
Binding methylarginines and methyllysines as free amino acids: a comparative study of multiple supramolecular host classes

Zoey Warmerdam,^{1,a} Bianca E. Kamba,^{1,b} My-Hue Le,^c Thomas Schrader,^c Lyle Isaacs,^{d*} Peter Bayer,^{b*} and Fraser Hof^{a*}

Dedicated to François Diederich, enthusiastic promotor of collaboration and partnerships

[a] 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] Department of Structural and Medicinal Biochemistry
Universität Duisburg Essen
Universitätsstrasse 2, 45141 Essen, Germany
* E-mail: peter.bayer@uni-due.de

[c] Department of Chemistry
Universität Duisburg Essen
Universitätsstrasse 7, 45117 Essen, Germany

[d] Department of Chemistry and Biochemistry
University of Maryland
College Park, MD 20742, United States
* E-mail: lisaacs@umd.edu

1 These authors contributed equally to the work in this paper

Supporting information for this article is given via a link at the end of the document.

Abstract: 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 cleft-like hosts. We determined the binding association constants for a panel of methylated amino acids using indicator displacement assays. The calixarene hosts show weak binding that favours the higher methylation states, with the strongest binding observed for trimethyllysine. The acyclic cucurbiturils display stronger binding to the methylated amino acids, and some unique patterns of selectivity. The cleft-like hosts follow two different trends, one shallow host following similar trends to the calixarenes, and the other more closed host binding certain less-methylated amino acids stronger than their per-methylated counterparts. Molecular modeling sheds some light on the different preferences of different hosts. The results identify hosts with selectivities that will be useful for certain biomedical applications. The overall selectivity patterns are explained by a common framework that considers the topology, 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. Amino acids are small and zwitterionic, and the relatively small hydrophobic surface area combined with relatively high charge means that they are strongly solvated by water. The common presence of salts and other co-solutes creates further challenges.¹ Although rare, there are dozens of diseases caused by disordered amino acid

metabolism.^{2,3} There are also multiple methylated amino acids whose concentrations are diagnostic of different medical conditions, and some of them are seen as causative agents in different pathologies. 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.^{4,5} Antibodies are known that bind amino acids but have some inherent shortcomings⁶⁻⁹ that could be overcome by the creation of synthetic organic binding tools for amino acids.

Supramolecular hosts are a great starting point to develop such new tools. The hosts have a concave binding pocket where the molecular recognition takes place via non-covalent interactions (e.g. electrostatics, hydrophobic effect, hydrogen bonding, van der Waals forces, π - π interactions, etc.) These binding pockets come with different shapes, sizes and chemical properties that influence the host-guest binding properties. The recognition of amino acids by supramolecular hosts has recently been reviewed by Basílio et al.¹⁰ Examples of supramolecular hosts binding to free amino acids include calixarenes, pillararenes, cucurbiturils, cyclodextrins and many other macrocyclic molecules. Although our ability to use hosts to target small molecules is improving, there are still challenges when designing supramolecular hosts to bind hydrophilic small molecules in water. In neutral aqueous solution, free amino acids are zwitterionic and strongly hydrated, 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. This hurdle becomes harder to overcome in a salty environment where, with an increasing ionic strength of the solution, a stronger competition arises through multiple mechanisms.¹

We studied a small set of 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.¹¹ We chose a parent anionic calixarene, **sCx4**, and the functionalized analogs **sCx4-NO₂** and **sCx4-CHO** to include in this study (Figure 1a-c). 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 analogs 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.¹² Another more rigid cleft-like host family includes the ‘tweezers’ and ‘clips’ introduced by Klärner and Schrader.^{13,14} 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).

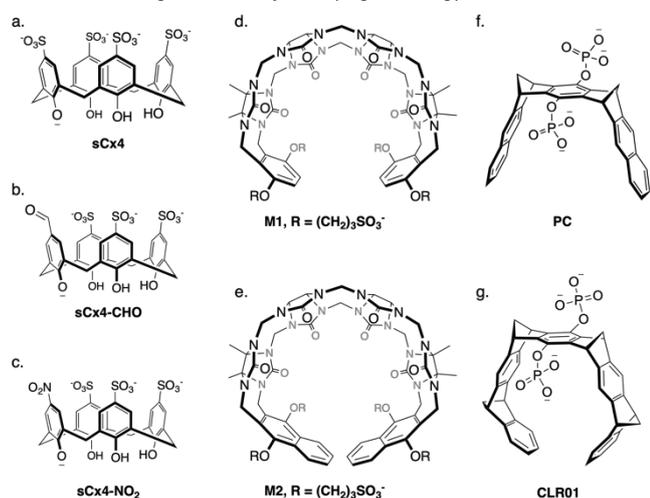


Figure 1. Hosts studied in this report. a) **sCx4**. b) **sCx4-CHO**. c) **sCx4-NO₂**. d) **M1**. e) **M2**. f) **PC**. g) **CLR01**.

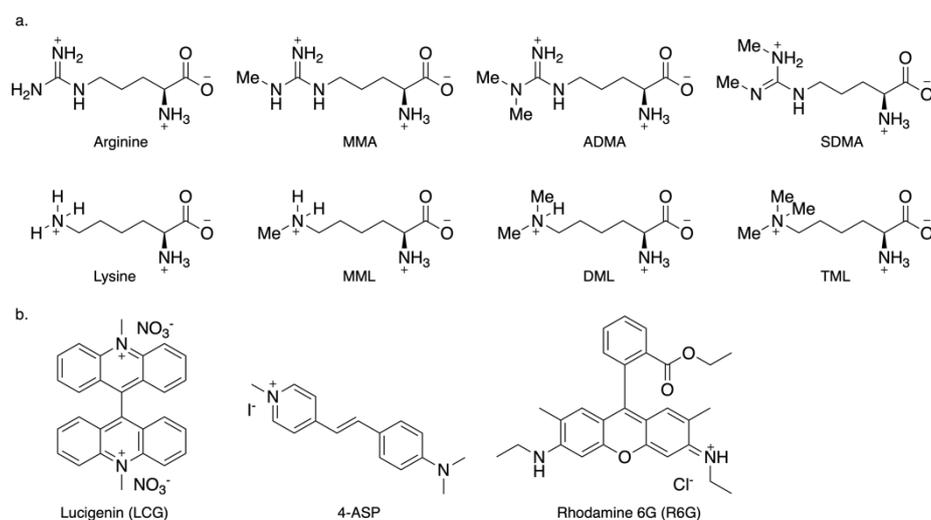


Figure 2. Guests and indicators studied in this report. a) Guests arginine, monomethylarginine (MMA), asymmetric dimethylarginine (ADMA), symmetric dimethylarginine (SDMA), Lysine, monomethyllysine (MML), dimethyllysine (DML), and trimethyllysine (TML), b) Indicators LCG, 4-ASP, and R6G.

For this study we have selected methylated amino acids as binding targets. They have very 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.¹⁵⁻²⁰ 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)). Lots of research has been focused on binding methylated lysines,²¹⁻²⁴ while the selective binding of methylated arginine has received relatively less attention.^{25,26} Most prior literature has focused on binding these methylated residues in the context of proteins and peptides. This study on free amino acids is motivated by a body of literature^{15,18,27-29} 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. As mentioned above, the existence of α -NH₃⁺ and CO₂⁻ groups on the free amino acids makes them challenging targets for binding in aqueous solutions. Against this backdrop, the presence of methyl groups in different numbers and arrangements is a subtle set of differences³⁰ for hosts to distinguish.

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 IDAs³¹ for **sCx4** (using lucigenin, Figure 2b)³² and **M1** and **M2** (using Rhodamine 6G, Figure 2b).^{33,34} 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

(K_{ind}) was determined by a direct titration of the host into indicator, and the host-guest K_d values 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 Na_2HPO_4 buffer at pH 7.4. This choice of buffer rules out the ability to measure very weak binding (which would require analyte concentrations that could 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 ($K_d < 200 \mu\text{M}$), medium to strong binding ($K_d 200\text{--}1500 \mu\text{M}$) and weak to no binding ($K_d > 1500 \mu\text{M}$). Exemplary primary data are presented in Figure 3 and all titrations are presented in the supportive information.

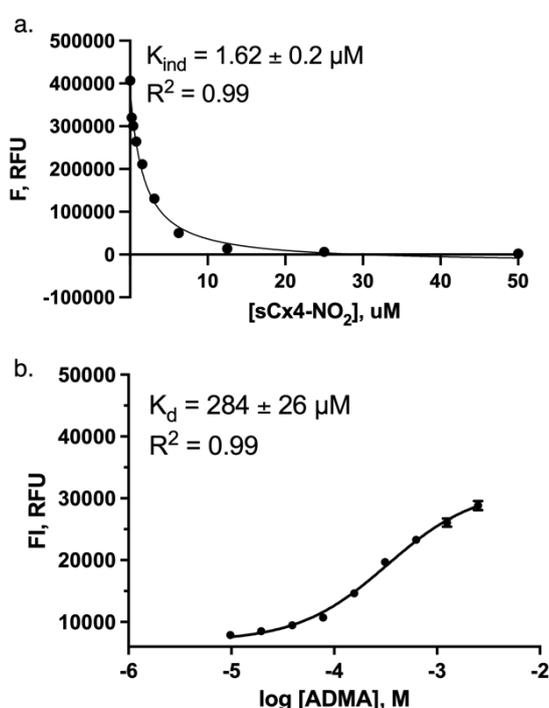


Figure 3. Primary IDA data. a) Direct titration between **sCx4-NO₂** and LCG. b) Competitive titration between **sCx4-NO₂** and ADMA. carried out in 10 mM NaH_2PO_4 buffer at pH 7.4. See the Supporting Information for titration curves, experimental details, and fitting details.

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 with each of the sulfonatocalixarenes. This trend is less straightforward for the arginine guests although the lower methylation states show no binding within the limits of this assay condition. Comparing the symmetric and asymmetric isomers of dimethylarginine, 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, which are generally low- or sub- μM .

Strong binding of the amino acids is consistently observed for **M1** and **M2**. **M1** binds all three methylated arginines, while ignoring unmethylated arginine. **M1** is slightly selective for ADMA ($K_d = 10 \mu\text{M}$) over MMA ($K_d = 45 \mu\text{M}$) and SDMA ($K_d = 35 \mu\text{M}$). It has a similar selectivity for the methylated lysines showing no binding to lysine and strong binding for the methylated lysines, being selective for TML ($K_d = 15 \mu\text{M}$) over DML ($K_d = 70 \mu\text{M}$) and MML ($K_d = 160 \mu\text{M}$). **M2** displays a sharp drop off between strongly binding highly methylated guests (ADMA, SDMA, DML, TML) and the weak binding of lower methylation states. CB[n]-type receptors are known to prefer quaternary over primary ammoniums.³⁵ Among the stronger binding guests, **M2** has a >3 -fold selectivity for SDMA ($K_d = 40 \mu\text{M}$) over ADMA ($K_d = 130 \mu\text{M}$), and a >10 -fold selectivity for TML ($K_d = 30 \mu\text{M}$) over DML ($K_d = 340 \mu\text{M}$).

Despite their chemical similarity, **PC** and **CLR01** have very different behaviors. **PC** has relatively weak binding across the set of guests, but some trends can be observed. The higher the methylation state of the arginine guests the stronger the binding gets. 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 unlike the other hosts this also includes $<100 \mu\text{M}$ binding of each unmethylated amino acid.³⁶ Lysine binds with a $K_d = 30 \mu\text{M}$, and the slightly more hydrophobic MML binds more strongly at $K_d = 15 \mu\text{M}$. 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.

Table 1. K_d values determined by IDA for each host-guest complex in 10 mM phosphate buffer. ^a

	sCx4^b	sCx4-CHO^b	sCx4-NO₂^b	M1^c	M2^c	PC^d	CLR01^d
	$K_d (\mu\text{M})$	$K_d (\mu\text{M})$	$K_d (\mu\text{M})$	$K_d (\mu\text{M})$	$K_d (\mu\text{M})$	$K_d (\mu\text{M})$	$K_d (\mu\text{M})$
Arginine	>1500	>1500	>1500	>1500	>1500	1100 ± 630	75 ± 20
MMA	>1500	>1500	>1500	45 ± 10	>1500	1230 ± 540	135 ± 30
ADMA	500 ± 50	230 ± 50	290 ± 60	10 ± 5	130 ± 40^e	650 ± 130	>1500
SDMA	1090 ± 300	370 ± 150	>1500	35 ± 10	40 ± 15	810 ± 270	>1500
Lysine	>1500	>1500	>1500	>1500	>1500	>1500	30 ± 5
MML	930 ± 120	470 ± 110	610 ± 130	160 ± 60	>1500	>1500	15 ± 5
DML	300 ± 40	240 ± 50	410 ± 60	70 ± 10	340 ± 150^e	980 ± 710	30 ± 5
TML	120 ± 20	100 ± 30	150 ± 30	15 ± 5	30 ± 10	1040 ± 360	65 ± 20

[a] All titrations were carried out in 10 mM NaH_2PO_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] Lucigenin (LCG) was used as indicator. [c] Rhodamine 6G (R6G) was used as indicator. [d] 4-(4-Diethylaminostyryl)-1-methylpyridinium iodide (4-ASP) was used as indicator. [e] $R^2 \geq 0.92$.

