala, E., Schiller, J.T., Thoresen, S., Youngman, L.D., Dillner, J., and Lehtinen, M. No excess risk of cervical carcinoma among women seropositive for both HPV 16 and HPV 6/11. International Journal of Cancer, 80: 818-822, 1999.

Lutzner MA, Blanchet-Bardon C and Orth G. Clinical observations, virologic studies, and treatment trials in patients with epidermodysplasia verruciformis, a disease induced by specific human papillomaviruses. Archives of Dermatology Research, 271:83-90, 1984.

Lutzner, M. Papillomavirus lesions in immunodepression and immunosuppression. Clinics in Dermatology, 3: 165-169, 1985.

Luxton, J.C., Rowe, A.J., Cridland, J.C., Coletart, T., Wilson, P., and Shepherd, P.S. Proliferative T cell responses to the human papillomavirus type 16 E7 protein in women with cervical dysplasia and cervical carcinoma and in healthy individuals. Journal of General Virology, 77: 1585-1593, 1996.

Luxton, J.C., Rose, R.C., Coletart, T., Wilson, P., and Shepherd, P.S. Serological and T-helper responses to human papillomavirus type 16 L1 in women with cervical dysplasia or cervical carcinoma and in healthy controls. Journal of General Virology, 78: 917-923, 1997.

Luxton, J., Mant, C., Greenwood, B., Derias, N., Nath, R., Shepherd, P., and Cason, J. HPV16 E6 oncogene variants in women with cervical intraepithelial neoplasia. Journal of Medical Virology, 60: 337-341, 2000.

Madrigal, M., Janicek, M.F., Sevin, B.U., Perras, J., Estape, R., Penalver, M., and Averette, H.E. *In vitro* antigene therapy targeting HPV-16 E6 and E7 in cervical carcinoma. Gynecologic Oncology, 64(1): 18-25, 1997.

Major, S. Human papillomavirus classified as carcinogenic. British Medical Journal, 313: 70, 1996.

Mandel, M., and Higa, A. Calcium-dependent bacteriophage DNA infection. Journal of Molecular Biology, 53(1): 159-162, 1970.

Mango, L.J. Reducing false negatives in clinical practice: The role of neural network technology. American Journal of Obstetrics and Gynecology, 175(4 Part 2): 1114-1119, 1996.

Manos, M.M., Kinney, W.K., Hurley, L.B., Sherman, M.E., Shieh-Ngai, J., Kurman, R.J., Ransley, J.E., Fetterman, B.J., Hartinger, J.S., McIntosh, K.M., Pawlick, G.F., and Hiatt, R.A. Identifying women with cervical neoplasia. The Journal of the American Medical Association, 291(17): 1605-1610, 1999.

Mantovani, F., and Banks, L. The interaction between p53 and papillomaviruses. Seminars in Cancer Biology, 9: 387-395, 1999.

Matlashewski, G., Schneider, J., Banks, N., Murray, A., and Crawford, L. Human papillomavirus type 16 DNA cooperates with activated ras in transforming primary cells. The EMBO Journal, 6(6): 1741-1746, 1987.

May, M., Dong, X.P., Beyer-Finkler, E., Stubenrauch, F., Fuchs, P.G., and Pfister, H. The E6/E7 promoter of extrachromosomal HPV 16 DNA in cervical cancers escapes from cellular repression by mutation of target sequences for YY1. The EMBO Journal, 13(6): 1460-1466, 1994.

McBride, A.A., Romanczuk, H., and Howley, P.M. The papillomavirus E2 regulatory proteins. The Journal of Biological Chemistry, 266(28): 18411-18414, 1991.

McBride, A and Myers, G. The E2 Proteins. *In:* Myers, G., Halpern, A., Baker, C., McBride, A., Wheeler, C., and Doorbar, J. (eds.), Human Papillomaviruses. pp III15-III57, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, 1996.

McGlennen, R.C., Ghai, J., Ostrow, R.S., LaBresh, K., Schneider, J.F., and Faras, A.J. Cellular transformation by a unique isolate of human papillomavirus type 11. Cancer Research, 52: 5872-5878, 1992.

McIntyre, M.C., Frattini, S.R., Grossman, S.R., and Laimins, L.A. Human papillomavirus type 18 E7 protein requires intact Cys-X-X-Cys motifs for zinc binding, dimerization, and transformation but not for Rb binding. Journal of Virology, 67(6): 3142-3150, 1993.

Meigs, J.V. Introduction. *In:* Meigs, J.V. (ed.), Surgical Treatment of Cancer of the Cervix. pp 1-24, Grune and Stratton, New York, 1954.

Meijer, C.J.L.M., Helmerhorst, T.J.M., Rozendaal, L., van der Linden, J.C., Voorhorst, F.J., and Walboomers, J.M.M. HPV typing and testing in gynecological pathology: Has the time come? Histopathology, 33: 83-86, 1998.

Meisels, A., Fortin, R., and Roy, M. Condylomatous lesions of the cervix and vagina I. Cytological Patterns. Acta Cytologica, 20: 505-509, 1976.

Meisels, A., Fortin, R., and Roy, M. Condylomatous lesions of the cervix II. Cytologic, colposcopic and histopathologic study. Acta Cytologica, 21(3): 379-390, 1977.

Meisels A. Cytological diagnosis of human papillomavirus. Influence of age and pregnancy state. Acta Cytolgica, 36: 480-482, 1992.

Meissner, J. Sequencing errors in reference HPV clones. *In:* Myers, G., Baker, C., Münger, K., Sverdrup, F., McBride, A., and Bernard, H.U. (eds.), Human Papillomaviruses. pp III110-III123, Los Alamos, New Mexico, USA, 1997.

Melero, I., Singhal, M.C., McGowan, P., Haugen, H.S., Blake, J., Hellström, K.E., Yang, G., Clegg, C.H., and Chen, L. Immunological ignorance of an E7-encoded cytolytic T-lymphocyte epitope in transgenic mice expressing the E7 and E6 oncogenes of human papillomavirus type 16. Journal of Virology, 71(5): 3998-4004, 1997.

Meyers, C. Organotypic (raft) epithelial tissue culture system for the differentiation-dependent replication of papillomaviruses. Methods in Cell Science, 18: 201-210, 1996.

Meyer, G., and Androphy, E. The E6 protein. *In:* Meyers, G., Delius, H., Icenogel, J., Bernard, H.U., Baker, C., Halpern, A., and Wheeler, C. (eds.), Human Papillomaviruses. pp III47-III57, Los Alamos, New Mexico, USA, 1995.

Minaguchi, T., Kanamori, Y., Matsushima, M., Yoshikawa, H., Taketani, Y., and Nakamura, Y. No evidence of correlation between polymorphism at codon 72 of p53 and risk of cervical cancer in Japanese patients with human papillomavirus 16/18 infection. Cancer Research, 58: 4585-4586, 1998.

Mincheva, A., Gissmann, L., and zur Hausen, H. Chromosomal integration sites of human papillomavirus DNA in three cervical cancer cell lines mapped by *in situ* hybridization. Medical Microbiology and Immunology, 176(5): 245-256, 1987.

Mitrani-Rosenbaum, S., Tsvieli, R., and Tur-Kaspa, R. Oestrogen stimulates differential transcription of human papillomavirus type 16 in SiHa cervical carcinoma cells. Journal of General Virology, 70: 2227-2232, 1989.

Mittal, R., Pater, A., and Pater, M.M. Multiple human papillomavirus type 16 glucocorticoid response elements functional for transformation, transient expression, and DNA-protein interactions. Journal of Virology, 67(9): 5656-5659, 1993.

Moy, R.L., Eliezri, Y.D., Nuovo, G.J., Zitelli, J.A., Bennett, R.G., and Silverstein, S. Human papillomavirus type 16 DNA in periungual squamous cell carcinomas. Journal of the American Medical Association, 261: 2669-2673, 1989.

