Binding Methylarginines and Methyllysines as Free Amino Acids: A Comparative Study of Multiple Host Classes

Zoey Warmerdam, Bianca E. Kamba, My-Hue Le, Thomas Schrader, Lyle Isaacs, Peter Bayer, Fraser Hof

November 2021

© 2021 Warmerdam et al. This is an open access article distributed under the terms of the Creative Commons Attribution License. https://creativecommons.org/licenses/by-nc/4.0/

This article was originally published at:

https://doi.org/10.1002/cbic.202100502

Citation for this paper:

Binding Methylarginines and Methyllysines as Free Amino Acids: A Comparative Study of Multiple Host Classes**


Dedicated to François Diederich, enthusiastic promoter of collaboration and partnerships.

Methylated free amino acids are an important class of targets for host-guest chemistry that have recognition properties distinct from those of methylated peptides and proteins. We present comparative binding studies for three different host classes that are each studied with multiple methylated arginines and lysines to determine fundamental structure-function relationships. The hosts studied are all anionic and include three calixarenes, two acyclic cucurbiturils, and two other cleft-like hosts, a clip and a tweezer. We determined the binding association constants for a panel of methylated amino acids using indicator displacement assays. The acyclic cucurbiturils display stronger binding to the methylated amino acids, and some unique patterns of selectivity. The two other cleft-like hosts follow two different trends, shallow host (clip) following similar trends to the calixarenes, and the other more closed host (tweezer) binding certain less-methylated amino acids stronger than their methylated counterparts. Molecular modeling sheds some light on the different preferences of the various hosts. The results identify hosts with new selectivities and with affinities in a range that could be useful for biomedical applications. The overall selectivity patterns are explained by a common framework that considers the geometry, depth of binding pockets, and functional group participation across all host classes.

Introduction

The selective binding of free amino acids in physiologically relevant solutions is difficult to achieve. In neutral aqueous solution arginine and lysine are zwitterionic and have a relatively small hydrophobic surface area, combined with relatively high charge, which means that they are strongly solvated by water. This results in unfavourable binding energy upon complexation with the host, as several water molecules need to be released from strong hydrogen bonds to the amino acid’s charged groups. The common presence of salts and other competing co-solutes creates further challenges.[1]

There are dozens of diseases caused by disordered amino acid metabolism.[2] Directly capturing and sequestering disease-related amino acids is a new approach that might be useful for diagnosis, disease monitoring, and possibly for direct therapeutics.[3] Antibodies are known that bind amino acids but have some inherent shortcomings[4] that could be overcome by the creation of synthetic organic binding tools for amino acids. Synthetic hosts are a great starting point to develop new tools. The hosts have a concave binding pocket where the molecular recognition takes place via non-covalent interactions (e.g., charge-charge interactions, hydrogen bonding, van der Waals forces, π-π interactions, etc.) and the hydrophobic effect. These binding pockets come with different shapes, sizes and chemical properties. The recognition of amino acids by hosts has recently been reviewed by Basilio et al.[5] Examples of hosts binding to free amino acids include calixarenes,[6] pillararenes,[7] cucurbiturils,[8] and cyclodextrins.[9]

We studied a set of seven hosts from three different classes to get a better fundamental understanding of the structure-function relationship for amino acid binding by multiple different kinds of hosts. Calixarenes are relative shallow but easily functionalized molecules, where the functionalization often directly lines the binding pocket.[6] We chose a parent anionic calixarene, sCx4, and the functionalized analogues sCx4-NO2 and sCx4-CHO (Figure 1a–c) to include in this study after a

---

[a] Z. Warmerdam,* F. Hof
Department of Chemistry and the Centre for Advanced Materials and Related Technology
University of Victoria
3800 Finnerty Rd, V8W 3V6 Victoria, BC (Canada)
E-mail: fhof@uvic.ca

[b] B. E. Kamba,* P. Bayer
Department of Structural and Medicinal Biochemistry
Universität Duisburg Essen
Universitätstraße 2, 45141 Essen (Germany)
E-mail: peter.bayer@uni-due.de

