



COMPUTATIONAL METHODS FOR THE ANALYSIS OF HIV DRUG RESISTANCE DYNAMICS

A thesis submitted in fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In

The Faculty of Science

School of Information Technologies

THE UNIVERSITY OF SYDNEY

MARCH 2007

ALI AL MAZARI

CERTIFICATION BY THE SUPERVISORS

We certify that the thesis entitled: “Computational Methods for the Analysis of HIV Drug Resistance Dynamics” submitted by Ali Al Mazari, for the Degree of Doctor of Philosophy, Biomedical Informatics Sciences, is ready for examination.

Supervisor:

Prof. Albert Y. Zomaya, Head of School
CISCO Systems Chair Professor of Internetworking
School of Information Technologies
The University of Sydney
Sydney, NSW, Australia

Co-supervisors:

Dr. Michael Charleston, Senior Lecturer in Bioinformatics
School of Information Technologies
The University of Sydney
Sydney, NSW, Australia

Dr. Roger J. Garsia, Clinical Senior Lecturer
Department of Medicine
The University of Sydney
Clinical Immunologist at RPAH
Clinical Director HIV/AIDS
Sydney South West Area Health Service

DECLARATION

I, the author of the thesis, declare that none of the material in this thesis has been previously submitted by me or any other candidate for any degree to this or any other university.

Ali A Al Mazari

ACKNOWLEDGEMENT

This thesis is dedicated to my family. I wish to thank my parents, Adnan and Farha and my wife, Om Kazem, for their encouragement and cheerful feeling during this quest, and my dear children, Kazem and Adnan. I thank my brothers, sisters and friends, who surround my heart.

Special thanks go to my supervisor, Professor Albert Zomaya, Head of the School of Information Technologies, for his skilled and supportive supervision and guidance, and to my cosupervisors, Dr Michael Charleston, School of Information Technologies, and Dr Roger Garsia, Department of Medicine. Thanks to Dr Tony Souter for his editorial assistance. Many thanks to Hanan Salem and Annamarie Maher, Molecular Biology Laboratory, Departments of Biochemistry and Clinical Immunology of Royal Prince Alfred Hospital, Sydney.

ETHICS REVIEW COMMITTEE APPROVAL

Ethics approval from the Sydney South West Area Health Service Ethics Review Committee (SSWAHS ERC X05-0205) was obtained for this project.

ABSTRACT

Despite the extensive quantitative and qualitative knowledge about therapeutic regimens and the molecular biology of HIV/AIDS, the eradication of HIV infection cannot be achieved with available antiretroviral regimens. HIV drug resistance remains the most challenging factor in the application of approved antiretroviral agents. Previous investigations and existing HIV/AIDS models and algorithms have not enabled the development of long-lasting and preventive drug agents. Therefore, the analysis of the dynamics of drug resistance and the development of sophisticated HIV/AIDS analytical algorithms and models are critical for the development of new, potent antiviral agents, and for the greater understanding of the evolutionary behaviours of HIV.

This study presents novel computational methods for the analysis of drug-resistance dynamics, including: viral sequences, phenotypic resistance, immunological and virological responses and key clinical data, from HIV-infected patients at Royal Prince Alfred Hospital in Sydney. The lability of immunological and virological responses is analysed in the context of the evolution of antiretroviral drug-resistance mutations. A novel Bayesian algorithm is developed for the detection and classification of neutral and adaptive mutational patterns associated with HIV drug resistance. To simplify and provide insights into the multifactorial interactions between viral populations, immune-system cells, drug resistance and treatment parameters, a Bayesian graphical model of drug-resistance dynamics is developed; the model supports the exploration of the interdependent associations among these dynamics.

TABLE OF CONTENTS

Certification by the supervisors	i
Declaration	ii
Acknowledgement	iii
Ethics Review Committee Approval	iv
Abstract	v
Table of contents	vi
List of figures	xiii
List of tables	xvi
Abbreviations	xvii
Publications from thesis	xx
<u>1 INTRODUCTION.....</u>	1
1.1 The key issue: HIV drug resistance dynamics	2
1.2 Research objectives	3
1.3 Significance of the study	4
1.4 Structure of the dissertation	5
<u>2 HIV BIOLOGY, THERAPY, DRUG RESISTANCE AND MODELLING.....</u>	7
2.1 Molecular biology of HIV/AIDS	8
2.1.1 HIV as a retrovirus	8
2.1.2 Target cells of HIV	9
2.1.3 Genetic structure of HIV	9
2.1.4 Life cycle	13
2.1.5 Associated diseases	16