Muir, C., Waterhouse, J., Mack, T., et al. Cancer incidence in five continents. IARC Publications, #88. Lyon, France: International Agency for Research on Cancer, 1987.

Müller, M., Viscidi, R.P., Sun, Y., Guerrero, E., Hill, P.M., Shah, F., Bosch, X., Muñoz, N., Gissman, L., and Shah, K.V. Antibodies to HPV-16 E6 and E7 proteins as markers for HPV-16-associated invasive cervical cancer. Virology, 187: 508-514, 1992.

Münger, K., Werness, B., Dyson, N., Phelps, W., Harlow, E., and Howley, P. Complex formation of the human papillomavirus E7 proteins with the retinoblastoma tumour suppressor gene product. The EMBO Journal, 8(13): 4099-4105, 1989a.

Münger, K., Phelps, W.C., Bubb, V., Howley, P.M., and Schlegel, R. The E6 and E7 genes of human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. Journal of Virology, 63(10): 4417-4421, 1989b.

Münger, K., Yee, C.L., Phelps, W.C., Pietenpol, J.A., Moses, H.L., and Howley, P.M. Biochemical and biological differences between E7 oncoproteins of the high- and low-risk human papillomavirus types are determined by amino-terminal sequences. Journal of Virology, 65(7): 3943-3948, 1991.

Münger, K., and Phelps, W.C. The human papillomavirus E7 protein as a transforming and trans-activating factor. Biochemica et Biophysica Acta, 5, 1155(1): 111-123, 1993.

Muñoz, N., and Bosch, FX. HPV and cervial neoplasia: Review of case-control and cohort studies. *In:* Muñoz, N., Bosch, FX., and Shah, K.V (eds.), Epidemiology of cervical cancer and human papillomaviruses, pp 251-261. IARC Scientific Publication Number 119, IARC, Lyon, France, 1992.

Muñoz, N., Bosch, FX., de Sanjosè, S., Tafur, L., Izarzugaza, I., Gili, M., Viladiu, P., Navarro, C., Martos, C., Ascunce, N., Gonzales, L.C., Kaldor, J.M., Guerrero, E., Lörincz, A., Santamaria, M., Alonso de Ruiz, P., Aristizabal, N., and Shah, K. The causal link between human papillomavirus and invasive cervical cancer, a population based study in Columbia and Spain. International Journal of Cancer, 52(5): 743-749, 1992a.

Muñoz, N., Bosch, FX., Shah, K.V., et al. The epidemiology of human papillomaviruses and cervical cancer (IARC scientific publication no. 119). Lyon, International Agency for Research on Cancer, pp 3-23, 1992b.

Murakami, M., Gurski, K.J., and Steller, M.A. Human papillomavirus vaccines for cervical cancer. Journal of Immunotherapy, 22(3): 212-218, 1999.

Myers, E.R., McCrory, D.C., Nanda, K., Bastian, L., and Matchar, D.B. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. American Journal of Epidemiology, 151(12): 1158-1171, 2000.

Myint, S.H., and Shaw, P.A.V. Laboratory identification of human papillomavirus Infection. *In:* Mindel, A. (ed.), Genital Warts - Human papillomavirus infection. pp 35-52, Edward Arnold- A member of the Hodder-Headline Group, London, UK, 1995.

Nagai, N., Oshita, T., Murakami, J., and Ohama, K. Semiquantitative analysis of telomerase activity in cervical cancer and precancerous lesions. Oncology Reports, 6(2): 325-328, 1999.

Nakshatri, H., Pater, M., and Pater, A. Ubiquitous and cell-type-specific protein interactions with human papillomavirus type 16 and type 18 enhancers. Virology, 178: 92-103, 1990.

Natesan, S., and Gilman, M.Z. DNA bending and orientation-dependent function of YY1 in the c-*fos* promoter. Genes and Development, 7: 2497-2509, 1993.

National Cervical Screening Program, 1998. http://www.cervical screen.health.gov.au

National Health and Medical Research Council (NH&MRC) of Australia. Cervical Cancer Screening: Interval of Screening Recommendations.
http://www.health.gov.au/nhmrc/publicat/statement/wh4stat.htm. 1991.

National Institutes of Health Consensus Development Conference Statement on Cervical Cancer (1997) Consensus Statement, Gynecologic Oncology, 66: 351-361, 1997.

Nawa, A., Nishiyama, Y., Kobayashi, T., Wakahara, Y., Okamoto, T., Kikkawa, F., Suganuma, N., Goto, S., Kuzuya, K., and Tomoda, Y. Association of human leukocyte antigen-B1*03 with cervical cancer in Japanese women aged 35 years and younger. Cancer, 75: 518-521, 1995.

New South Wales Cancer Council. Cancer in New South Wales, Incidence and Mortality 1996. http://www.nswcc.org.au/pages/ccic/ccr/1996stat/cervix/htm, 1996.

Nindl, I., Rindfleisch, K., Lotz, B., Schneider, A., and Durst, M. Uniform distribution of HPV 16 E6 and E7 variants in patients with normal histology, cervical intraepithelial neoplasia and cervical cancer. International Journal of Cancer, 82: 203-207, 1999.

Nindl, I., Zumbach, K., Pawlita, M., Teller, K., Schneider, A., and Dürst, M. Absence of antibody against human papillomavirus type 16 E6 and E7 in patients with cervical cancer is independent of sequence variations. The Journal of Infectious Diseases, 181: 1764-1767, 2000.

Nurnberg, W., Artuc, M., Vorbruggen, G., Kalkbrenner, F., Moelling, K., Czarnetzki, B.M., and Schadendorf, D. Nuclear proto-oncogene products transactivate the human papillomavirus type 16 promoter. British Journal of Cancer, 71: 1018-1024, 1995.

Nuss Parker, J., Zhao, W., Askins, K.J., Broker, T.R., and Chow, L.T. Mutational analyses of differentiation-dependent human papillomavirus type 18 enhancer elements in epithelial raft cultures of neonatal foreskin keratinocytes. Cell Growth and Differentiation, 8: 751-762, 1997.

O'Connor, M., and Bernard, H.U. Oct-1 activates the epithelial-specific enhancer of human papillomavirus type 16 via a synergistic interaction with NFI at a conserved composite regulatory element. Virology, 207: 77-88, 1995.

O'Connor, M., Chan, S.Y., and Bernard, H.U. Transcription factor binding sites in the long control region of genital HPVs. *In:* Myers, G., Delius, H., Icenogel, J., Bernard, H.U., Baker, C., Halpern, A., Wheeler, C. (eds.), Human Papillomaviruses. pp III21-III40, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, 1995.

O'Connor, M.J., Tan, S.H., Tan, C.H., and Bernard, H.U. YY1 represses human papillomavirus type 16 transcription by quenching AP-1 activity. Journal of Virology, 70(10): 6529-6539, 1996.

O'Connor, M.J., Stünkel, W., Zimmermann, H., Koh, C.H., and Bernard, H.U. A novel YY1-independent silencer represses the activity of the human papillomavirus type 16 enhancer. Journal of Virology, 72(12): 10083-10092, 1998.

O'Connor, M., Stünkel, W., Koh, C.H., Zimmermann, H., and Bernard, H.U. The differentiation-specific factor CDP/Cut represses transcription and replication of human papillomaviruses through a conserved silencing element. Journal of Virology, 74(1): 401-410, 2000.

Odunski, K., Terry, G., Ho, L., Bell, J., Cuzick, J, and Ganesan, T.S. Susceptibility to human papillomavirus-associated cervical intra-epithelial neoplasia is determined by specific HLA DR-DQ alleles. International Journal of Cancer, 67: 595-602, 1996.

Oelze, I., Kartenbeck, J., Crusius, K., and Alonso, A. Human papillomavirus type 16 E6 protein affects cell-cell communication in an epithelial cell line. Journal of Virology, 69: 4489-4492, 1995.