[c] M.-H. Le, T. Schrader
Department of Chemistry
Universität Duisburg Essen
Universitätstraße 7, 45117 Essen (Germany)

[d] L. Isaacs
Department of Chemistry and Biochemistry
University of Maryland, College Park
MD 20742 College Park (USA)
E-mail: lisaacs@umd.edu

[∗] These authors contributed equally to the work.
[**] A previous version of this manuscript has been deposited on a preprint server (https://doi.org/10.33774/chemrxiv-2021-2b4b).

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202100502

This article belongs to a Joint Special Collection dedicated to François Diederich.

preliminary screen (not shown) ruled out some weaker-binding calixarenes. Cucurbituril (CB) hosts have a deeper and more rigid binding pocket, but the functional group additions to CBs often do not directly influence the binding properties as they happen on the outside of the host. Acyclic CB analogues such as M1 and M2 (Figure 1d,e) can be functionalized along the edge of their binding surfaces, and are increasingly being used in biomedical applications. Another more rigid cleft-like host family includes the ‘tweezers’ and ‘clips’ introduced by Klärner and Schrader. The “phosphate clip” PC carries planar aromatic sidewalls ideal for aromatic cations, whereas the “tweezers” CLR01 form a torus-shaped cavity that is selective for arginine and lysine (Figure 1f,g).

For this fundamental, comparative study we have selected methylated amino acids as binding targets. Our test set (Figure 2) includes arginine and lysine, and each of their physiologically relevant methylated states: (monomethyl arginine (MMA), asymmetric dimethyl arginine (ADMA), symmetric dimethyl arginine (SDMA), monomethyl lysine (MML) dimethyl lysine (DML) and trimethyl lysine (TML)). They have subtle structural differences among them, which make them an interesting test case for molecular recognition and selectivity. They participate in diverse biological pathways that are relevant to multiple pathological states. Methylated arginines have been linked to a wide array of diseases such as cardiovascular disease, renal disease, and others. While the exact values are sometimes controversial, meta-analyses show that the ADMA plasma concentration in healthy adults is ~0.7 μM, and higher in various disease states (with some studies showing plasma [ADMA] for disease conditions as high as 5–7 μM). SDMA plasma levels are lower (<1 μM). The plasma concentration of the methylated lysines has not significantly been linked to disease, but studying them will provide us insight into this family of complexes.

Lots of research has been focused on binding methylated lysines, while the selective binding of methylated arginine has received relatively less attention. Most prior literature has focused on binding these methylated residues in the context of proteins and peptides. The current study on free amino acids is motivated by a body of literature demonstrating that some of the free methylated amino acids are metabolites that play critical biological roles in pathways that are distinct from those involving whole proteins, protein tails, and peptides with post-translationally methylated residues.

Figure 1. Hosts studied in this report. Calixarenes a) sCx4, b) sCx4-CHO, and c) sCx4-NO₂. Cleft-like hosts d) M1, e) M2, f) PC, and g) CLR01.

Figure 2. Guests and indicators studied in this report. a) Guests arginine, monomethylarginine (MMA), asymmetric dimethylarginine (ADMA), symmetric dimethylarginine (SDMA), lysine, monomethyllysine (MML), dimethyllysine (DML), and trimethyllysine (TML), b) Indicators LCG, 4-ASP, and R6G.
Results and Discussion