To determine the effect of a higher salt concentration, IDAs 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 2–4-fold weaker for each guest, but still remains in the “strong” range of K_d values.

Table 2. K_d values determined by IDA for each host-guest complex in 50 mM phosphate buffer. ^a

	M1 ^b	M2 ^b	CLR01 ^c
	K_d (μ M)	K_d (μ M)	K_d (μ M)
Arginine	>1500	>1500	93 ± 26
MMA	177 ± 63	>1500	157 ± 28
ADMA	30 ± 11	124 ± 42	>1500
SDMA	56 ± 14	42 ± 11	>1500
Lysine	>1500	>1500	51 ± 14
MML	297 ± 145	>1500	33 ± 6
MML	205 ± 52	188 ± 112 ^d	55 ± 9
TML	30 ± 6	24 ± 5	38 ± 5

[a] All titrations were carried out in 50 mM NaH_2PO_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] $R^2 \geq 0.93$.

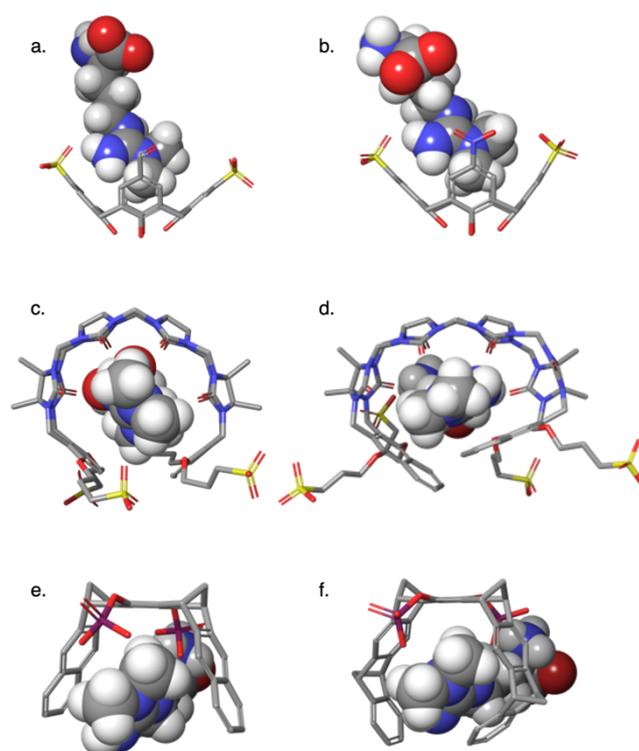
We did molecular modeling for six of the hosts to provide a general view of the differences and similarities between the hosts for certain key complexes. Molecular modeling was done using minimization in explicit water for all the indicated complexes using Maestro (Schrodinger, Inc). Host **sCx4-CHO**, **sCx4-NO₂**, **M1**, **M2**, **PC**, **CLR01** were modeled in complex with ADMA (Figure 4) and hosts **M1**, **M2**, **PC**, **CLR01** were also modeled 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 topology hosts that engage only the charged side chain of the guests (the calixarenes and **PC**), and closed topology hosts that almost completely surround their guests (**M1**, **M2**, 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 topological 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 strongly²², where the tweezer-type host do have a precedent as strong binding host.^{36,37} 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 guests have a zwitterionic α -amino-acid component, and a net charge of +1. The methylated guests spread the positive charge around to more distributed regions of the guest’s surface area. 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 also generate K_d values that vary by almost 100-fold

between different guests. **M1** and **M2** have eight carbonyl groups that can form ion-dipole interaction, and two sulfonate groups that can form ion-ion interactions. Molecular modeling 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 of the hosts can form cation-pi interactions with their cationic guests, although the geometric details vary.

Figure 4. Molecular modelling of each host in complex with ADMA. a) **sCx4-**



CHO, front view, b) **sCx4-NO₂**, front view/ c) **M1**, top view. d) **M2**, top view. e) **PC**, front view. f) **CLR01**, front view. Molecules were energy-minimized in explicit water (not shown) using OPLS_2005 as implemented in Maestro (Schrodinger, Inc). See Supporting information for more views and for other host-guest complexes.

Guest hydrophobicity is the main determinant for guest selectivity in hosts with more open topologies, 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.³⁸ 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-topology hosts that don't constrain for molecular shape, this results in a stronger engagement of ADMA in the host pockets relative to SDMA.

Topological shape matching also contributes to the binding strengths and selectivity of the host-guest complexes. The relative openness of topology and is a key determinant.³⁹ 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 isn't a large complementary overlap of the surface area, the binding is weak. Hosts with closed topologies 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 very 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

In conclusion, we have studied a host library containing calixarenes, acyclic cucurbiturils, and cleft-like hosts. From a series of guest binding studies with methylated amino acids we see that the relatively open hosts, calixarenes and **PC**, have an overall weaker binding than the host with a deeper binding pocket. 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_d values the host **sCx4-NO₂** 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**. 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 methylarginines, and the novel selectivity for SDMA over ADMA 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 **sCx4** hosts than like its close chemical relative **CLR01**. In this

work we can tie that similar behaviour across many guests to similarities in host topology that mostly override the more obvious differences in functional group identity and arrangement that typically dominate our thinking about host-guest binding. We think that additional non-dogmatic, collaborative, open comparisons of host molecules will lead to more such insights in the near future.

Acknowledgements

We thank Hendrik Kirschner and Christine Beuck for carrying out preliminary studies. We thank Rebecca Hof for the training and Allison Selinger for helpful discussions on the data analysis for indicator displacement assays.

Funding

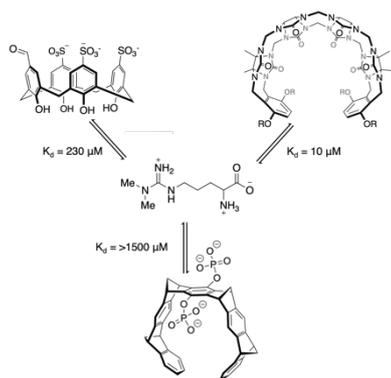
We thank the National Science Foundation (CHE-1807486), NSERC (RGPIN-2019-04806), National Science Foundation (CHE-1807486), Deutsche Forschungsgemeinschaft (DFG) to TS (Schr 604/19-1) and PB (BA1624/15-1) for their financial support.

Keywords: macrocycles • methylated amino acids • host-guest chemistry • indicator displacement assays

- (1) Beatty, M. A.; Hof, F. Host-guest binding in water, salty water, and biofluids: general lessons for synthetic, bio-targeted molecular recognition. *Chem Soc Rev* **2021**.
- (2) Schlesinger, S.; Sonntag, S. R.; Lieb, W.; Maas, R. Asymmetric and Symmetric Dimethylarginine as Risk Markers for Total Mortality and Cardiovascular Outcomes: A Systematic Review and Meta-Analysis of Prospective Studies. *PLoS One* **2016**, *11*, e0165811.
- (3) DeBerardinis, R. J.; Thompson, C. B. Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* **2012**, *148*, 1132-1144.
- (4) Gowda, G. A.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. Metabolomics-based methods for early disease diagnostics. *Expert Rev Mol Diagn* **2008**, *8*, 617-633.
- (5) Ruigrok, V. J.; Levisson, M.; Eppink, M. H.; Smidt, H.; van der Oost, J. Alternative affinity tools: more attractive than antibodies? *Biochem J* **2011**, *436*, 1-13.
- (6) Egelhofer, T. A.; Minoda, A.; Klugman, S.; Lee, K.; Kolasinska-Zwierz, P.; Alekseyenko, A. A.; Cheung, M. S.; Day, D. S.; Gadel, S.; Gorchakov, A. A.; Gu, T.; Kharchenko, P. V.; Kuan, S.; Latorre, I.; Linder-Basso, D.; Luu, Y.; Ngo, Q.; Perry, M.; Rechtsteiner, A.; Riddle, N. C.; Schwartz, Y. B.; Shanower, G. A.; Vielle, A.; Ahringer, J.; Elgin, S. C.; Kuroda, M. I.; Pirrotta, V.; Ren, B.; Strome, S.; Park, P. J.; Karpen, G. H.; Hawkins, R. D.; Lieb, J. D. An assessment of histone-modification antibody quality. *Nat Struct Mol Biol* **2011**, *18*, 91-93.
- (7) Kungulovski, G.; Jeltsch, A. Quality of histone modification antibodies undermines chromatin biology research. *F1000Res* **2015**, *4*, 1160.
- (8) Rothbart, S. B.; Dickson, B. M.; Raab, J. R.; Grzybowski, A. T.; Krajewski, K.; Guo, A. H.; Shanle, E. K.; Josefowicz, S. Z.; Fuchs, S. M.; Allis, C. D.; Magnuson, T. R.; Ruthenburg, A. J.; Strahl, B. D. An Interactive Database for the Assessment of Histone Antibody Specificity. *Mol Cell* **2015**, *59*, 502-511.
- (9) Acharya, P.; Quinlan, A.; Neumeister, V. The ABCs of finding a good antibody: How to find a good antibody, validate it, and publish meaningful data. *F1000Res* **2017**, *6*, 851.
- (10) Martins, J. N.; Lima, J. C.; Basilio, N. Selective Recognition of Amino Acids and Peptides by Small Supramolecular Receptors. *Molecules* **2020**, *26*.