Ong, C.K., Chan, S.Y., Campo, M.S., Fujinaga, K., Mavromara-Nazos, M., Labropoulou, V., Pfister, H., Tay, S.K., ter Meulen, J., Villa, L.L., and Bernard, H.U. Evolution of human papillomavirus type 18: An ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. Journal of Virology, 67(11): 6424-6431, 1993.

Papanicolaou, G.N., and Traut, H.F. The diagnostic value of vaginal smears in carcinoma of the uterus. American Journal of Obstetrics and Gynecology, 42(2): 193-206, 1941.

Papavassiliou, A.G. Transcription factors. The New England Journal of Medicine, 332(1): 45-47, 1995.

Park, J.S., Hwang, E.S., Park, S.N., Ahn, H.K., Um, S.J., Kim, C.J., Kim, S.J., and Namkoong, S.E. Physical state and expression of HPV genes in cervical cancers. Gynecologic Oncology, 65: 121-129, 1997.

Park, J.S., Hwang, E.S., Lee, C.J., Kim, C.J., Rha, J.G., Kim, S.J., Namkoong, S.E., and Um, S.J. Mutational and functional analysis of HPV-16 URR derived from Korean cervical neoplasia. Gynecologic Oncology, 74: 23-29, 1999.

Park, P., Copeland, W., Yang, L., Wang, T., Botchan, M.R., and Mohr, I.J. The cellular DNA polymerase α-primase is required for papillomavirus DNA replication and associates with the viral E1

helicase. Proceedings of the National Academy of Science of the United States of America, 91: 8700-8704, 1994.

Parker, S.L., Tong, T., and Bolden, S. Cancer statistics 1996. A Cancer Journal for Clinicians. 65: 5-27, 1996.

Patel, D., Huang, S.M., Baglia, L.A., and McCance, D.J. The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. The EMBO Journal, 18(18): 5061-5072, 1999.

Pater, A., Bayatpour, M., and Pater, M.M. Oncogenic transformation by human papillomavirus type 16 deoxyribonucleic acid in the presence of progesterone or progestins from oral contraceptives. American Journal of Obstetrics and Gynecology. 162(4): 1099-1103, 1990.

Pater, M.M., and Pater, A. Human papillomavirus types 16 and 18 sequences in carcinoma cell lines of the cervix. Virology, 145(2): 313-318, 1985.

Pater, M.M., Hughes, G.A., Hyslop, D.E., Nakshatri, H., and Pater, A. Glucocorticoid-dependent oncogenic transformation by type 16 but not type 11 human papilloma virus DNA. Nature, 335: 832-835, 1988.

Pattison, S., Skalnik, D.G., and Roman, A. CCAAT displacement protein, a regulator of differentiation-specific gene expression, binds a negative regulatory element within the 5' end of the human papillomavirus type 6 long control region. Journal of Virology, 71(3): 2013-2022, 1997.

Penn, I. Cancers of the anogenital region in renal transplant recipients. Analysis of 65 cases. Cancer, 58: 611-616, 1986.

Peto, M., Tolle-Ersü, I., Kreysch, H.G., and Klock, G. Epidermal growth factor induction of human papillomavirus type 16 E6/E7 mRNA in tumour cells involves two AP-1 binding sites in the viral enhancer. Journal of General Virology, 76: 1945-1958, 1995.

Pfeifer, D., Stellwag, B., Pfeifer, A., Borlinghaus, P., Meier, W., and Scheidel, P. Clinical implications of the epidermal growth factor receptor in the squamous cell carcinoma of the uterine cervix. Gynecologic Oncology, 33: 146-150, 1989.

Pfister, H., and Fuchs, P.G. Papillomaviruses: particles, genome organisation and proteins. *In:* Syrjänen, K., Gissman, L., Koss, L.G. (eds.), Papillomaviruses and Human Disease. pp 1-18, Springer-Verlag, Berlin, Germany, 1987.

Pfister, H., and Fuchs, P.G. Anatomy, taxonomy and evolution of the papillomaviruses. Intervirology, 37: 143-149, 1994.

Pfister, H. The role of human papillomavirus in anogenital cancer. Obstetrics and Gynecology Clinics of North America, 23(3): 579-595, 1996.

Phelps, W.C., Yee, C.L., Münger, K., and Howley, P.M. The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A. Cell, 53(4): 539-547, 1988.

Phelps, W.C., Münger, K., Yee, C.L., Barnes, J.A., and Howley, P.M. Structure-function analysis of the human papillomavirus type 16 E7 oncoprotein. Journal of Virology, 66(4): 2418-2427, 1992.

Pim, D., Collins, M., and Banks, L. Human papillomavirus type 16 E5 gene stimulates the transforming activity of the epidermal growth factor receptor. Oncogene, 7: 27-32, 1992.

Pim, D., Storey, A., Thomas, M., Massimi, P., and Banks, L. Mutational analysis of HPV-18 E6 identifies domains required for p53 degradation *in vitro*, abolition of p53 transactivation *in vivo* and immortalisation of primary BMK cells. Oncogene, 9: 1869-1876, 1994.

Pinion, S.B., Kennedy, J.H., Miller, R.W., and MacLean, A.B. Oncogene expression in cervical intraepithelial neoplasia and invasive cancer of the cervix. The Lancet, 337: 819-820, 1991.

Pirami, L., Giachè, V., and Becciolini, A. Analysis of HPV 16, 18, 31, and 35 DNAs in pre-invasive and invasive lesions of the uterine cervix. Journal of Clinical Pathology, 50: 600-604, 1997.

Pirisi, L., Yasumoto, S., Fellery, M., Doninger, J.K., and DiPaolo, J.A. Transformation of human fibroblasts and keratinocytes with human papillomavirus type 16 DNA. Journal of Virology, 61(4): 1061-1066, 1987.

Piver, M.S. Handbook of gynecologic oncology. Little Brown, Boston, USA, 1996.

Pontén, J., and Guo, Z. Precancer of the human cervix. Cancer Surveys, 32: 201-229, 1998.

Reed, K.C., and Mann, D.A. Rapid transfer of DNA from agarose to nylon membranes. Nucleic Acids Research, 13: 7207-7221, 1985.

Reithmuller, D., Gay, C., Bertrand, X., Bettinger, D., Schaal, J.P., Carbillet, J.P., Lassabe, C., Arveux, P., Seilles, E., and Mougin, C. Genital human papillomavirus infection among women recruited for

routine cervical cancer screening or for colposcopy determined by Hybrid Capture II and polymerase chain reaction. Diagnostic Molecular Pathology, 8(3): 157-164, 1999.

Ressing, M.E., Sette, A., Brandt, R.M.P., Ruppert, J., Wentworth, P.A., Hartman, M., Oseroff, C., Grey, H.M., Melief, C.J.M., and Kast, W.M. Human CTL epitopes encoded by human papillomavirus type 16 E6 and E7 identified through *in vivo* and *in vitro* immunogenicity studies of HLA-A*0201-binding peptides. The Journal of Immunology, 154: 5934-5943, 1995.

Richart, R.M., Masood, S., Syrjänen, K.J., Vassilakos, P., Kaufman, R.H., Meisels, A., Olszewski, W.T., Sakamoto, A., and Stoler, M.H., Vooijs, G.P., and Wilbur, D.C. Human papillomavirus. International Academy of Cytology Task Force summary. Diagnostic cytology towards the 21st century: An international expert conference and tutorial. Acta Cytologica, 42: 50-58, 1998.

Ries, E. Modern treatment of carcinoma of the cervix. Chicago Medical Record, 9: 284-289, 1895.

Rigoni-Stern, D.fatti statistici relativi alle malattie cancerose. Giornale Service Progr Pathol Terap Ser, 2: 507-517, 1842.

Roberts, S., Ashmole, I., Johnson, G.D., Kreider, J.W., and Gallimore, P.H. Cutaneous and mucosal human papillomavirus E4 proteins form intermediate filament-like structures in epithelial cells. Virology, 197: 176-187, 1993.

