Structurally diverse functionalized hosts for the binding of cationic drugs and metabolites

By

Zoey Warmerdam

B.Sc., University of Applied Science Leiden, 2016

A Dissertation Submitted in Partial Fulfilment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Chemistry

©Zoey Warmerdam, 2023 University of Victoria All rights reserved. This dissertation may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

We acknowledge and respect the lək^{w}əŋən peoples on whose traditional territory the university stands and the Songhees, Esquimalt and WSÁNEĆ peoples whose historical relationships with the land continue to this day.

Structurally diverse functionalized hosts for the binding of cationic drugs and metabolites

By

Zoey Warmerdam

B.Sc., University of Applied Science Leiden, 2016

Supervisory Committee

Dr. Fraser Hof, Department of Chemistry
Supervisor

Dr. Peter Wan, Department of Chemistry
Departmental Member

Dr. Peter Constabel, Department of Biology
Internal External Member

Abstract

The goal for my thesis is to learn about structure-activity relationships between sulfonatocalix[4]arene host and small molecule targets, to extend on the fundamental knowledge that is present in today's literature. This was achieved using diverse synthetic approaches and supramolecular binding studies.

Chapter 1 introduces the diversity-oriented functionalization of host molecules. The topics of supramolecular chemistry, molecular recognition and diversity-oriented synthesis are introduced, and it also reviews literature on diversity-based covalent functionalization of hosts. Extra attention is spent on the calix[n]arene hosts as they form the basis for this thesis. In Chapter 2, my work on regioselective modifications of the calix[4]arene upper rim is presented. This work is fundamental going forward in my thesis, as in this chapter I establish the synthetic firsts steps of the hosts that are used throughout. I then studied the binding of nicotine and analogues by a small library of modified hosts through indicator displacement assays and NMR, to learn lessons about binding strength and geometry. I conclude that calix[4]arenes show different patterns of binding strength, geometry and selectivity depending on the upper-rim substitutions. In Chapter 3 this approach was extended to study methylated arginines and lysines and to include a wider variety of host structures. This chapter is part of a collaboration in which our collaborators contributed non-calixarene hosts with different shapes, allowing us to explain selectivity patterns by a common framework that considers the geometry, depth of binding pockets, and functional group participation across different host classes. For this I utilized indicator displacement assays and visualized our findings

with molecular modelling of the complexes. I found that a more enclosed geometry of the host with a bigger binding pocket allowed for stronger interactions.

In Chapter 4 I establish a method using phage displayed peptides to synthesize a large library of calix[4]arenes with more enclosed binding pockets. The general design was to make calixarenes that are strapped by peptidic elements, I used phage-displayed peptides to bring diversity to the host library. I establish the creation of calixarene-peptide hybrids on the surface of phage and made first efforts to use an affinity pull-down to select a host from this library. A set of hosts selected through the affinity pull-down were re-synthesized and their binding constants for targeted guests were determined *via* indicator displacement assays. I found that the affinity pull-down did not result in strong binding host and that improvements are needed. To get a better understanding of why the hosts form weak complexes, I studied one of the complexes with 1D and 2D NMR and learned that the peptide undergoes a hydrophobic collapse creating a competition for the target of interest.

Table of Contents

List of Tables

List of Figures

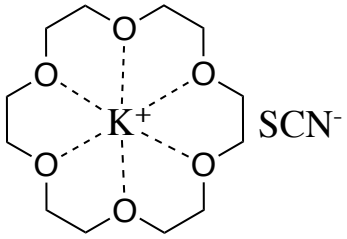
List of abbreviations

[D]	molar concentration of dye
[H]	molar concentration of host
4-ASP	4-(4-Diethylaminostyryl)-1-methylpyridinium iodide
ADMA	asymmetric dimethyl arginine
BIA	biotin iodoacetamide
br	broad
CAD	covalent antibody displays
CB	cucurbituril
CD	cyclodextrin
CDCl3	deuterated chloroform
CGTase	cyclodextrin glucotransferase
CH2Cl2	dichloromethane
CIS	*cis*-activity based
COSY	correlated spectroscopy
DCC	dynamic combinatorial chemistry
DCL	dynamic combinatorial library
DIPEA	N,N-diisopropylethylamine
DMF	dimethylformamide
DML	dimethyl lysine
DNA	deoxyribonucleic acid
DOS	diversity-oriented synthesis
e.g.	exempli gratia, for example
ESI	electrospray ionization
F	fitted data point
FBDD	fragment based drug design
FDA	food and drug administration
Fmax	maximum signal
Fmin	minimum signal
FT-IR	Fourier-transform infrared spectroscopy
g	gram
GDCCs	Gibb deep cavity cavitands
H	enthalpy
h	hour
HA	hydrogen bond acceptor
HBTU	1-[Bis(dimethylamino)methylene]-^{1}H-benzotriazolium 3-Oxide Hexafluorophosphate.
HD	hydrogen bond donor
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

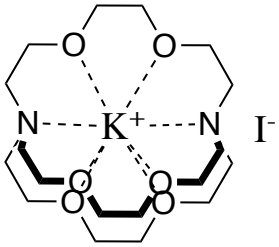
HMTA	hexamethylenetetramine
HPLC	high-performance liquid chromatography
HPβCD	hydroxypropyl-β-cyclodextrin
HR-MS	high resolution mass spectrometry
Hsp90	heat shock protein 90
Hz	hertz
i.e.	id est, in other words
IDAs	indicator displacement assays
IMPDH	inosine 5′-monophosphate dehydrogenase
IPTG	isopropyl β- d-1-thiogalactopyranoside
IR	infrared
ITC	isothermal titration calorimetry
IUPAC	International Union of Pure and Applied Chemistry
K_{assoc}	association constant
K_d	dissociation constant
K_i	equilibrium dissociation constant in Molar
K_{ind}	host-indicator dissociation constant
KMe3	trimethylated lysine
LCG	lucigenin
logEC50	log of the concentration of the competitor binding half-way between Fmin and Fmax
logK_d	log K_d
M	molar
m	medium
MHz	mega hertz
mL	milliliter
mM	micromolar
mM	millimolar
mm	millimeter
MMA	monomethyl arginine
MML	monomethyl lysine
mmol	millimol
MS	mass spectroscopy
msec	millisecond
n	number of experiments
Na_2HPO_4	sodium phosphate
nM	nanomolar
nm	nanometer
NMR	nuclear magnetic resonance
NO	nitric oxide

NOESY	nuclear overhauser effect spectroscopy
NOS	nitric oxide synthase
OBOC	one-bead-one-compound
PBS	phosphate-buffered saline
PFUs	plaque forming units
pH	potential of hydrogen
ppm	parts per million
R2	coefficient of determination
RAMEβCD	randomly methylated β-cyclodextrin
Rhodamine 6G	R6G
RMSE	root mean squared error
RNA	ribonucleic acid
S	entropy
s	strong
SAMPL	statistical assessment of modeling of proteins and ligands
SD1	standard derivation of the first triplicate
SD2	standard derivation of the second triplicate
SDK_d	standard error of K_d
SDlogK_d	standard error of the log K_d
SDMA	symmetric dimethyl arginine
SDtotal	standard derivation of both experiments
SPPS	solid-phase peptide synthesis
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
TLC	tin layer chromatography
TML	trimethyl lysine
UV	ultraviolet
w	weak
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
δΔ	delta
μL	microliter
μM	micromolar
μm	micrometer

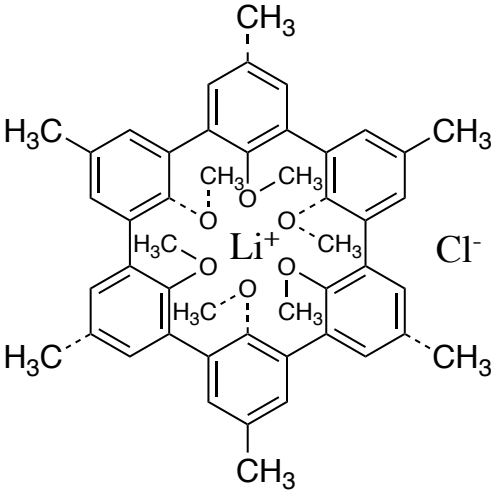
List of chemical structures in order of which they appear in the thesis



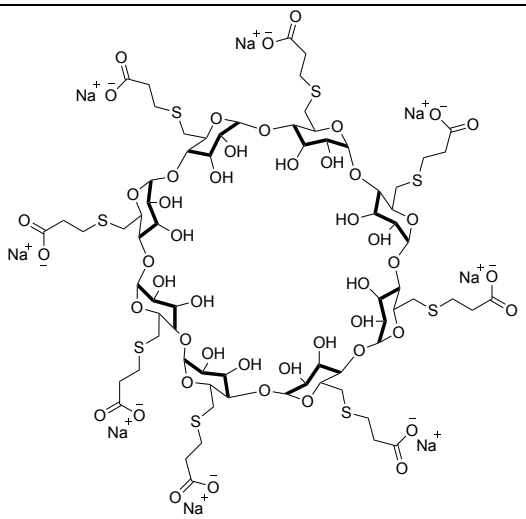
Crown ether in complex with Potassium



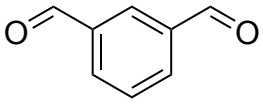
Cryptand in complex with Potassium



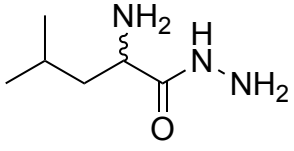
Spherand in complex with Lithium



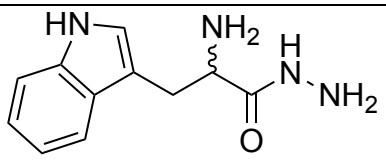
Sugammadex



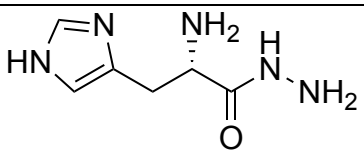
1



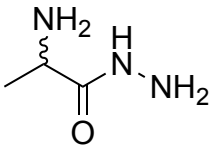
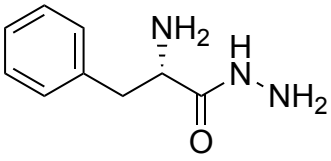
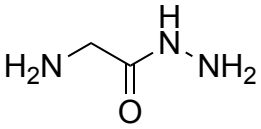
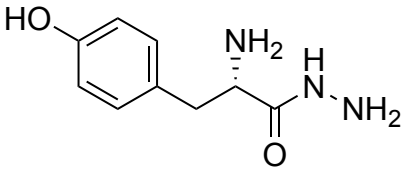
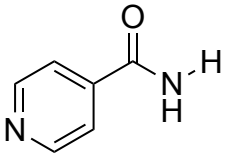
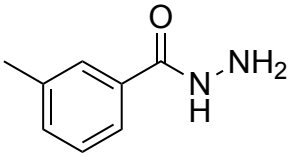
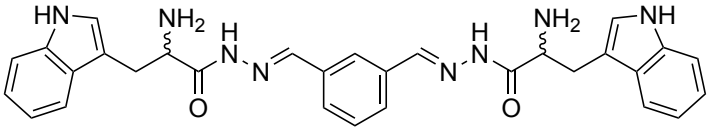
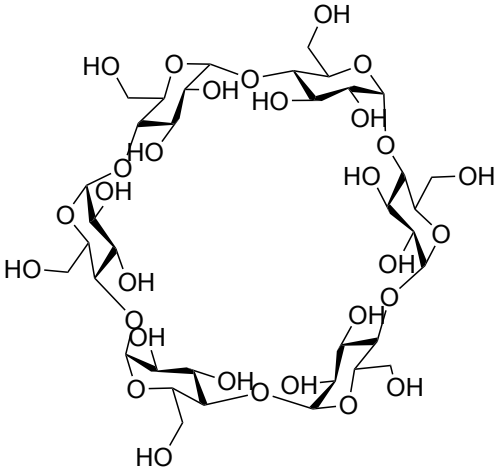
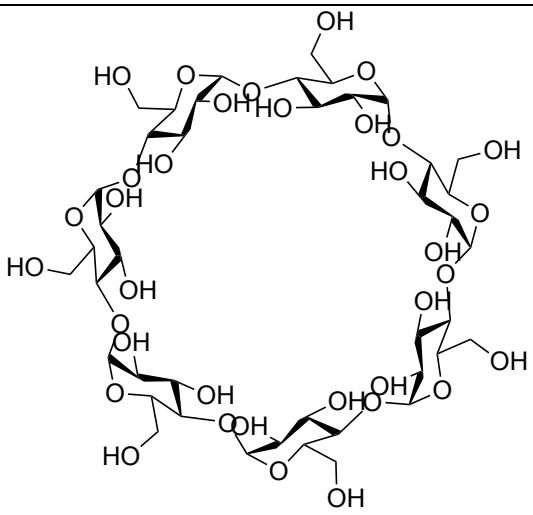
2

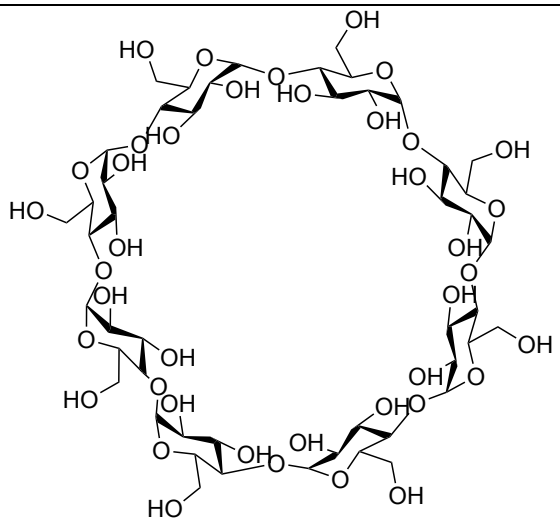


3

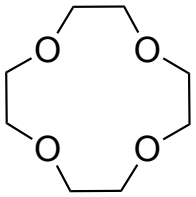


4

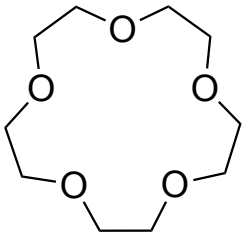
 **5**	 **6**
 **7**	 **8**
 **9**	 **10**
 **11**	
 **α-Cyclodextrin**	 **β-cyclodextrin**