- (11) Shinkai, S.; Araki, K.; Manabe, O. Does the Calixarene Cavity Recognize the Size of Guest Molecules - on the Hole-Size Selectivity in Water-Soluble Calixarenes. *J Chem Soc Chem Comm* **1988**, 187-189.
- (12) Zhang, B.; Isaacs, L. Acyclic Cucurbit[n]uril-type Molecular Containers: Influence of Aromatic Walls on their Function as Solubilizing Excipients for Insoluble Drugs. *J Med Chem* **2014**, *57*, 9554-9563.
- (13) Hadrovic, I.; Rebmann, P.; Klärner, F. G.; Bitan, G.; Schrader, T. Molecular Lysine Tweezers Counteract Aberrant Protein Aggregation. *Front Chem* **2019**, *7*, 657.
- (14) Talbiersky, P.; Bastkowski, F.; Klärner, F. G.; Schrader, T. Molecular clip and tweezer introduce new mechanisms of enzyme inhibition. *J Am Chem Soc* **2008**, *130*, 9824-9828.
- (15) Cooke, J. P. ADMA: its role in vascular disease. *Vasc Med* **2005**, *10*, S11-S17.
- (16) Cooke, J. P.; Ghebremariam, Y. T. DDAH says NO to ADMA. *Arterioscler Thromb Vasc Biol* **2011**, *31*, 1462-1464.
- (17) Tsikas, D.; Bollenbach, A.; Hanff, E.; Kayacelebi, A. A. Asymmetric dimethylarginine (ADMA), symmetric dimethylarginine (SDMA) and homoarginine (hArg): the ADMA, SDMA and hArg paradoxes. *Cardiovascular diabetology* **2018**, *17*, 1.
- (18) Emrich, I. E.; Zawada, A. M.; Martens-Lobenhoffer, J.; Fliser, D.; Wagenpfeil, S.; Heine, G. H.; Bode-Boger, S. M. Symmetric dimethylarginine (SDMA) outperforms asymmetric dimethylarginine (ADMA) and other methylarginines as predictor of renal and cardiovascular outcome in non-dialysis chronic kidney disease. *Clin Res Cardiol* **2018**, *107*, 201-213.
- (19) Rowe, E. M.; Xing, V.; Biggar, K. K. Lysine methylation: Implications in neurodegenerative disease. *Brain Res* **2019**, *1707*, 164-171.
- (20) Hamamoto, R.; Saloura, V.; Nakamura, Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat Rev Cancer* **2015**, *15*, 110-124.
- (21) Daze, K. D.; Pinter, T.; Beshara, C. S.; Ibraheem, A.; Minaker, S. A.; Ma, M. C. F.; Courtemanche, R. J. M.; Campbell, R. E.; Hof, F. Supramolecular hosts that recognize methyllysines and disrupt the interaction between a modified histone tail and its epigenetic reader protein. *Chem Sci* **2012**, *3*, 2695-2699.
- (22) Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F. A Simple Calixarene Recognizes Post-translationally Methylated Lysine. *Chembiochem* **2010**, *11*, 63-66.
- (23) Lindsey A. Ingerman; Cuellar, M. E.; Waters, M. L. A small molecule receptor that selectively recognizes trimethyl lysine in a histone peptide with native protein-like affinity. *Chem. Commun.* **2010**, *46*, 1839-1841.
- (24) Gruber, T. Synthetic Receptors for the Recognition and Discrimination of Post-Translationally Methylated Lysines. *Chembiochem* **2018**, *19*, 2324-2340.
- (25) Mullins, A. G.; Pinkin, N. K.; Hardin, J. A.; Waters, M. L. Achieving High Affinity and Selectivity for Asymmetric Dimethylarginine by Putting a Lid on a Box. *Angew Chem Int Ed Engl* **2019**, *58*, 5282-5285.
- (26) James, L. I.; Beaver, J. E.; Rice, N. W.; Waters, M. L. A synthetic receptor for asymmetric dimethyl arginine. *J. Am. Chem. Soc.* **2013**, *135*, 6450-6455.
- (27) Bode-Boger, S. M.; Scalera, F.; Ignarro, L. J. The L-arginine paradox: Importance of the L-arginine/asymmetrical dimethylarginine ratio. *Pharmacology & therapeutics* **2007**, *114*, 295-306.
- (28) Adler-Abramovich, L.; Vaks, L.; Carny, O.; Trudler, D.; Magno, A.; Caffisch, A.; Frenkel, D.; Gazit, E. Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria. *Nat Chem Biol* **2012**, *8*, 701-706.
- (29) Blackburn, P. R.; Gass, J. M.; Vairo, F. P. E.; Farnham, K. M.; Atwal, H. K.; Macklin, S.; Klee, E. W.; Atwal, P. S. Maple syrup urine disease: mechanisms and management. *Appl Clin Genet* **2017**, *10*, 57-66.
- (30) McQuinn, K.; McIndoe, L. S.; Hof, F. Insights into the post-translational methylation of arginine from studies of guanidinium-water nanodroplets. *Chem-Eur J* **2008**, *14*, 6483-6489.
- (31) Dsouza, R. N.; Pischel, U.; Nau, W. M. Fluorescent Dyes and Their Supramolecular Host/Guest Complexes with Macrocycles in Aqueous Solution. *Chemical Reviews* **2011**, *111*, 7941-7980.
- (32) Warmerdam, Z.; Kamba, B. E.; Shaurya, A.; Sun, X. X.; Maguire, M. K.; Hof, F. Calix[4]arene sulfonate hosts selectively modified on the upper rim: a study of nicotine binding strength and geometry. *Supramol Chem* **2021**.
- (33) Ma, D.; Zhang, B.; Hoffmann, U.; Sundrup, M. G.; Eikermann, M.; Isaacs, L. Acyclic cucurbit[n]uril-type molecular containers bind neuromuscular blocking agents in vitro and reverse neuromuscular block in vivo. *Angew Chem Int Ed Engl* **2012**, *51*, 11358-11362.
- (34) Ma, D.; Zavalij, P. Y.; Isaacs, L. Acyclic cucurbit[n]uril congeners are high affinity hosts. *J Org Chem* **2010**, *75*, 4786-4795.
- (35) Cao, L.; Sekutor, M.; Zavalij, P. Y.; Mlinaric-Majerski, K.; Glaser, R.; Isaacs, L. Cucurbit[7]urilguest pair with an attomolar dissociation constant. *Angew Chem Int Ed Engl* **2014**, *53*, 988-993.
- (36) Fokkens, M.; Schrader, T.; Klärner, F. G. A molecular tweezer for lysine and arginine. *J Am Chem Soc* **2005**, *127*, 14415-14421.
- (37) Ndendjio, S. A. Z.; Isaacs, L. Molecular recognition properties of acyclic cucurbiturils toward amino acids, peptides, and a protein. *Supramol Chem* **2019**, *31*, 432-442.
- (38) Evich, M.; Stroeve, E.; Zheng, Y. G.; Germann, M. W. Effect of methylation on the side-chain pKa value of arginine. *Protein Sci* **2016**, *25*, 479-486.
- (39) Beaver, J. E.; Peacor, B. C.; Bain, J. V.; James, L. I.; Waters, M. L. Contributions of pocket depth and electrostatic interactions to affinity and selectivity of receptors for methylated lysine in water. *Org Biomol Chem* **2015**, *13*, 3220-3226.

Entry for the Table of Contents



We present comparative binding studies for three different supramolecular host classes (calixarenes, acyclic cucurbiturils and cleft-like hosts), each studied with multiple methylated arginines and lysines to determine fundamental structure-function relationships. Molecular modeling was used to give an insight on the preferences of the different hosts.

Supporting information

for

Binding methylarginines and methyllysines as free amino acids: a comparative study of multiple supramolecular host classes

Table of Contents

1. General information and materials	2
2. IDA titrations	2
2.1 Equations	3
2.1.1. Outliers	3
2.1.2. The standard error	3
2.1.3. The total standard error	3
2.1.4. Curve fit for the direct titration	3
2.1.5. Curve fit for the competitive titration	4
2.2. Fluorescence based studies 10 mM buffer	5
2.2.1. Fluorescence based studies of sCx4	5
2.2.2. Fluorescence based studies of sCx4-CHO	6
2.2.3. Fluorescence based studies of sCx4-NO2	7
2.2.4. Fluorescence based studies of CLR01	8
2.2.5. Fluorescence based studies of PC	9
2.2.6. Fluorescence based studies of M1	10
2.2.7. Fluorescence based studies of M2	11
2.3. Fluorescence based studies 50 mM buffer	
2.3.1. Fluorescence based studies of M1	12
2.3.2. Fluorescence based studies of M2	13
2.3.4. Fluorescence based studies of CLR01	14
3. Molecular Modeling	15
3.1. Each host modeled with ADMA. Front, back, top, and bottom views for all complexes.	15
3.2. Hosts M1 , M2 , PC , CLR01 each modeled with MML and with TML.	18
4. References	22

1. General information and materials

The host used in these studies have all been previously published or are commercially available. **sCx4** (4-sulfocalix[4]arene Hydrate) was purchased from Tokyo Chemical Industry (TCI) (CAS 112269-92-8, >94.0%). **sCx4-CHO**¹, **sCx4-NO2**², **M1**³, **M2**³, **CLR01**⁴ and **PC**⁵ have been previously published. The dyes used in these studies have been purchased from Sigma-Aldrich, lucigenin (CAS 2315-97-1), 4-ASP (CAS 68971-03-9, 98%), R6G (CAS 989-38-8, ~95 %). The guests were purchased from different suppliers. Arginine was purchased from Calbiochem (CAS 1119-34-2, 99.7%). Lysine was purchased from USB Corporation (CAS 56-87-1). MMA (53308-83-1, ≥99%), and SDMA (30344-00-4, ≥97%) were purchased from ENZO. ADMA (CAS 220805-22-1) and MML (CAS 7622-29-9) were purchased from Toronto research chemicals. DML (CAS 79416-87-8, ≥96%) and TML (55528-53-5, ≥97%) were purchased from Sigma-Aldrich.

2. IDA titrations

The IDAs were conducted in 384 well plates (Nunc™ 384-Well, Non-Treated, Flat-Bottom Microplate). The fluorescent signal was read on the Cytation 5 (Software Version 3.05.11) at room temperature. All wells had a final total volume of 50 µL. The IDA data was analysed in GraphPad Prism Version 8.3.0 (328). One exemplary replica is shown for each host. All of the experiments for the 10 mM phosphate buffer were done as duplicates of triplicates. All of the experiments for the 50 mM phosphate buffer were done in triplicate.

The fluorescent signal for the calix[4]arene host titrations was read as a fluorescence endpoint measurement. Lucigenin (0.25 µM) was used as the indicator. The settings were as follows: Excitation: 369/20, Emission: 475/20. Optics: Top, Gain: extended. Light Source: Xenon Flash, Lamp Energy: High, Extended Dynamic Range. Read Speed: Normal, Delay: 100 msec, Measurements/Data Point: 10. Read Height: 10.5 mm.

The fluorescent signal for **CLR01** and **PC** titrations was read as a fluorescence endpoint measurement. 4-ASP (0.20 µM) was used as the indicator. The settings were as follows: Excitation: 490/10, Emission: 610/10. Optics: Bottom, Gain: extended. Light Source: Xenon Flash, Lamp Energy: High, Extended Dynamic Range. Read Speed: Normal, Delay: 100 msec, Measurements/Data Point: 10. Read Height: 10.5 mm.

The fluorescent signal for **M1** and **M2** titrations was read as a fluorescence endpoint measurement. R6G (0.10 µM) was used as the indicator. The settings were as follows: Excitation: 510/10, Emission: 550/10. Optics: Bottom, Gain: extended. Light Source: Xenon Flash, Lamp Energy: High, Extended Dynamic Range. Read Speed: Normal, Delay: 100 msec, Measurements/Data Point: 10. Read Height: 10.5 mm.

2.1 Equations

2.1.1. Outliers

Outliers are determined by the Dixon's Q-test, Equation 1, where the gap is the absolute difference between the outlier in question and the closest number to it. With three observations and at 95% confidence, $Q > 0.970 = Q_{95\%,n=3}$, we conclude the data point is an outlier.

Equation 1. Dixon's Q-test.

$$Q = \frac{gap}{range}$$

2.1.2. The standard error

The standard error is calculated to each triplicate, Equation 2.

Equation 2. Standard error.

$$SD_{K_i} = (\ln \ln (10) * 10^{\log K_i}) * SD_{\log K_i}$$

SD_{K_i} = Standard error of K_i

$\log K_i$ = Value of the log

$SD_{\log K_i}$ = Standard error of the $\log K_i$

2.1.3. The total standard error

The total standard error between two the triplicates is calculated by Equation 3.

Equation 3. Total standard error.

$$SD_{total} = \sqrt{\frac{(SD_1^2 + SD_2^2)}{n}}$$

SD_{total} = Standard derivation of both experiments

SD_1 = Standard derivation of the first triplicate

SD_2 = Standard derivation of the second triplicate

n = Number of experiments

2.1.4. Curve fit for the direct titration

To curve fit for the direct titration Equation 4, Equation 5 are used.

Equation 4. Curve fit for the direct titration for a turn on signal.

$$F = F_{min} + \frac{(F_{max} - F_{min}) * ([D] + [H] + K_d) - \sqrt{([D] + [H] + K_d)^2 - 4 * [H] * [D]}}{2 * [D]}$$

Equation 5. Curve fit for the direct titration for a turn off signal

$$F = F_{max} - \frac{(F_{max} - F_{min}) * ([D] + [H] + K_d) - \sqrt{([D] + [H] + K_d)^2 - 4 * [H] * [D]}}{2 * [D]}$$

- F = Fitted data point
- F_{max} = Maximum signal
- F_{min} = Minimum signal
- [D] = Molar concentration of dye in μM
- [H] = Molar concentration of host (titrant)
- K_d = Dissociation constant

2.1.5. Curve fit for the competitive titration

$$\log_{EC50} = \log \left(10^{\log K_i} * \left(1 + \frac{[D]}{K_d} \right) \right)$$

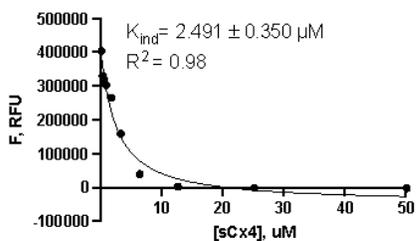
$$F = F_{min} + (F_{max} - F_{min}) / (1 + 10^{(X - \log_{EC50})})$$

- \log_{EC50} = log of the concentration of the competitor binding half-way between F_{min} and F_{max}
- K_i = Equilibrium dissociation constant in Molar
- [D] = Concentration of dye in nM
- K_d = Equilibrium dissociation constant of the direct titration
- F = Fitted data point
- F_{max} = Maximum signal
- F_{min} = Minimum signal

2.2. Fluorescence based studies 10 mM buffer

2.2.1. Fluorescence based studies of sCx4

a.



b.