Roberts, S., Ashmole, I., Gibson, L.J., Rookes, S.M., Barton, G.J., and Gallimore, P.H. Mutational analysis of human papillomavirus E4 proteins: Identification of structural features important in the formation of cytoplasmic E4/ cytokeratin networks in epithelial cells. Journal of Virology, 68: 6432-6445, 1994.

Roberts, S., Ashmole, L., Rookes, S.M., and Gallimore, P.H. Mutational analysis of the human papillomavirus type 16 E1^E4 protein shows that the C terminus is dispensable for keratin cytoskeleton association but is involved in inducing disruption of the keratin filaments. Journal of Virology, 71: 3554-3562, 1997.

Roden, R.B.S., Greenstone, H.L., Kirnbauer, R., Christensen, N.D., Lowy, D.R., and Schiller, J.T. *In vitro* generation and type-specific neutralization of a human papillomavirus type 16 virion pseudotype. Journal of Virology, 70: 5875-5883, 1996.

Rogel-Gaillard, C., Breiburd, F., and Orth, G. Human papillomavirus type 1 E4 proteins differing by their N-terminal ends have distinct cellular localisations when transiently expressed *in vitro*. Journal of Virology, 66: 816-823, 1992.

Rogel-Gaillard, C., Pehau-Arnaudet, G., Breitburd, F., and Orth, G. Cytopathic effect in human papilloamvirus type 1-induced inclusion warts: *In vitro* analysis of the contribution of two forms of the viral E4 protein. Journal of Investigative Dermatology, 101: 843-851, 1993.

Rogozinski, T.T., Jablonska, S., and Jarzabek-Chorzelska, M. Role of cell-mediated immnuity in spontaneous regression of plane warts. International Journal of Dermatology, 27: 322-326, 1988.

Romanczuk, H., Thierry, F., and Howley, P.M. Mutational analysis of *cis* elements involved in E2 modulation of human papillomavirus type 16 $P_{97}$ and type 18 $P_{105}$ promoters. Journal of Virology, 64(6): 2849-2859, 1990.

Romanczuk, H., Villa, L.L., Schlegel, R., and Howley, P.M. The viral transcriptional regulatory region upstream of the E6 and E7 genes is a major determinant of the differential immortalization activities of human papillomavirus types 16 and 18. Journal of Virology, 65(5): 2739-2744, 1991.

Romanczuk, H., and Howley, P.M. Disruption of either the E1 or E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. Proceedings of the National Academy of Sciences of the United States of America, 89: 3159-3163, 1992.

Rose, B.R., Thompson, C.H., Chantrill, L.A., Tattersall, M.H.N., and Cossart, Y.E. Prevalence and distribution of human papillomavirus type-16 DNA in pelvic lymph nodes of patients with cervical cancer and in women with no history of cervical abnormality. International Journal of Cancer, 53: 1-4, 1992.

Rose, B.R. Human papillomavirus prognostic indicators in the primary tumours and pelvic lymph nodes of women with cervical cancer. PhD. Sydney University, Sydney, Australia, 1993.

Rose, B.R., Thompson, C.H., Simpson, J.M., Jarrett, C.S., Elliott, P.M., Tattersall, M.H.N., Dalrymple, C., and Cossart, Y.E. Human Papillomavirus deoxyribonucleic acid as a prognostic indicator in early-stage cervical cancer - A possible role for type 18. American Journal of Obstetrics and Gynecology, 173(5): 1461-1468, 1995.

Rose, B.R., Thompson, C.H., Zhang, J., Stoeter, M., Stephen, A., Pfister, H., Tattersall, M.H.N., and Cossart, Y.E. Sequence variation in the upstream regulatory region of HPV 18 isolates from cervical cancers. Gynecologic Oncology, 66(2): 282-289, 1997.

Rose, B., Steger, G., Dong, X.P., Thompson, C., Cossart, Y., Tattersall, M., and Pfister, H. Point mutations in Sp1 motifs in the upstream regulatory region of HPV 18 isolates from cervical cancers increase promoter activity. Journal of General Virology, 79: 1659-1663, 1998.

Rose, P.G., and Lappas, P.T. Analysis of the cost effectiveness of concurrent Cisplatin-based chemoradiation in cervical cancer: Implications from five randomised trials. Gynecologic Oncology, 78: 3-6, 2000.

Rosen, M., and Auborn, K. Duplication of the upstream regulatory sequences increases the transformation potential of human papillomavirus type 11. Virology, 185: 484-487, 1991.

Rosenthal, A.N., Ryan, A., Al-Jehani, R.M., Storey, A., Harwood, C.A., and Jacobs, I.J. p53 codon 72 polymorphism and risk of cervical cancer in UK. The Lancet, 352: 871-872, 1998.

Rösl, F., Arab, A., Klevenz, B., and zur Hausen, H. The effect of DNA methylation on gene regulation of human papillomaviruses. Journal of General Virology, 74: 791-801, 1993.

Rous, P., and Beard, J.W. The progression to carcinoma of virus-induced rabbit papillomas (Shope). Journal of Experimental Medicine, 62: 523, 1934.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. Enzymatic amplification of β globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. Science, 230: 1350-1354, 1985.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science, 239: 487-491, 1988.

Sailaja, G., Watts, R.M., and Bernard, H.U. Many different papillomaviruses have low transcriptional activity in spite of strong epithelial specific enhancers. Journal of General Virology, 80: 1715-1724, 1999.

Sakai, H., Yasugi, T., Benson, J.D., Dowhanick, J.J., and Howley, P.M. Targeted mutageneis of the human papillomavirus type 16 E2 transactivation domain reveals separable transcriptional activation and DNA replication functions. Journal of Virology, 70(3): 1602-1611, 1996.

Sakamuro, D., Sabbatini, P., White, E., and Prendergast, G. The polyproline region of p53 is required to activate apoptosis but not growth arrest. Oncogene, 15: 887-898, 1997.

Sambrook, J., Fritsch, E.F., and Maniatis, T. Molecular Cloning, A Laboratory Manual, Second Edition. Cold Spring Harbour Laboratory Press, New York, USA, 1989.

Sang, B.C., and Barbosa, M.S. Single amino acid substitutions in low risk human papillomavirus (HPV) type 6 E7 protein enhance features characteristic of the "high-risk" HPV E7 oncoproteins.

Proceedings of the National Academy of Sciences of the United States of America, 89: 8063-8067, 1992.

Sanger, F., Nicklen, S., and Coulson, A.R. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America, 74(12): 5463-5467, 1977.

Santoro, C., Mermod, N., Andrews, P.C., and Tjian, R. A family of human CCAAT-box-binding proteins active in transription and DNA replication: Cloning and expression of multiple cDNAs. Nature, 334: 218-224, 1988.

Sastre-Garau, X., Loste, M-N., Vincent-Salomon, A., Favre, M., Mouret, E., de la Rochefordiere, A., Durand, J.C., Tartour, E., Lepage, V., and Charron, D. Decreased frequency of HLA-DRB1*13 alleles in Frechwomen with HPV-positive carcinoma of the cervix. International Journal of Cancer (Predictive Oncology), 69: 159-164, 1996.

Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J., and Howley, P.M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell, 63: 1129-1136, 1990.

Scheffner, M., Münger, K., and Byrne, J.C. The state of the p53 and retinoblastoma genes in human cervical carcinoma cell lines. Proceedings of the National Academy of Sciences of the United States of America, 88: 5523-5527, 1991.

Schiffman, M.H., Haley, N.J., Felton, J.S., Andrews, A.W., Kaslow, R.A., Lancaster, W.D., Kurman, R.J., Brinton, L.A., Lannom, L.B., and Hoffmann, D. Biochemical epidemiology of cervical neoplasia: Measuring cigarette smoke constituents in the cervix. Cancer Research 47: 3886-3888, 1987.