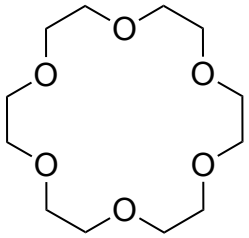
γ-cyclodextrin



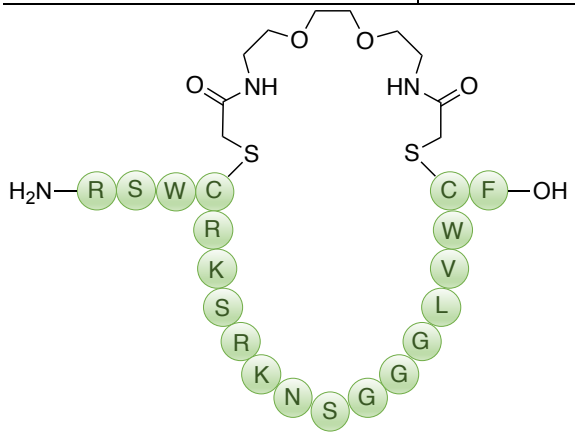
12-crown-4



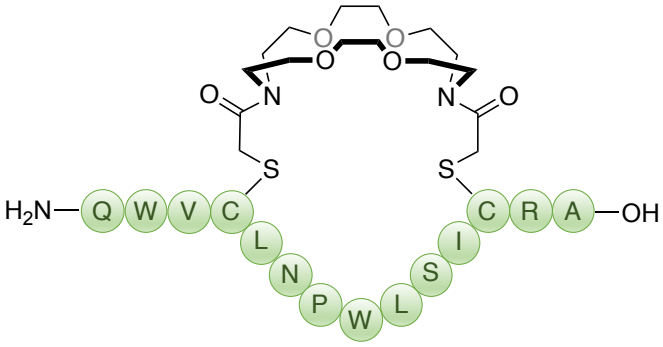
15-crown-5



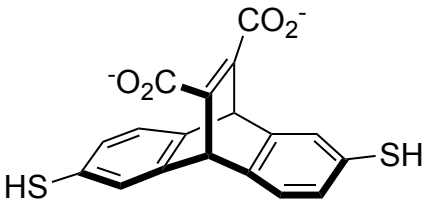
18-crown-6



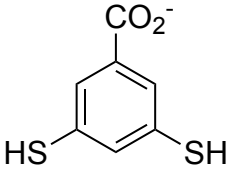
12



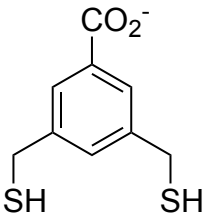
13



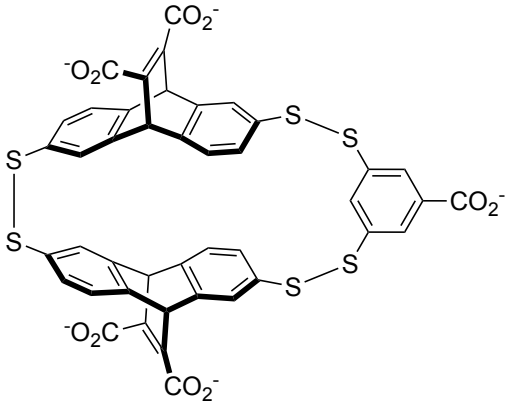
14



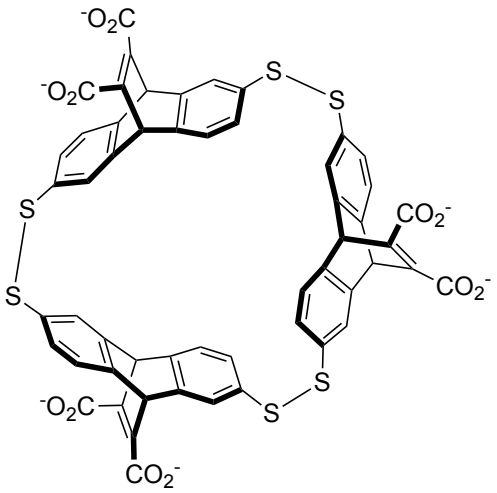
15



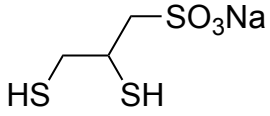
16



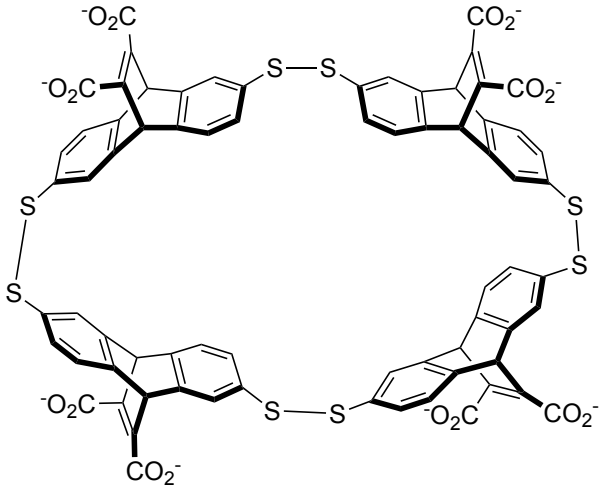
17



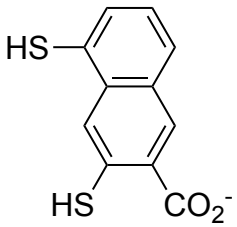
18



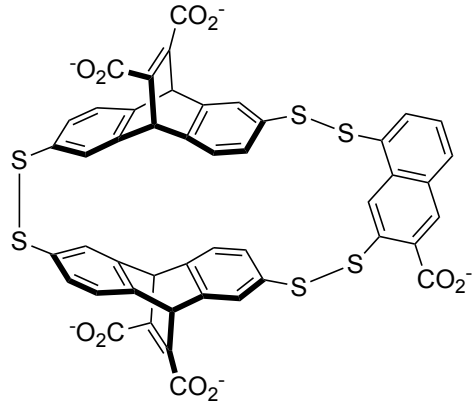
19



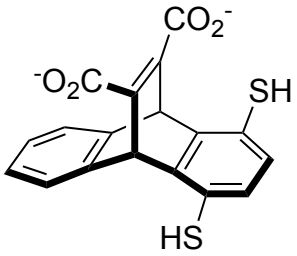
20



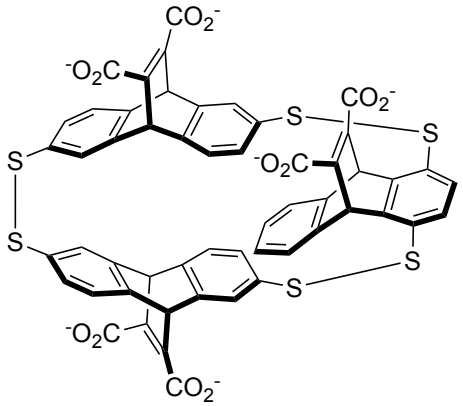
21



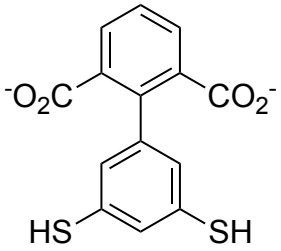
22



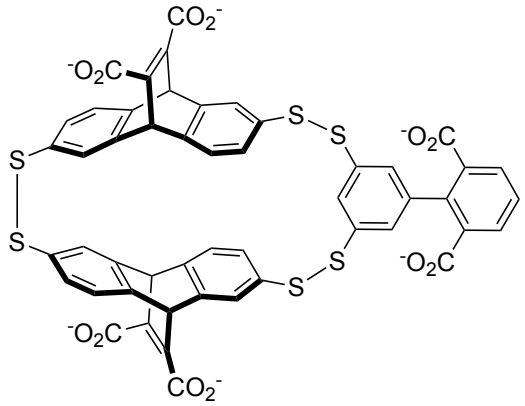
23



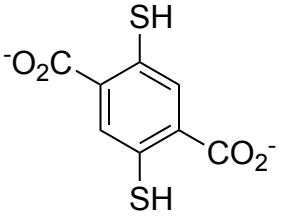
24



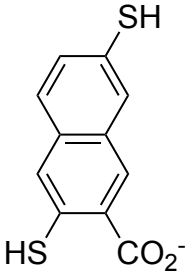
25



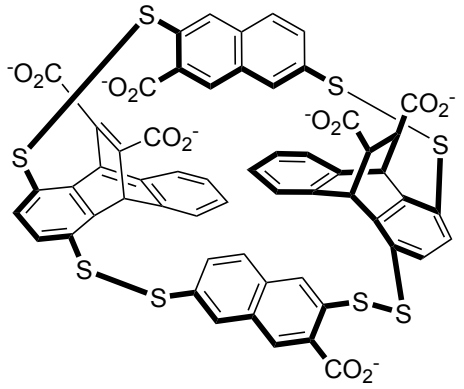
26



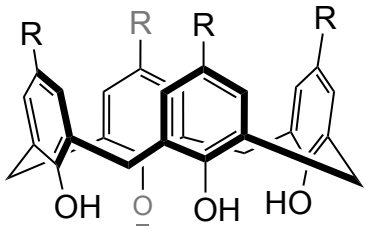
27



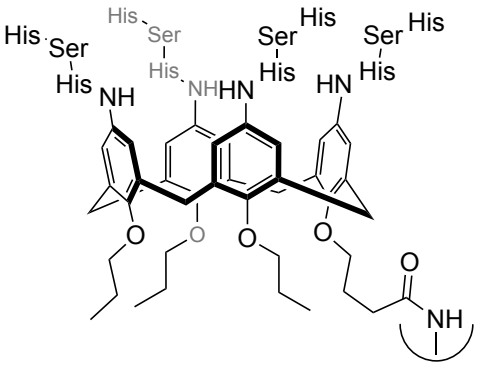
28



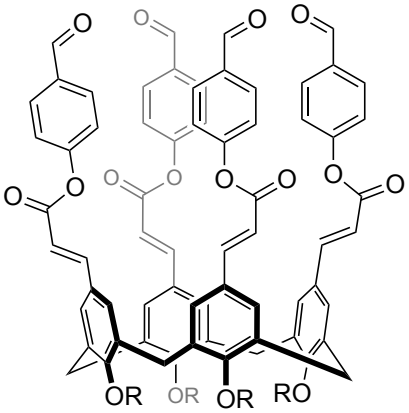
29



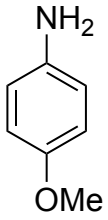
Calix[4]arene



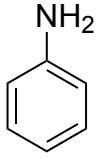
30



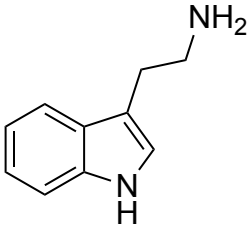
31



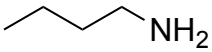
32



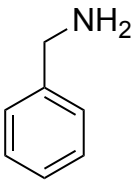
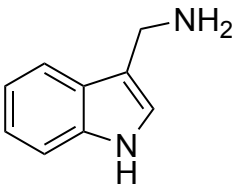
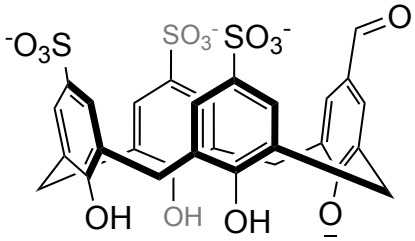
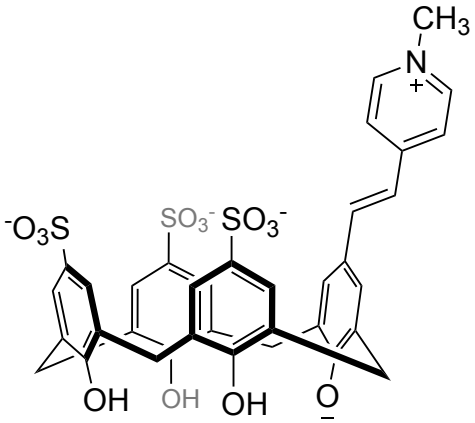
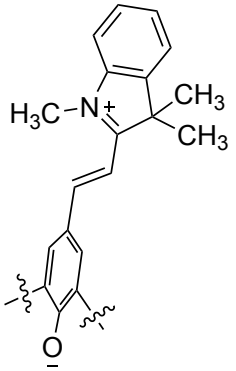
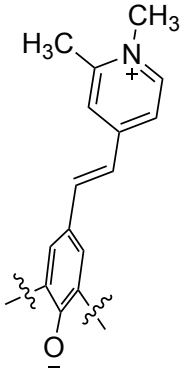
33

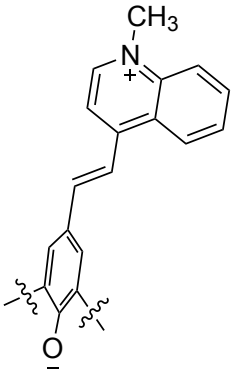


34

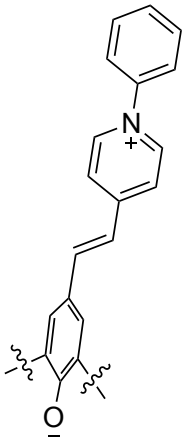


35

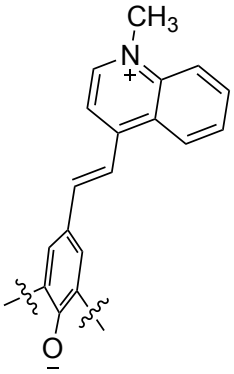
 **36**	 **38**
 **sCx4-CHO**	 **38**
 **39 (see 38)**	 **40 (see 38)**



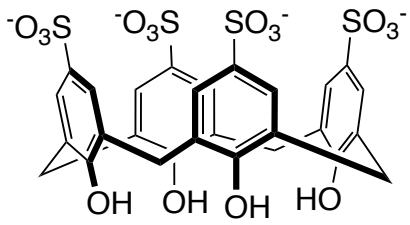
41 (see 38)



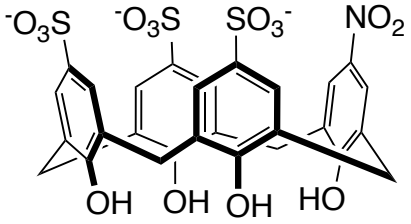
42 (see 38)



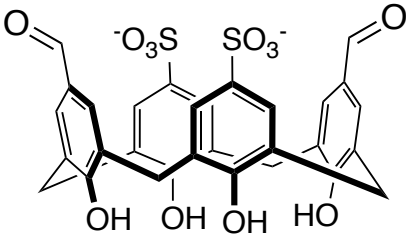
43 (see 38)



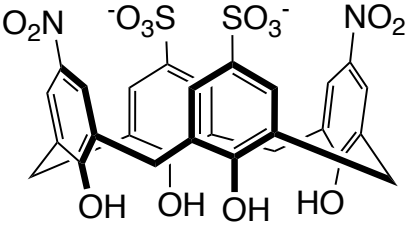
sCx4



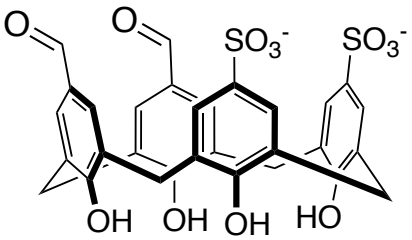
sCx4-NO_2



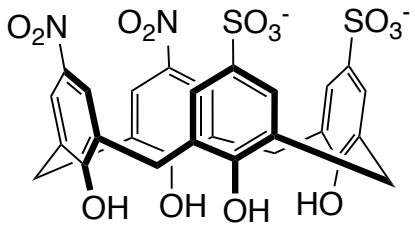
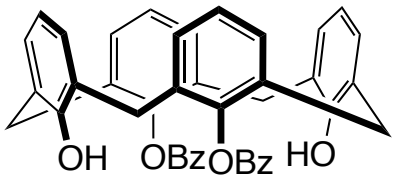
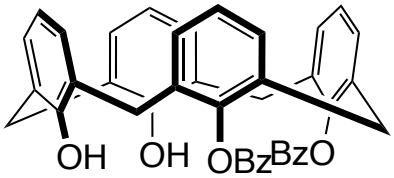
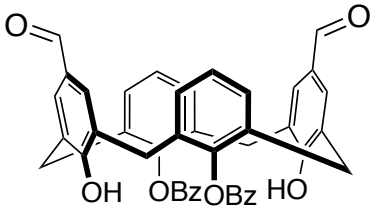
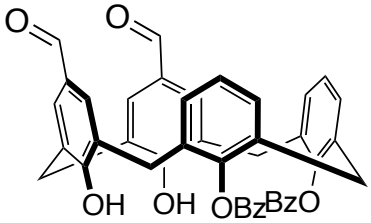
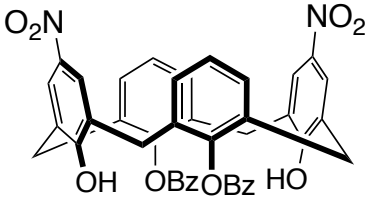
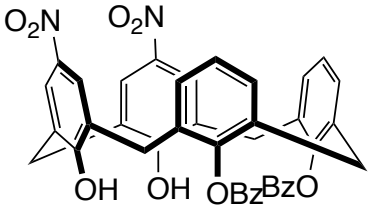
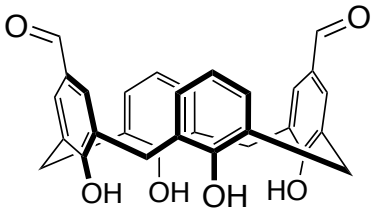
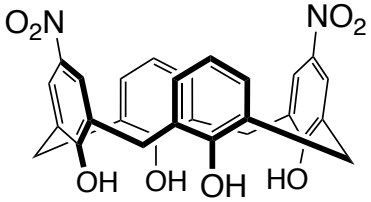
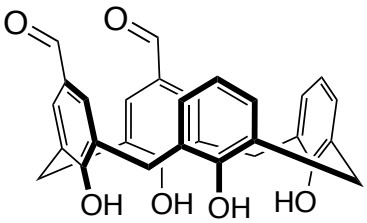
sCx4-1,3-CHO

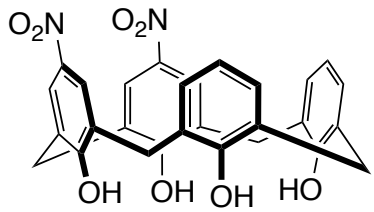


sCx4-1,3-NO_2

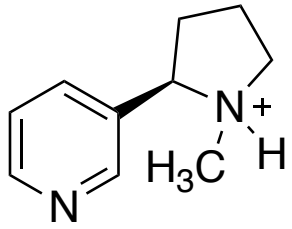


sCx4-1,2-CHO

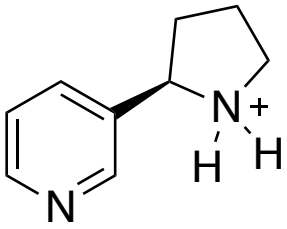
 **sCx4-1,2-NO_2**	 **Cx4-1,3-OBz**
 **3,4-dibenzoylcalix[4]arene**	 **1,3-diformyl-2,4-dibenzoylcalix[4]arene**
 **1,2-diformyl-3,4-dibezoylcalix[4]arene**	 **1,3-dinitro-2,4-dibezoylcalix[4]arene**
 **1,2-dinitro-3,4-dibezoylcalix[4]arene**	 **1,3-diformyl-calix[4]arene**
 **1,3-dinitro-calix[4]arene**	 **1,2-diformylcalix[4]arene**