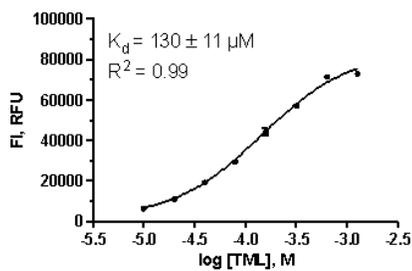
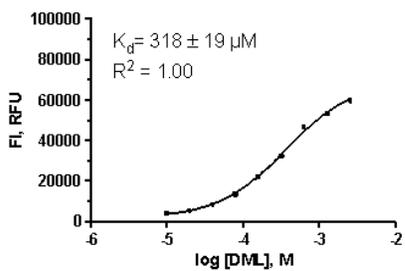
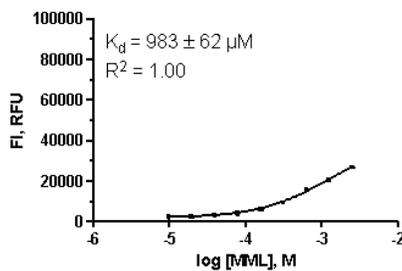
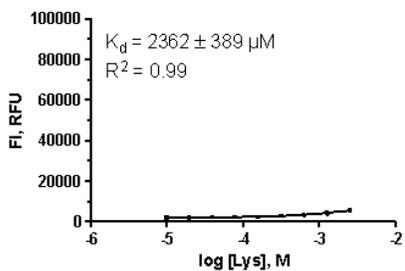
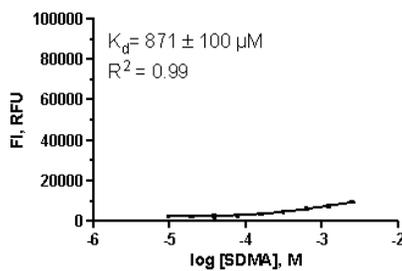
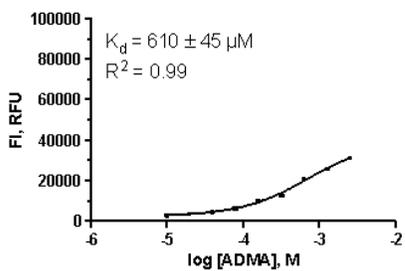
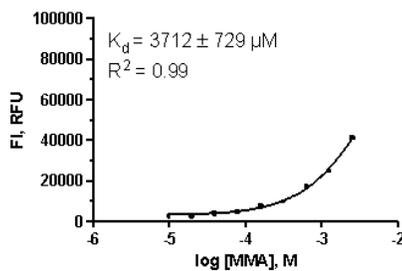
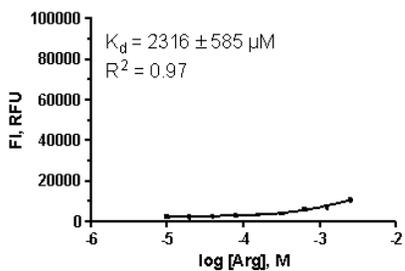


Figure S1: Fluorescence based studies of sCx4. a) Direct titration of LCG (0.25 μM) with sCx4 (0 – 50 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the sCx4-LCG (5 μM sCx4, 0.25 μM LCG) complex.

2.2.2. Fluorescence based studies of sCx4-CHO

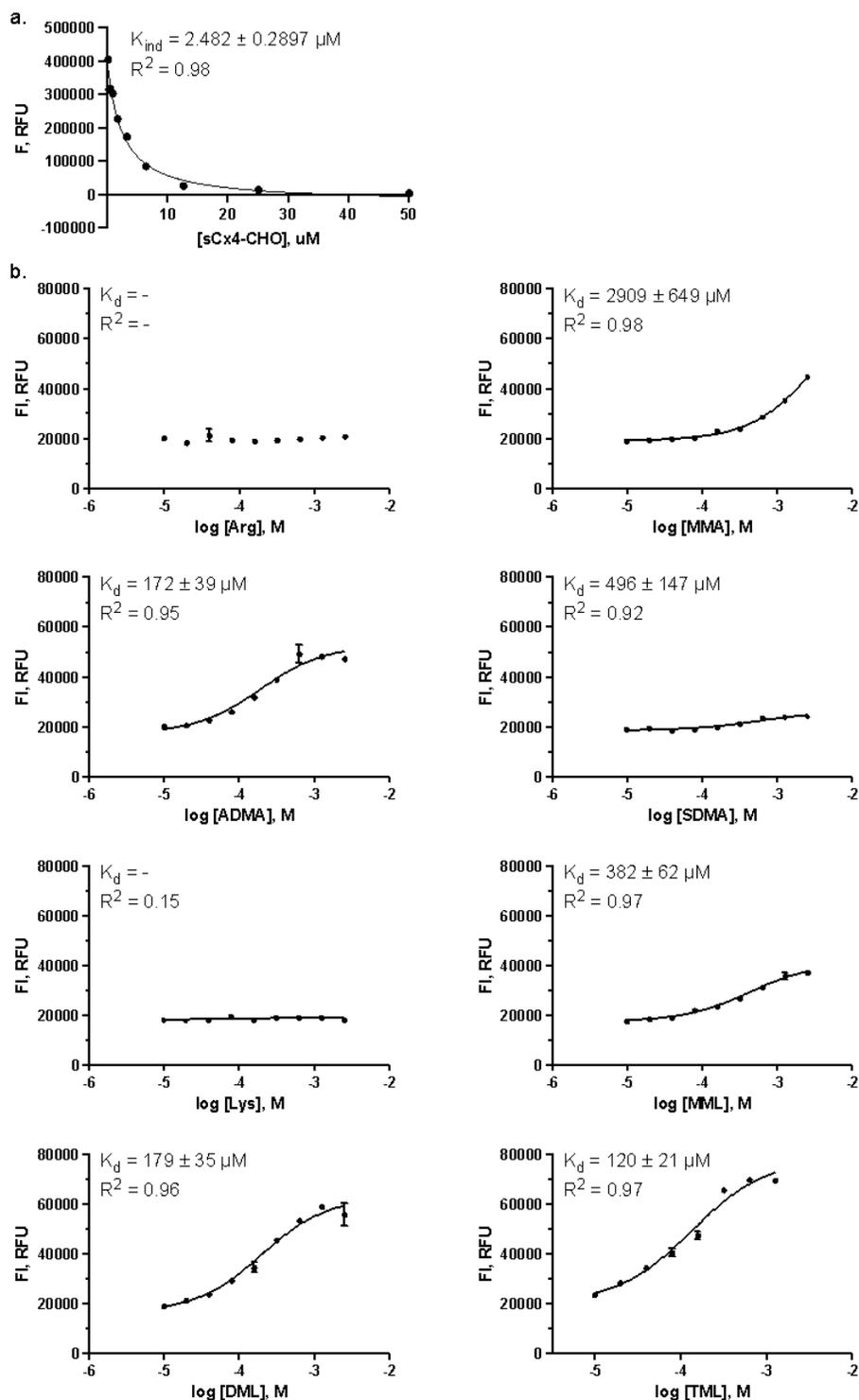


Figure S2: Fluorescence based studies of sCx4-CHO. a) Direct titration of LCG (0.25 μM) with sCx4-CHO (0 – 50 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the sCx4-CHO-LCG (5 μM sCx4-CHO, 0.25 μM LCG) complex.

2.2.3. Fluorescence based studies of sCx4-NO2

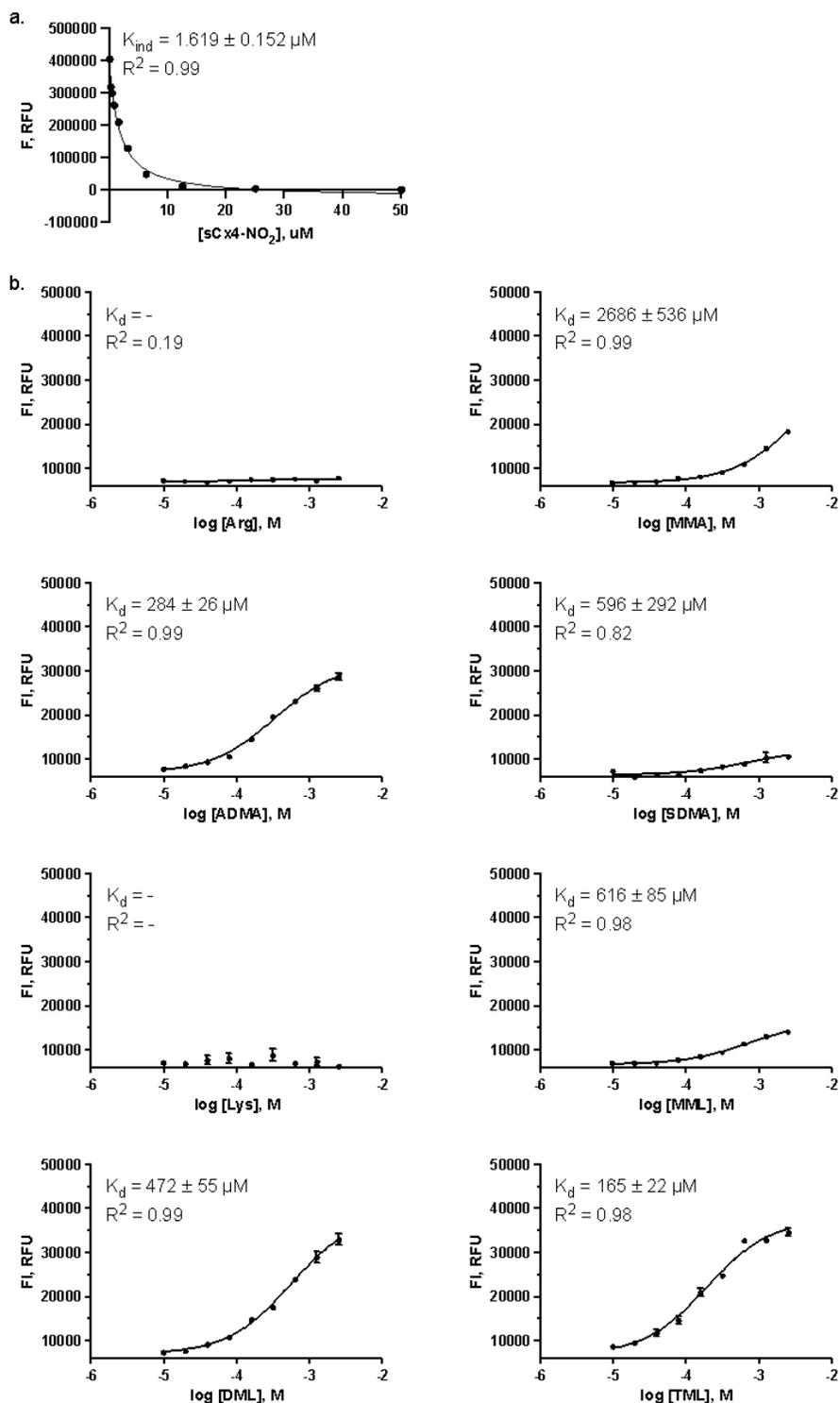


Figure S3: Fluorescence based studies of sCx4-NO2. a) Direct titration of LCG (0.25 μM) with sCx4-NO2 (0 – 50 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the sCx4-NO2-LCG (5 μM sCx4-NO2, 0.25 μM LCG) complex.