The binding constants were determined for the complexes formed by each member of the host library with each of the guests, using indicator displacement assays (IDAs). We adapted previously reported IDAs[22] for sCx4 (using lucigenin, Figure 2b)[23] and M1 and M2 (using Rhodamine 6G, Figure 2b).[24] We established IDAs for PC and CLR01 during this work. Initial experiments using Rhodamine 6G, Proflavine, and Neutral Red revealed either non-ideal stoichiometries of binding or inadequate intensity changes upon binding. Studies using 4-ASP (Figure 2b) as the indicator gave reliable results for PC and CLR01 with the whole panel of guests. Each host-indicator dissociation constant (Kd) was determined by a direct titration of the host into indicator, and the host-guest equilibrium dissociation constants (Kd) were then determined using competitive titrations of guests into a pre-formed host-indicator complex. The calixarene and the acyclic CB IDAs gave a turn-on signal, where displacement of the indicator by the guest results in an increase of fluorescence emission. The PC and CLR01 host provided a turn-off signal where displacement of 4-ASP quenches its fluorescence emission. All the titrations were optimized to work in a 10 mM NaH2PO4 buffer at pH 7.4. This choice of buffer rules out the ability to measure weak (> mM) binding (which would require analyte concentrations that would overwhelm the buffer), but it ensures that the trends for stronger-binding guests can be compared across different host types. The studied host-guests show a range of affinities (Table 1), which we categorize as follows for convenient presentation: strong binding (Kd < 200 μM), medium to strong binding (Kd 200–1500 μM) and weak to no binding (Kd > 1500 μM). We also presented the data as a 3D bar graph (Figure 3), to help visualise the binding trends.

For the calixarene hosts, strong binding is only observed for a few host-guest combinations. A binding trend is observed for the calixarene-guest complexes in which higher methylation states generally result in stronger binding. Most notably, trimethyllysine (TML) is the guest that shows strongest binding states generally result in stronger binding. Most notably, trimethyllysine (TML) is the guest that shows strongest binding. A binding trend is observed for Figure 3, to help visualise the binding trends.

Among the stronger binding, ADMA displays a ~2-fold stronger binding. Interestingly, sCx4-NO2 selectively binds ADMA over the other arginine guests, for which no binding is observed. Overall, the sCx-hosts bind the free amino acids weakly compared to physiological concentration ranges for these amino acids.

Table 1. Kd values determined by IDA for each host-guest complex in 10 mM phosphate buffer.[25]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>75 ± 20</td>
</tr>
<tr>
<td>MAA</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>45 ± 10</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>135 ± 30</td>
</tr>
<tr>
<td>ADMA</td>
<td>500 ± 50</td>
<td>230 ± 50</td>
<td>290 ± 60</td>
<td>10 ± 5</td>
<td>60 ± 20</td>
<td>&gt; 1500</td>
<td>650 ± 130</td>
</tr>
<tr>
<td>SDMA</td>
<td>1090 ± 300</td>
<td>370 ± 150</td>
<td>&gt; 1500</td>
<td>35 ± 10</td>
<td>20 ± 5</td>
<td>&gt; 1500</td>
<td>810 ± 270</td>
</tr>
<tr>
<td>Lysine</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
</tr>
<tr>
<td>MML</td>
<td>930 ± 120</td>
<td>470 ± 110</td>
<td>610 ± 130</td>
<td>160 ± 60</td>
<td>130 ± 90</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
</tr>
<tr>
<td>DML</td>
<td>300 ± 40</td>
<td>240 ± 50</td>
<td>410 ± 60</td>
<td>70 ± 10</td>
<td>130 ± 90</td>
<td>30 ± 5</td>
<td>980 ± 710</td>
</tr>
<tr>
<td>TML</td>
<td>120 ± 20</td>
<td>100 ± 30</td>
<td>150 ± 30</td>
<td>15 ± 5</td>
<td>12 ± 5</td>
<td>65 ± 20</td>
<td>1040 ± 360</td>
</tr>
</tbody>
</table>

[a] All titrations were carried out in 10 mM NaH2PO4 buffer at pH 7.4. See the Supporting Information for titration curves experimental details, and fitting details. All Kd values arise from fits with R² ≥ 0.95 except where indicated. [b] Lucigenin (LCG) was used as indicator. [c] Rhodamine 6G (R6G) was used as indicator. [d] 4-(4-Diethylamino styryl)-1-methylpyridinium iodide (4-ASP) was used as indicator. [e] R² ≥ 0.92. [f] Value arises from one triplicate.
binding. When looking at the lysine guests, the binding is weak with relatively large uncertainties, and no strong conclusions can be drawn relating to binding trends. CLR01 shows different binding trends than the other hosts. The binding is strong, and we see that the strong binding and binding strength for each unmethylated amino acid.