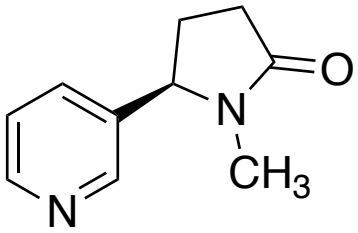
1,2-dinitrocalix[4]arene



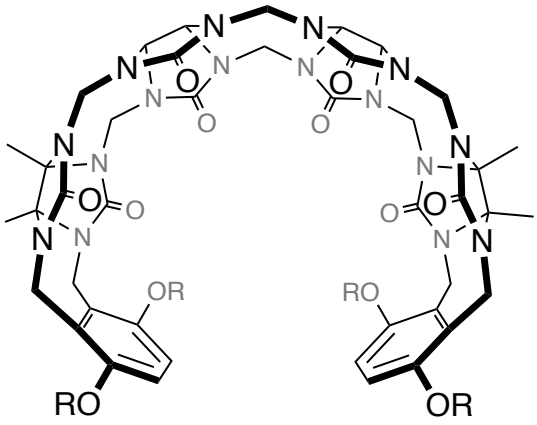
Nicotine



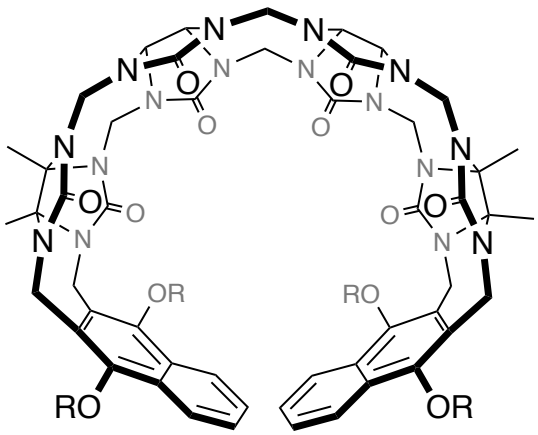
Nornicotine



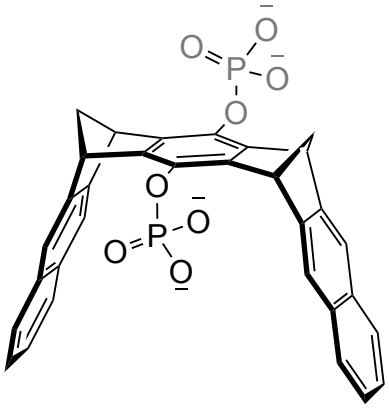
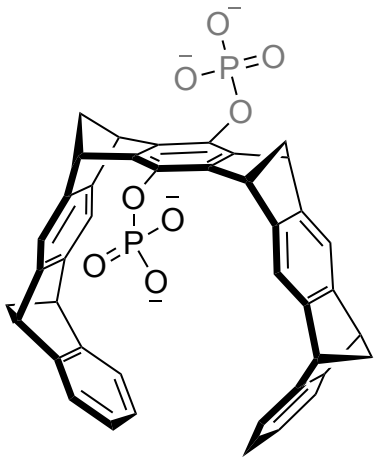
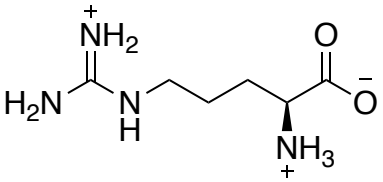
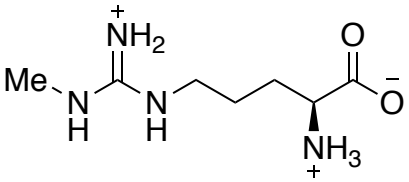
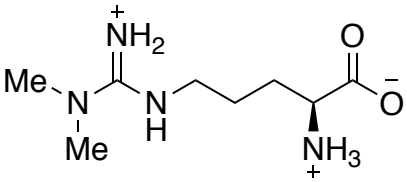
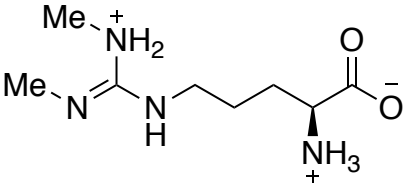
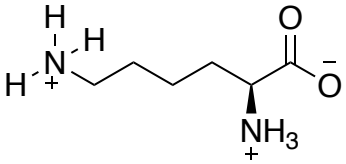
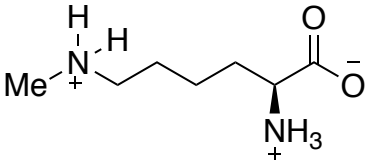
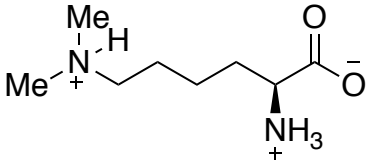
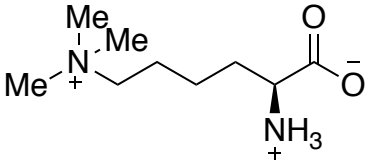
Cotinine

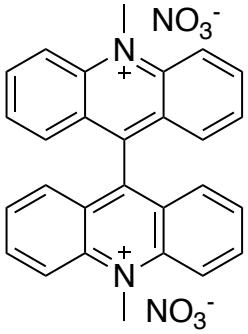
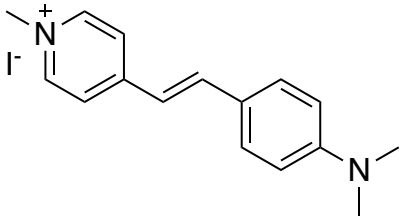
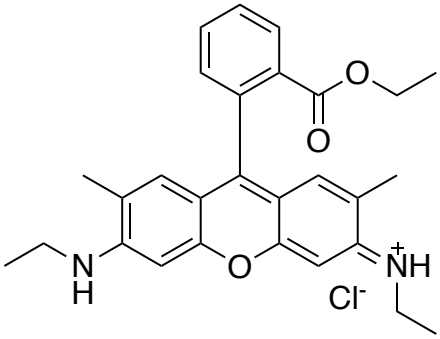
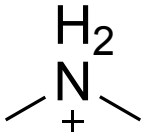
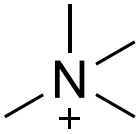
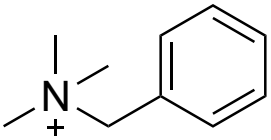
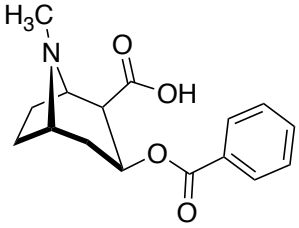
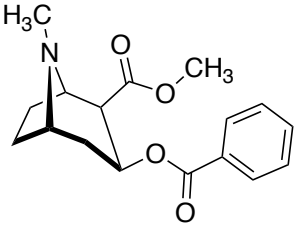


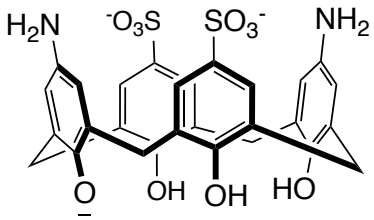
M1, R = $(CH_2)_3SO_3^-$



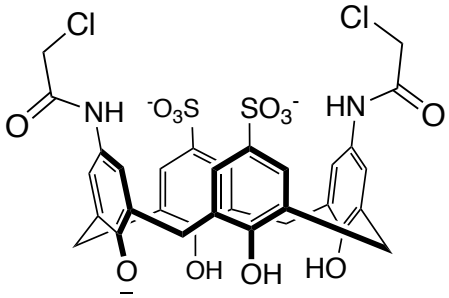
M2, R = $(CH_2)_3SO_3^-$

 **PC**	 **CLR01**
 **Arginine**	 **MMA**
 **ADMA**	 **SDMA**
 **Lysine**	 **MML**
 **DML**	 **TML**

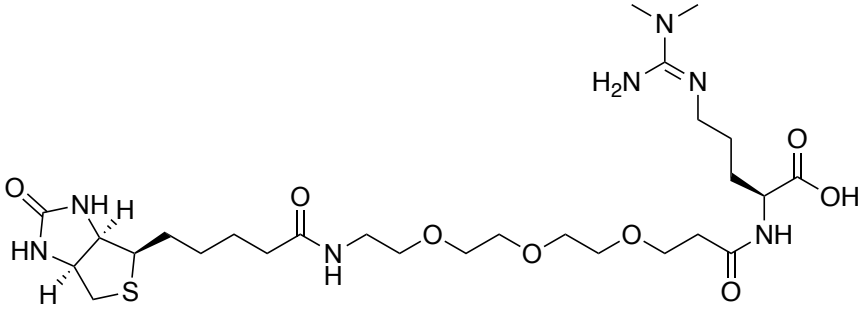
 **Lucigenin**	 **4-ASP**
 **Rhodamine 6G**	 **Dimethylammonium**
 **Tetramethylammonium**	 **Benzyl trimethylammonium**
 **Benzoylecgonine**	 **Cocaine**



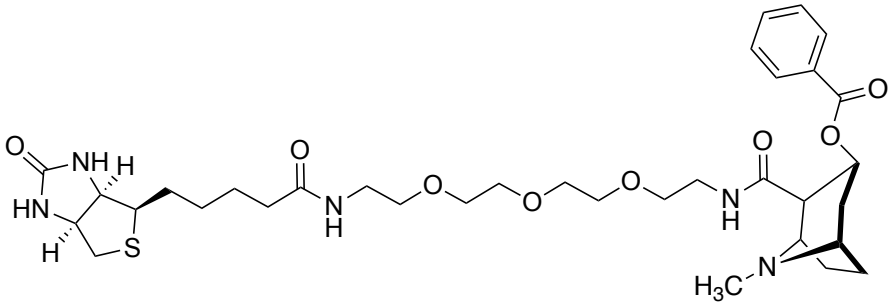
sCx4-1,3-NH_2



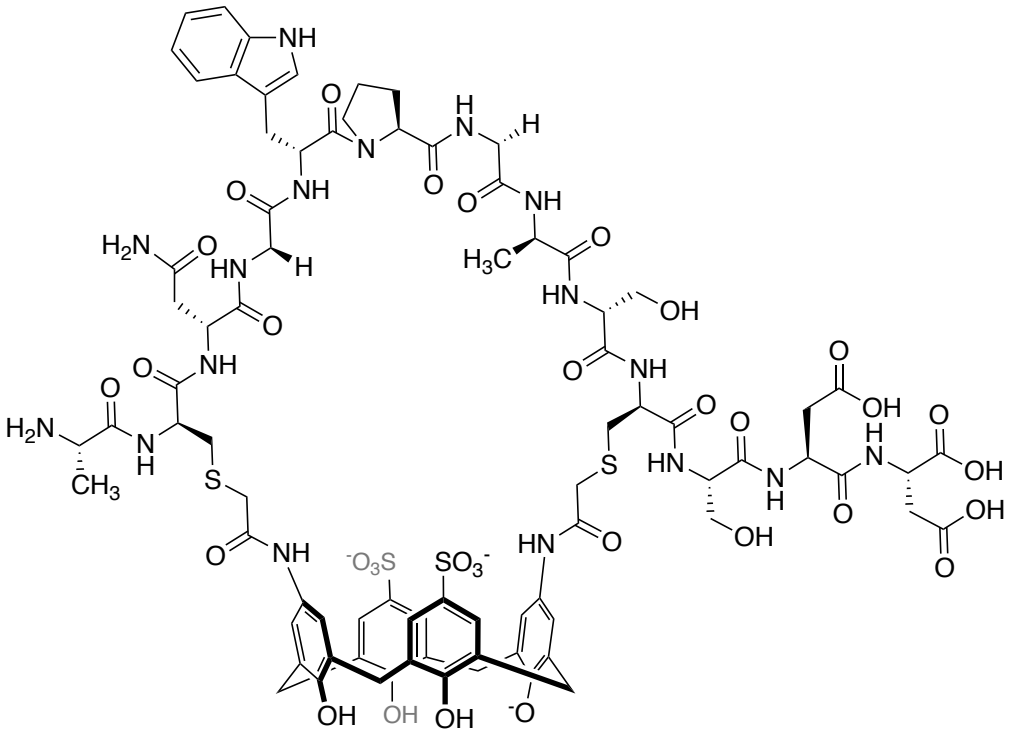
sCx4-1,3-$NHCOCH_2Cl$



Biotin-PEG3-ADMA



Biotin-PEG3-benzoylecgonine



sCx4-P1

Dedication

Aan mama en papa

Acknowledgements

Throughout my PhD work I have received a lot of support. First, I would like to thank Dr. Fraser Hof, for being a kind supportive mentor. Your patience, guidance and witty comments have thought me a lot. Secondly, my work could have not been completed if it was not for the support, experience and knowledge of the technical staff, Becky Hof, Chris Barr, Mark Evans, Ori Granot, Sean Adams, and Tyler Trefz you all were fundamental to my research.

A lot of support came from my friends and family. I especially want to thank Leah Gajecki and Nick Richard, for always being there to grab a drink and to give me an "outsiders" view on my science. Allison Selinger and Chelsea Wilson, you for all the walks, talks and coffee breaks. Finally, I want to thank my parents for always being there for me, and fully supporting my decision to move far away from home. The many video calls and cards have made the distance feel way smaller than it is.

Chapter 1: Diversity-oriented functionalization of host molecules

Creating hosts with high selectivity and strong binding is one of the goals in supramolecular chemistry.[1] The term supramolecular chemistry was first proposed by Jean-Marie Lehn, to combine the chemistry of molecular recognition, molecular assembly, and molecular association.[2] Molecular recognition is at the core of supramolecular chemistry and looks at systems where two or more molecules are selectively interacting through non-covalent interactions. On their own, non-covalent interactions are weak, but when combined, they can have major effects on molecular behaviours. Non-covalent interactions are found throughout nature where they play a critical role in receptor-ligand binding (e.g. antibody-antigen, protein-DNA, protein-inhibitors etc.), and are crucial for the folded structures of biomacromolecules such as proteins, DNA, and RNA.[3]

In supramolecular chemistry, the molecule that recognizes and binds another molecule is referred to as the host when it has a concave binding pocket that can accommodate a guest. "Host-guest" chemistry is a major area of study under the general umbrella of molecular recognition. In 1987, host-guest chemistry was acknowledged with the Nobel prize in chemistry when it was awarded to Charles J. Pedersen, Jean-Marie Lehn, and Donald J. Cram for their work on the development of molecules that achieve selective recognition, mimicking the complexes of important biological processes. In 1967, Charles J. Pedersen published on crown ether hosts, he showed that they can selectively bind different ions depending on the size of the host. In 1969, Jean-Marie Lehn developed cryptands, which are bicyclic crown ether-like structures that showed an

improved selectivity over crown ethers. Subsequently Jean-Marie Lehn and Donald J. Cram developed more intricate hosts which were able to bind small molecules (Figure 1.1).

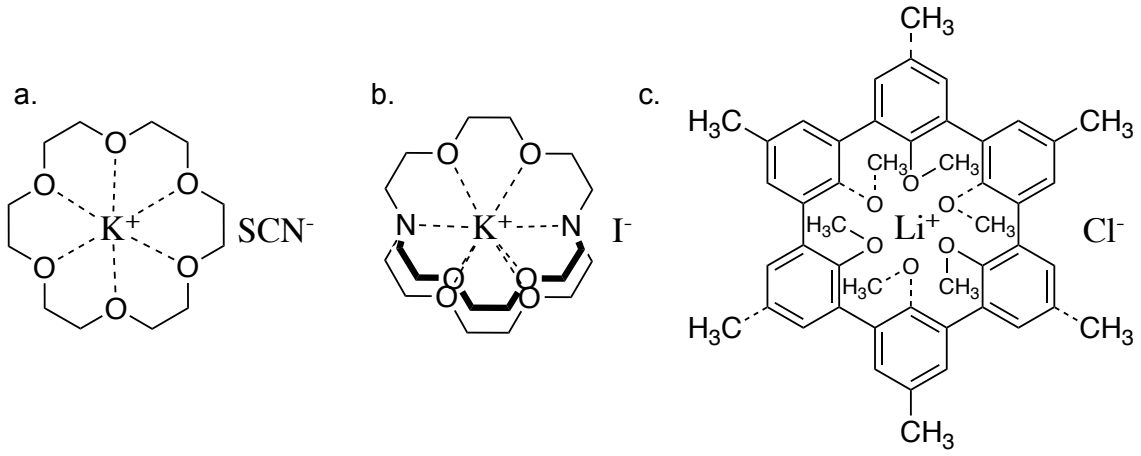


Figure 1.1. Host-guest complexes awarded with the Nobel price. a) Crown ether in complex with Potassium according to Pedersen. b) Cryptand in complex with Potassium according to Lehn. c) Spherand in complex with Lithium according to Cram.

The study of molecular recognition is traditionally based on individual non-covalent interactions. It is challenging to assign or understand any one single non-covalent interaction that is occurring at any individual moment within complex molecular structures. Despite this, host molecules have been designed for decades by using a reductionist approach, wherein individual functional groups are positioned in a way that will introduce one specific non-covalent interaction to achieve a certain molecular recognition outcome. This approach and its limits are described in more detail below.

In all cases, the study of a novel host molecule or other artificial recognition system starts with the synthesis of a new candidate molecule. Synthetic chemistry can be

a very time consuming endeavour;[4] the design, synthesis, purification, and characterization of new compounds can take anywhere from days to months, with no guarantee that a functional molecule will be produced. Many design-driven synthetic efforts are wasted when molecular hosts do not perform as expected. Many efforts in host-guest chemistry aim to mimic or perturb biological systems, but unfortunately the shortcomings of molecular designs are especially problematic in water, due to the unpredictability of solvation effects in this medium because of the strong hydrogen bond network and polarity.

Although host-guest chemistry has been studied since the 1960's, there are still relatively few examples that have made a significant impact in medicinal and biomedical chemistry. Cyclodextrins are widely used in pharmaceutical applications to increase the solubility of poorly soluble drugs and to increase their bioavailability and stability[5], an example of this is the melarsoprol–cyclodextrins inclusion complexes. Melarsoprol is a poorly water soluble drug used to treat trypanosomiasis (sleeping sickness), when in complex with randomly methylated β-cyclodextrin (RAMEβCD) or hydroxypropyl-β-cyclodextrin (HPβCD) the solubility enhancement factor is about 7.2×10^3 compared to the inherent solubility of Melarsoprol.[6] The only host that is an FDA-approved drug is Sugammadex (Figure 1.2). Sugammadex is a reversal agent of neuromuscular blockades, and is currently used as a reversal agent in anaesthetic practice.[7] There are advances being made toward other host molecules as drugs[8] and reversal agents[9,10].

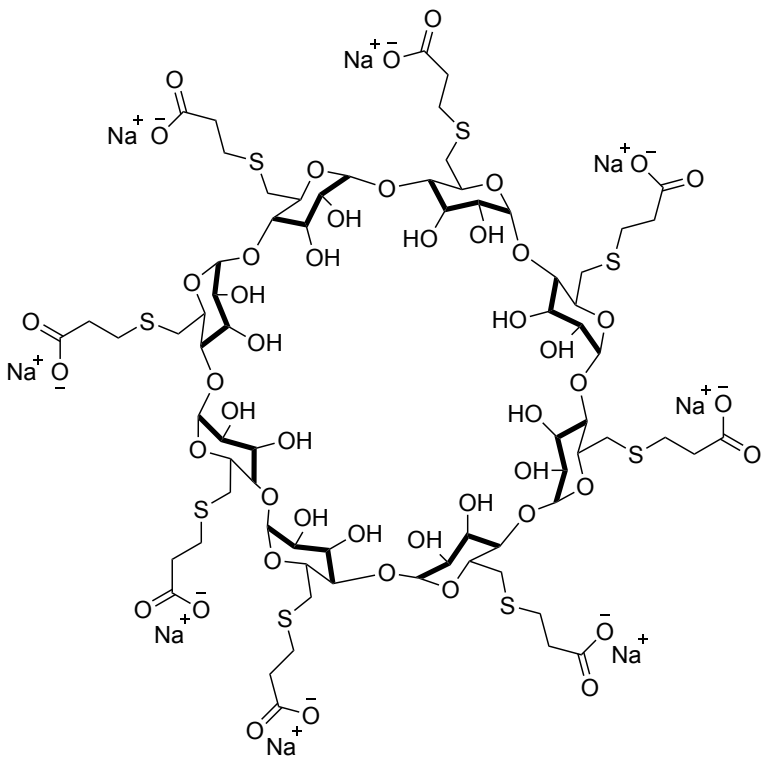


Figure 1.2. Sugammadex

Better methods are needed to design new host-guest systems that have a significant biomedical impact. The development of hosts in biomedicine has been inhibited by poor binding potencies and/or selectivity in biologically relevant conditions, such as salty warm water.[11, 12] Challenges associated with host-guest chemistry in biologically relevant conditions are the solubility of the hosts and competitive interactions between the intended guest, co-solutes, and water. Modification to the binding pockets of hosts provides a structural alteration to explore novel functionality and overcome the challenges associated with host-guest binding in water. The introduction of functional groups is often used to increase the solubility of hosts in water. These can be polar functional group

(carbonyls, polyethylene glycol tails, etc.) or charged functional groups (e.g., sulfonates, phosphonates, ammonium, etc.).[13] Other modifications directly lining the binding pocket can alter the host's selectivity and strength of binding for its targeted guest.