2.2.4. Fluorescence based studies of CLR01

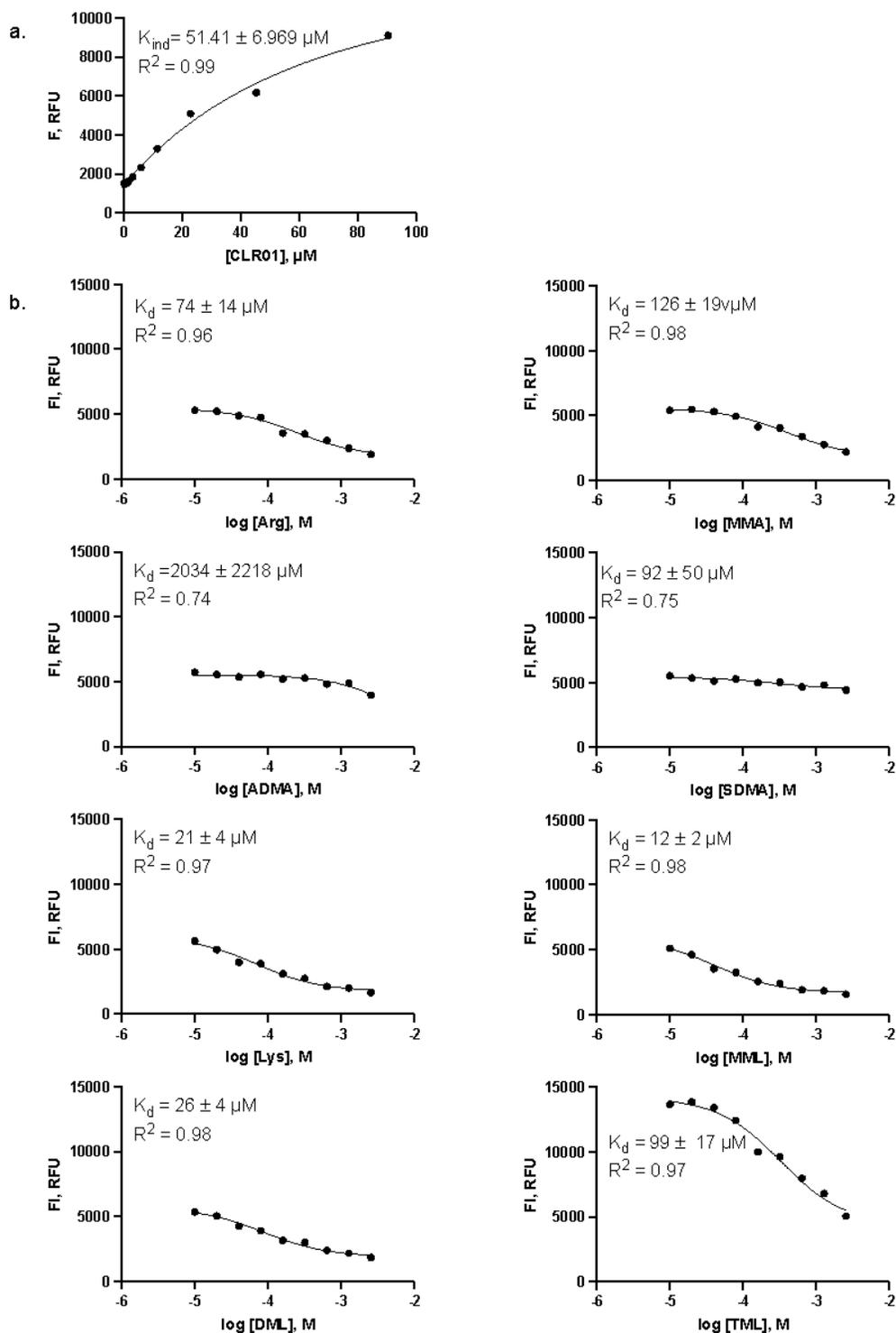


Figure S4: Fluorescence based studies of CLR01. a) Direct titration of 4-ASP (20 μM) with CLR01 (0 – 50 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the CLR01-4-ASP (70 μM CLR01, 20 μM 4-ASP) complex.

2.2.5. Fluorescence based studies of PC

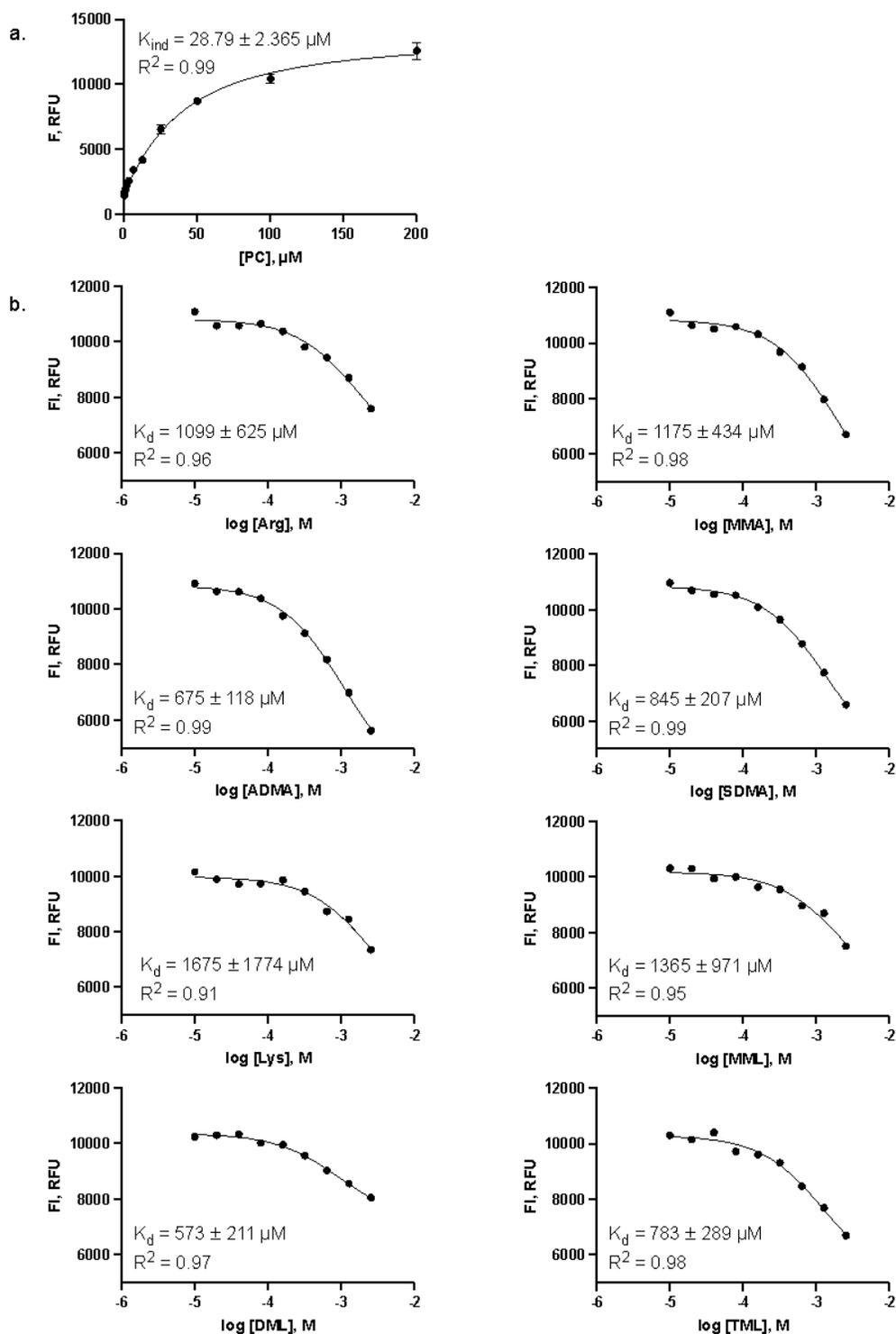


Figure S5: Fluorescence based studies of PC. a) Direct titration of 4-ASP (20 μM) with PC (0 – 200 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the PC-4-ASP (30 μM PC, 20 μM 4-ASP) complex.

2.2.6. Fluorescence based studies of **M1**

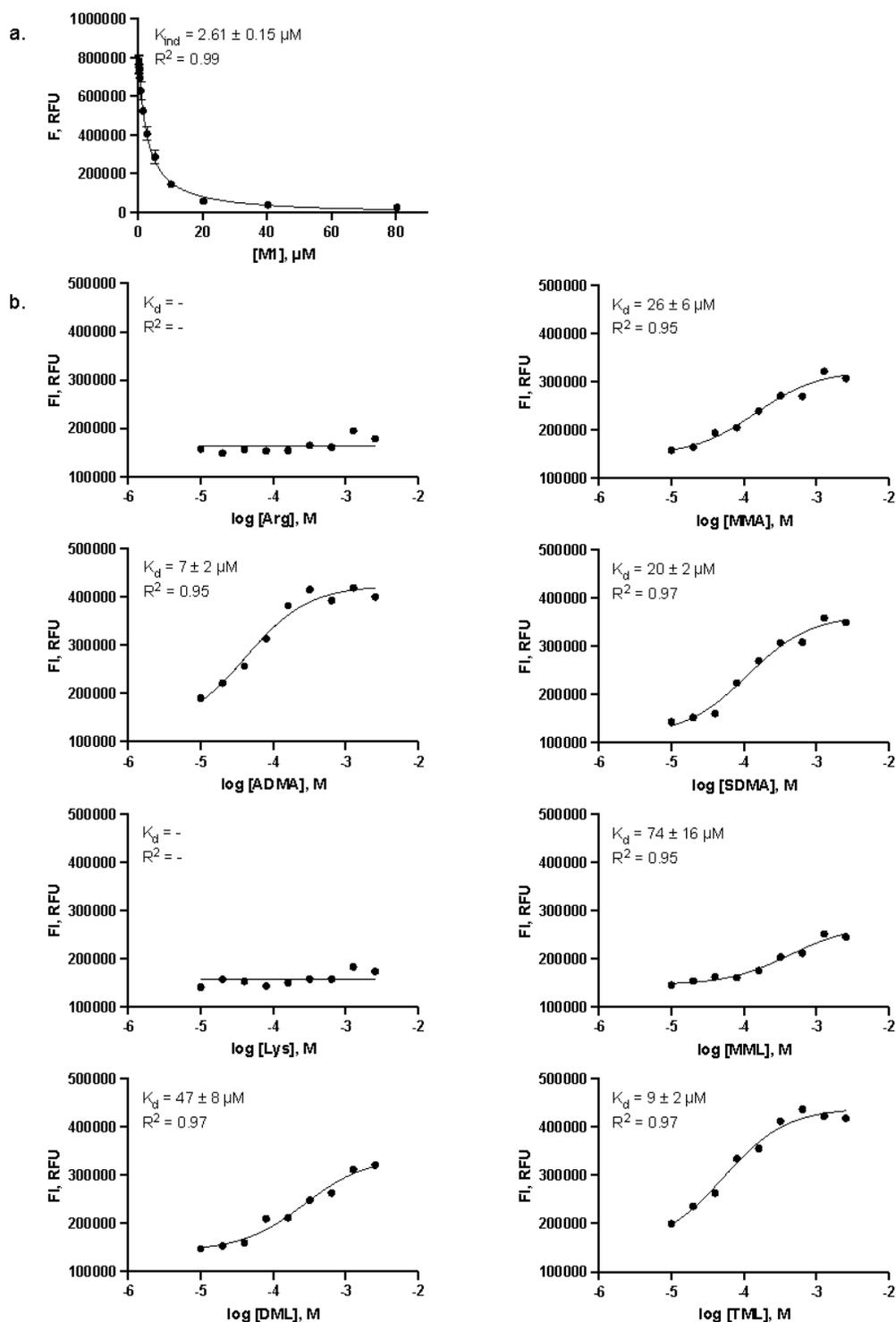


Figure S6: Fluorescence based studies of **M1**. a) Direct titration of R6G (10 μM) with **M1** (0 – 80 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **M1**-R6G (10 μM **M1**, 10 μM R6G) complex.

2.2.7. Fluorescence based studies of M2

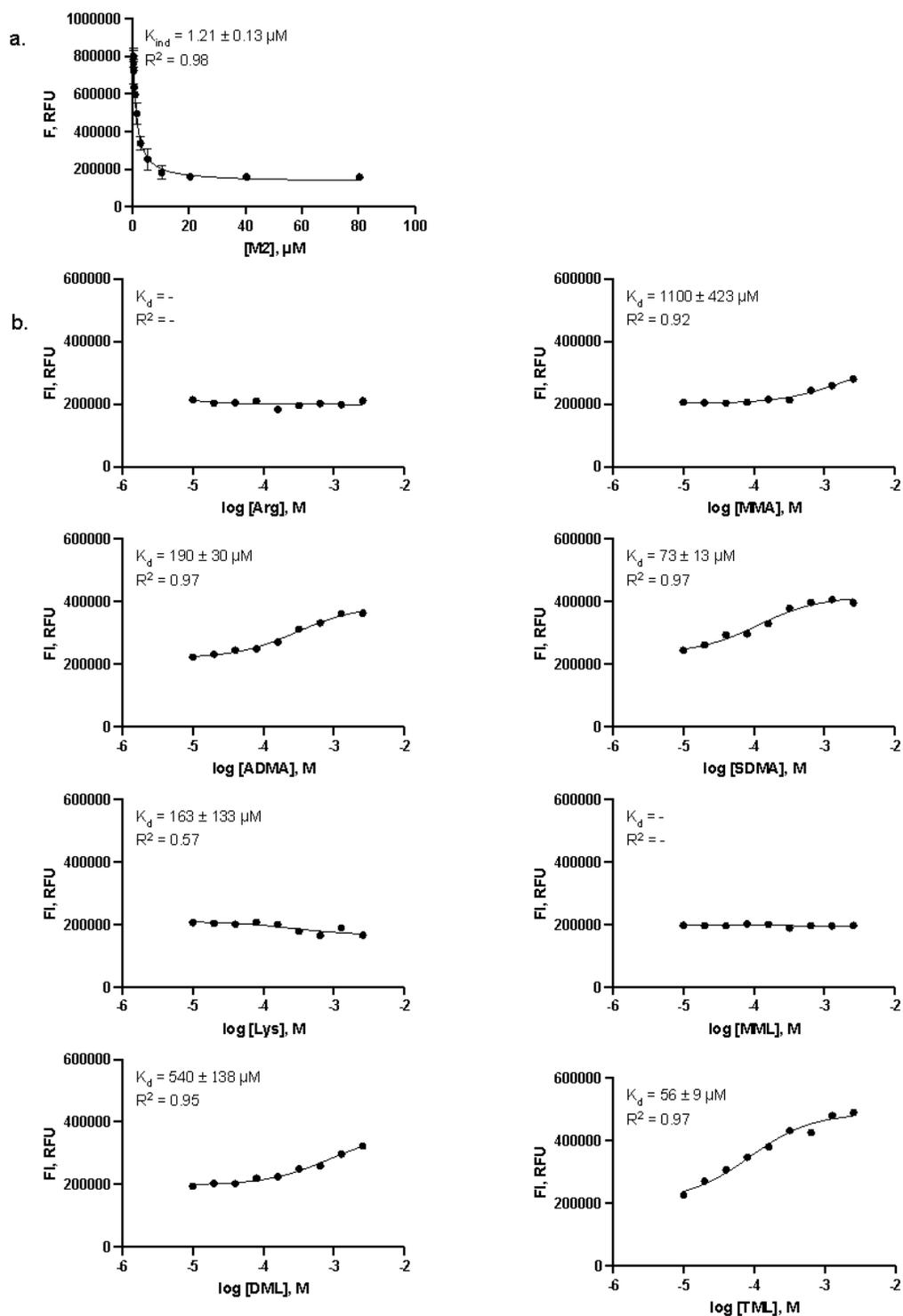


Figure S7: Fluorescence based studies of **M2**. a) Direct titration of R6G (10 μM) with **M2** (0 – 80 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **M2**-R6G (10 μM **M2**, 10 μM R6G) complex.