2.1.6	Diagnosis and stages	16
2.1.7	Genetic variability of HIV	20
2.1.8	Epidemiology	22
2.1.9	Clinical laboratory tests	24
2.1.9.1	CD4 T-Cell count test	24
2.1.9.2	Viral load test	24
2.1.9.3	Drug resistance tests	25
2.1.10	Course of HIV and CD4 interaction	25
2.2	Drug resistance	28
2.2.1	Conceptual basis	28
2.2.2	Drug resistance levels	29
2.2.3	Resistance emergence and evolution of resistance	29
2.2.4	Drug resistance assays	30
2.2.4.1	Genotypic assays	30
2.2.4.2	Phenotypic assays	31
2.2.4.3	Virtual phenotypic assays	32
2.2.5	Genotypic sequence interpretation systems	33
2.2.5.1	Rule-based interpretation systems	33
2.2.5.2	Machine-learning interpretation systems	33
2.2.5.3	Inter-algorithm comparisons and algorithm validation systems	34
2.2.6	CREST Trial algorithm	34
2.2.7	Drug resistance databases	37
2.2.8	Database analysis methods	37
2.3	Antiretroviral therapy and associated resistance mutations and mechanisms	38
2.3.1	Antiretroviral agents and combinations	38
2.3.1.1	Protease inhibitors (PIs)	39
2.3.1.2	Nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs)	39
2.3.1.3	Non-nucleoside reverse transcriptase inhibitors (NNRTIs).....	40
2.3.1.4	Fusion inhibitors (FIs)	40
2.3.1.5	Combination antiretroviral therapy	43
2.3.2	Mechanisms of resistance to ARV agents	46
2.3.2.1	Mechanisms of resistance to PIs	46
2.3.2.2	Mechanisms of resistance to NRTIs	46

2.3.2.3 Mechanisms of resistance to NNRTIs	47
2.3.2.4 Mechanisms of resistance to FIs	47
2.3.3 Mutations associated with antiretroviral agents	50
2.3.3.1 Mutations associated with PIs	51
2.3.3.2 Mutations associated with NRTIs	53
2.3.3.3 Mutations associated with NNRTIs	55
2.3.4 Treatment considerations	57
2.3.4.1 Cross-resistance	57
2.3.4.2 Adherence	57
2.3.4.3 Toxicity	57
2.3.4.4 Initiation	58
2.3.4.5 Supporting Treatment strategies	58
2.4 Modelling of HIV/AIDS	60
2.4.1 HIV/AIDS modelling	60
2.4.2 The basic model of HIV	60
2.4.3 Modelling HIV/AIDS from an epidemiological and pathological perspective	62
2.4.3.1 Epidemiology models	62
2.4.3.2 Pathogenetic models	62
2.4.4 Modelling HIV/AIDS dynamics	62
2.4.4.1 Virology and immunology models	62
2.4.4.2 Drug resistance models	63
2.4.4.3 Treatment models	63
2.4.5 Modelling based on analytical methods	63
2.4.5.1 Mathematical modelling	63
2.4.5.2 Graphical modelling	64
2.4.5.3 Pharmacokinetic modelling	65
2.4.6 Modelling approaches	65
2.4.6.1 Stochastic modelling	65
2.4.6.2 Deterministic modelling	66
2.4.7 Significance and limitations of model types	66
2.4.7.1 Significance of graphical models	66
2.4.7.2 Limitations of mathematical models	67
<u>3 HIV GENOTYPING DATABASES.....</u>	69
3.1 Introduction	70

3.2 Data sources and analysis	71
3.2.1 Data on drug resistance mutations	71
3.2.2 The systematic analysis of genotypic databases	74
3.3 Related drug-resistance databases	75
3.3.1 Stanford University HIV Drug Resistance Database (HIVdb)	75
3.3.2 Arevir Database	76
3.3.3 UK Drug Resistance Database	76
3.3.4 Los Alamos HIV Databases	77
3.4 HIV-1 Genotyping Database	78
3.4.1 Tests	78
3.4.2 Database model	79
3.4.3 Database Dictionary	83
3.4.4 Database Schema	86

4 IMMUNOLOGICAL AND VIROLOGICAL RESPONSES IN THE EVOLUTION OF RESISTANCE..... 87

4.1 Introduction	88
4.2 Classification of resistance and immunological and virological responses	89
4.2.1 Resistance classification	89
4.2.2 Immunological and virological responses classification	91
4.2.3 Tracking trends of CD4 T-cell counts and VL associated with EoR	92
4.3 Experimental materials and methods	95
4.3.1 Study population: sampling and baseline	95
4.3.2 Mutation genotyping and interpretation	95
4.3.3 Immunological and virological parameters	96
4.3.4 Data analysis and computations	97