To determine the effect of a higher salt concentration, IDAs on the amino acids were also done in a more concentrated 50 mM phosphate buffer with the three strongest binding hosts: M1, M2 and CLR01. We see that the strong binding and the binding trends are maintained for M2 and CLR01. The binding strength for M1 becomes 2- to 4-fold weaker for each guest, but still remains in the “strong” range of $K_d$ values, see Table 2.

Studies with a small set of simpler guests show how the impact of hydrophobicity manifests differently in the different host classes. We did IDA studies with dimethylammonium chloride, tetramethylammonium chloride, and benzyltrimethylammonium chloride as controls that display increasing hydrophobicity without neighbouring polar functionality and with relatively simple changes in size and shape (Supporting Information). Dimethylammonium is the smallest, least hydrophobic, resulting in weak binding ($K_d > 100 \mu M$); the slightly more hydrophobic MML binds more strongly at $K_d = 15 \mu M$. Besides this one exception, each other amino acid binds progressively weaker with increasing hydrophobicity. Unlike for all other hosts, dimethylarginines are not measurably bound by CLR01 under the conditions of the experiment.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>90 ± 25</td>
</tr>
<tr>
<td>MMA</td>
<td>180 ± 60</td>
<td>&gt; 1500</td>
<td>155 ± 30</td>
</tr>
<tr>
<td>ADMA</td>
<td>30 ± 10</td>
<td>125 ± 45</td>
<td>&gt; 1500</td>
</tr>
<tr>
<td>SDMA</td>
<td>55 ± 15</td>
<td>40 ± 10</td>
<td>&gt; 1500</td>
</tr>
<tr>
<td>Lysine</td>
<td>&gt; 1500</td>
<td>&gt; 1500</td>
<td>50 ± 15</td>
</tr>
<tr>
<td>MML</td>
<td>300 ± 150</td>
<td>&gt; 1500</td>
<td>35 ± 5</td>
</tr>
<tr>
<td>DML</td>
<td>210 ± 50</td>
<td>190 ± 110</td>
<td>55 ± 10</td>
</tr>
<tr>
<td>TML</td>
<td>30 ± 5</td>
<td>25 ± 5</td>
<td>40 ± 5</td>
</tr>
</tbody>
</table>

[a] All titrations were carried out in 50 mM NaH$_2$PO$_4$ buffer at pH 7.4. See the Supporting Information for titration curves, experimental details, and fitting details. All $K_d$ values arise from fits with $R^2 \geq 0.95$ except where indicated. [b] Rhodamine 6G (R6G) was used as indicator. [c] 4-(4-Diethylaminostyryl)-1-methylpyridinium iodide (4-ASP) was used as indicator. [d] $R^2 \geq 0.93$.

The hosts studied in this paper (except sCx4-CHO, sCx4-NO2 and PC) have previously been studied via ITC with lysine and arginine. From those studies, we see that enthalpy is the dominant driving force across all hosts (Table S2). ITC studies for sCx4 with methyllysine and methylarginine analogues show that increasing methylation is accompanied by significant increases in enthalpic driving force, and smaller changes in entropy. As in all similar work, we caution against drawing specific mechanistic conclusions for host-guest ITC studies that are done in salty aqueous solution, because salt effects are often large and unpredictable contributors to the heats of binding.