In the following sections, I will introduce non-covalent interactions from a holistic point of view, highlighting in places the features of non-covalent interactions that make it hard to design molecular recognition systems. I will then review examples of diversity-oriented synthesis (DOS) of host libraries by introducing each host class and the literature approaches to library experiments. Finally, I will state my thesis goals and give a preview of what will be presented in the following chapters.

1.1 Non-covalent interactions

Opposites attract. The recurring theme when studying non-covalent interactions is the central role of electrostatic interactions, in particular the attraction between positive and negative charges and charge distributions. These attractions can be observed for different subclasses of electrostatic interactions, which occur between formally charged species (ions), polar charge distributions (dipoles), and induced dipoles. An induced dipole is created when two molecules approach each other, and a temporary uneven charge distribution is formed within each molecule.

Proteins offer a good framework for discussing a variety of electrostatic interactions (Figure 1.3). The structure of proteins can be broken down into four levels of structure. The primary structure is the sequence of the peptide chain; the secondary structure is the folding of portions of the polypeptide into alpha helix, beta sheet, or other organized structural elements; the tertiary structure is the three-dimensional folding of the

polypeptide into an organized structure; The quaternary structure is where multiple molecules' tertiary structures assemble to form higher level arrangements. Protein folding is largely governed by different kinds of electrostatic interactions between the different charges and local charge distributions within the protein's structure.

Electrostatic interactions can arise in different ways and are often divided into separate classes (ion-ion, ion-dipole, and dipole-dipole). Ion pairs are the strongest type of electrostatic interaction, but, like all non-covalent interactions are heavily dependent on their environment. Examples of ion pairs are endless, from something as simple as sodium and chloride (Na^+Cl^-) making contact with each other, to more complex structures like salt bridges between anionic and cationic side chains buried within folded proteins.[14]

In proteins, ion-dipole and dipole-dipole interactions have a wide range of important functions, such as structural stabilization, fixing geometry, upholding catalytic activity and ion channel selectivity.[15] An important feature of peptide and protein structures is the dipole moment in a peptide bond between the amide nitrogen and carbonyl group. In an alpha-helical secondary structure, many of these local dipoles are electrostatically aligned and almost parallel to the long axis of the helix.[16] An example of such ion-dipole interaction being part of an allosteric catalysis regulation is potassium accelerating the conformational change in inosine 5′-monophosphate dehydrogenase (IMPDH).[17] Potassium interacts with the dipole of six carbonyl oxygens in a geometrically flexible way that accelerates the conformation change and mobilizes the segments in IMPDH needed for catalysis. Increasing the catalysis rate by a 100-fold, compared the basal rate without the potassium cation present.

Hydrogen bonds are an important sub-type of dipole-dipole interactions. They are formed between a hydrogen atom on a polarized X-H bond (hydrogen bond donor, HD) and the lone pairs on an electronegative atom like oxygen or nitrogen (hydrogen bond acceptor, HA). When water is used as a solvent the hydrogen bond effect becomes one of the most important interactions. Having two hydrogen bond donors and acceptors, water is able to form on average 3.6 hydrogen bonds in bulk water.[18] However, hydrogen bonds are not purely electrostatic in nature. Steiner and others have shown that the HD and HA overlapping orbitals share electrons, giving rise to specific geometric constraints that are most easily understood as arising from the hybridization of orbitals at each functional group.[19] Yet, a hydrogen bond is much weaker than a covalent bond. A classic example of hydrogen bonding is the duplex structure of DNA. DNA consists of the base pairs, adenine with thymine and guanine with cytosine, and hydrogen bonds between each pair drive the base pairing at the core of the DNA double helix.

Interactions between pairs of aromatic rings strongly involve the aromatic π-electrons.[20] In one extreme aromatic interactions can be viewed in an "electrostatic" way, in which an aromatic ring is pictured as a quadrupolar distribution of charges, where the hydrogen atoms are partially positively charged and the π-electrons are partially negatively charged. Aromatic rings can interact with each other through π-π interactions, which can occur in several different orientations (T-shaped, stacked, or slipped parallel). Other aromatic interactions that are well explained by electrostatic models are cation-π and anion-π interactions. Anion-π interactions are rare in nature, however cation-π interactions play an important role in molecular recognition and catalysis.[21] Another way of understanding aromatic interactions is that, regardless of the charge distribution,

aromatic compounds are attracted to each other. Aromatic rings are polarizable and can form strong van der Waals interactions, regardless of their relative orientation or electrostatic charge distributions. Considering both these models is important when thinking about aromatic interactions.

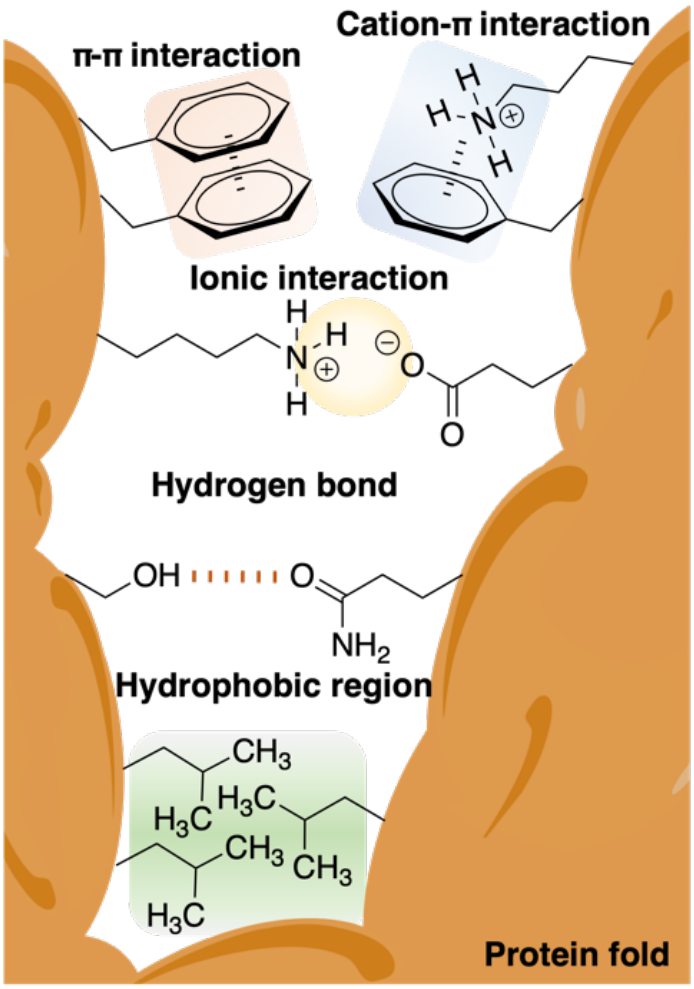


Figure 1.3. Schematic overview of non-covalent interactions displayed in a protein.

1.2 The hydrophobic effect

The studies that are covered in this thesis all take place in biologically relevant solutions (i.e., salty water). Under biologically relevant conditions, the hydrophobic effect plays a critical role for many molecular processes. The hydrophobic effect describes the tendency of non-polar molecules to aggregate in water, it itself is not a non-covalent interaction but a consequence of the energy differences of non-covalent interactions upon aggregation.[22] There are countless examples of systems that show the importance of the hydrophobic effect in biological systems ranging from protein folding[23] to microbial infections.[24] However, there is still discussion on how best to depict and understand the mechanisms underpinning the hydrophobic effect.[25] The hydrophobic effect is often defined by two distinct mechanistic models: the classical and the non-classical hydrophobic effect.[26,27]

The classical model of the hydrophobic effect is derived from classic experiments on the separation of oil from water, which is driven mainly by entropy. The model thus is used to explain systems of all kinds in which the aggregation of nonpolar molecules is driven by the change in entropy (ΔS) (Figure 1.4). Entropy describes the inherent disorder of a system, where a greater degree of disorder is more favourable. Water molecules have rotational, translational, and vibrational degrees of freedom. Water at an interface (e.g., next to non-miscible organic solvents, non-polar molecules, or air), has reduced mobility compared to water in the bulk, and is therefore in an entropically unfavourable environment.

The clathrate model is one way of depicting the structures thought to contribute to the classical hydrophobic effect (Figure 1.4a). Methane-ice clathrates occur in nature (Figure 1.4b), where methane is caged by highly ordered, hydrogen bonding water. These clathrates are formed under high pressure and low temperatures.[28] In the more general clathrate model of hydrocarbons in solution, water is depicted as being similarly ordered around aliphatic chains, restricting its degrees of freedom. It is entropically favourable when the chains come together and release some of their solvating water molecules into the bulk.

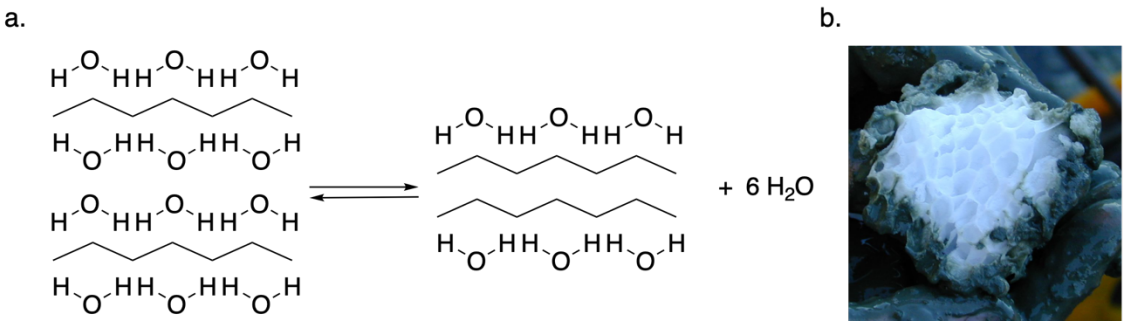


Figure 1.4. Clathrates. a) Clathrate model demonstrating the classical hydrophobic effect. b) Methane clathrate block embedded in the sediment of hydrate ridge, Oregon, USA (Credit: Wussel 007 Wikimedia (CC BY-SA 3.0)).

The non-classical model of the hydrophobic effect explains systems in which the aggregation of nonpolar molecules is driven by a change in enthalpy (ΔH) (Figure 1.5). Enthalpy describes the internal energy of the system, often represented as a total sum of attractive interactions, where a lower enthalpy is more favourable. Water at the interface makes fewer hydrogen bonds than bulk water, and therefore has a higher enthalpy. When

the interface gets reduced in size, more of the water will be in the bulk water, which results in more degrees of freedom and increased amounts of hydrogen-bond formation. This model is often invoked when water is in small molecular cavities (as in many host molecules) because water molecules at curved interfaces are especially prone to being deprived of forming the optimal number of hydrogen bonds until it is released into the bulk.

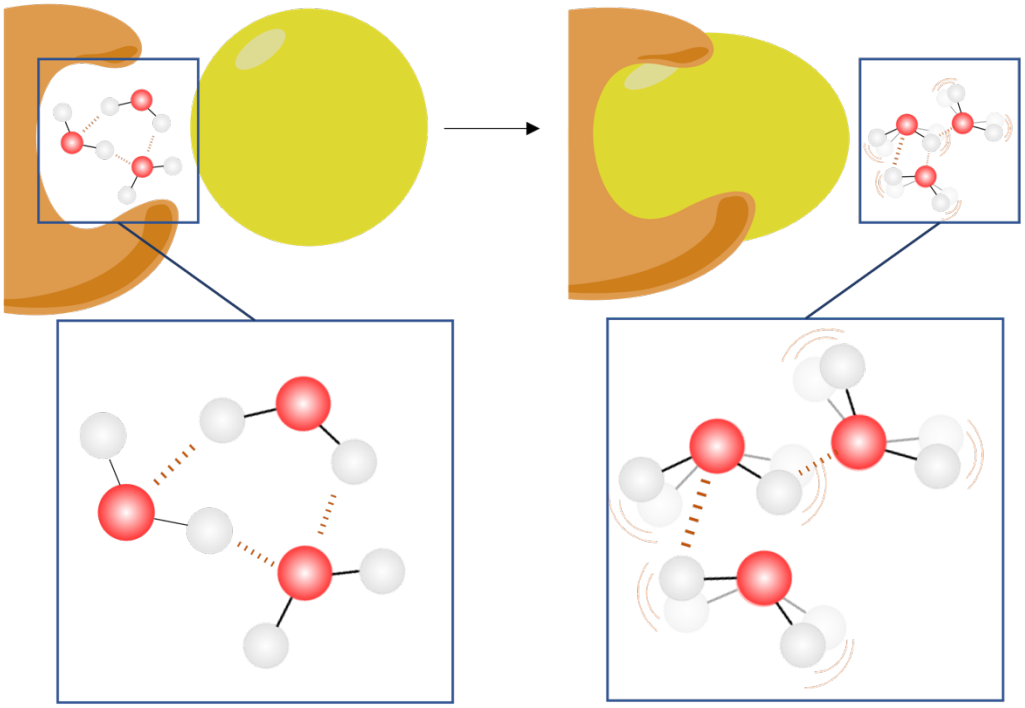


Figure 1.5. Schematic view of the non-classical model of the hydrophobic effect. Water restricted in a cavity is being released by a guest binding into the cavity, leading to a decrease in enthalpy.

Although the classical and non-classical models describe aspects of the hydrophobic effect, there are still features that are not fully explained. A more advanced way of describing the origins of the hydrophobic effect is to view it more holistically as the result of energy differences of noncovalent interactions, general effects that are common to all interfaces, and cavitation energies. These favourable energy differences come from water molecules at the interface in different states of the binding system (bound versus unbound).[22] The hydration mechanism of the solute significantly depends on the size of the solute.[29] For small hydrophobic solutes (< 1 nm) the water density around the solute is increased, adapting a pattern similar to the hydrogen-bonding patterns in bulk water, and the solute is considered "wet". The hydrophobic effect does not have a significant impact on small solutes. Bigger hydrophobic solutes (> 1 nm) make it impossible for interfacial water molecules to stay in an ordered hydrogen-bonding network. The water molecules "move away" and the density around the solute is decreased, the solute is considered "dewetted."[30,31]

When discussing the hydrophobic effect, the Hofmeister series should also be considered. The Hofmeister series is a series of salts that influence the solubility of a solute by either "salting-out" the solute, making it less soluble, or "salting-in" the solute, making it more soluble. When first studied by Franz Hofmeister it was thought that the salts altered the bulk property of water thereby altering the hydrophobic effect, [32] but more recent studies has disproven this. Now it is thought that the salts interact with the solute, changing the solubility. The current understanding is that salting-out salts require too many water molecules to solvate, simply not leaving enough water molecules to solvate

the solute causing it to crash out. Salting-in salts are thought to directly interact with the solute, changing it solubility.[33,34]

1.3 Diversity-oriented synthesis

DOS is used to create diverse libraries of biologically active molecules. The principle of DOS is achieving greater diversity through modifying a basic scaffold, while still keeping the key concepts of appendage-, functional group-, stereochemical- and skeletal- diversity in mind.[35]

In combinatorial chemistry a single scaffold is irreversibly reacted with multiple reagents to create a library of small molecules. Libraries of up to millions of possible biologically active compounds are created in a single reaction, after which they can be screened for their corresponding biological activity against target molecules. These libraries often lack coverage of the chemical space because they are limited by the linking functional groups that are used.[36]

Libraries for a specific target (e.g., a protein) can be optimized by dynamic combinatorial library (DCL) synthesis. In a DCL the individual library members bind reversibly and exchange constantly, resulting in an adaptive library of molecules. Upon the introduction of a target molecule the library re-equilibrates because of the thermodynamic driving force of tight-binding library members, resulting in an increased concentration of hit molecules (Figure 1.6). Finding biologically active compounds is not just a numbers game, the introduction of structural diversity is required for it to be successful.[4,35]

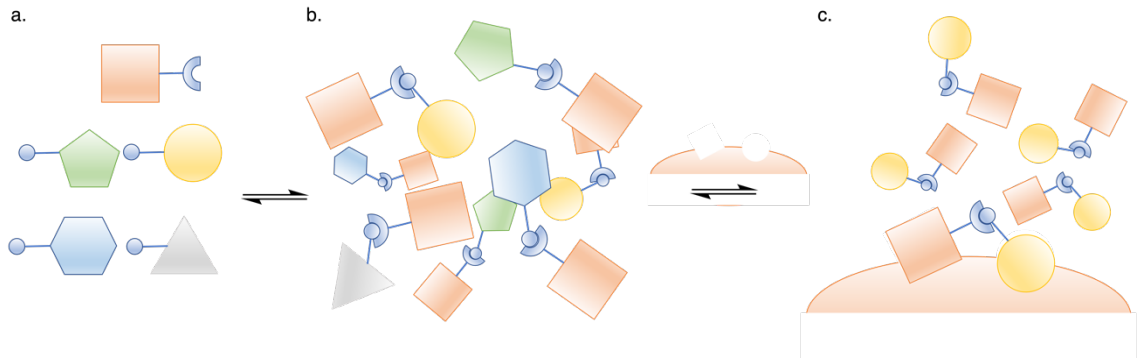


Figure 1.6. Schematic overview of a dynamic combinatorial library. a) Library building blocks. b) Library building blocks are in constant exchange. c) Introduction of the target re-equilibrated the library increasing the concentrations of hit combinations.