4.3.5 LOESS Normalisation curves	97
4.4 Experimental results	98
4.4.1 Population and treatment baselines	98
4.4.2 Evolutionary mutational patterns associated with ARV	98
4.4.3 Immunological changes associated with EoR	103
4.4.4 Virological changes associated with EoR	108
4.5 Discussion	112
4.5.1 Patterns of EoR	112
4.5.2 Predictors of EoR	113
4.6 Conclusion	114

5 ADAPTIVE AND NEUTRAL EVOLUTIONARY PATTERNS ASSOCIATED WITH HIV DRUG RESISTANCE **116**

5.1 Introduction	117
5.2 Evolutionary patterns associated with HIV drug resistance	118
5.2.1 Emergence and evolution of drug resistance	118
5.2.2 Classes of the evolutionary patterns associated with drug resistance	119
5.2.3 Current evolutionary models and their limitations	122
5.3 An algorithm for HIV evolutionary patterns	123
5.3.1 The formalism of Bayes' theorem	123
5.3.2 Data parameters	125
5.3.3 Algorithm and procedure	125
5.4 Experimental data and methods	130
5.4.1 Study population	130
5.4.2 Mutational patterns associated with antiretroviral agents	131
5.4.3 Application of the algorithm	132

5.5 Results and discussion	137
5.5.1 Prevalence of neutral, positively and negatively selected patterns	137
5.5.2 Evolutionary patterns consisting of two mutations	139
5.5.3 Evolutionary patterns consisting of three mutations	140
5.5.4 Significance of the algorithm	141
5.5.5 Limitations	141
5.6 Further applications and conclusions	141
5.6.1 Applications based on mutation types	141
5.6.2 Applications based on the site of mutations	142
5.6.3 Applications based on the number and combination of mutations	142
5.6.4 Applications in relation to other HIV variables	142
5.6.5 Conclusion	143

**6 A MULTIVARIATE BAYESIAN MODEL OF HIV
DRUG RESISTANCE DYNAMICS.....** **144**

6.1 Introduction	145
6.2 HIV drug resistance	146
6.3 HIV/AIDS modelling	148
6.3.1 The need for modelling	148
6.3.2 Mathematical modelling of HIV/AIDS	149
6.3.3 Limitations of mathematical modelling of HIV	150
6.3.4 Graphical modelling of HIV/AIDS	151
6.4 BNs and modelling	152
6.4.1 Bayesian belief networks	152
6.4.2 BN modelling	156
6.5 The model of HIV drug resistance dynamics	157
6.5.1 Methods of model construction	157
6.5.2 Learning model parameters and structure	158
6.5.3 Making inferences	167
6.5.4 Message-Passing Algorithm	169

6.6 Discussion	176
6.6.1 Significance of BNs	176
6.6.2 Limitations of BNs	176
6.7 Conclusion	176
<u>7 CONCLUSION.....</u>	178
7.1 Summary of the dissertation	178
7.2 Summary of results	179
7.3 Research limitations	180
7.4 Future directions	181
<u>APPENDICES.....</u>	184
Appendix A: Web resources for HIV drug resistance testing in clinical and research Settings	185
Appendix B: HIV-1 PR and RT Sequencing Kits	188
Appendix C: Sample reports used for clinical decision-making at RPAH	190
<u>REFERENCES.....</u>	196

LIST OF FIGURES

Figure	Title	Page
Figure 2.1	: Flow of biological sequence information in living organisms	8
Figure 2.2	: HIV genomic structure	11
Figure 2.3	: HIV life cycle	15
Figure 2.4	: Generalized time course of HIV infection and disease	18
Figure 2.5	: Phylogenetic tree of the SIV and HIV	21
Figure 2.6	: Adults and children estimated to be living with HIV in 2006	22
Figure 2.7	: Adults and children estimated to be newly infected with HIV in 2006	23
Figure 2.8	: Estimated adult and child deaths from AIDS in 2006	23
Figure 2.9	: Generalised relationship between HIV viral copies and CD4 T-cell counts during the typical course of the disease in untreated individuals	27
Figure 2.10	(A): Prognosis According to CD4 Cell Count and Viral Load in the Pre-HAART Eras (so-called CART)	45
	(B): Prognosis According to CD4 Cell Count and Viral Load in the HAART Eras	45
Figure 2.11	(A): Resistance by interference with the incorporation of NRTIs	48
	(B): Resistance by the ATP-mediated excision of NRTIs	48
Figure 2.12	: Mechanism of resistance associated with NNRTIs	49
Figure 2.13	: Structural model of HIV-1-PR, labelled with PI-associated resistance mutations	52
Figure 2.14	: Structural model of HIV-1-RT, labelled with NRTI- associated resistance mutations	54
Figure 2.15	: Structural model of HIV-1-RT, labelled with NNRTI- associated resistance mutations	56
Figure 2.16	: The basic model of HIV viral dynamics	61

