Investigating the effects of structure on sulfate recognition by neutral dipeptide receptors

Nikki A. Tzioumis, Karen K.Y. Yuen and Katrina A. Jolliffe*

*aSchool of Chemistry, The University of Sydney, Sydney, 2006, NSW, Australia

* kate.jolliffe@sydney.edu.au

Keywords: anion recognition, sulfate, squaramide, thiourea

Abstract:
A small library of neutral peptide-based anion receptors was synthesised, where changes were made to the scaffold structure to investigate the effect these structural features have on the anion binding ability of these receptors. These changes included shortening the peptide side chain lengths, increasing the number of electron withdrawing substituents present on the squaramide phenyl substituents and increasing the length and flexibility of the peptide backbone. An effort was also made to increase the aqueous solubility of these receptors by functionalising the N-terminus of the peptide with a hydrophilic moiety. All the receptors displayed strong affinity and selectivity for sulfate in 20% v/v H₂O/DMSO-\textit{d₆} and a 5-fold increase in the affinity of the thiourea receptors was observed upon shortening the side chains by one methylene unit. Overall, the squaramide derivatives displayed much stronger association, in this competitive media, than the thiourea based receptors.
**Introduction**

Anions play an integral role in many biological and chemical processes and recent interest in the potential applications of molecules capable of the selective recognition of anionic species in biology, medicine and catalysis has spurred a surge of research in this field. This interest has also been heightened by the need to detect and separate the abundant anionic pollutants that exist in industrial wastes before disposal.\(^{(1-2)}\)

Sulfate is an anion of particular interest as it is required for proper cell growth and development; is involved in biosynthesis and detoxification; and imbalances in sulfate homeostasis are known to be caused by many human diseases.\(^{(3-4)}\) It is also a large source of contamination in aquatic ecosystems, *via* coal mining, aerial deposition and leakage from agricultural systems, which leads to increased acidification, deoxygenation and the mobilization of heavy metals in the affected ecosystems.\(^{(5-6)}\) High concentrations of sulfate can also cause problems in radioactive waste remediation, resulting in a product that exhibits increased leach rates and low chemical durability.\(^{(7-9)}\) However, due to the high solvation energy of sulfate in water (\(\Delta G = -1080 \text{ kJ mol}^{-1}\))\(^{(10)}\) binding to sulfate in these aqueous systems is very challenging as a potential receptor needs to be able to overcome the strong sulfate-water interactions to be effective.\(^{(11-12)}\)

Although hydrogen bonding interactions with anions are generally weaker in aqueous solution than other binding interactions, such as electrostatic and metal-ligand interactions, hydrogen bonds are highly directional. This allows for the design of receptors with sizes and geometries that can be tailored to selectively bind to different anionic species in solution.\(^{(13)}\) This property has been utilised highly effectively in natural anion binding proteins such as the sulfate binding protein which utilises seven hydrogen bond donors to completely desolvate sulfate.\(^{(14-15)}\)

A variety of H-bond motifs including ureas, thioureas, amides and squaramides have been utilised to bind to anions,\(^{(16)-(17)}\) but very few of these are effective in water, unless combined with electrostatic interactions.\(^{(18-21)}\) However, Kubik and co-workers have recently reported a water soluble bis-macrocyclic peptide that binds to chloride, bromide, iodide and sulfate ions through hydrogen bonding interactions in water, although this receptor does not discriminate between these anions (sulfate \(\log K = 3.31\); iodide \(\log K = 3.62\)).\(^{(22)}\)
Squaramides were first employed as an anion binding motif in 1997 by Costa and co-workers (23) and since then have been shown to have improved anion affinities when compared to analogous urea and thiourea derivatives (24-27) and also to be potent transmembrane anion transporters (26). We have recently described a number of squaramide based receptors including a macrocyclic receptor which exhibited strong affinity and selectivity for sulfate in 33% v/v H₂O/DMSO (28). We have also previously observed that neutral dipeptides 1 and 2, bearing side chains functionalised with thioureas or squaramides (Figure 1), bind to anions in mixtures of DMSO and water, and show surprisingly high affinity and selectivity for sulfate ions in competitive media (29). This was attributed to binding of this anion by both the side chain H-bond donor groups and the peptide backbone amide protons, as had previously been observed for thiourea-functionalised cyclic peptides (30-32). In the case of these cyclic peptide receptors, shortening the side chains to reduce the distance between the cyclic peptide scaffold and the side chain H-bond donors resulted in significantly increased binding affinity for sulfate (33). To investigate whether a similar effect might be observed with the linear dipeptide receptors, we aimed to prepared analogues of 1 and 2 in which the lysine (Lys) side chains were replaced by shorter sidechains derived from ornithine (Orn) and 2,4-diaminobutyric acid (Dab) to evaluate their anion recognition properties.
Results and Discussion