An example of this type of diversity oriented synthesis was reported by Hirsch and co-workers on the optimization of an aspartic protease endothiapepsin inhibitor, which combined fragment based drug design (FBDD) and DCL.[37] In FBDD, an initial fragment is identified that has a favourable binding to a region of the target protein.[38] This fragment is then modified in order to have stronger binding and therefore a more viable lead molecule. This process has the drawback that the design, synthesis, and validation of each respective modified fragment is still needed.

Dynamic combinatorial chemistry (DCC) can be used to overcome some of the challenges associated with FBDD and accelerate the process. By using a DCL consisting of a compound **1** and nine hydrazides (compound **2**-**10**) (Figure 1.7), they produced a library containing 78 bis-acylhydrazones and 12 mono-acylhydrazones. Hirsch and co-workers re-equilibrated the library by introducing the target protein alongside to select the strong binding library members and analyzed them. Five molecules were amplified in the

library. They found that their best compound **11** had a 240-fold increase in binding, IC_{50} = 54 nM, compared to the original identified fragments, IC_{50} = 12.8-14.5 μM.

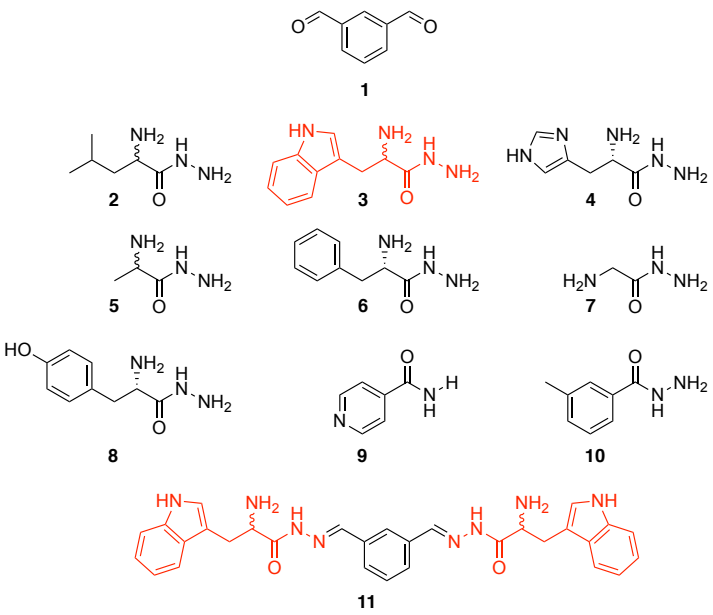


Figure 1.7. Fragment based dynamic combinatorial chemistry

1.4 Diversity-based covalent functionalization of hosts

Design still plays a big part in almost all host functionalization projects, even though rationally designing hosts to achieve high specificity, potency, and/or sensing functions in water is a complex and often unrewarding task. Using diversity-based covalent functionalization of hosts allows for the diverse synthesis of host libraries. Following is a review of diversity-oriented covalent synthesis of host libraries, arranged by host class.

Cyclodextrins (CDs) are naturally occurring hosts that are produced from starch through enzymatic conversion.[39,40] The hosts consist of glucose subunits that are joined

by α-1,4 glycosidic bonds. Native CDs consist of 6, 7 or 8 D-glucopyranose unites, α-, β- or γ-cyclodextrin respectively (Figure 1.8) and have a cone-like shape. CDs are notoriously hard to modify, although great progress has been made towards out-of-equilibrium enzymatic conversion through template directed synthesis,[41,42] light-controlled assembly,[43] chaotropic and kosmotropic anion regulated synthesis[44] and even through synthesising CDs from scratch.[45]

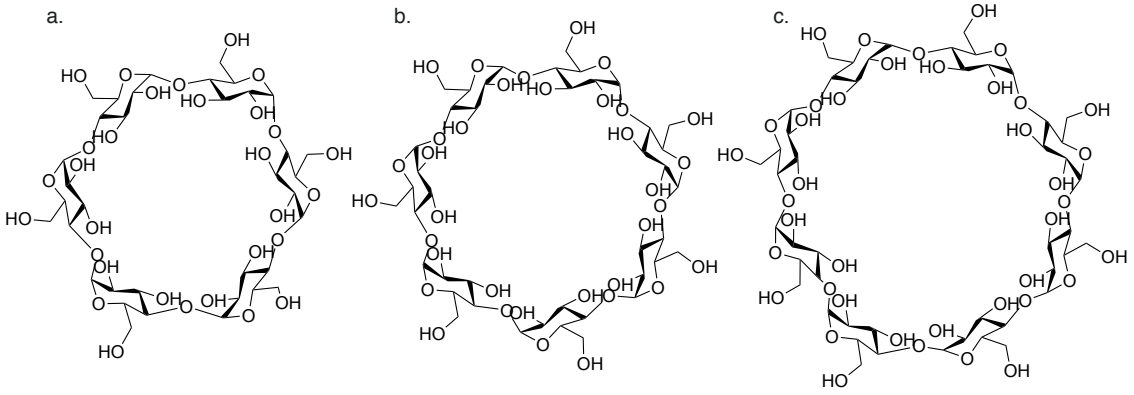


Figure 1.8. Naturally occurring cyclodextrins. a) α-Cyclodextrin. b) β-cyclodextrin. c) γ-cyclodextrin

Beeren and co-workers accessed a DCL of unnatural cyclodextrins, by exploiting the promiscuity of cyclodextrin glucotransferase (CGTase).[46] CGTase is an enzyme that catalyzes the reversible inter- and intramolecular trans-glycosylation of a-1,4-glucans to produce cyclodextrins and linear products. They introduced unnatural cyclodextrins (mono-6-deoxy-α-CD, mono-6-deoxy-β-CD, mono-6-azido-6-deoxy-glucopyranose and mono-6-deoxy-6-iodo-glucopyranose) to CGTase where they were accepted as substrates. This led to different DCL with incorporated modified D-glucopyranose units. They also

showed that they could shift the equilibrium by introducing templates since these libraries are under pseudo-thermodynamic control.

The first synthetic hosts that were fully characterized and studied were crown ethers,[47,48] initially reported by C. J. Pedersen in 1967.[49] Crown ethers are cyclic oligomers of ethylene oxide, named after their resemblance of a crown when bound to a cation (Figure 1.9).

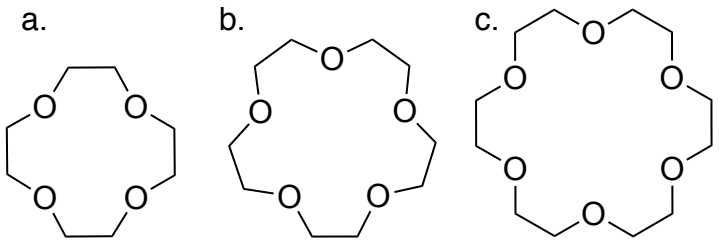
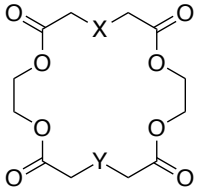


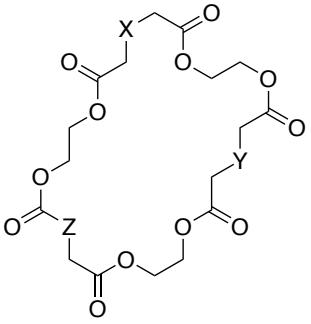
Figure 1.9. Examples of basic crown ether scaffolds. a) 12-crown-4 b) 15-crown-5 c) 18-crown-6

Hodge and co-workers reported three combinatorial libraries of crown ether-ester analogues.[50] The first library consisted of 16 different 18 or 27-membered rings, achieved by cyclodepolymerization of three polymers. For the second library an extra polymer was added resulting in 30 different 18 or 27-membered rings. The third library contained only two different starting polymers, where approximately 66% of the library were 24-membered ring dimers (n = 2) and 33% were 36-membered ring trimers (n = 3) (Figure 1.10). Although these libraries of hosts have the potential to be involved in guest binding no further studies were conducted.



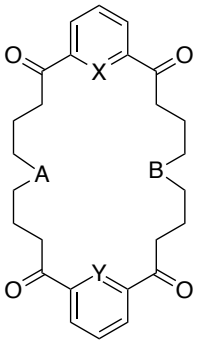
Library 1:
6 combinations of X and/or Y = CH_2, O, S

Library 2:
All 10 combinations of X and/or Y = CH_2, O, S or NMe

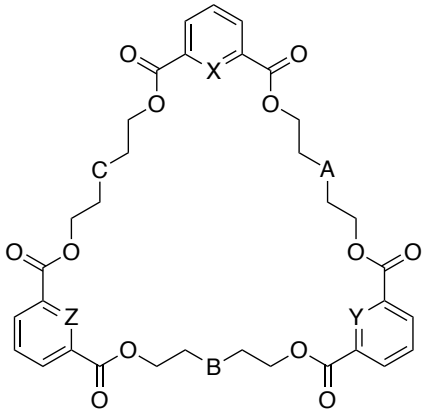


Library 1:
10 combinations of X and/or Y and/or Z = CH_2, O, S

Library 2:
All 20 combinations of X and/or Y and/or Z = CH_2, O, S or NMe



Library 3:
9 combinations of X and/or Y = O or pyridine and A and/or B = O or S



Library 3:
20 combinations of X and/or Y and/or Z = O or pyridine and A and/or B and/ or C = O or S

Figure 1.10. Combinatorial libraries of crown ether-ester analogues via the cyclodepolymerisation of linear polyesters.

Taki and co-workers reported a crown ether-like library where diversity was introduced using a library of peptides displayed on a bacteriophage viral coat protein.[51] Phage display is a technique that introduces molecular diversity through large peptide libraries, up to 10^{10} unique sequences, that are expressed on the surface of bacteriophage.[52]

Taki and co-workers synthesized this library, by cyclizing an oligoethylene glycol unit bearing electrophilic alkyl halides at each end to randomized phage peptides. The phage peptides were designed to display peptides of the general sequence -X3-C-X7-C-X3, where X represents any random amino acid and C represents the two cysteine residues that were used to make linkages to the oligoethylene glycol reagent. The library was used to find a binder for the N-terminal domain of heat shock protein 90 (Hsp90). After six rounds of bio-panning against Hsp90 they unexpectedly found an amplification of the one specific peptide (-R-S-W-C-R-K-S-R-K-N-S-G-G-G-L-V-W-C-F) of which trace amounts had been present in the library (Figure 1.11a). They showed with independent experiments that the amplified crown ether-peptide hybrid (compound **12**) binds the N-terminal region of Hsp90 ($K_d = 1.7 \pm 0.5\ \mu M$).

Taki and co-workers continued their quest to find strong binding Hsp90 drug-candidates in a subsequent paper by deploying a similar strategy with cryptand-like structures.[53] They reasoned that the crown ether hosts are too flexible, which results in an unfavourable entropy loss upon binding Hsp90. The cryptand-like structures are less flexible, which minimizes the conformational entropy loss upon binding, while the increased number of hydrogen bond acceptors results in a favourable contribution to enthalpy. After seven rounds of bio-panning against Hsp90, they found that the peptide-cryptand hybrid (compound **13)** derived from an amplified sequence, QWV-C-

LNPWLSI-C-RA, had a K_d value of 62 ± 10 nM, a 27-fold improvement over the best result from the crown-ether library (Figure 1.11b).

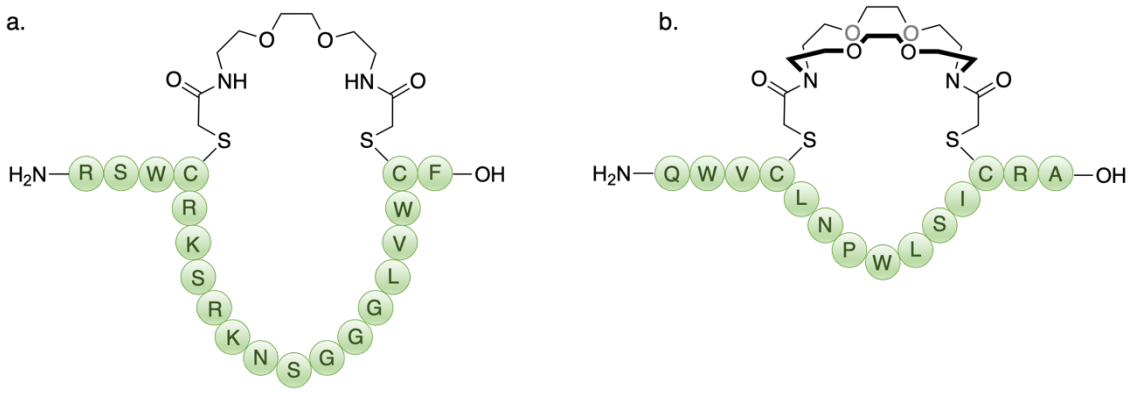


Figure 1.11. Library construction through genetically encoded peptide linkers displayed on bacteriophage T7. a) Selected crown ether-like host (**12**) for Hsp90. b) Selected cryptand-like host (**13**)for Hsp90.

Thia-cyclophanes are another class of macrocycles that have been extensively explored for DOS of hosts. Cyclophanes are molecules that have an aromatic unit that is bridged by an aliphatic chain. The first report of cyclophanes being used as hosts comes from Dougherty and co-workers. They used cyclophanes to catalyze the dealkylation of a simple sulfonium compound.[54] This work inspired Otto, Sanders and co-workers to use these structurally similar compounds for DCL using sulfide bridges as reversible linkers.

The first library they created consisted out of three building blocks, compound **14**-**16** (Figure 1.12). Without a guest this library formed 45 different macrocyclic disulfides of unique mass. They individually introduced two different targets, 2-methylisoquinolinium iodide and N-methylated morphine, to the library. For each target

a unique host emerged compound **17** and **18** (Figure 1.12), showing that these libraries can be used to discover artificial receptors.[55] Using the same combination of building blocks in a different study they obtained catalysts for the Diels–Alder reaction between acridizinium bromide and cyclopentadiene. Interestingly amplifying the exact same hosts as in their previous work. Although modest, compound **18**, showed catalytic activity.[56] In a short communication they also reported diastereoselective amplification of an induced-fit host out of a DCL containing compound **14** and compound **19** with NMe_4 as target. Of the four possible stereoisomeric tetramers that could be formed, compound **20** was amplified 400-fold.[57, 58]

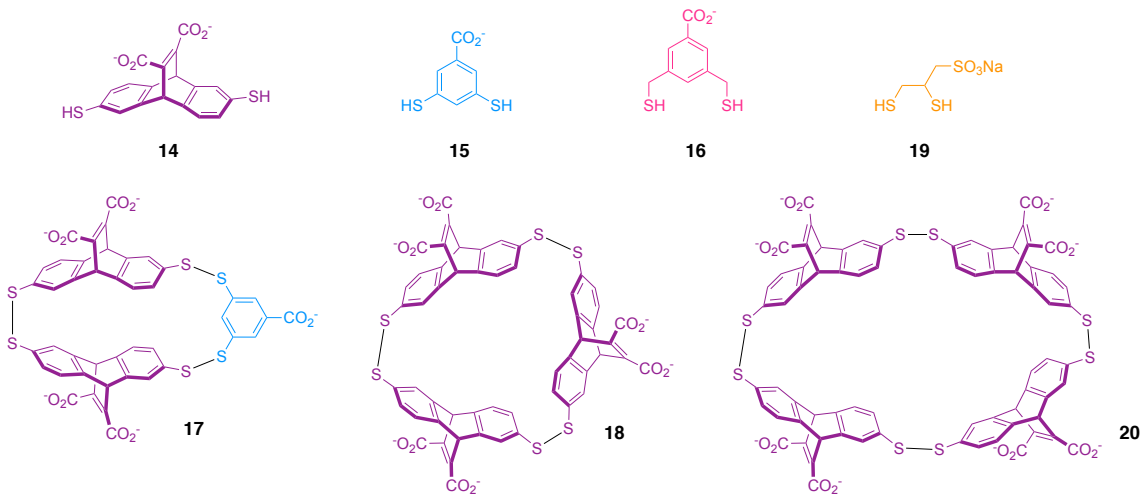


Figure 1.12. Sanders and Otto building blocks and hosts

Waters and co-workers have published a body of work targeting the methylated states of arginine and lysine on histone protein with these macrocycles using dynamic combinatorial chemistry (DCC). In their first publication they targeted trimethylated lysine (KMe_3) with a DCL consisting of eight building blocks, and capable of reversibly

forming macrocyclic hosts.[59] The DCLs were prepared by mixing 3-4 building blocks at a time and introducing Ac-KMe_3-G-NH_2 as a template. In one initial example, receptor compound **17** was amplified in the presence of the KMe_3 guest and showed comparable binding (K_d = 25 μM) and selectivity (~50 fold) to the native protein (Figure 1.13). Continuing with three building blocks from the original library (compound **14, 15** and **21**) they targeted asymmetric dimethyl arginine (ADMA).[60] Waters *et al* identified a new receptor compound **22**, by templating the DCLs with Ac-RMe_x-G-NH_2 (x = 0–2). Compound **22** has low micromolar affinity (K_d = 5 μM) for ADMA and ~7-fold selectivity for ADMA over symmetric dimethyl arginine (SDMA) (Figure 1.13). This is the first selective ADMA synthetic receptor reported; however, it is still overall selective for KMe_3 over ADMA.