We did molecular modeling for six of the hosts to provide a general view of the differences and similarities between the hosts for certain key complexes. Molecular modeling was done using minimization in explicit water for all the indicated complexes using Maestro (Schrodinger, Inc.). Host sCx4-CHO, sCx4-NO2, M1, M2, PC, CLR01 were modelled in complex with ADMA (Figure 4) and hosts M1, M2, PC, CLR01 were also modelled each in complex with MML and TML (Supporting Information) in order to gain further insight into selectivity among these guests. The models reveal the extent and nature of interactions between hosts and guests and show qualitatively that the different hosts generally fit into two classes: open geometry hosts that engage only the charged side chain of the guests (the calixarenes and PC), and closed geometry hosts that almost completely surround their guests (M1, M2, ADMA, tetramethylammonium chloride, and benzyltrimethylammonium chloride as controls that display increasing hydrophobicity without neighbouring polar functionality and with relatively simple changes in size and shape (Supporting Information). Dimethylammonium is the smallest, least hydrophobic, resulting in weak binding ($K_d > 100 \mu M$); the slightly more hydrophobic MML binds more strongly at $K_d = 15 \mu M$. Besides this one exception, each other amino acid binds progressively weaker with increasing methylation. Unlike for all other hosts, dimethylarginines are not measurably bound by CLR01 under the conditions of the experiment.

To determine the effect of a higher salt concentration, IDAs on the amino acids were also done in a more concentrated 50 mM phosphate buffer with the three strongest binding hosts: M1, M2 and CLR01. We see that the strong binding and the binding trends are maintained for M2 and CLR01. The binding strength for M1 becomes 2- to 4-fold weaker for each guest, but still remains in the “strong” range of $K_d$ values, see Table 2.

Studies with a small set of simpler guests show how the impact of hydrophobicity manifests differently in the different host classes. We did IDA studies with dimethylammonium chloride, tetramethylammonium chloride, and benzyltrimethylammonium chloride as controls that display increasing hydrophobicity without neighbouring polar functionality and with relatively simple changes in size and shape (Supporting Information). Dimethylammonium is the smallest, least hydrophobic, resulting in weak binding ($K_d > 100 \mu M$) across the set of hosts. In shallow hosts like the calixarenes and PC, tetramethylammonium generally binds in the range of $K_d = 300–900 \mu M$, with approximately the same affinity as benzyltrimethylammonium in each case. The deeper hosts M1, M2, and CLR01 show clear preference for benzyltrimethylammonium over tetramethylammonium, as would be expected from the guest’s increased size and hydrophobicity. We rationalise that this difference in trends for benzyltrimethylammonium arises because the shallow hosts do not have any additional hydrophobic surface area available to interact with the larger guest.
and CLR01). They also reveal differences in host-guest interactions among the different host classes. The connections between these structural features and guest-binding selectivities are discussed below.

By comparing across host classes, we can derive some general lessons about the contributions from electrostatics, hydrophobicity, and geometric shape matching.

While electrostatics are undoubtedly important for molecular recognition, our binding data show that they are not the key determining factor for guest selectivity. We know from other literature that neutral guests do not bind the calixarene hosts as strongly,[6,4] where the tweezer-type host do have a precedent as strong binding host.[26,27] Yet, the comparison across all three host classes makes it clear that the selective binding of charged species in this relatively salty environment is not strongly controlled by the charges on hosts or guests. All amino acids studied have a zwitterionic α-amino-acid component, and a net charge of +1. The methylated guests have positive charges that are distributed significantly onto their methyl CH atoms, changing the nature of electrostatic interactions with the hosts.[28] The calixarenes and acyclic CBs have net charges of between −4 and −5, and yet bind guests with affinities that vary over >2 orders of magnitude. The clip-type hosts have net charges of between −2 and −4 (depending on the degree of second ionization of the phosphate groups) and generate $K_\text{b}$ values that vary by almost 100-fold between different guests. M1 and M2 have eight carbonyl groups that can form ion-dipole interactions, and two sulfonate groups that can form ion-ion interactions. Molecular modelling corroborates that the ion-dipole interactions are the main electrostatic interaction that is taking place for M1 and M2, whereas charged interactions between the guests and the hosts’ sulfonate arms are not as prominently seen in the energy-minimized structures (Figure 4c-d). The calixarenes also have multiple sulfonate groups, and in these complexes host-guest salt bridges are prominent features of the complexes (Figure 4a,b and Supporting Information). PC and CLR01 both have two phosphate groups that can have ion-ion interactions between the charged phosphate groups and the guanidinium, see Figure 4e,f. All the hosts can form cation-pi interactions with their cationic guests, although the geometric details vary.