For our initial studies, we prepared receptors 3 and 4 (Figure 1), in which the Lys sidechains of 1 and 2 were replaced by the shorter Orn sidechains. Receptors 3 and 4 were prepared using Fmoc solid-phase peptide synthesis (Fmoc-SPPS) on Rink amide resin with PyBOP/DIPEA as the coupling agent and utilising on-resin side chain functionalisation as described previously for the synthesis of 1 and 2.(29)

In a previous investigation of the anion binding properties of receptors 1 and 2 it was found that these compounds exhibit strong binding to SO$_4^{2-}$, with association constants ($K_a$) of 1282 M$^{-1}$ and $>10^4$ M$^{-1}$, respectively in 20% v/v H$_2$O/DMSO-$d_6$. Of the other anions screened (AcO$^-$, BzO$^-$, H$_2$PO$_4^-$, Br$^-$, HSO$_4^-$, NO$_3^-$, TsO$^-$, and Cl$^-$ added as their tetrabutylammonium (TBA) salts) only AcO$^-$ and BzO$^-$ showed measurable binding affinities to these receptors and binding of both 1 and 2 to these carboxylates was significantly weaker than their binding to sulfate (Table 1). Notably, the squaramide based receptor 2 showed higher binding affinity for sulfate when compared to the thiourea
based receptor 1, demonstrating the potential of the squaramide moiety as a strong anion binding motif.

To directly compare the effect that shortening receptor side chain length had on their binding affinity, quantitative $^1$H NMR spectroscopic titration experiments were conducted on receptors 3 and 4 with SO$_4^{2-}$, AcO$^-$ and BzO$^-$ in 20% v/v H$_2$O/DMSO-$d_6$ by adding aliquots of the anions as TBA salts to a solution of receptor. Due to the high content of non-deuterated water that is present in solution, we employed water suppression by gradient-tailored excitation (WATERGATE)(34) during the $^1$H NMR spectroscopic experiments. The observed changes in chemical shift of the signals attributable to the thiourea, squaramide and/or peptidic amide protons were then fitted to a 1:1 binding model using HypNMR® software(35) to determine the association constants summarised in Table 1 (see Supporting Information for fitted titration isotherms and stack plots).

Table 1. Association constants ($K_a$, M$^{-1}$) for receptors 1-4 determined by $^1$H NMR spectroscopic titrations, with various anions (as their TBA salts), in 20% v/v H$_2$O/DMSO-$d_6$.

<table>
<thead>
<tr>
<th>Receptor</th>
<th>SO$_4^{2-}$</th>
<th>AcO$^-$</th>
<th>BzO$^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>1282</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>2*</td>
<td>$&gt; 10^4$</td>
<td>45</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>6866</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>$&gt; 10^4$</td>
<td>168</td>
<td>97</td>
</tr>
</tbody>
</table>

Notes: Titrations were performed at 300 K. Proton shifts were fitted to a 1:1 binding model and $K_a$ values are an average of the obtained fits with errors $\leq$ 15%. * Values obtained from Elmes et. al. (29)

The association constants obtained (Table 1) show that the Orn derived thiourea receptor 3 exhibits a 5-fold increase in binding affinity for SO$_4^{2-}$ ($K_a = 6866$ M$^{-1}$) when compared to Lys derived thiourea receptor 1 ($K_a = 1282$ M$^{-1}$), supporting our theory that side chain length can be used to tune the H-bond interactions of the backbone amide groups with the anion. In all cases, there is an observable shift of the peaks attributable to the peptidic N-H protons in the presence of the above anions, supporting the hypothesis that they contribute significantly towards binding to these anions. Interestingly, shortening the side
chain length had negligible effect on the binding affinity of receptor 3 to AcO$^-$ and BzO$^-$.