Schiffman, M.H. Recent progress in defining the epidemiology of human papillomavirus infection and cervical neoplasia. Journal of the National Cancer Institute, 84: 394-398, 1992.

Schiffman, M.H. Epidemiology of cervical human papillomavirus infections. *In:* Human Papillomaviruses. pp 55-82. Springer-Verlag, Berlin, 1994.

Schiffman, M.H. and Brinton, L.A. The epidemiology of cervical carcinogenesis. Cancer, 76: 1888-1901, 1995.

Schiller, J.T., and Okun, M.M. Papillomavirus vaccines: Current status and future prospects. Advanced Dermatology, 11: 355-381, 1996.

Schmauz, R., and Owor, R. Epidemiology of malignant degeneration of condylomata acuminata in Uganda. Pathology Research and Practice, 170: 91-103, 1980.

Schneider, A., Hotz, M., and Gissman, L. Increased prevalence of human papillomavirus in the lower genital tract of pregnant women. International Journal of Cancer, 40: 198-201, 1987.

Schneider, A., and Shah, K. The role of vitamins in the etiology of cervical neoplasia: An epidemiological review. Archives of Gynecology and Obstetrics, 246(1): 1-13, 1989.

Schreiber, E., Matthias, P., Muller, M.M., and Schaffner, W. Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells. Nucleic Acids Research, 17: 6419, 1989.

Schwarz, E., Freese, U.K., Gissmann, L., Mayer, W., Roggenbuck, B., Stremlau, A., and zur Hausen, H. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. Nature, 314(6006): 111-114, 1985.

Sedman, S.A., Barbosa, M.S., Vass, W.C., Hubbert, N.L., Haas, J.A., Lowy, D.R., and Schiller, J.T. The full-length E6 protein of human papillomavirus type 16 has transforming and *trans*-activating activities and cooperates with E7 to immortalise keratinocytes in culture. Journal of Virology, 65(9): 4860-4866, 1991.

Seedorf, K., Krämmer, G., Dürst, M., Suhai, S., and Röwekami, W.G. Human papillomavirus type 16 DNA sequence. Virology, 145: 181-185, 1985.

Shamanin, V, zur Hausen, H., Lavergne, D., Proby, C.M., Leigh, I.M., Neumann, C., Hamm, H., Goos, M., Haustein, U.F., Jung, E.G., Plewig, G., and De Villiers, E.M. Human papillomavirus infections in nonmelanoma skin cancers from renal transplant recipients and nonimmunosupressed patients. Journal of the National Cancer Institute, 88: 802-811, 1996.

Sherman, L., Alloul, N., Golan, I., Durst, M., and Baram, A. Expression and splicing patterns of human papillomavirus type-16 mRNAs in pre-cancerous lesions and carcinomas of the cervix, in human keratinocytes immortalized by HPV 16, and in cell lines established from cervical cancers. International Journal of Cancer, 50: 356-364, 1992.

Sherman, L, and Schlegel, R. Serum- and calcium-induced differentiation of human keratinocytes is inhibited by the oncoprotein of human papillomavirus type 16. Journal of Virology, 70(5): 3269-3279, 1996.

Sherman, L., Jackman, A., Itzhaki, H., Conrad Stöppler, M., Koval, D., and Schlegel, R. Inhibition of serum- and calcium-induced differentiation of human keratinocytes by HPV 16 E6 oncoprotein: Role of p53 inactivation. Virology, 237: 296-306, 1997.

Shibata, D.K., Arnheim, N., and Martin, W. Detection of human papillomavirus in paraffin-embedded tissue using the polymerase chain reaction. Journal of Experimental Medicine, 167: 225-230, 1988.

Shope, R.E., and Hurst, E.W. Infectious papillomatosis of rabbits. With a note on histopathology. Journal of Experimental Medicine, 58: 607-624, 1933.

Sibbet, G.J., and Campo, M.S. Multiple interactions between cellular factors and the non-coding region of human papillomavirus type 16. Journal of General Virology, 71: 2699-2707, 1990.

Sibbet, G.J., Cuthill, S., and Campo, M.S. The enhancer in the long control region of human papillomavirus type 16 is up-regulated by PEF-1 and down-regulated by Oct-1. Journal of Virology, 69(7): 4006-4011, 1995.

Sillman, F., Stanek, A., Sedlis, A., Rosenthal, J., Lanks, K.W., Buchhagen, D., Nicastri, A., and Boyce J. The relationship between human papilloma virus and lower genital intraepithelial neoplasia in immunosuppressed women. American Journal of Obstetrics and Gynecology, 150: 300-308, 1984.

Singer, A., Ho,L., Terry, G., and Kwie, T.S. Association of human papillomavirus with cervical cancer and precancer. *In:* Mindel, A. (ed.), Genital Warts- Human Papillomavirus Infection. pp 105-129, Edward Arnold - A member of the Hodder Headline Group, London, UK, 1995.

Smits, H.L., Cornelissen, M.T.E., Jebbink, M.F., van den Tweel, J.G., Struyk, A.P.H.B., and Schegget, J.T. Human papillomavirus type 16 transcripts expressed from viral-cellular junctions and full-length viral copies in CaSki cells and in a cervical carcinoma. Virology, 182: 870-873, 1991.

Song, Y.S., Kee, S.H., Kim, J.W., Park, N.H., Kang, S.B., Chang, W.H., and Lee, H.P. Major sequence variants in E7 gene of human papillomavirus type 16 from cervical cancerous and noncancerous lesions of Korean women. Gynecologic Oncology, 66: 275-281, 1997.

Southern, E. Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology, 98: 503-517, 1975.

Spalholz, B.A., Yang, Y.C., and Howley, P.M. Transactivation of a bovine papilloma virus transcriptional regulatory element by the E2 gene product. Cell, 42: 183-191, 1985.

Steger, G., Ham, J., Lefebvre, O., and Yaniv, M. The bovine papillomavirus 1 E2 protein contains two activation domains: One that interacts with TBP and another that functions after TBP binding. The EMBO Journal, 14(2): 329-340, 1995.

Steger, G., and Corbach, S. Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. Journal of Virology, 71(1): 50-58, 1997.

Steller, M.A., and Schiller, JT. Human papillomavirus immunology and vaccine prospects. Journal of the National Cancer Institute, Monographs, (21): 145-148, 1996.

Stephen, A. L. Thompson, C.H., Tattersall, M.H., Cossart, Y.E., and Rose, B.R. Analysis of mutations in the URR and E6/E7 oncogenes of HPV 16 cervical cancer isolates from central China. International Journal of Cancer, 86: 695-701, 2000.

Stewart, A.C., Eriksson, A.M., Manos, M.M., Munoz, N., Bosch, F.X., Peto, J., and Wheeler, M.C. Intratype variation in 12 human papillomavirus types: A worldwide perspective. Journal of Virology, 70: 3127-3136, 1996.

Stoler, M.H., Rhodes, C.R., Whitbeck, A., Wolinski, S.M., Chow, L.T., and Broker, T.R. Human papillomavirus type 16 and 18 gene expression in cervical neoplasias. Human Pathology, 23: 117-128, 1992.

Storey, A., Pim, D., Murray, A., Osborn, K., Banks, L., and Crawford, L. Comparison of the *in vitro* transforming activities of human papillomavirus types. The EMBO Journal, 7: 1815-1820, 1988.

Storey, A., Greenfield, I., Banks, L., Pim, D., Crook, T., Crawford, L., and Stanley, M. Lack of immortalizing activity of a human papillomavirus type 16 variant DNA with a mutation in the E2 gene isolated from normal human cervical keratinocytes. Oncogene, 7: 459-465, 1992.

Storey, A., Thomas, M., Kalita, A., Harwood, C., Gardiol, D., Mantovani, F., Breuer, J., Leigh, I.M., Matlashewski, G., and Banks, L. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. Nature, 393: 229-234, 1998a.