2.3. Fluorescence based studies 50 mM buffer

2.3.1. Fluorescence based studies of M1

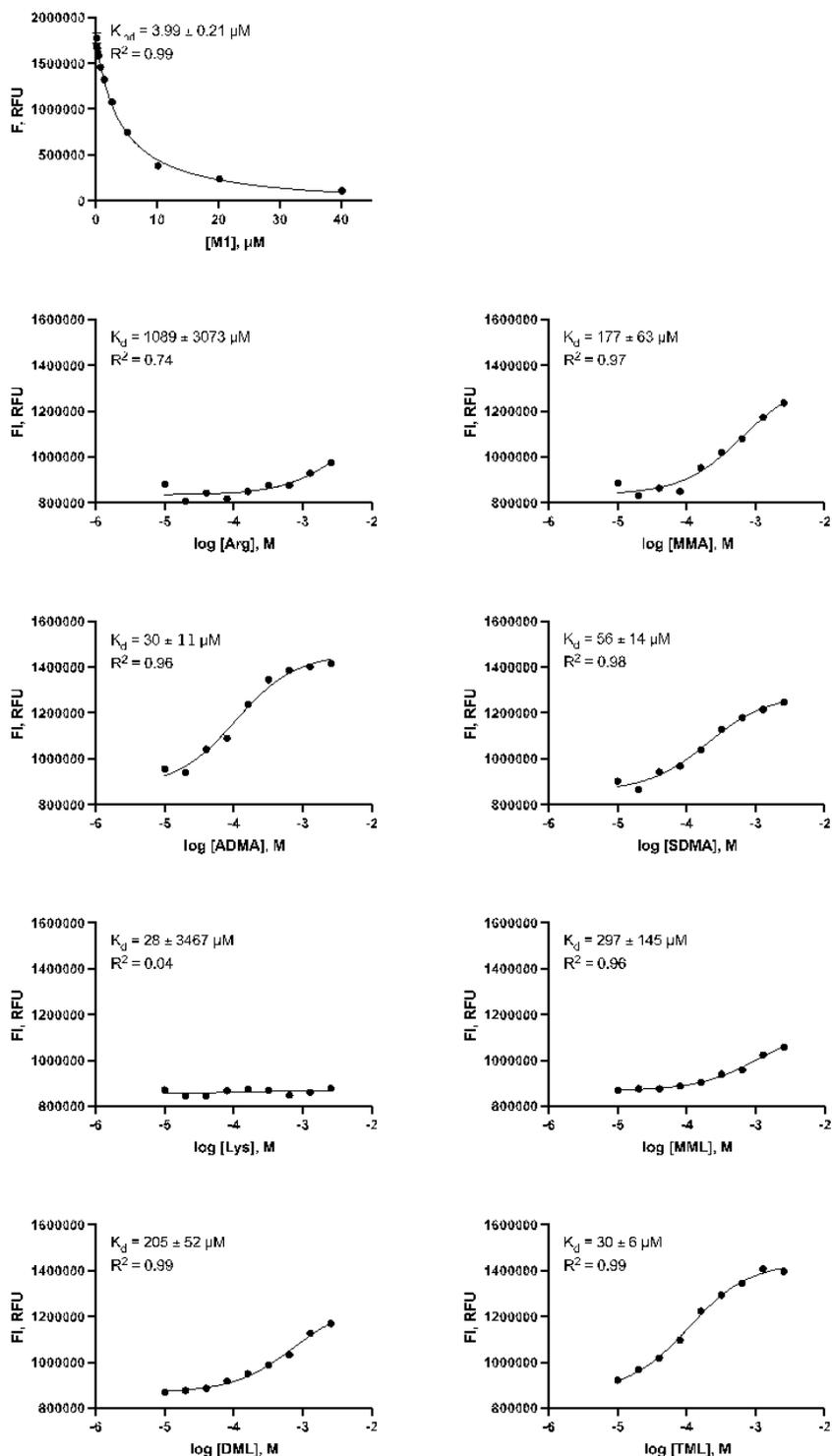


Figure S8: Fluorescence based studies of M1. a) Direct titration of R6G (10 μM) with M1 (0 – 40 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the M1-R6G (10 μM M1, 10 μM R6G) complex.

2.3.2. Fluorescence based studies of M2

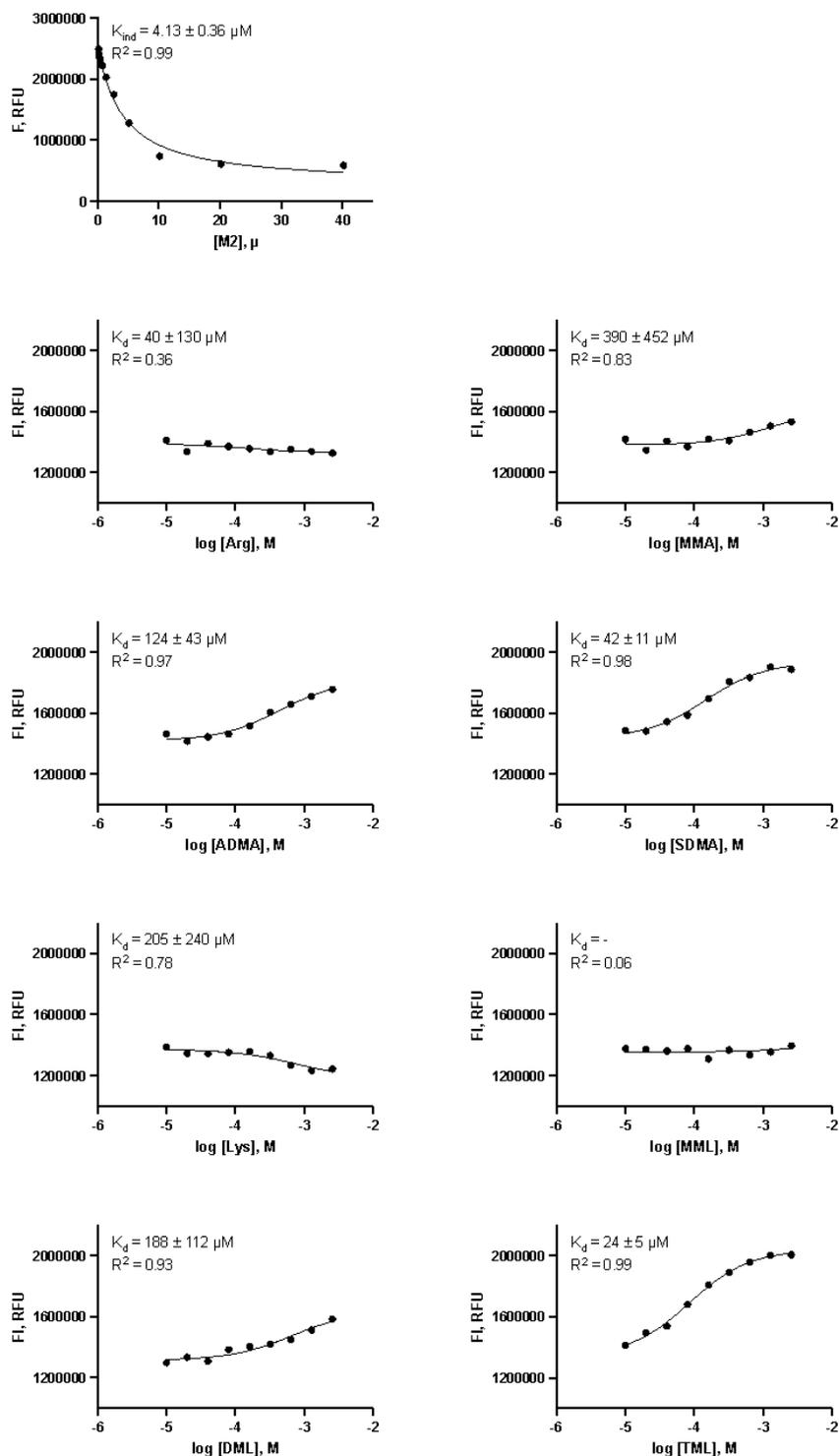


Figure S9: Fluorescence based studies of **M2**. a) Direct titration of R6G (10 μM) with **M2** (0 – 40 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **M2**-R6G (10 μM **M2**, 10 μM R6G) complex.