Waters and co-workers redesigned the library by optimizing their hit molecule and replacing one of the monomers **15** with **23**. This introduced an additional cation–π interaction into the binding pocket and provided a deeper binding pocket. This was expected to improve the selectivity due to the weaker solvation of KMe_3 versus KMe_2. Waters and co-workers templated the DCLs with Ac-KMe_xGGL-NH_2 (x = 0–3). Using this strategy they improved the KMe_3 binding affinity 10-fold and improved the selectivity of KMe_3 5-fold over dimethylated lysine (KMe_2), with receptor **24** (Figure 1.13).[61] In a further optimization Waters and co-workers redesigned the library by replacing **15** with **25**.[62] With **25**, biphenyl-2,6-dicarboxylate, the ability to have an inwardly directed carboxylate was created in the hope of creating a buried salt bridge with the NH^+ on the side chain of KMe_2. They templated the DCLs with methylated butylamine guests as methylated lysine side chain mimics. Compound **26** was selected as a selective KMe_2

receptor (Figure 1.13). Compared to compound **17** they increased the binding affinity 32-fold (K_d = 0.2 μM) and the KMe_2 selectivity 2.4-fold over KMe_3.

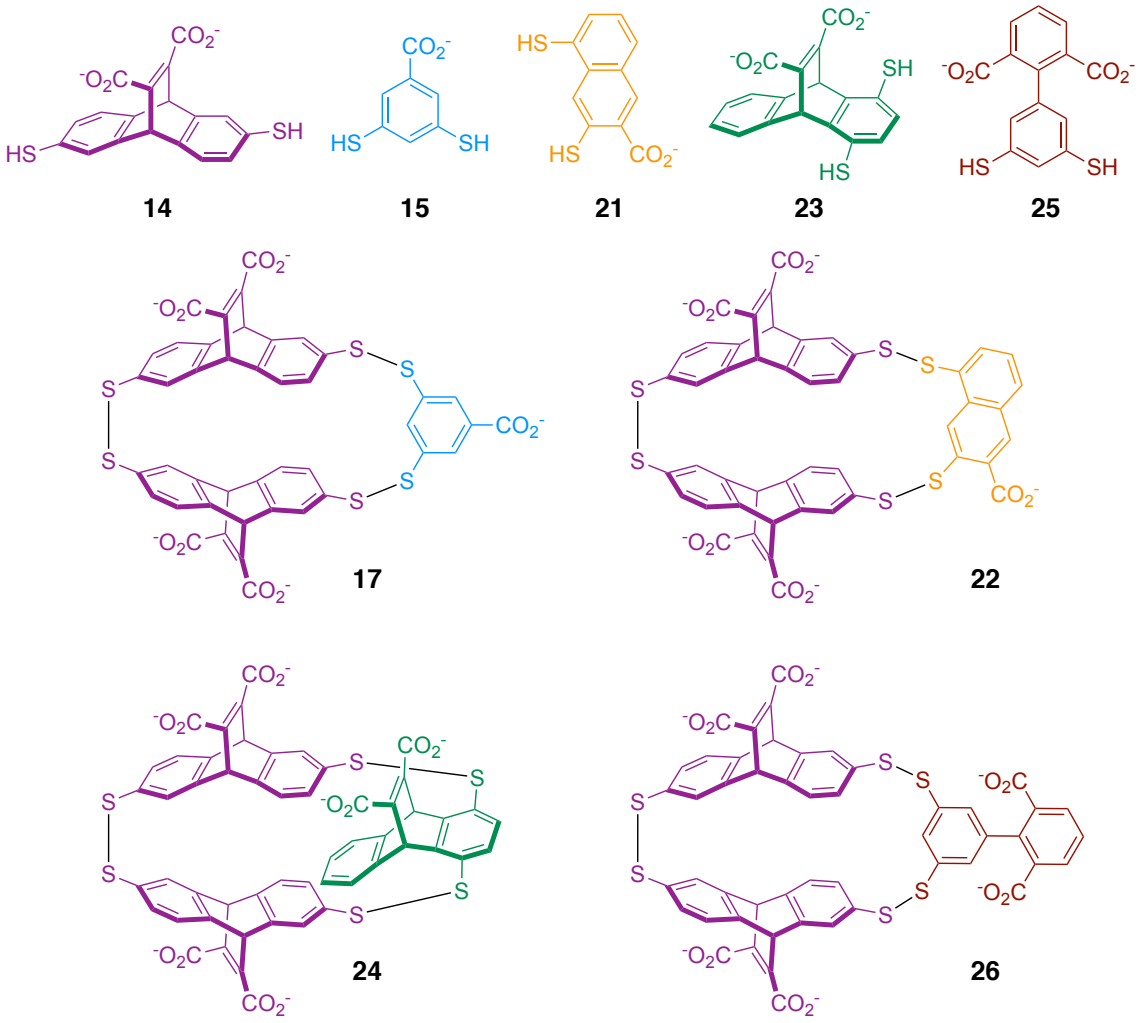


Figure 1.13. The building blocks and hit compounds from the Water's DCL.

In later work Waters and co-workers re-designed the library to more closely resemble the native protein.[63] They employed three new building blocks **23**, **27** and **28**, which mimic a box-like "aromatic cage" structure found in many proteins by creating a

5-sided box-like shape (Figure 1.14). Using this library, they found an ADMA selective receptor, **29**, with a stronger affinity (K_d = 1.2 μM) than the native reader protein. They applied this library making it into a sensor array that can differentiate between arginine, lysine, and their methylation states.[64,65]

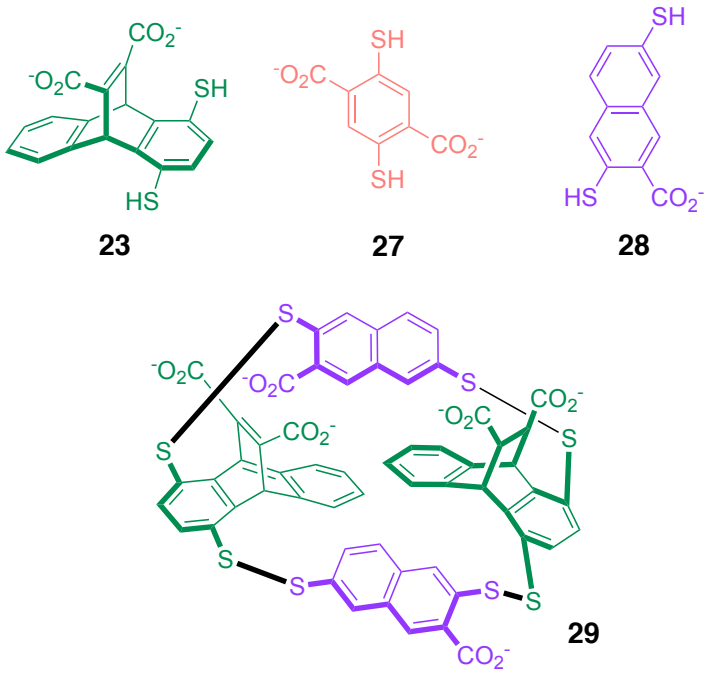


Figure 1.14. The building blocks and hit compound from the Water's DCL for ADMA.

1.5 Calix[4]arenes

The last host class to be introduced are the calix[*n*]arenes, which are the focus of this dissertation. The word 'calixarene' is derived from Greek, where "calix" means vase or chalice, and "arene" indicates the incorporation of aromatic rings.[66,67] They were first discovered by Alfred von Baeyer in 1872, however they were not characterized until the early twentieth century. Leo Baekeland synthesized a resin, Bakelite, based on the phenol-formaldehyde reaction, which he patented in 1907.[68] In the 1940's Zinke and Ziegler were

the first to assign the by-products from Bakelite synthesis as cyclic tetrameric structures, now known to be calix[4]arenes, and to recognize their potential to form complexes with small molecules. In 1970, C. D. Gutsche named these macrocyclic phenol-derived compounds calix[*n*]arenes.

Calixarenes are cyclophanes with their phenolic units linked together by methylene groups at the *meta* position. Having, a ring size '*n*' varying from 4-8 aromatic residues, our focus will be on the calix[4]arenes. Calix[4]arenes can adopt four different conformations: cone, partial cone, 1,2-alternate and 1,3-alternate (Figure 1.15). When the lower-rim oxygen substituents are present as free OH groups, calix[4]arenes prefer the cone conformation. This is due to the hydroxy groups having strong intramolecular hydrogen bonding interactions.[69] The "calix" name for this family of macrocycles refers to the cone shape of this conformation of these molecules.

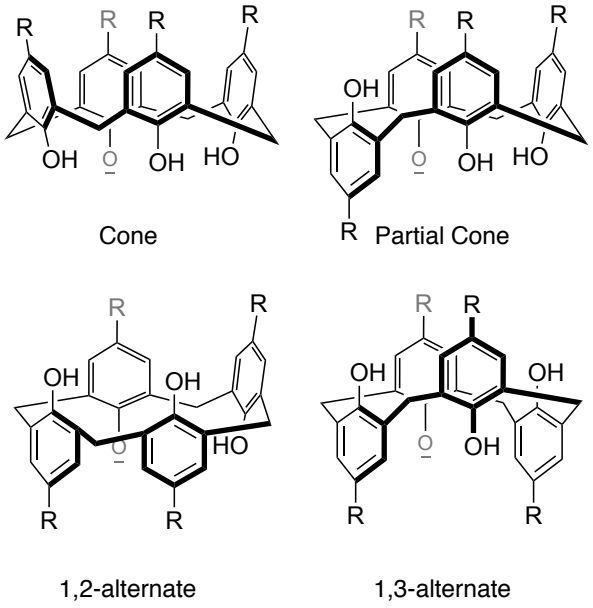


Figure 1.15. Calix[4]arene conformations

Calix[*n*]arenes have multiple characteristics that make them effective hosts. Besides the ability to alter the binding pocket size by synthesizing calix[*n*]arenes with a different number '*n*' of phenolic units, they are easily modified on both the upper and lower rims of the phenolic units (Figure 1.16).[70] The new groups installed on the calixarene directly influence the binding pocket and can change the binding strength and selectivity of the hosts.[71-74] Due to their varied functionality, molecular recognition within calixarenes takes place via many kinds of non-covalent interactions which include electrostatic, hydrophobic, hydrogen bonding, van der Waals forces, π-π interactions, and cation-π interactions.[75]

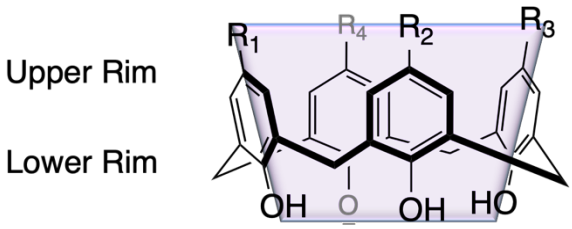


Figure 1.16. Calix[4]arene

Water solubility of calix[*n*]arenes was first achieved in 1984 by Ungaro and co-workers through the functionalization of the lower rim with esters.[76] In the same year, Shinkai and co-workers synthesized water soluble calix[*n*]arene by sulfonating the upper rim.[77] The strategy developed by Shinkai is the basis of producing most water soluble calix[*n*]arenes today. Sulfonated calix[*n*]arenes are well suited towards biological applications.[78] The parent, *p*-sulfonatocalix[4]arene, shows low to no toxicity in animal models (100 mg/kg^{-1} i.v. in mice)[79,80,81] and site-selective functionalization gives rise to new binding properties. The small hydrophobic cavity (an inner diameter = 3.8 Å[82]) and

facile sulfonation of the upper rim make them suited to interact with small cationic hydrophobic guests.

Calixarenes are easily modified, with multiple diversity libraries having been reported. Hioki and co-workers described the development of peptidocalix[4]arene libraries for binding peptides[83] and subsequently developed them into chemosensors.[84,85] To achieve the original library, a calixarene was attached to polystyrene resin beads via the upper-rim. The lower rim was modified using combinatorial split-and-pool peptide synthesis with 15 amino acids to create a library of approximately 50,000 peptidocalixarenes, (Figure 1.17a). Hioki and co-workers screened this library against four dye-labelled oligopeptides and observed binding for 13 hosts, which they decoded to identify the amino acid sequences. Further studies were inhibited by poor solubility.

To synthesize a chemosensor for peptides they used the same method but opted for tripeptidocalixarenes instead, modifying the terminal nitrogen with pyrenyl groups to create a library of ~3,300 potential sensors (Figure 1.17b). The tripetidocalixarenes had some success in sensing and were found to display a change in fluorescence upon addition of analytes. Hioki and co-workers further modified their chemosensors by synthesizing tripeptide/dye arms attached to the upper rim of the calix[4]arene, and in doing so further increased the binding selectivity. All the initial studies by Hioki and co-workers were performed in chloroform. The next step was to produce water-soluble hosts that can sense peptides in biologically relevant conditions using a similar strategy.[86] They achieved this by using charged and polar amino acids and extending the structural modifications to all four upper-rim positions thereby creating a library consisting of 1,000 members (Figure

1.17c). Although they achieved water solubility, the corresponding selectivity of the analogues produced decreased. However it was shown that this water-soluble library could be used to find hydrolytic catalysts.[87]

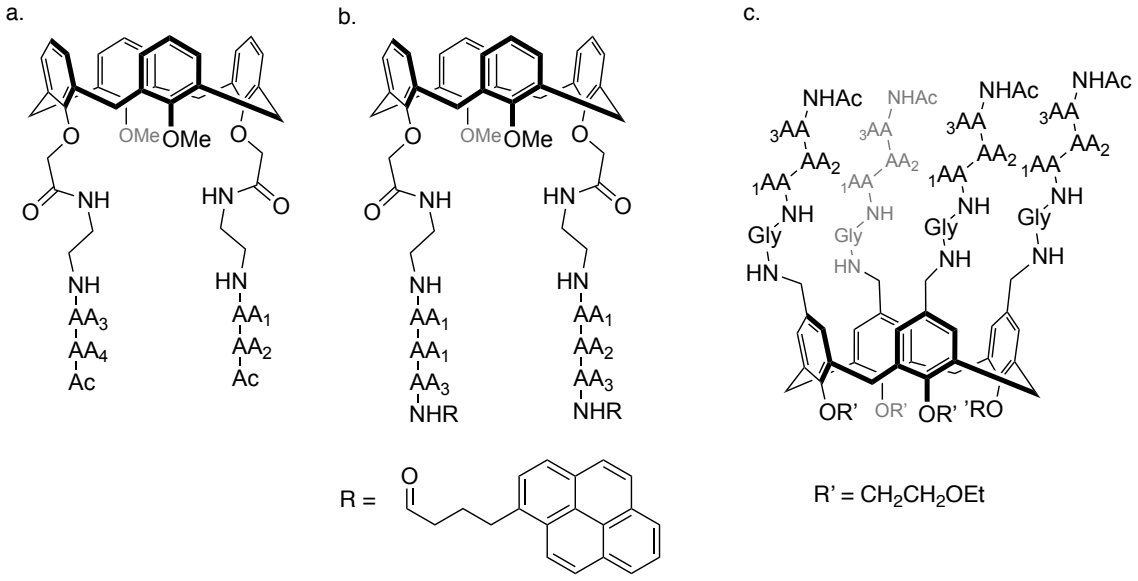


Figure 1.17. Hoiki's peptidocalix[4]arenes. a) The original library. b) Lower-rim modified fluorescence-labelled library. c) Hydrophilic peptidocalix[4]arene library.

Shuker and co-workers were interested in calix[4]arenes as receptors and enzyme mimics. They used parallel, solid phase synthesis to create a 120-member library of upper-rim modified peptidocalix[4]arenes.[88] The peptidocalix[4]arene were modified on all four upper-rim positions using tri-peptides. These peptides were selected to contain the amino acids serine and histidine, because they are present in the catalytic triad of serine proteases and esterases. The last amino acid was one of the 20 common amino acids, in any of the sequence positions. This library was screened for the catalytic hydrolysis of *p*-nitrophenyl acetate. From the 120-member library six derivatives were identified that catalyze the

hydrolysis, with the most efficient catalyst being His-Ser-His-calix[4]arene, **30** (Figure 1.18). Compound **30** has a 30% increase in the initial rate of reaction over that of its related peptide (His-Ser-His), and a 1520% increase over the uncatalyzed hydrolysis.

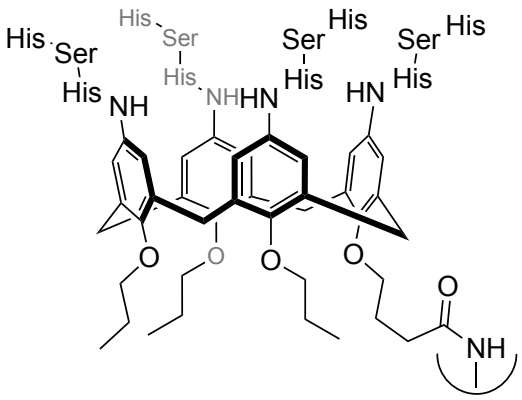


Figure 1.18. His-Ser-His-calix[4]arene (**30**)

Le-Gresley and co-workers reported the synthesis of static and dynamic libraries based on tetra-formyl calix[4]arenes. In 2005 they published on two different tetra-formyl calix[4]arene acrylates and 16 structurally diverse aromatic, aliphatic and functionalized amines.[89] For the static libraries, they reacted the amines individually with the two tetra-formyl calix[4]arene acrylates via an imine condensation whereby they found 26 potential combinations of which 21 were successfully prepared.

They moved toward the dynamic library, using tetra-formyl calix[4]arene acrylates **31** and four aryl and alkyl amines (compound **32-35),** for potentially 146 library members (Figure 1.19). After the library was sufficiently equilibrated, it was analyzed using liquid secondary ion mass spectrometry-MS. This revealed four dominant species. They hypothesized that the library is templated by one of the amines. As a control they

templated the library with barbituric acid and biotin, each resulting in a different dominant species.

In their 2010 publication, Le-Gresley and co-workers achieved similar libraries.[90] For their DCL they combined **31** with **32**, **35-37,** for potentially 157 library members (Figure 1.19). After the library was sufficiently equilibrated, four dominant species were revealed. The library was templated with barbituric acid and biotin, once again each resulting in a different dominant species.