(30-32)

Analysis of the effect that side chain length had on the binding affinity of the squaramide based receptors to SO$_4^{2-}$ could not be determined from this data because the binding affinities for these receptors were too strong ($K_a > 10^4$ M$^{-1}$) in 20% v/v H$_2$O/DMSO-$d_6$ to accurately quantify using $^1$H NMR spectroscopic experiments. We reasoned that increasing the competitive nature of the solvent by increasing the water content would allow this comparison, but unfortunately the solubility of these compounds in solvent mixtures with higher concentrations of water was too low to enable a quantitative determination of the differences in binding affinity of these receptors to SO$_4^{2-}$.

**Design and Synthesis of 5-10**

In an attempt to improve the aqueous solubility of these receptors we next prepared a series of analogues (Figure 2) of the previous squaramide receptors where a hydrophilic “water-solubilising” group, the 3,4,5-tris(2-(2-(2-methoxyethoxy)ethoxy)ethoxy) (TriTEG) benzoyl moiety, was appended to the $N$-terminus of the peptide. We envisaged that appending a group at the peptide $N$-terminus would not significantly affect the anion binding behaviour of these compounds, making it a convenient site for functionalisation. Only derivatives of the squaramide based receptors were synthesised as they exhibited much stronger binding affinities than the thiourea based receptors.

We therefore prepared Lys based receptor 5 to use as a direct comparison with receptor 2 as well as the Dab based receptor 7, which has a shorter side chain length, to further investigate the effect the distance between the squaramide H-bond donors and the backbone H-bond donors has on the binding ability of the receptors. We chose to use Dab rather than Orn in these receptors because our previous studies with cyclic peptide based sulfate receptors indicated that further reducing the side chain length from Orn to Dab resulted in increased SO$_4^{2-}$ binding affinity.(33)

In addition, we prepared compound 9 which has a glycine (Gly) moiety inserted between the functionalised Lys amino acids to investigate whether additional space and flexibility between the functional squaramide side chains would increase binding affinity and/or solubility. Receptors 6, 8, and 10, which are direct analogues of 5, 7 and 9, bearing an
additional trifluoromethyl group on the squaramide aromatic substituent, were also prepared as it has previously been observed that the incorporation of an additional electron withdrawing group on the aryl ring of the squaramide leads to increased anion affinities.(26)

Figure 2. Structure of squaramide receptors 5-9.

The synthesis of receptors 5-10 (Scheme 1) was achieved via Fmoc-SPPS as described above but with two major differences to our previous syntheses. Firstly, the initial resin loading was performed using Fmoc protected Gly (Fmoc-Gly-OH), since it has previously been shown that including a C-terminal Gly spacer increases the efficiency of subsequent functionalisation reactions on the solid phase.(36) Secondly, instead of an acetylation capping step, the N-terminus of the peptide was functionalised with the water solubilising moiety using TriTEG-benzoic acid 11 using PyBOP as the coupling agent. Side-chain functionalisation and cleavage of the peptide from the resin proceeded under the same conditions as previously to give squaramide receptors 5-10 in yields ranging from 35-58%
Scheme 1. General synthetic route for the preparation of 5-10.