2.3.4. Fluorescence based studies of CLR01

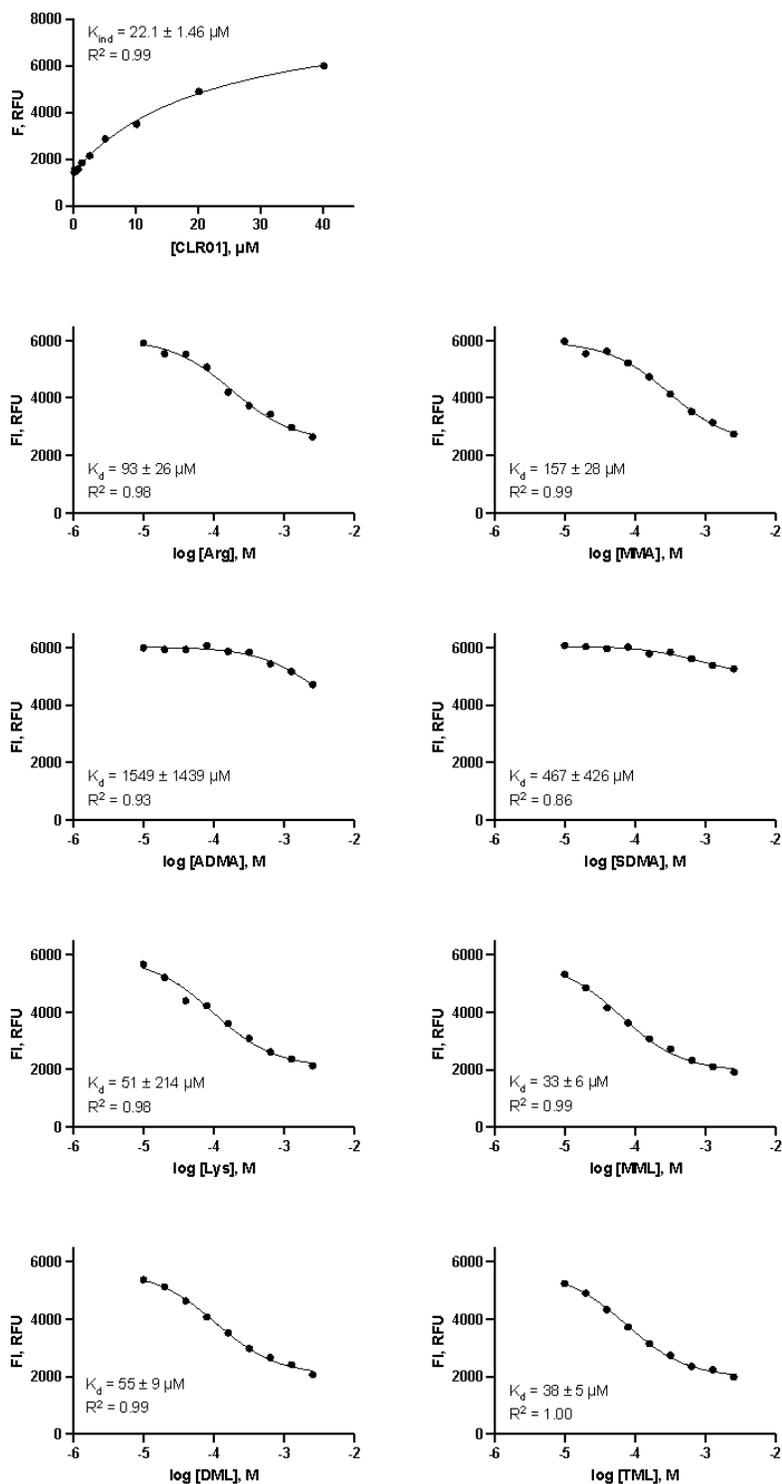


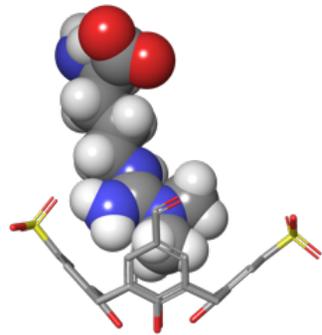
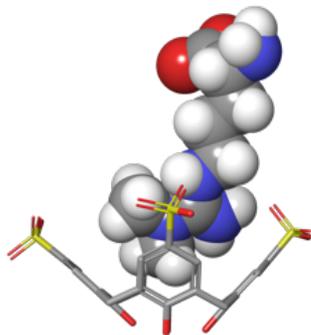
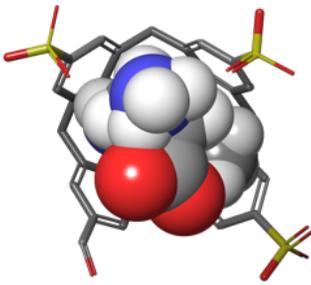
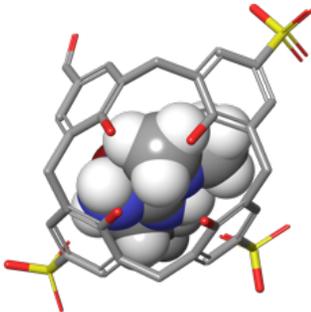
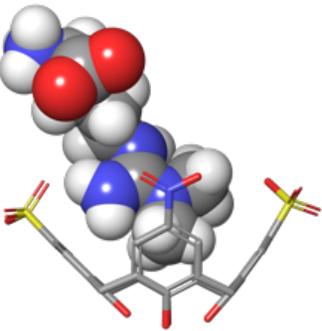
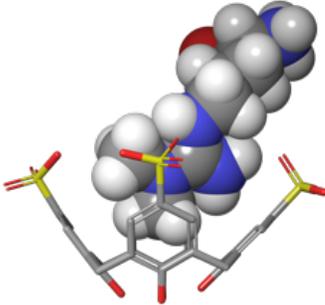
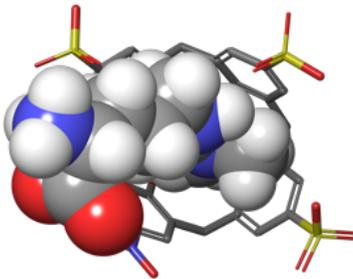
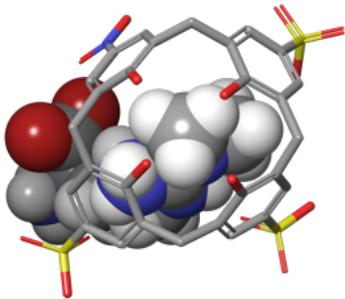
Figure S10: Fluorescence based studies of **CLR01**. a) Direct titration of 4-ASP(20 μM) with **CLR01** (0 – 40 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **CLR01**-4-ASP (70 μM **CLR01**, 20 μM 4-ASP) complex.

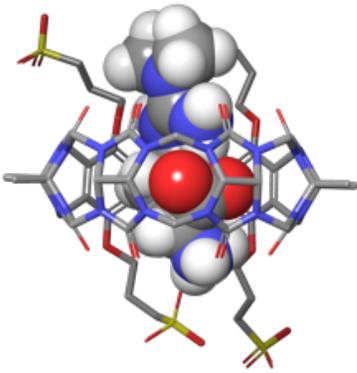
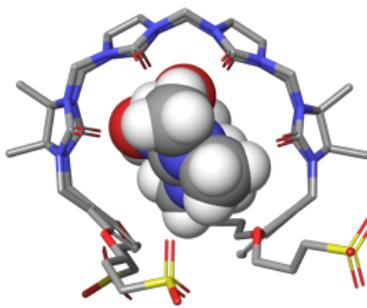
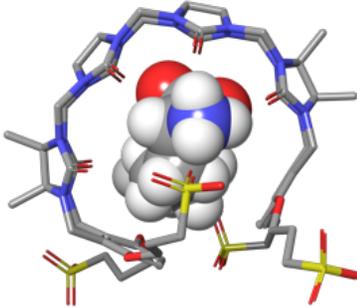
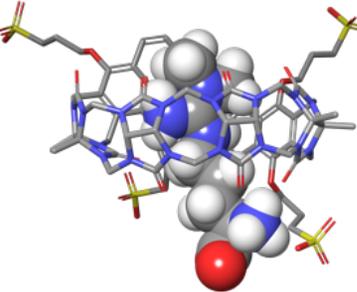
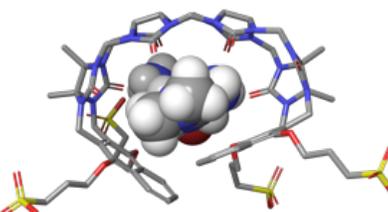
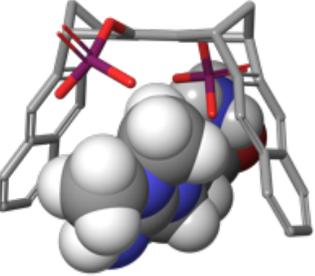
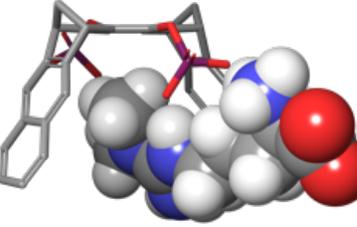
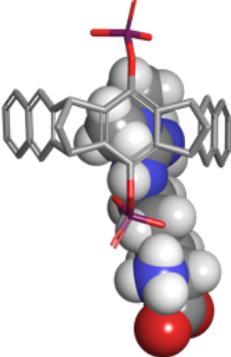
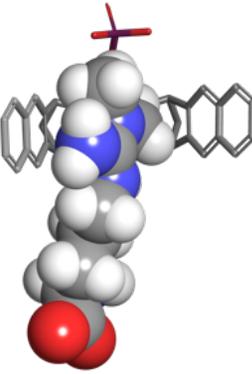
3. Molecular Modeling

Modeling was done using minimization in explicit water for all the complexes (Force Field: OPLS_2005), in Maestro. Hosts are shown without hydrogen atoms and guests are shown with H atoms. Color Code for atoms: C = gray, N = blue, H = white, O = red, P = purple, S = yellow. Each host is modeled with ADMA (**sCx4-CHO**, **sCx4-NO2**, **M1**, **M2**, **PC**, **CLR01**). Hosts **M1**, **M2**, **PC**, **CLR01** are also modeled with MML and with TML.

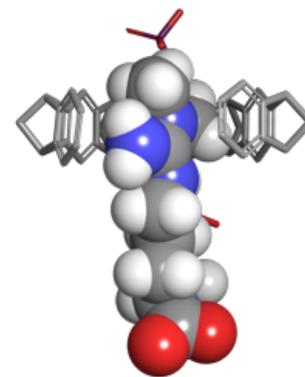
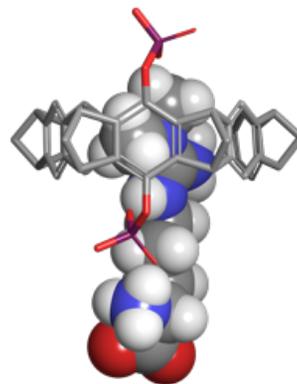
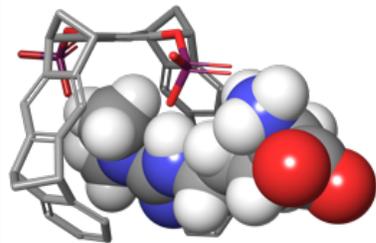
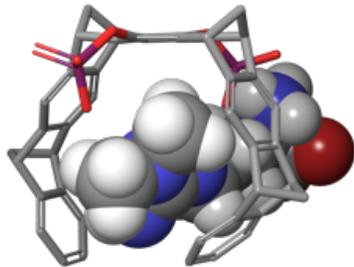
3.1. Each host modeled with ADMA. Front, back, top, and bottom views for all complexes.

Table S1: Energy-minimized complexes of hosts with ADMA (OPLS_2005, explicit water)

	Front	Back	Top	Bottom
sCx4-CHO				
sCx4-NO2				

M1				
M2				
PC				

CLR0
1



3.2. Hosts **M1**, **M2**, **PC**, **CLR01** each modeled with MML and with TML.

Table S2: Energy-minimized complexes of host **M1** with MML and TML (OPLS 2005, explicit water)

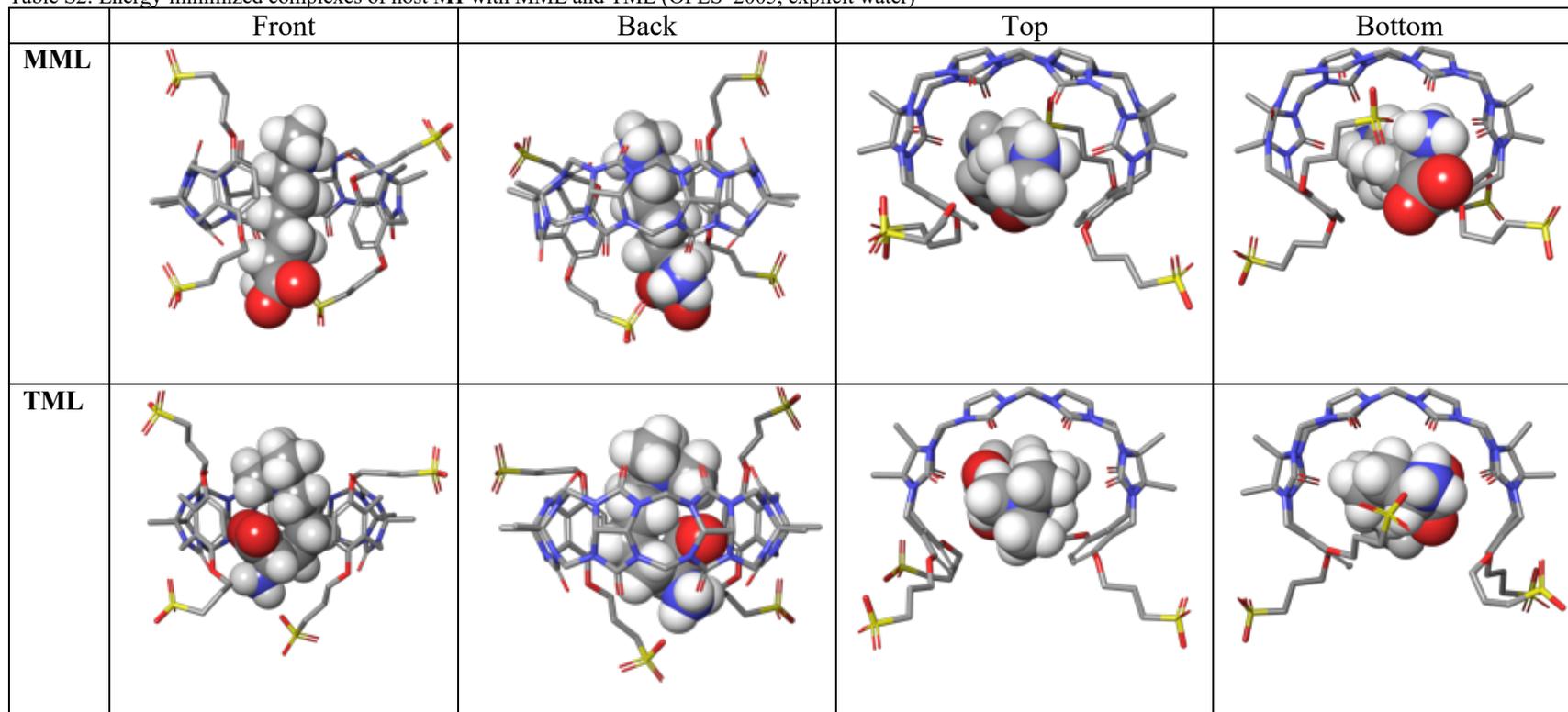


Table S3: Energy-minimized complexes of host **M2** with MML and TML (OPLS 2005, explicit water)