Figure 3.1	: Schematic of showing the distribution of drug-resistance mutations within the PR and RT genes	73
Figure 3.2	: HIV-1 Genotyping Database Model	82
Figure 3.3	: HIV-1 Genotyping Database Schema	86
Figure 4.1	: The classes of EoR associated with NRTIs, NNRTIs or PIs	90
Figure 4.2	: (A): Tracking the trends of CD4 T-cell counts in patients with and without EoR associated with the NRTI class of drugs	93
	(B): Tracking the trends of VL in patients with and without EoR associated with the PI class of drugs	94
Figure 4.3	: (A): Frequencies of mutations associated with NRTIs	101
	(B): Frequencies of mutations associated with NNRTIs	102
	(C): Frequencies of mutations associated with PIs	102
Figure 4.4	: (A): LOESS curves of the immunological responses correlated with EoR associated with primary-NRTIs mutations	104
	(B): LOESS curves of the immunological responses correlated with EoR associated with secondary-NRTIs mutations	105
	(C): LOESS curves of the immunological responses correlated with EoR associated with primary-PIs mutations	106
Figure 4.5	: (A): LOESS curves of the virological responses correlated with EoR associated with NRTIs associated mutations	109
	(B): LOESS curves of the virological responses correlated with EoR associated with PIs associated mutations	110
Figure 5.1	: Evolutionary patterns associated with HIV drug resistance ..	121
Figure 5.2	: Multiple sequential evidence	124
Figure 5.3	: The distribution of V118I+T215x and other patterns	135

List of Figures

Figure 5.4	: The prevalence of two mutational patterns	138
Figure 5.5	: Prevalence of the patterns that consist of two mutations	139
Figure 5.6	: Prevalence of the patterns that consist of three mutations ...	140
Figure 6.1	: A Simple Bayesian Network	155
Figure 6.2	: A Bayesian model of HIV drug resistance dynamics; asterisked parameters are multiple states	166
Figure 6.3	: Inductive and deductive reasoning	168
Figure 6.4	: Different types of network topologies	170
Figure 6.5	: A Multiply Connected network partitioned into two networks for making inference	170
Figure 6.6	: A four-node BN with the four corresponding CPTs	172
Figure 6.7	: Initialised version of the BN in Figure 6.6, illustrating the values and messages of π and λ probability vectors	174
Figure 6.8	: An updated version of the BN in Figure 6.7	175

LIST OF TABLES

Table	Title	Page
Table 2.1	: HIV genes and proteins and their functions	12
Table 2.2	: HIV/AIDS classification systems provided by WHO and CDC	19
Table 2.3	: Comparison of the GTA, PTA and VPA	33
Table 2.4	: Primary/secondary resistance mutations associated with PRIs, NRTIs and NNRTI as reported by CREST study	35
Table 2.5	: Algorithms for interpreting HIV-1 PR and RT sequences ..	36
Table 2.6	: ARV agents used in the treatment of HIV infection	41
Table 2.7	: CART options following virological failure	44
Table 3.1	: HIV-1 Genotyping Database Dictionary	83
Table 4.1	: CD4 T-cell counts and severity immunosuppression	92
Table 4.2	: Frequencies of mutations associated with PIs, NRTIs and NNRTIs	100
Table 4.3	: The mean values of CD4 T-cell counts	107
Table 4.4	: The mean values of VL	111
Table 5.1	: Categories and prevalence of the evolutionary pattern p_x ...	130
Table 5.2	: Evolving mutational patterns in the HIV-RT gene	133
Table 6.1	: Primary and secondary resistance mutations associated with PRIs, NRTIs and NNRTI	162
Table 6.2	: CD4 levels and the severity of immunosuppression	163
Table 6.3	: Variables and states of HIV drug resistance dynamics model	164