To investigate the binding ability of receptors 5-10, we first performed screening experiments by adding 12 equivalents of anions (AcO\textsuperscript{–}, BzO\textsuperscript{–}, Cl\textsuperscript{–}, H\textsubscript{2}PO\textsubscript{4}\textsuperscript{–} and SO\textsubscript{4}\textsuperscript{2–} as their TBA salts), to a 2.5 mM solution of receptor in 20% v/v H\textsubscript{2}O/DMSO-d\textsubscript{6}. We chose to screen this selection of anions as they had previously shown affinity to receptor 2 in 0.5% v/v H\textsubscript{2}O/DMSO-d\textsubscript{6}. (29) The spectra for receptor 5 (Figure 3) show large downfield shifts of the signals attributable to the squaramide protons in the presence of SO\textsubscript{4}\textsuperscript{2–} and a modest downfield shift in the presence of Cl\textsuperscript{–}. In the presence of AcO\textsuperscript{–}, BzO\textsuperscript{–} and H\textsubscript{2}PO\textsubscript{4}\textsuperscript{–} the signals attributable to the squaramide protons exhibited peak broadening behaviour, however significant downfield shifts in the signals attributable to the peptidic amide protons as well as distinct shifts of the signals attributable to the aromatic protons were also exhibited, which could be monitored to calculate the association constants of the receptors to these anions. This behaviour is consistent with that observed previously for receptor 2, (29) and indeed this general behaviour was observed for all the receptors 5-10.
Figure 3. $^1$H NMR spectra from the anion screening of 5 in 20% v/v H$_2$O/DMSO-$d_6$ ($\blacktriangleleft$ = NH$_A$ (squaramide), $\blacklozenge$ = NH (peptide backbone) and $\blackcircle$ = ArH (aromatic proton)). Symbols identify signals employed for data fitting to the 1:1 binding model.

To analyse and compare the differences in anion binding affinity of receptors 5-10, quantitative titration experiments were then undertaken in 20% v/v H$_2$O/DMSO-$d_6$. The association constants obtained (Table 2) indicate that while binding to Cl$^-$ was weak for all receptors, the 3,5-bis-CF$_3$-squaramide functionalised receptors (6, 8 and 10) generally exhibit stronger binding to AcO$^-$, BzO$^-$ and H$_2$PO$_4$$^-$ than the p-CF$_3$-squaramide receptors (5, 7 and 9). This is likely due to the increased acidity of the squaramide proton (NH$_A$) resulting from increased electron withdrawing groups on the phenyl ring appended to the squaramide.(24, 26) All receptors exhibited high affinity and selectivity for SO$_4$$^{2-}$ in this solvent system, with $K_a >10^4$ M$^{-1}$ regardless of structural differences in the peptidic scaffold or the number of electron withdrawing groups on the phenyl substituent appended to the squaramide.
Table 2. Association constants ($K_a$, M$^{-1}$) for receptors 5-10 determined by $^1$H NMR spectroscopic titrations, with various anions (as their TBA salts), in 20% v/v H$_2$O/DMSO-$d_6$.

<table>
<thead>
<tr>
<th>Receptor</th>
<th>$SO_4^{2-}$</th>
<th>Cl$^-$</th>
<th>AcO$^-$</th>
<th>BzO$^-$</th>
<th>H$_2$PO$_4^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$&gt;10^4$</td>
<td>15</td>
<td>39</td>
<td>33</td>
<td>47</td>
</tr>
<tr>
<td>6</td>
<td>$&gt;10^4$</td>
<td>26</td>
<td>294</td>
<td>99</td>
<td>239</td>
</tr>
<tr>
<td>7</td>
<td>$&gt;10^4$</td>
<td>13</td>
<td>60</td>
<td>86</td>
<td>196</td>
</tr>
<tr>
<td>8</td>
<td>$&gt;10^4$</td>
<td>26</td>
<td>111</td>
<td>131</td>
<td>206</td>
</tr>
<tr>
<td>9</td>
<td>$&gt;10^4$</td>
<td>26</td>
<td>33</td>
<td>23</td>
<td>84</td>
</tr>
<tr>
<td>10</td>
<td>$&gt;10^4$</td>
<td>15</td>
<td>58</td>
<td>65</td>
<td>437</td>
</tr>
</tbody>
</table>

Notes: Titrations were performed at 300 K at 2.5mM receptor concentration. Proton shifts were fitted to a 1:1 binding model and $K_a$ values are an average of the obtained fits with errors $\leq$ 15%.