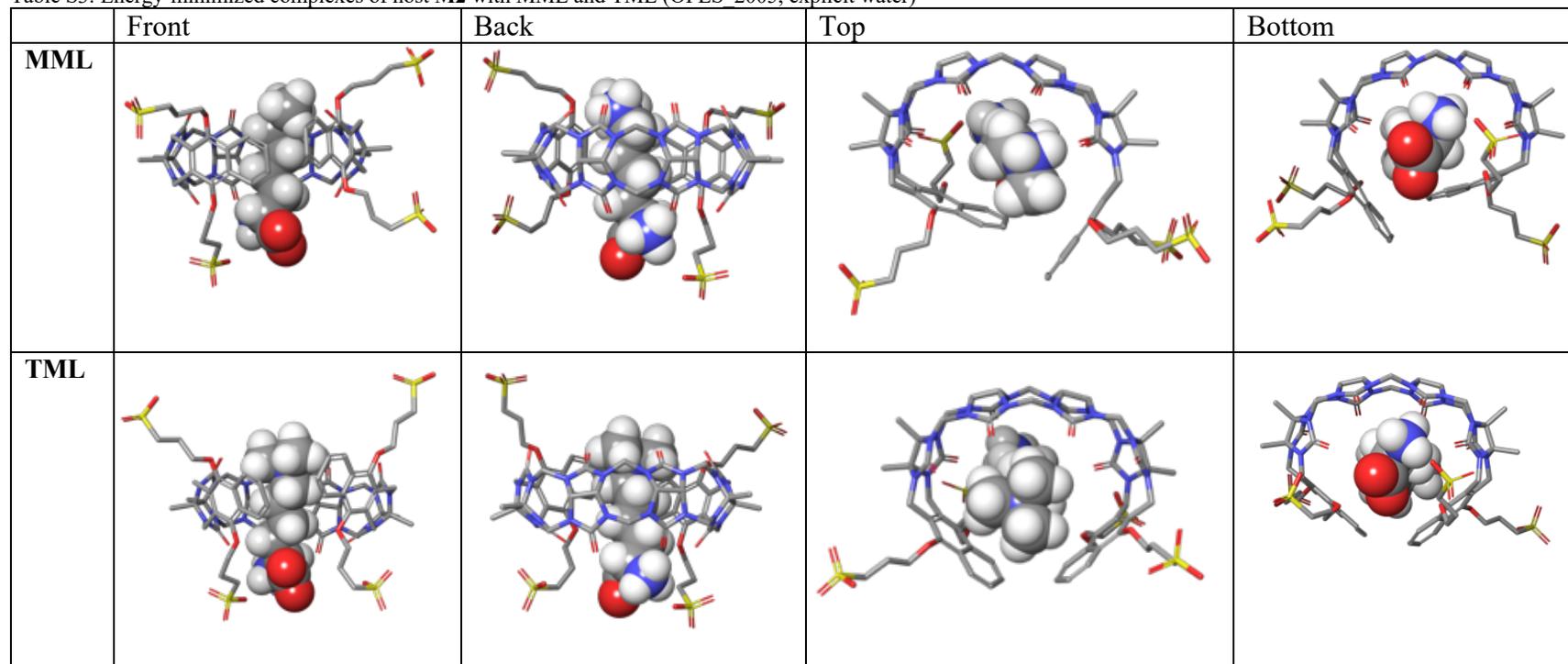


Table S4: Energy-minimized complexes of host PC with MML and TML (OPLS 2005, explicit water)

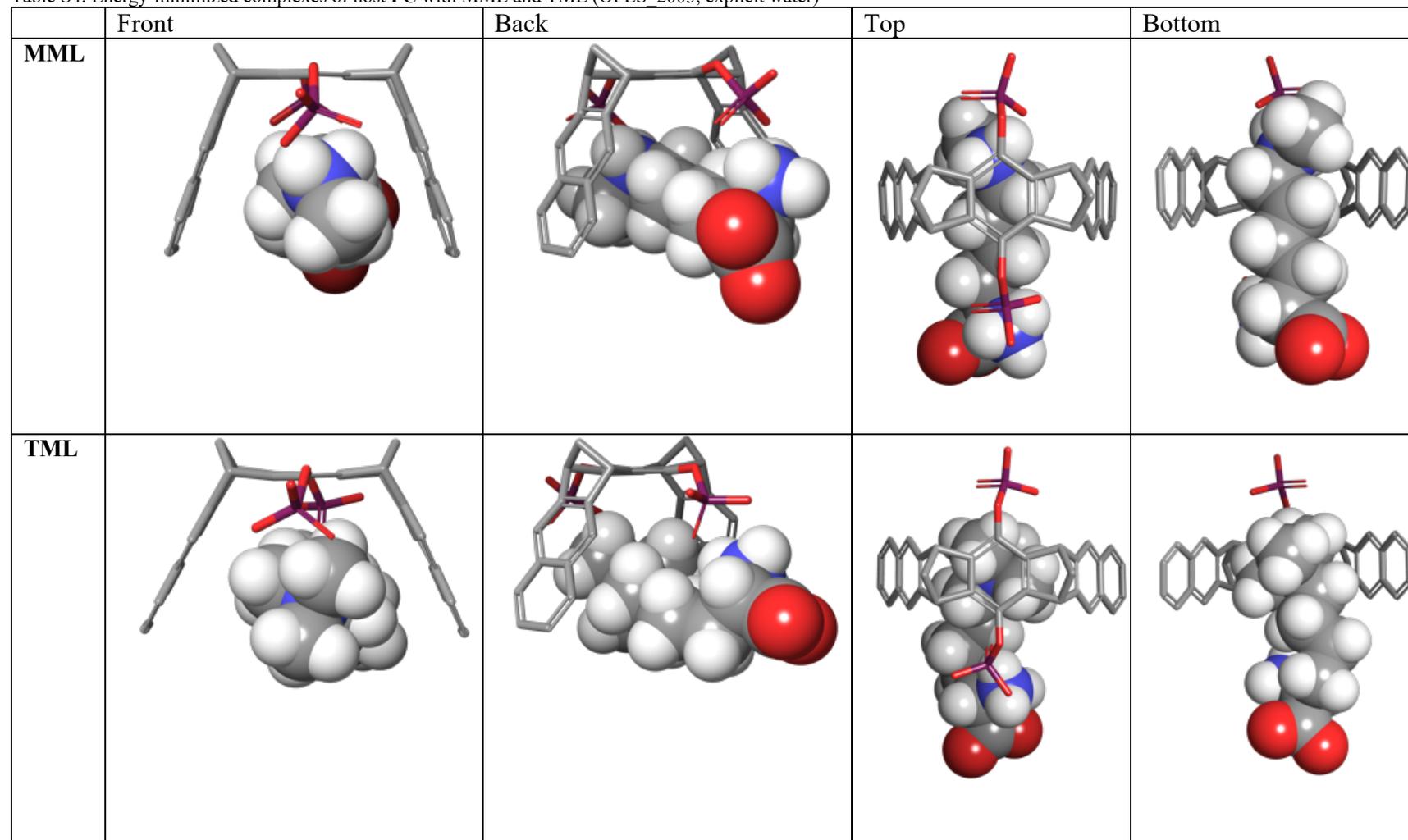
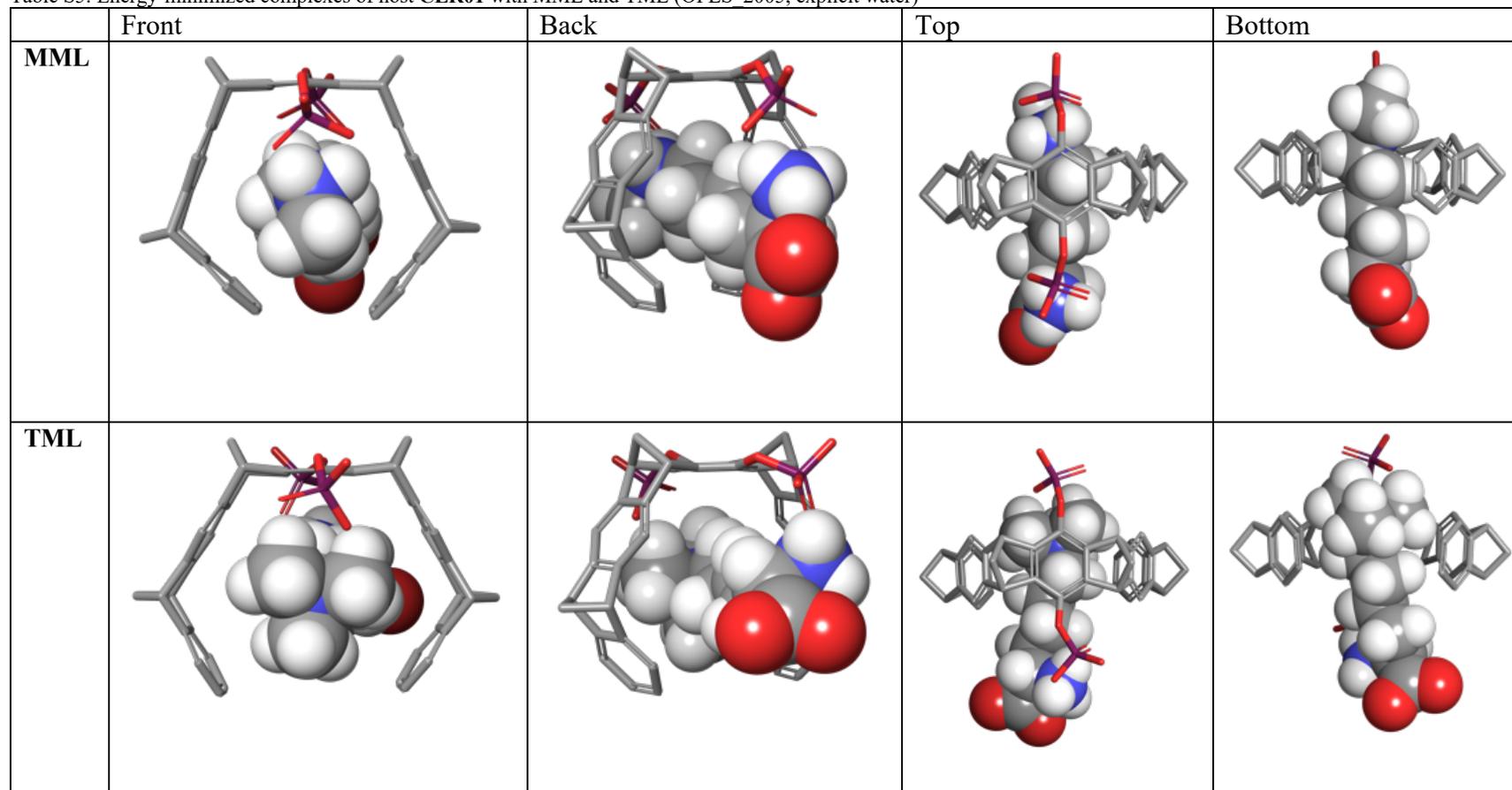


Table S5: Energy-minimized complexes of host **CLR01** with MML and TML (OPLS_2005, explicit water)



4. References

- (1) Beatty, M. A.; Borges-Gonzalez, J.; Sinclair, N. J.; Pye, A. T.; Hof, F. Analyte-Driven Disassembly and Turn-On Fluorescent Sensing in Competitive Biological Media. *J Am Chem Soc* **2018**, *140*, 3500-3504.
- (2) Daze, K. D.; Ma, M. C.; Pineux, F.; Hof, F. Synthesis of new trisulfonated calix[4]arenes functionalized at the upper rim, and their complexation with the trimethyllysine epigenetic mark. *Org Lett* **2012**, *14*, 1512-1515.
- (3) Ma, D.; Hettiarachchi, G.; Nguyen, D.; Zhang, B.; Wittenberg, J. B.; Zavalij, P. Y.; Briken, V.; Isaacs, L. Acyclic cucurbit[n]uril molecular containers enhance the solubility and bioactivity of poorly soluble pharmaceuticals. *Nat Chem* **2012**, *4*, 503-510.
- (4) Hadrovic, I.; Rebmann, P.; Klarner, F. G.; Bitan, G.; Schrader, T. Molecular Lysine Tweezers Counteract Aberrant Protein Aggregation. *Front Chem* **2019**, *7*, 657.
- (5) Talbiersky, P.; Bastkowski, F.; Klarner, F. G.; Schrader, T. Molecular clip and tweezer introduce new mechanisms of enzyme inhibition. *J Am Chem Soc* **2008**, *130*, 9824-9828.