Storey, A., Thomas, M., Kalita, A., Harwood, C., Gardiol, D., Mantovani, F., Breuer, J., Leigh, I.M., Matlashewski, G., and Banks, L. p53 polymorphism and risk of cervical cancer. Nature, 396: 532, 1998b.

Strang, G., Hickling, J.K., McIndoe, A.J., Howland, K., Wilkinson, D., Ikeda, H., and Rothbard, J.B. Human T cell responses to human papillomavirus type 16 L1 and E6 synthetic peptides: Identification

of T cell determinants, HLA-DR restriction and virus type specificity. Journal of General Virology, 71: 423-431, 1990.

Strauss, M.J., Shaw, E.W., Bunting, H., and Melnick, J.L. "Crystalline" virus-like particles from skin papillomas characterised by intranuclear inclusion bodies. Proceedings of the Society for Experimental Biology Medicine, 72: 46, 1949.

Stubenrauch, F., and Laimonis, L.A. Human papillomavirus life cycle: Active and latent phases. Seminars in Cancer Biology, 9: 379-386, 1999.

Sun, Q., Tang, S.C., Pater, M.M., and Pater, A. Different HPV16 E6/E7 oncogene expression patterns in epithelia reconstructed from HPV16-immortalized human endocervical cells and genital keratincoytes. Oncogene, 15: 2399-2408, 1997.

Swan, D.C., Tucker, R.A., Tortolero-Luna, G., Mitchell, M.F., Wideroff, L., Unger, E.R., Nisenbaum, R.A., Reeves, W.C., and Icenogle, J.P. Human papillomavirus (HPV) copy number is dependent on grade of cervical disease and HPV type. Journal of Clinical Microbiology, 37(4): 1030-1034, 1999.

Syrjänen, K., Hakama, M., Saarikoski, S, Vayrynen, M., Yliskoski, M., Syrjanen, S., Kataja, V., and Castren, O. Prevalence, incidence and estimated life-time risk of cervical human papillomavirus infection in a nonselected Finnish female population. Sexually Transmitted Diseases, 17: 15-19, 1990.

Syrjanen, S., and Puranen, M. Human papillomavirus infections in children: The potential role of maternal transmission. Critical Reviews in Oral Biology and Medicine, 11(2): 259-274, 2000.

Tachezy, R., Mikyšková, I., Saláková, M., and Van Ranst, M. Correlation between human papillomavirus-associated cervical cancer and p53 codon 72 arginine/proline polymorphism. Human Genetics, 105: 564-566, 1999.

Tagami, H., Aiba, S., and Rokugo, M. Regression of flat warts and common warts. Clinics in Dermatology, 3: 170-178, 1985.

Takami, Y., Sasagawa, T., Sudiro, T.M., Yutsudo, M., and Hakura, A. Determination of the functional difference between human papillomavirus type 6 and 16 E7 proteins by their 30 N-terminal amino acid residues. Virology, 186(2): 489-495, 1992.

Tan, S.H., Gloss, B., and Bernard, H.U. During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element. Nucleic Acids Research, 20(2): 251-256, 1991.

Tan, S.H., Gloss, B., and Bernard, H.U. During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element. Nucleic Acids Research, 20(2): 251-256, 1992.

Tan, S.H., Leong, L.E.C., Walker, P.A., and Bernard, H.U. The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. Journal of Virology, 68(10): 6411-6420, 1994.

Tan, S.H., Bartsch, D., Schwarz, E., and Bernard, H.U. Nuclear matrix attachment regions of human papillomavirus type 16 point toward conservation of these genomic elements in all genital papillomaviruses. Journal of Virology, 72(5): 3610-3622, 1998.

Tan, T.M.C., and Ting, R.C.Y. *In vitro* and *in vivo* inhibition of human papillomavirus type 16 E6 and E7 genes. Cancer Research, 55: 4599-4605, 1995.

Terry, G., Ho, L., and Cuzick, J. Analysis of E2 amino acid variants of human papillomavirus types 16 and 18 and their association with lesion grade and HLA DR/DQ type. International Journal of Cancer, 73: 651-655, 1997.

Thomas, M., and Banks, L. Inhibition of Bak-induced apoptosis by HPV-18 E6. Oncogene, 17: 2943-2954, 1998.

Thomas, M., Kalita, A., Labreque, S., Pim, D., Banks, L., and Matlashewski, G. Two polymorphic variants of wild-type p53 differ biochemically and biologically. Molecular and Cellular Biology, 19: 1092-1100, 1999a.

Thomas, M., Pim, D., and Banks, L. The role of the E6-p53 interaction in the molecular pathogenesis of HPV. Oncogene, 18: 7690-7700, 1999b.

Thring, E.T. On the treatment of uterine cancer. Intercolonial Quarterly Journal of Medicine and Surgery, pp 120-129, 1895.

Tindle, R.W. Human papillomavirus vaccines for cervical cancer. Current Opinions in Immunology, 8: 643-650, 1996.

Tommasino, M., Adamczewski, J.P., Carlotti, F, Barth, C.F., Manetti, R., Contorni, M., Cavalieri, F., Hunt, T., and Crawford, L. HPV 16 E7 protein associates with the protein kinase $p33^{CDK2}$ and cyclin A. Oncogene, 8: 195-202, 1993.

Tornesello, M.L., Buonaguro, F.M., Bethgiraldo, E., and Giraldo, G. Human immunodeficiency virus type-1 tat gene enhances human papillomavirus early gene expression. Intervirology, 36(2): 57-64, 1993.

Tortolero-Luna G. Epidemiology of Genital Human Papillomavirus. Hematology/Oncology Clinics of North America, 13(1): 245-257, 1999.

Turek, L.P and Smith, E.M. The genetic program of genital human papillomaviruses in infection and cancer. *In:* Lörincz, A.T., and Reid, R. (eds.), Human papillomavirus II. Second Edition. pp 735-758, Saunders, Philadelphia, USA, 1996.

Ullman, C.G., and Emery, V.C. Transforming proteins of human papillomaviruses. Reviews in Medical Virology, 6: 39-55, 1996.

Unger, E.R., Vernon, S.D., Thoms, W.W., Nisenbaum, R., Spann, C.O., Horowitz, I.R., Icenogle, J.P., and Reeves, W.C. Human papillomavirus and disease-free survival in FIGO stage Ib cervical cancer. Journal of Infectious Diseases, 172: 1184-1190, 1995.

Ushikai, M., Lace, M.J., Yamkawa, Y., Kono, M., Anson, J., Ishiji, T., Parkkinen, S., Wicker, N., Valentine, M.E., Davidson, I, Turek, L.P., and Haugen, T.H. Trans activation by the full-length E2 proteins of human papillomavirus type 16 and bovine papillomavirus type 1 *in vitro* and *in vivo*: Cooperation with activation domains of cellular transcription factors. Journal of Virology, 68: 6655-6666, 1994.

van Belkum, A., Juffermans, L., Schrauwen, L., van Doornum, G., Burger, M., and Quint, W. Genotyping human papillomavirus type 16 isolates from persistently infected promiscuous individuals and cervical neoplasia patients. Journal of Clinical Microbiology, 33(11): 2957-2962, 1995.

Vance, K.W., Campo, M.S., and Morgan, I.M. An enhanced epithelial response of a papillomavirus promoter to transcriptional activators. Journal of Biological Chemistry, 274(39): 27839-27844, 1999.

van Duin, M., Snijders, P.J.F., Vossen, M.T.M., Klaassen, E., Voorhorst, F., Verheijen, R.H.M., Helmerhorst, T.J., Meijers, C.J.L.M., and Walboomers, J.M.M. Analysis of human papillomavirus type 16 E6 variants in relation to p53 codon 72 polymorphism genotypes in cervical carcinogenesis. Journal of General Virology, 81: 317-325, 2000.

Veress, G., Szarka, K., Dong, X.P., Gergely, L., and Pfister, H. Functional significance of sequence variation in the E2 gene and the long control region of human papillomavirus type 16. Journal of General Virology, 80: 1035-1043, 1999.