ABBREVIATIONS

3TC	Lamivudine
AA	Amino acid
ABC	Abacavir sulfate
AD	Adherence
ADRA	Antiviral drug resistance analysis
AE	AIDS event
AIDS	Acquired immune deficiency syndrome
ANRS	French national agency for AIDS research
APV	Amprenavir
ARV	Antiretroviral
ATV	Atazanavir sulfate
AZT or ZDV	Zidovudine or Azidothymidine
BMI	biomedical informatics
BNs	Bayesian belief networks
CA	Capsid
CART	Combination antiretroviral therapy
CDC	Centres for disease control and prevention
CI	Confidence interval
CM	Codon of mutations
CMP	Category of mutational pattern
CPT	Conditional probability table
CREST	Can resistance enhance selection of therapy
d4T	Stavudine
ddC	Zalcitabine or Dideoxycytidine
ddI	didanosine or dideoxyinosine
DI	Death incidence
DLV	Delavirdine
DNA	Deoxyribonucleic acid
EFV	Efavirenz

EM	Expectation maximisation
EoR	Evolution of resistance
FDA	Food and Drug Administration
FIs	Fusion inhibitors
FPV	Fosamprenavir Calcium
FTC	Emtricitabine
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HIVdb	Stanford university's HIV database
IDV	Indinavir
IN	Integrase
JPD	Joint probability distribution
LOESS	LOcally wEighted Scatter-plot Smoothing
MA	matrix
MDR	Multi-drug resistance
mRNA	messenger RNA
NA	Nucleotide acid
NC	Nucleocapsid
NE	Neutral evolution
NFV	Nelfinavir mesylate
NIH	National Institutes of Health
NIR	Non-responding immunological response
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRL	National Serology Reference Laboratory
NRTI	Nucleotide reverse transcriptase inhibitor
NS	Negative selection
NVP	Nevirapine
NVR	Non-responding virological response
NWT	non-wild-type
OI	Opportunistic infection
PCR	Polymerase chain reaction
PI	Protease inhibitor
PIR	Positive immunological response

Abbreviations

PMCC	Pearson product moment correlation coefficient
Pol	Polymerase
PR	Protease
PS	Positive selection
PTA	Phenotypic assays
PVR	Positive virological response
RCG	Resistance collaborative group
RE	Resistance emergence
RFLP	Restriction fragment-length polymorphism
RNA	Ribonucleic acid
RPAH	Royal Prince Alfred Hospital
RT	Reverse transcriptase
RT-PCR	Reverse transcriptase-polymerase chain reaction
RTV	Ritonavir
SARS	Severe acute respiratory syndrome
SPSS	Statistical Package for the Social Sciences
SQV	Saquinavir mesylate
STI	Structured treatment interruption
T-20	Enfuvirtide
TA	Toxicity of CART
TAMs	Thymidine analogue mutations
TDF	Tenofovir disoproxil fumarate
TMC114	Darunavir
TPV	Tipranavir
UNAIDS	United Nations Programme on HIV/AIDS
vDNA	viral DNA
VL	Viral load
WHO	World Health Organisation
WT	wild-type

PUBLICATIONS FROM THESIS

1. A. A. Mazari, A. Y. Zomaya, M. Charleston, H. Salem, A. Maher, and R. J. Garsia, "Immunological and virological responses correlated with evolution of resistance in patients treated with antiretroviral agents. Abstract No: 95," presented at 18th Annual Conference of the Australasian Society for HIV Medicine, Melbourne, VIC, Australia, 2006.
2. A. A. Mazari and A. Y. Zomaya, "A Brief Overview of Grid Activities for Bioinformatics and Health Applications (Chapter 20)," in *Parallel Computing for Bioinformatics and Computational Biology*, A. Y. Zomaya, Ed. NY, USA: John Wiley & Sons, 2005.
3. A. A. Mazari, A. Y. Zomaya, M. Charleston, and R. J. Garsia, "A Novel Algorithm for Adaptive and Neutral Evolutionary Patterns Associated with HIV Drug Resistance," The fifth ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-07), [Accepted], Amman, Jordan, 2007.
4. A. A. Mazari, A. Y. Zomaya, M. Charleston, and R. J. Garsia, "Bayesian Networks for Modelling HIV Drug Resistance Dynamics, [Extended Abstract]," The fifth ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-07) [Accepted], Amman, Jordan, 2007.
5. A. A. Mazari, A. Y. Zomaya, M. Charleston, H. Salem, A. Maher, and R. J. Garsia, "Lability of Antiretroviral Drug Resistance Mutations—Correlates with Immunological and Virological Responses," *Current HIV Research* [Accepted], 2007.
6. A. A. Mazari, A. Y. Zomaya, M. Charleston, and R. J. Garsia, "Adaptive and Neutral Evolutionary Mutations at the HIV-RT Gene Associated with Resistance to NNRTIs," the 4th IAS Conference on HIV Pathogenesis, Treatment and Prevention (IAS 2007), [Submitted], Sydney, Australia, 2007.