To study the effect the differences in design have on the binding ability of the receptors for $SO_4^{2-}$, investigations in water-DMSO mixtures with concentrations of water greater than 20% need to be undertaken. Unfortunately, it was found that although these receptors are functionalised with a highly hydrophilic moiety, this did not provide a significant increase in aqueous solubility when compared to receptors 1-4, with receptors 5-10 exhibiting solubility thresholds ranging from 20-25% v/v H$_2$O/DMSO at 2.5 mM receptor concentration. Therefore, further investigations in solvent systems with higher water concentrations could not be undertaken with these receptors.

Although further investigations into the anion binding ability of these receptors were not able to be performed, density functional theory (DFT) calculations were performed on the Dab based receptor 7 and glycine spaced receptor 9 complexed to $SO_4^{2-}$ (Figure 4), at the B3LYP/6-31G* level of theory, to investigate the possible differences in the binding modes of these receptors to $SO_4^{2-}$. The modelled structures show that both the backbone amide protons and squaramide protons can participate in the binding by both receptors. They also show that additional H-bonds are donated by the C-terminal Gly residue that was included as a spacer from the resin, which acts as a potential third “arm” donating potential H-bonds to $SO_4^{2-}$. This indicates that not only is the Gly spacer advantageous for efficient synthesis, but it could also be an effective binding motif in
these receptors. This is also evidenced by the $^1$H NMR spectroscopic titrations with SO$_4^{2-}$, in which the triplet peak attributed to the Gly amide proton exhibits a significant downfield shift, consistent across all receptors. While the tri-TEG benzyol moiety was replaced by 3,4,5-trimethoxybenzyol in the modelled structures, it is clear from these structures that functionalisation at the N-terminus of the peptide does not appear to interfere with the binding of the receptor to SO$_4^{2-}$. Although these binding models are not conclusive evidence of the effectiveness of the receptors, they provide an indication of the possible binding modes of the receptors and are a useful tool for determining whether some design aspects are worth investigating further.

Figure 4. Left: DFT calculated structure of Dab based receptor 7 and SO$_4^{2-}$ complex showing the anion in the binding cavity with 9 NH-O hydrogen bonds interactions; Right: DFT calculated structure of glycine spaced receptor 9 and SO$_4^{2-}$ complex showing the anion in the binding cavity with 10 NH-O hydrogen bonds interactions.

**Conclusion**

We have synthesised a small library of linear peptide based receptors to investigate the effect that different design features in the receptors have on their binding affinity to SO$_4^{2-}$. These features include: shortening the side chain length, increasing length and flexibility of receptor backbone, increasing the acidity of the squaramide proton NH$_A$ and introducing a water-solubilising group to the N-terminus of the peptide scaffold. Overall, the receptors maintained strong affinity and selectivity for SO$_4^{2-}$ in 20% v/v H$_2$O/DMSO-$d_6$, and while the receptor affinity and selectivity to other anions (AcO$^-$, BzO$^-$ and H$_2$PO$_4^-$) varied slightly, no significant changes in selectivity were observed in this solvent
mixture. Unfortunately, the incorporation the highly hydrophilic Tri-TEG group, appended to the N-terminus of the peptide scaffold, did not increase receptor solubility by a significant amount and we were unable to test the effect the differences in receptor design had on binding affinity to $\text{SO}_4^{2-}$ in $\text{H}_2\text{O/}D\text{MSO-}d_6$ mixtures with higher concentrations of water. Future work will focus on the synthesis and evaluation of derivatives with improved water solubility.
References:

(2) Gale, P. A., Chem. Commun. 2011, 47, 82-86.
(22) Sommer, F.; Marcus, Y.; Kubik, S., ACS Omega 2017, 2, 3669-3680.