Vernon, S.D., Hart, C.E., Reeves, W.C., and Icenogle, J.P. The HIV-1 tat protein enhances E2-dependent human papillomavirus 16 transcription. Virus Research, 27(2): 133-145, 1993.

Vernon, S.D., Unger, E.R., Miller, D.L., Lee, D.R., and Reeves, W.C. Association of human papillomavirus type 16 integration in the E2 gene with poor disease-free survival from cervical cancer. International Journal of Cancer, 74: 50-56, 1997.

Vernon, S.D., Unger, E.R., and Reeves, W.C. Human Papillomavirus and Cervical Cancer. Current Problems in Obstetrics, Gynecology and Fertility, July/Aug, 104-124, 1998.

Viscidi, R.P., Sun, Y., Tsuzaki, B., Bosch, F.X., Muñoz, N., and Shah, K.V. Serologic response in human papillomavirus-associated invasive cervical cancer. International Journal of Cancer, 55: 780-784, 1993.

von Knebel Doeberitz, M., Oltersdorf, T., Schwarz, E., and Gissmann, L. Correlation of modified human papilloma virus early gene expression with altered growth properties in C4-1 cervical carcinoma cells. Cancer Research, 48(13): 3780-3786, 1988.

von Knebel Doeberitz, M., Bauknecht, T., Bartsch, D., and zur Hausen, H. Influence of chromosomal integration on glucocorticoid-regulated transcription of growth-stimulating papillomavirus genes E6 and E7 in cervical carcinoma cells. Proceedings of the National Academy of Sciences of the United States of America, 88: 1411-1415, 1991.

von Knebel Doeberitz, M., Rittmüller, C., Geitz, D., zur Hausen, H., and Dürst, M. Inhibition of tumourigenicity of cervical cancer cells in nude mice by HPV-18 E6-E7 antisense RNA. International Journal of Cancer, 51(5): 831-834, 1992.

von Knebel Doeberitz, M., Rittmüller, C., Aengeneyndt, F., Jansendurr, P., and Spitkovsky, D. Reversible repression of papillomavirus oncogene expression in cervical carcinoma cells – consequences for the phenotype and E6-p53 and E7-pRb interactions. Journal of Virology, 68: 2811-2821, 1994.

Wagatsuma, M., Hashimoto, K., and Matsukura, T. Analysis of integrated human papillomavirus type 16 DNA in cervical cancers: Amplification of viral sequences together with cellular flanking sequences. Journal of Virology, 64: 813-821, 1990.

Waggoner, SE, and Wang, X. Effect of nicotine on proliferation of normal, malignant and human papillomavirus-transformed human cervical cells. Gynecologic Oncology, 55: 91-95, 1994.

Walboomers, JMM, Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J.F., Meijers, C.J.L.M., and Munoz, N. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. Journal of Pathology, 189(1): 12-19, 1999.

Walder, J. Antsense DNA and RNA: Progress and prospects. Genes and Development, 2: 502-504, 1988.

Walker, J., Bloss, J.D., Liao, S.Y., Berman, M., Bergen, S., and Wilczynski, S.P. Human papillomavirus genotype as a prognostic indicator in carcinoma of the uterine cervix. Obstetrics and Gynecology, 74(5): 781-785, 1989.

Walker, K.K., and Levine, A.J. Identification of a novel p53 functional domain that is necessary for efficient growth suppression. Proceedings of the National Academy of Sciences of the United States of America, 93: 15335-15340, 1996.

Watanabe, S., Kanda, T., Sato, H, Furuno, A., and Yoshiike, K. Mutational analysis of human papillomavirus type 16 E7 functions. Journal of Virology, 64: 207-214, 1990.

Watanabe, S., Sato, H, Komiyama, N., Kanda, T., and Yoshiike, K. The E7 functions of human papillomaviruses in rat 3YI cells. Virology, 187: 107-114, 1992.

Watts, K.J. Sequence variation within the upper regulatory region of human papillomavirus type 16 and its effects on the clinical behaviour of cervical cancer. Honours thesis, The University of Sydney, Sydney, Australia, 1996.

Welter, J.F., and Eckert, R.L. Differential expression of the fos and jun family members c-fos, fosB, Fra-1, Fra-2, c-jun, junB and junD during epidermal keratincoyte differentiation. Oncogene, 11: 2681-2687, 1995.

Werness, B.A., Levine, A.J., and Howley, P.M. Association of human papillomavirus types 16 and 18 E6 proteins with p53. Science, 248(4951): 76-79, 1990.

Wertheim, E. The extended abdominal operation for carcinoma of the cervix. American Journal of Obstetrics, 66: 169-232, 1912.

Wheeler, C., and Icenogle, J. A closer look at papillomavirus variants. *In:* Myers, G., Delius, H., Icenogel, J., Bernard, H.U., Baker, C., Halpern, A., Wheeler, C. (eds.), Human Papillomaviruses. pp III41-III46, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, 1995.

Wheeler, C. Human papillomavirus type-specific prevalence. *In:* Myers, G., Halpern, A., Baker, C., McBride, A., Wheeler, C., and Doorbar, J. (eds.), Human Papillomaviruses. pp III112-III124, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, 1996.

Wheeler, C.M., Yamada, T., Hildesheim, A., and Jenison, S.A. Human papillomavirus type 16 sequence variants: Identification by E6 and L1 lineage-specific hybridisation. Journal of Clinical Microbiology, 35(1): 11-19, 1997.

White, A.E., Livanos, E.M., and Tlsty, T.D. Differential disruption of genomic integrity and cell cycle regulation in normal human fibroblasts by the HPV oncoproteins. Genes and Development, 8: 666-677, 1994.

Wieland, U., and Pfister, H. Papillomaviruses in human pathology: Epidemiology, pathogenesis and oncogenic role, Chapter 1. *In:* Gross, G.E., Barrasso, R. (eds.), Human Papilloma Virus Infection. A clinical atlas. pp 1-20, Ullstein Mosby, Berlin, Germany, 1997.

Wigle, DT., Mao, Y., and Grace, M. Re: smoking and cancer of the uterine cervix : Hypothesis. American Journal of Epidemiology, 111: 125-127, 1980.

Wilczynski, S.P., Pearlman, L., and Walker, J. Identification of HPV 16 early genes retained in cervical carcinomas. Virology, 166: 624-627, 1988.

Woodworth, C.D., Doniger, J., and DiPaolo, J.A. Immortalization of human foreskin keratinocytes by various human papillomavirus DNAs correspond to their association with cervical carcinoma. Journal of Virology, 63: 159-164, 1989.

Woodworth, C.D., and Simpson, S. Comparative lymphokine secretion by cultured normal human cervical keratinocytes, papillomavirus-immortalized, and carcinoma cell lines. American Journal of Pathology, 142(5): 1544-1555, 1993.

Wright, T.C., Sun, X.W., and Koulos, J. Comparison of management algorithms for the evaluation of women with low-grade cytologic abnormalities. Obstetrics and Gynecology, 85: 202-210, 1995.

Xi, L.F., Demers, G.W., Kiviat, N.B., Kuypers, J., Beckman, A.M., and Galloway, D.A. Sequence variation in the noncoding region of human papillomavirus type 16 detected by single-strand conformation polymorphism analysis. The Journal of Infectious Diseases, 168: 610-617, 1993.

Xi, L.F., Demers, G.W., Koutsky, L.A., Kiviat, N.B., Kuypers, J., Watts, D.H., Holmes, K.K., and Galloway, D.A. Analysis of human papillomavirus type 16 variants indicates establishment of persistent infection. The Journal of Infectious Diseases, 172: 747-755, 1995.

Xi, L.F., Koutsky, L.A., Galloway, D.A., Kuypers, J., Hughes, J.P., Wheeler, C.M., Holmes, K.K., and Kiviat, N.B. Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. Journal of the National Cancer Institute, 89(11): 796-802, 1997.