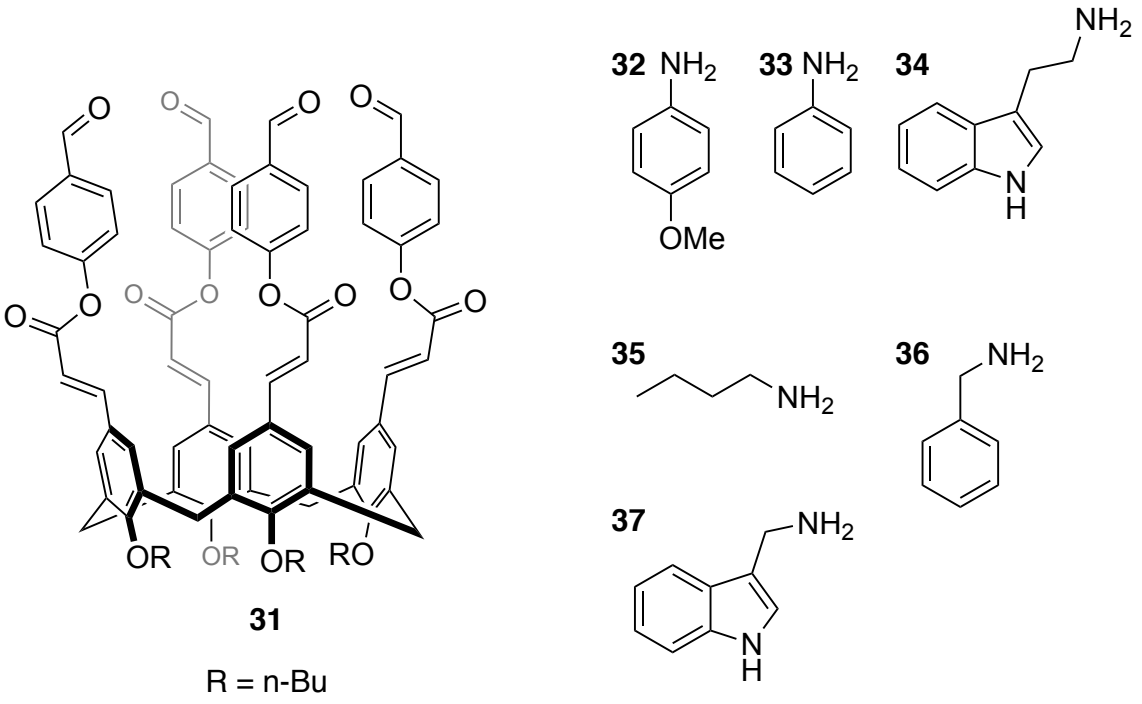


Figure 1.19. Components for dynamic libraries

Hof and co-workers have described the parallel synthesis of calix[4]arene based chemosensors that can detect drugs in water and in saliva.[91] The parallel synthesis described is driven by the condensation of a calix[4]arene-aldehyde (**sCx4-CHO)** and 16 heterocyclic nucleophiles that form merocyanine-functionalized hosts called DimerDyes. The condensation reactions resulted in 13 successful DimerDye syntheses. The formation of the DimerDyes was simply observed via a colour change and eliminates the need for purification in the early stages. When the DimerDyes are present alone in solution they form homodimers and their fluorescence is quenched. Upon the introduction of a guest the complex disassembles, and fluorescence can be observed. The crude dyes were screened against nicotine, a model metabolite, where an increase in fluorescence indicates the host-guest complex forming.

Five DimerDyes (compound **38-42)** as well as a control DimerDye **(43**) were selected to be re-synthesized, purified, and studied (Figure 1.20). The dyes were studied with nicotine, methylenedioxymethamphetamine (MDMA) and cocaine via fluorescence titrations and NMR in both water and saliva. They show that the DimerDyes have a low micromolar limit of detection, for the drugs in water and saliva. The five selected DimerDyes were then put to the test in a sensor array, where they were able detect and differentiate between closely related drugs and their metabolites across multiple drug families.

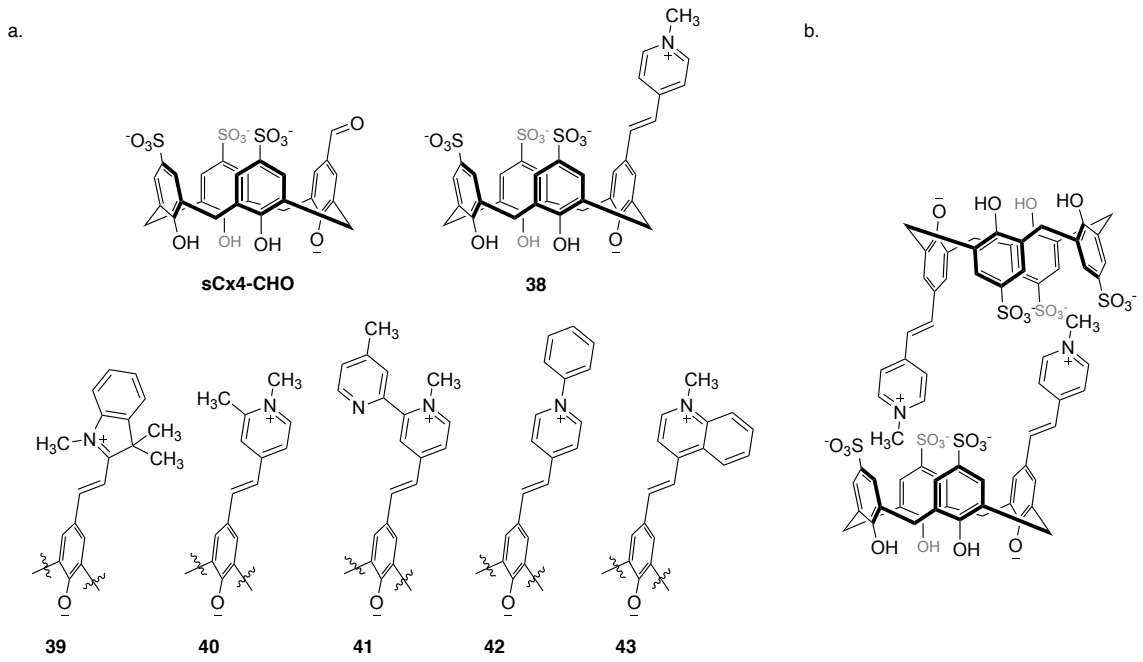


Figure 1.20. DimerDyes. a) parent calixarene **sCx4**-CHO and selected DimerDyes. b) Dimerization of **38**

While great efforts have been made towards diversity-based covalent functionalization of hosts, this field is still in the early stages. The difficulty of synthesis of some of the projects (e.g., Waters cyclophanes) is a barrier to successful libraries. Late-stage introduction of diversity is important for feasibility and efficiency (e.g., DimerDye library), as well as extensive amounts of diversity can give major improvements in properties (e.g., phage cryptand). Among the different host classes, calixarenes are the easiest to modify synthetically and therefore have been selected for the work in this thesis.

1.6 Goals of this dissertation

This thesis will report on calix[4]arene libraries, the fundamental lessons we can learn from them and applying them in assay design.

The goals of this dissertation are:

1. Synthesis of novel calix[4]arene hosts
2. Learning fundamental lessons about host-guest interactions
3. Combining knowledge into a new high-throughput functionalization approach

This chapter reviewed non-covalent interactions, host-guest chemistry, and high throughput functionalization of hosts. It took a deeper look at calixarenes, what makes them promising host molecules and what has currently been achieved in library synthesis.

Chapter 2 looks at a small library of calix[4]arenes that are systematically modified on the upper-rim to investigate structure-activity relationships. The binding strength and geometry of binding between the library members and nicotine (and its two main metabolites nornicotine and cotinine) are studied. From this we learned that upper-rim substitution patterns influence the binding mode of nicotine.

In Chapter 3 we extended the fundamental binding studies, in a collaboration with the Universität Duisburg Essen and the University of Maryland. Besides the aforementioned calix[4]arene library we looked at two cleft-like hosts and at two acyclic cucurbiturils. As guests, we studied the amino acids arginine and lysine and each of their methylated

analogues and gained fundamental lessons about molecular recognition from this comparative study of different host classes.

In Chapter 4 we take a jump forward in host complexity by creating a large library of macrobicyclic-calix[4]arene-peptide hybrids. Using phage display, we achieved massive diversity through crafting a calix[4]arene onto the phage peptides. First efforts towards affinity selections against ADMA and cocaine were made, but further improvements are needed. The weak binding of the selected host was further investigated with NMR studies.

Chapter 2: Calix[4]arene sulfonate hosts selectively modified on the upper rim: a study of nicotine binding strength and geometry

This work was adapted from a previously published paper.

Z. Warmerdam, B. E. Kamba, A. Shaurya, X. X. Sun, M. Maguire, F. Hof, *Supramol. Chem.,* **2021**, 33, 88-96

Alok Shaurya established the synthesis of **sCx4-1,3-NO_2**. Zoey Warmerdam established the synthesis of **sCx4-1,2-NO_2** and **sCx4-1,3-CHO**. XuXin Sun established the synthesis of **sCx4-1,2-CHO**. Zoey Warmerdam established the indicator displacement assays. Bianca E. Kamba repeated the IDA experiments, her data is used in this chapter. Zoey Warmerdam conducted the NMR titrations. Zoey Warmerdam analyzed all data sets. The data sets from the IDA experiments presented in the publication are reanalyzed, in order to be consistent with the best practices during development of this thesis. The data trends stayed the same, with values within error of the original reported values. Mary Maguire tabulated the NMR and IR characterization data. Zoey Warmerdam wrote the paper.

2.1 Foreword

Chapter 1 covers the importance of diversity-based covalent functionalization of hosts. To achieve this on the calix[4]arenes, I first needed to synthesize calix[4]arenes with regioselective modifications on the upper rim. Chapter 2 covers the synthesis of a small calix[4]arene library with regioselective modifications that are achieved through the deactivation of upper-rim positions toward electrophilic aromatic substitutions by using benzoyl ester protecting groups on the lower rim. The binding characteristic of this library

towards nicotine and its metabolites were studied to gain a better fundamental understanding of binding strengths and motifs of modified calixarene-type hosts.

2.2 Introduction

This chapter reports the nicotine binding properties of novel calix[4]arenes that are modified on the upper rim with sulfonates and one or two other functional groups. We address the selective di-functionalisation of the neighbouring or opposing phenolic units, resulting in 1,2- or 1,3-disubstituted regioisomers (Figure 2.1). We also explore how changes between relatively similar functional groups can drive unexpected differences in the binding orientations for a given guest. Although this work specifically focuses on nicotine and its metabolites[92], the new hosts set the stage for further elaborations in order target a broad range of other small cationic molecules including amino acids, metabolites, drugs, and drug metabolites.

We have chosen nicotine as a representative model drug in this work because of its cationic amine and overall hydrophobicity, which make it likely to be a good guest for calix[4]arene hosts. It has structural features and metabolic relatives similar to many alkaloid-type drug molecules, making it well suited for obtaining broadly applicable lessons in drug recognition. Using nicotine in binding studies with novel calix[4]arene hosts provides fundamental insight into binding strength, stoichiometry, and binding motifs of small molecule host-guest complexes that could provide insight into to other drugs. In the literature, binding studies with nicotine have been completed with cucurbit[n]urils, β-Cyclodextrin, and an upper-rim bridged calix[4]arene possessing a seven membered ring. Depending on the modification of the hosts, and pH and

temperature of the binding assays, the cucurbituril hosts have association constants (K_{assoc}) that range from $3.6x10^2$–$6.5x10^4$ M^{-1},[93,94] while β-Cyclodextrin has shown K_{assoc} values of 14–260 M^{-1}.[95] The aforementioned calix[4]arene was studied only in organic solvents, where it showed K_{assoc} values of 6–60 M^{-1} for a quaternized analogue of nicotine.[96]

2.3 Results and Discussion

2.3.1 Synthesis

We synthesized a small set of structurally related sulfonatocalix[4]arenes with regioselective introduction of different functional groups. We adapted an approach that uses regioselective protection of lower-rim phenols,[97] followed by sequences of electrophilic aromatic substitutions at the upper rim, deprotection of the lower rim, and ultimately sulfonation of the remaining upper-rim sites (Figure 2.1). The commercially available *p*-sulfonatocalix[4]arene (**sCx4**) is used in all studies for comparison. The targets for synthesis included mono-substituted sulfonatocalix[4]arenes **(sCx4-NO_2** and **sCx4-CHO),** which come from a tribenzoylated intermediate[98,99] and were previously reported.[100-102] The synthetic targets also included novel di-substituted calix[4]arenes (**sCx4-1,2-CHO**, **sCx4-1,3-CHO**, **sCx4-1,2-NO_2** and **sCx4-1,3-NO_2**).

The synthesis of the disubstituted hosts proceeds efficiently by using different dibenzoyl-protected calixarenes as key intermediates. The 1,3-dibenzoyl protected calix[4]arene [97] can be carried forward to make 1,3-disubstituted hosts (Figure 2.1). We optimized a previously reported procedure[103] for benzoyl rearrangement of 1,3-dibenzoyl protected calix[4]arene to yield the1,2-dibenzoyl protected calix[4]arene. The nitro- or formyl-groups are then introduced onto each unsubstituted phenol ring, with the benzoyl

protection groups on the lower rim offering remote protection by deactivating rings toward electrophilic aromatic substitutions. The formylated calix[4]arenes were prepared using a Duff formylation with hexamethylenetetramine (HMTA) and trifluoroacetic acid. For the nitration we used potassium nitrate and boron trifluoride diethyl etherate,[104] to acquire nitrated calix[4]arenes. Each of these conditions were selected to substitute the calix[4]arenes under mild conditions to minimize over-modification. The lower rim was then deprotected by basic hydrolysis, some of which were carried forward without purification (see materials and methods, and the supporting information). The deprotected calix[4]arenes were sulfonated under mild conditions using eight equivalents of sulfuric acid in dichloromethane to provide novel **sCx4-1,2-CHO**, **sCx4-1,3-CHO**, **sCx4-1,2-NO_2** and **sCx4-1,3-NO_2**.

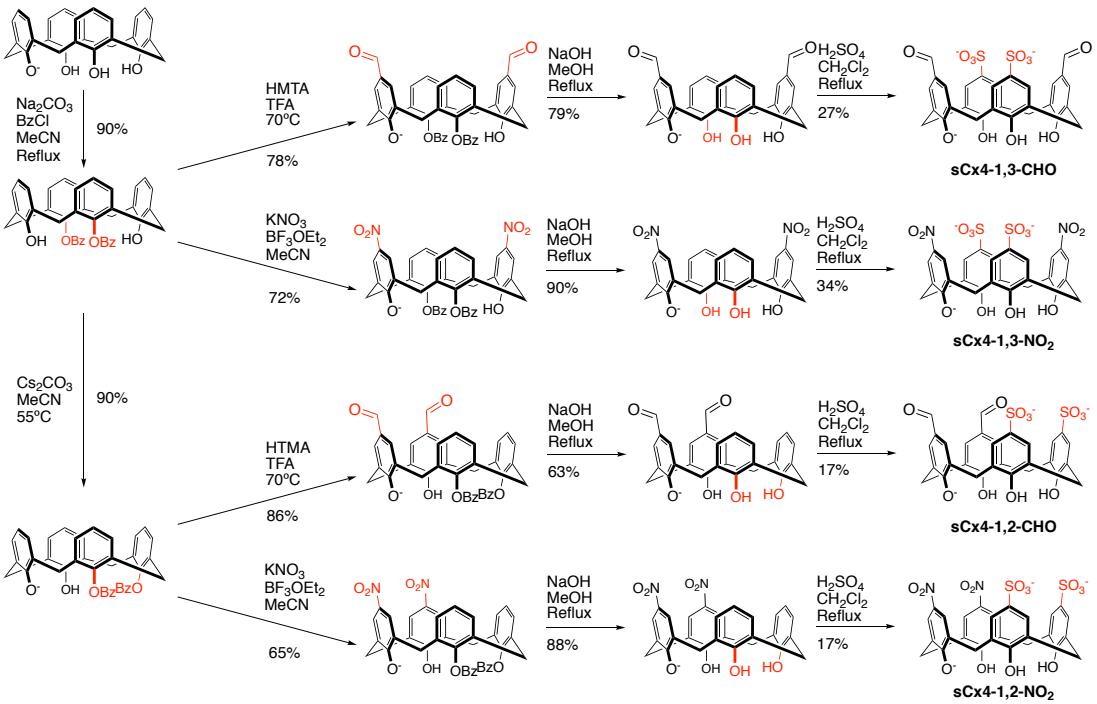


Figure 2.1. Synthesis of new regioselectivity modified calixarene sulfonates used in this study.

2.3.2 Binding studies

The binding constants were determined for the complexes formed by a set of seven calix[4]arene hosts (**sCx4, sCx4-CHO sCx4-1,2-CHO**, **sCx4-1,3-CHO**, **sCx4-NO$_2$**,, **sCx4-1,2-NO$_2$** and **sCx4-1,3-NO$_2$**) with each of the guest's nicotine, nornicotine, and cotinine. We used Na_2HPO_4 buffer at pH 7.4 to ensure consistent host and guest protonation states throughout all titrations. The binding constants were initially determined *via* indicator displacement assays (IDAs) based on an IDA previously reported for **sCx4**.[97] Each calixarene was first titrated directly into lucigenin (LCG) in order to determine the calixarene-dye K_{ind} values, which range from 1.78–6.80 μM (Figure 2.2a and supporting information). LCG is a fluorescent dye whose fluorescence is quenched upon complexation with a calix[4]arene host. During the ensuing competitive titrations, the indicator LCG was displaced from the calix[4]arene binding pocket by added guests and increased its fluorescence emission.[105,106] The fluorescence was measured and plotted against the concentration of guest, and the resulting curve was fitted to determine the binding constant for each host-guest pair (Figure 2.2b–d).