Guest hydrophobicity is the main determinant for guest selectivity in hosts with more open geometries, including all calixarenes and PC. Methylation of arginine and lysine increases the volume of the head groups. This results in an increased hydrophobic surface area, a change to a more diffuse charge distribution, and decrease in the guest’s potential to form strong NH hydrogen bonds.[28] This modification is favourable when looking at the calixarenes and PC, where an increased number of methyl groups results in a stronger binding. This trend can clearly be observed for the lysine guests, where unmethylated lysine displays weak binding and TML displays strong binding to the hosts (Table 1). When looking at the dimethylated arginines we also see that position of the methyl groups has an influence. The models show that the hydrophobic surface area of ADMA is localized in one patch made up of two geminal methyl groups, whereas the hydrophobic surface area of SDMA is separated with methyls on two distal nitrogen atoms. In open-geometry hosts that do not constrain for molecular shape, this results in a stronger engagement of ADMA in the host pockets relative to SDMA.

Geometric shape matching also contributes to the binding strengths and selectivity of the host-guest complexes. The relative openness of geometry is a key determinant.[29] A shallower or open binding pocket is unfavourable for binding the free amino acids, as would be expected from the arguments made above. This can be clearly observed for the calixarenes and PC, where there is not a large complementary overlap of the surface area, the binding is weak. Hosts with closed geometries can display selectivities that run against the underlying trends caused by hydrophobicity. In general, increasing methylation increases the tendency to bind, but binding can be discouraged when shape and fit are incompatible. This can most clearly be observed for CLR01 and M2. CLR01 is in the middle of the “openness range” of this library, it has nine rings forming its tweezer shape. Its binding pocket is a perfect fit to bind medium-sized hydrophobic molecules and is the only host in this library with strong binding for both unmethylated arginine and lysine. When looking at the dimethylated arginines, ADMA’s geminal methyls can fit into the pocket of most hosts (see Figure 4). The exception is M2, where ADMA binding is 3-fold weaker compared to SDMA. This makes M2 a rare example of an SDMA-selective host molecule. The structural difference between M1 and M2 is the naphthalene of M2. This creates a bigger, more hydrophobic, and somewhat more closed-off binding pocket. This small change makes a significant difference in the binding properties of the host-guest complexes. When looking at MML versus TML we can see that M1 keeps the same formation where M2 must accommodate for its bigger structure and is slightly askew (see Supporting Information). PC and CLR01 are more rigid molecules, and do not have the option to flex to accommodate guests in the same way as the acyclic cucurbiturils.

**Conclusion**

Our studies have revealed some interesting individual complexes, as well as some broader trends that arise from comparisons. We found that the CLR01 host was the best size for the non-methylated amino acids, showing a strong binding for arginine and lysine which the other hosts do not. The methylated arginine guests are bound the strongest by the acyclic CB hosts, but in a weaker range of $K_b$ values the host sCx4-NO$_2$ shows good selectivity for ADMA over all other arginines. TML is bound strongly by all hosts, except PC, but is only bound with good selectivity over other related guests by M2. Although most of the hosts bind their guests outside of the targeted physiological concentration ranges, the $K_b$ for the complex of M1-ADMA is at the upper limit of the target concentration for that amino acid, which motivates further work on this class of hosts. While this study is fundamental and not targeted at applied science, we do identify useful new selectivities, such as the complete selectivity of CLR01 for...
methyllysines over histidinines, and the novel selectivity for SDMA over ADMT that is displayed by M2. This study also allows us to see some interesting trends emerge from direct comparison of different host classes under identical conditions. For example, in most ways PC behaves more like the sC4x hosts than like its close chemical relative CLR01. In this work we can...