Xi, L.F., Critchlow, C.W., Wheeler, C.M., Koutsky, L.A., Galloway, D.A., Kuypers, J., Hughes, J.P., Hawes, S.E., Surawicz, C., Goldbaum, G., Holmes, K.K., and Kiviat, N.B. Risk of anal carcinoma *in situ* in relation to human papillomavirus type 16 variants. Cancer Research, 58: 3839-3844, 1998.

Yaegashi, N., Xi, L.F., Batra, M., and Galloway, D.A. Sequence and antigenic diversity in two immunodominant regions of the L2 protein of human papillomavirus types 6 and 16. The Journal of Infectious Diseases, 168: 743-747, 1993.

Yamada, T., Wheeler, C.M., Halpern, A.L., Stewart, A.C.M., Hildesheim, A., and Jenison, S.A. Human papillomavirus type 16 variant lineages in United States populations characterised by nucleotide sequence analysis of the E6, L2, and L1 coding segments. Journal of Virology, 69(12): 7743-7753, 1995.

Yamada, T., Manos, M.M., Peto, J., Greer, C.E., Munoz, N., Bosch, F.X., and Wheeler, C.M. Human papillomavirus type 16 sequence variation in cervical cancers: A worldwide perspective. Journal of Virology, 71(3): 2463-2472, 1997.

Yasugi, T., Benson, J.D., Sakai, H., Vidal, M., and Howley, P.M. Mapping and characterisation of the interaction domains of human papillomavirus type 16 E1 and E2 proteins. Journal of Virology, 71(2): 891-899, 1997.

Yasumoto, S., Burchardt, A,L., Doniger, J., and DiPaolo, J.A. Human papillomavirus type 16 DNA induced malignant transformation of NIH3T3 cells. Journal of Virology, 57(2): 572-577, 1986.

Yasumoto, S., Taniguchi, A., and Sohma, K. Epidermal growth factor (EGF) elicits down-regulation of human papillomavirus type 16 (HPV-16) E6/E7 mRNA at the transcriptional level in an EGF-stimulated human keratinocyte cell line: Functional role of EGF-responsive silencer in the HPV-16 long control region. Journal of Virology, 65(4): 2000-2009, 1991.

Yee, C., Krishnan-Hewlett, Z., Baker, C.C., Schlegel, R., and Howley, P.M. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. American Journal of Pathology, 119: 361-366, 1985.

Ylitalo, N., Sorensen, P., Josefsson, A.M., Magnusson, P.K.E., Andersen, P.K., Pontén, J., Adami, H.O., Gyllensten, U.B., and Melbye, M. Consistent high viral load of human papillomavirus 16 and

risk of cervical carcinoma in situ: a nested case-control study. The Lancet, 355(9222): 2194-2198, 2000.

Yutsudo, M., Okamoto, Y., and Hakura, A. Functional dissociation of transforming genes of human papillomavirus type 16. Virology, 166: 594-597, 1988.

Zahm, D.M., Ninal, I., and Schneider, A. General principles of detection, Chapter 2. *In:* Gross, G.E., Barrasso, R. (eds.), Human Papilloma Virus Infection. A clinical atlas. pp 21-45, Ullstein Mosby, Berlin, Germany, 1997.

Zhan, D., Santin, A.D., Liu, Y., Parham, G.P., Li, C., Meyers, C., and Hermonat, P.L. Binding of the human papillomavirus type 16 p97 promoter by the adeno-associated virus Rep78 major regulatory protein correlates with inhibition. Journal of Biological Chemistry, 274(44): 31619-31624, 1999.

Zhang, D.K., Ngan, H.Y.S., Cheng, R.Y.S., Cheung, A.N.Y., Liu, S.S., and Tsao, S.W. Clinical significance of telomerase activation and telomeric restriction fragment (TRF) in cervical cancer. European Journal of Cancer, 35(1): 154-160, 1999.

Zhao, W., Chow, L.T., and Broker, T.R. Transcriptional activities of human papillomavirus type 11 E6 promoter-proximal elements in rafts and submerged cultures of foreskin keratinocytes. Journal of Virology, 71(11): 8832-8840, 1997.

Zhebe, I., and Wilander, E. Human papillomavirus infection and invasive cervical neoplasia: A study of prevalence and morphology. Journal of Pathology, 181: 270-275, 1997.

Zhebe, I., Voglino, G., Delius, H., Wilander, E., and Tommasino, M. Risk of cervical cancer and geographical variations of human papillomavirus 16 E6 polymorphisms. The Lancet, 352: 1441-1442, 1998a.

Zhebe, I., Wilander, E., Delius, H., and Tommasino, M. Human papillomavirus 16 E6 variants are more prevalent in invasive cervical carcinoma then the prototype. Cancer Research, 58: 829-833, 1998b.

Zhou, J., and Fraser, I.H. Papovaviridae: Capsid structure and capsid protein function. Papillomavirus Rep., 6: 59-64, 1995.

Zhou, J., and Fraser, I.H. Papovaviridae: Capsid structure and capsid protein function. *In:* Lacey, C. (ed.), Papillomavirus Reviews: Current Research on Papillomaviruses. pp 93-100, Leeds University Press, Leeds, UK, 1996.

Zimmermann, H., Degenkolbe, R., Bernard, H.U., and O'Connor, M.J. The human papillomavirus type 16 E6 oncoprotein can down-regulate p53 activity by targeting the transcriptional coactivator CBP/p300. Journal of Virology, 73(8): 6209-6219, 1999.

Ziol, M., Di Tomaso, C., Biaggi, A., Tepper, M., Piquet, P., Carbillon, L., Uzan, M., and Guettier, C. Virological and biological characteristics of cervical intraepithelial neoplasia grade I with marked koilocytotic atypia. Human Pathology, 29(10): 1068-1073, 1998.

Zondervan, K.T., Carpenter, L.M., Painter, R., and Vessey, M.P. Oral contraceptives and cervical cancer – further findings from the Oxford Family Planning Association contraceptive study. British Journal of Cancer, 73: 1291-1297, 1996.

zur Hausen, H, Meinhof, W., Scheiber, W., and Bornkamm, G.W. Attempts to detect virus specific DNA sequences in human tumors. I. Nucleic acid hybridisations with complementary RNA of human wart virus. International Journal of Cancer, 13: 650-656, 1974.

zur Hausen, H. Condylomata acuminata and human genital cancer. Cancer Research, 36: 794, 1975.

zur Hausen, H. Condylomata acuminata and human genital cancer. Cancer Research, 36: 530, 1976.

zur Hausen, H. Human papillomaviruses and their possible role in squamous cell carcinomas. Current Topics in Microbiology and Immunology 78, 78: 1-30, 1977.

zur Hausen, H. Human genital cancer: Synergism between two virus infections or synergism between a virus infection and initiating events? The Lancet, 2(8312): 1370-1372, 1982.

zur Hausen, H. Intracellular surveillance of persisting viral infections. Human genital cancer results from deficient control of papillomavirus gene expression. The Lancet, 2: 489-491, 1986.

zur Hausen, H. Papillomaviruses in anogenital cancer as a model to understand the role of viruses in human cancers. Cancer Research, 49: 4677-4681, 1989.

zur Hausen, H. Human papillomaviruses in the pathogenesis of anogenital cancer. Virology, 184: 9-13, 1991.

zur Hausen., H. Molecular pathogenesis of cancer of the cervix and its causation by specific human papillomavirus types. *In:* Human Papillomaviruses. pp 131-156, Springer-Verlag, Berlin, Germany, 1994.

zur Hausen, H., and de Villiers, E.M. Human Papillomaviruses. Annual Review of Microbiology, 48: 427-447, 1994.

zur Hausen, H. Roots and perspectives of contemporary papillomavirus research. Journal of Cancer Research and Clinical Oncology, 122: 3-13, 1996.