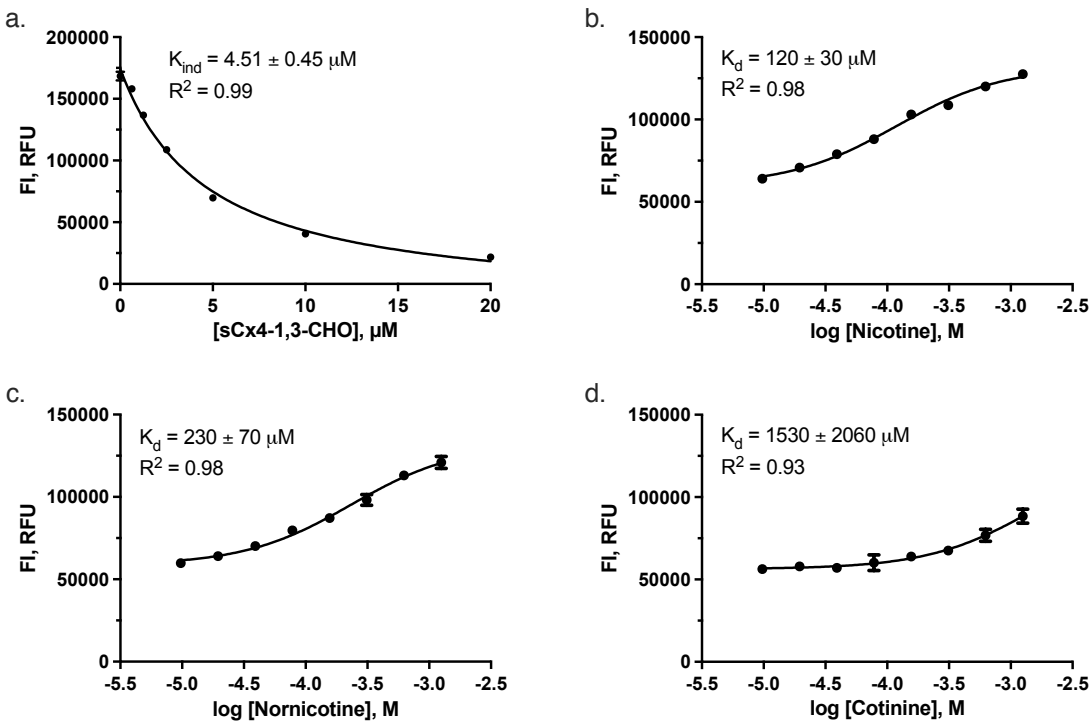


Figure 2.2. Fluorescence-based indicator displacement assays provide ready access to host-guest binding constants. Exemplary data for fluorescence-based studies of **sCx4-1,3-CHO** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-1,3-CHO** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

The studied guests display distinct differences in their binding constants. Nicotine and nornicotine display medium (K_d 200–1000 μM) to strong binding (K_d < 200 μM), whereas cotinine shows very weak to no binding (K_d > 1500 μM) to any host studied. Rather than the secondary amines present in nicotine and nornicotine, cotinine contains an amide bond and is uncharged at neutral pH. Overall, nicotine binds the hosts in general stronger than its slightly less hydrophobic metabolite nornicotine, but they are in a similar binding range (Table 2.1). The lack of measurable binding for neutral cotinine shows that a cationic group on the guest is absolutely required for binding. Based on the IDA data, nicotine was selected for further NMR-based binding and structural studies.

Table 2.1. K_d values determined by IDA for each host with nicotine, nornicotine and cotinine.[a,b]

	Nicotine K_d (μM)	Nornicotine K_d (μM)	Cotinine K_d (μM)
sCx4	180 ± 40	260 ± 110	>1500
sCx4-CHO	120 ± 20	200 ± 60	>1500
sCx4-1,2-CHO	510 ± 170	630 ± 120	1010 ± 1140
sCx4-1,3-CHO	140 ± 50	260 ± 120	>1500
sCx4-NO_2	200 ± 60	350 ±100	>1500
sCx4-1,2-NO_2	570 ± 180	640 ± 160	>1500
sCx4-1,3-NO_2	780 ± 420	880 ± 360	>1500

a) 0.01 M phosphate buffer pH 7.4, see the supporting information for titration curves and experimental details.

b) the values for nicotine and nornicotine arise from duplicates of triplicates, the IDAs for cotinine arise from duplicates of duplicates.

The binding strengths and structural motifs of nicotine with the same set of calix[4]arene hosts were further studied using NMR titrations.[107,108] The titration data was analyzed by using Bindfit as implemented on the open source website http://supramolecular.org in order to determine binding constants (see supporting information). The binding constants (Table 2.2) are similar to those determined by IDA and reveal similar trends among related family members in most cases, except for **sCx4-**

CHO. The IDA data showed stronger binding for **sCx4-CHO** than do the NMR data. The NMR data is a direct measurement, where the data arises from a direct observation of many changing chemical shifts, while the IDA arises from an indirect measurement from two competing equilibria. Where the two methods disagree, we take the NMR titration as the more reliable source for the binding constants.

In addition to K_d values, the NMR curve fitting provides structural insight by providing maximum binding-induced chemical shift ($\Delta\delta_{max}$) values. The nicotine protons involved in complexation with the interior of the host become shielded, causing chemical shifts to move upfield. The difference in chemical shift upon binding tells us how deep each proton is within the host's pocket, which we can interpret to determine how the guest is positioned inside of each host.[109,110]

There are three possible binding motifs for calix[4]arene host-guest complexes of this type.[111,112] In binding motif *I*, the biggest chemical shift is observed for the *N*-methyl, because the pyrrolidine ring is deeply bound in the pocket. For binding motif *II*, the biggest chemical shift is observed for the pyridine protons due to the pyridine being deeply bound in the pocket. In binding motif *III*, a similar (but relatively small) maximum chemical shift is observed for all protons as the protons associate with the upper rim but do not enter the binding pocket (Figure 2.3). This kind of non-insertion binding mode is less common for sulfonatocalixarenes, but has been seen, for example, when lysine binds to **sCx4**.[111,113]

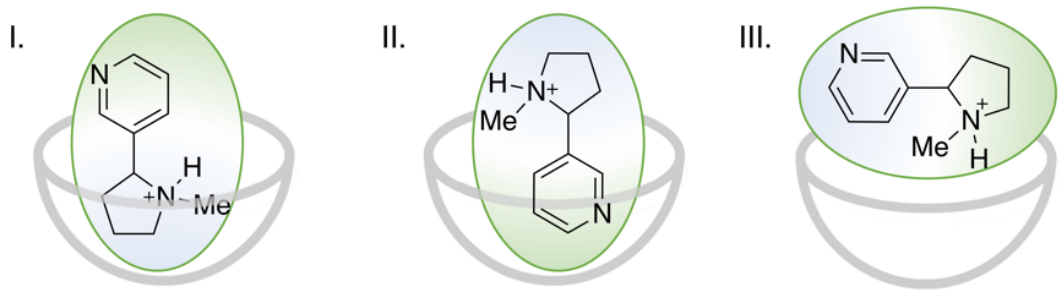


Figure 2.3. Possible binding motifs of nicotine with calixarene-type hosts

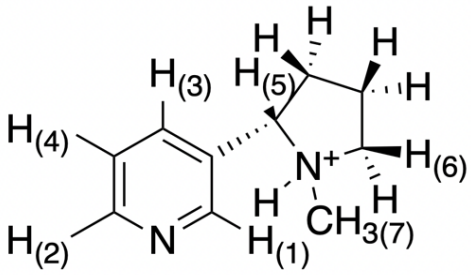
The formyl-substituted hosts bind nicotine *via* binding motif *I*. The biggest chemical shift is observed for the pyrrolidine protons, as they bind the deepest into the formylated binding pocket. This is consistent with the relative values of dissociation constants. **sCx4-CHO** has three anionic charges at the upper rim. When the pyrrolidine binds deeply in the pocket these charged groups are adjacent to the electron-poor pyridine heterocycle, forming less favourable host-guest contacts compared to **sCx4-1,2-CHO** and **sCx4-1,3-CHO**, which each have two anionic charges at the upper rim. This is reflected in the weaker binding of nicotine (Table 2.2).

The nitro-substituted hosts bind nicotine via binding motif *II*. Here, the biggest chemical shift is observed for the pyridine protons, and again we see trends in the binding constants that are consistent with this host-guest structure. When the pyridine is deep in the pocket, the positively charged pyrrolidine interacts favorably with the anionic charges of the hosts. **sCx4-NO$_2$** has three anionic charges at the upper rim and shows the strongest binding to nicotine, whereas compounds **sCx4-1,2-NO$_2$** and **sCx4-1,3-NO$_2$** have two anionic charges and display weaker binding than compound **sCx4-NO$_2$**. The close proximity of the anionic charges of **sCx4-1,2-NO$_2$** are favorable for binding compared to

when the charges are located further apart in **sCx4-1,3-NO_2**. Overall, the control, **sCx4**, (K_d = 100 μM), shows the strongest binding to nicotine while **sCx4-NO_2** (K_d = 150 μM) is the strongest among the other hosts.

Table 2.2. NMR summary, maximum chemical shift of nicotine protons upon binding with host library and K_d values.[a]

	δΔ max (ppm)							
	H1	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**	K_d (μM)
sCx4	0.37	0.31	0.62	0.37	0.97	n/a	0.89	96 ± 6
sCx4-CHO	0.62	0.64	1.12	0.83	1.45	1.74	1.12	404 ± 17
sCx4-1,2-CHO	0.66	0.97	1.40	1.21	1.35	1.29	0.88	249 ± 6
sCx4-1,3-CHO	0.55	0.81	1.25	1.06	1.22	1.35	0.98	235 ± 4
sCx4-NO_2	0.68	1.19	1.60	1.55	1.28	1.37	0.92	152 ± 5
sCx4-1,2-NO_2	0.72	1.49	1.75	1.93	1.05	0.80	0.60	245 ± 3
sCx4-1,3-NO_2	0.66	1.85	2.15	2.40	1.01	0.81	0.67	390 ± 6



a) See supporting information for NMR titration stacked plots and data fitting details.

These studies show how the number and positions of charged sulfonates can modulate sulfonatocalix[4]arene host-guest complexation, and how substituents can cause structural differences in guest binding that are not immediately obvious from the binding

constants themselves. A bigger chemical shift tells us how deeply a proton is buried in the binding pocket but does not directly indicate that stronger binding is occurring. Even though the biggest maximum chemical shift perturbation (2.40 ppm) is observed for **sCx4-1,3-NO_2**, it is one of the weaker binding hosts. In some cases, a very small chemical shift can arise from strong complexation. This is most clearly observed for **sCx4**, where all the chemical shifts are smaller than 1 ppm. The chemical shift profile for **sCx4** is hard to interpret definitively and suggests that the guest is bound by some mixture of modes *I*, *II*, and/or *III*.

We can learn some lessons by comparing these results to the previously mentioned nicotine-host complexes. Lhoták et al. showed how their chiral calix[4]arene receptor binds the pyridine moiety of *N*-methylnicotinium salt, corresponding to binding mode *II*.[96] They observed a difference in binding constants for their two host enantiomers, which showed K_{assoc} values of 81 ± 2 M^{-1} and 94 ± 5 M^{-1}. Their use of the quaternized *N*-methyl nicotinium analogue as a guest presumably arose because their binding studies occurred in organic solvent, where binding to native nicotine was too weak to measure. In water, the β-Cyclodextrin studies showed association constants between 10^1 and 10^2 M^{-1}, whereas the cucurbituril hosts have association constants in the range of 10^2–10^4 M^{-1}.[93-95] The calix[4]arenes presented in this paper have an association constant range of $2.5x10^3$–$1.0x10^4$ M^{-1}, showing that, even without optimization, the transfer of upper-rim substituted calixarene-type hosts into water leads to binding strengths that are on par with other nicotine binding supramolecular hosts.

2.4 Conclusions

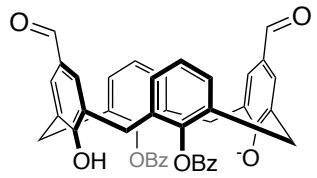
We have synthesized a set of novel calix[4]arene hosts and studied their binding to nicotine, nornicotine and cotinine. Calix[4]arenes show different patterns of binding selectivity depending on the identity and position of the upper-rim substitutions. From among the guests, nicotine forms the strongest host-guest complexes because it is cationic at neutral pH, unlike cotinine, and it is slightly more hydrophobic than nornicotine. From among the substituted hosts, **sCx4-NO_2** shows the strongest binding to nicotine. In the future, this fundamental knowledge of binding strengths and motifs can be applied further to a broad range of guest molecules, and the functional groups we have installed offer easy routes to further elaboration of the regioselectivity functionalized host scaffolds.

2.5 Supporting information

2.5.1 General

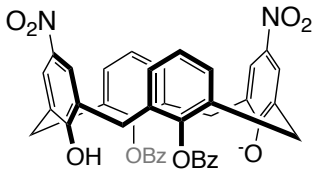
^{1}H- and ^{13}C-NMR were recorded on a Bruker Avance Neo 500 MHz spectrometer or Bruker AV300 MHz spectrometer and processed with MestReNova by Mestrelab Research S.L and TopSpin 3.5. Deuterated solvents were purchased from Sigma Aldrich. Melting points are uncorrected and were collected on a Gallenkamp Melting Point apparatus. High resolution mass spectra of novel compounds were collected on a Thermo Scientific Ultimate 3000 ESI-Orbitrap Exactive. Infrared (IR) spectra were acquired using a Perkin Elmer 1000 FT-IR spectrometer. Data are reported as follows: frequency of absorption (cm^{-1}), intensity of absorption (s = strong, m = medium, w =weak, br = broad). Compound nicknames below are numbered based on the four benzene positions around the macrocycle, rather than by IUPAC naming schemes, in order to give an accessible description of different regioismeric structures. The indicator displacement assays were conducted in a 10 mM Na_2HPO_4 buffer at pH 7.4. The NMR titrations were conducted in a 50 mM Na_2HPO_4 buffer prepared in D_2O at pH 7.4. Titrations and dilutions were conducted in a Nunc™ 384-Well, Non-Treated, Flat-Bottom, Optical Polymer Base Microplates.

2.5.2 Synthesis



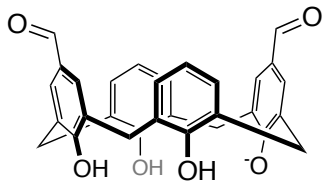
1,3-diformyl-2,4-dibenzoylcalix[4]arene

2,4-dibenzoylcalix[4]arene (1.0 g, 1.58 mmol) and HMTA (1.99 g, 14.2 mmol) were added to TFA (20 mL), and the reaction was heated to 70°C overnight. The reaction was diluted with water and the precipitate collected by vacuum filtration, yielding **1,3-diformyl-2,4-dibenzoylcalix[4]arene** as a pale-yellow solid (0.849 g, 78%) that was carried forward without further purification.



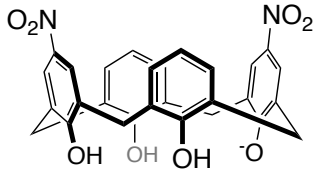
1,3-dinitro-2,4-dibezoylcalix[4]arene

2,4-dibenzoylcalix[4]arene (1.0 g, 1.58 mmol) and potassium nitrate (0.399 g, 3.95 mmol) were added to dry round-bottom flask. The system was evacuated and flushed three times with nitrogen. Dry acetonitrile (100 mL) was added via cannula. Boron trifluoride diethyl etherate (0.561 g, 3.95 mmol) was added via syringe and the reaction was stirred under nitrogen atmosphere at room temperature for 48 h. The reaction was quenched with hydrochloric acid (1 M) and the precipitate was collected by vacuum filtration, yielding **1,3-dinitro-2,4-dibezoylcalix[4]arene** as a yellow solid (0.822 g, 72%) that was carried forward without further purification.



1,3-diformyl-calix[4]arene

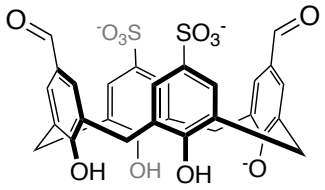
1,3-diformyl-2,4-dibenzoylcalix[4]arene (0.85 g, 1.23 mmol) was added to methanol (50 mL). Sodium hydroxide (0.74 g, 18.4 mmol) was added to the suspension. The solution was heated at 70°C for 4 hours. The reaction was quenched with hydrochloric acid (1 M) and the precipitate was collected by vacuum filtration. The crude product was triturated in hot hexanes to remove the benzoic acid, yielding **1,3-diformyl-calix[4]arene** as a dark brown/purple solid (0.468 g, 79%). ^{1}H-NMR (300 MHz, $CDCl_3$): δ 10.18 (s, 4H), 9.73 (s, 2H), 7.61 (s, 4H), 7.13 (d, J = 7.5 Hz, 4H), 6.80 (t, J = 7.5 Hz, 2H), 4.23 (s, 4H), 3.69 (s, 4H). ^{13}C-NMR (300 MHz, $CDCl_3$): 190.6, 154.6, 148.2, 131.1, 129.5, 128.9, 128.4, 127.8, 127.3. Melting point >230°C. HR-MS ($(M–H)^{-1}$ m/z): Calculated for $C_{30}H_{23}O_6^-$, 479.15001, Found 479.14987. FT-IR (cm^{-1}): 3179 (br), 1679 (m), 1596 (m), 1454 (m), 1280 (m), 752 (m).



1,3-dinitro-calix[4]arene

1,3-dinitro-2,4-dibezoylcalix[4]arene (0.82 g, 1.13 mmol) was added to methanol (50 mL). Sodium hydroxide (0.68 g, 17.1 mmol) was added to the suspension, which was then heated at 70°C for 4 hours. The reaction was quenched with hydrochloric acid (1 M) and the precipitate was collected by vacuum filtration. The crude product was triturated in hot hexanes to remove the benzoic acid, yielding **1,3-dinitro-calix[4]arene** as a dark

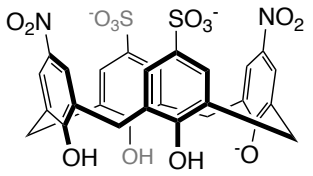
brown/purple solid (0.529 g, 90%). ^{1}H-NMR (300 MHz, $(CD_3)_2SO$): δ 8.02 (s, 4H), 7.12 (d, J = 7.5 Hz, 4H), 6.63 (t, J = 7.5, 2H), 3.93 (s, 8H). ^{13}C-NMR (300 MHz, D_2O): 161.8, 151.2, 138.0, 130.3, 128.61, 128.55, 124.0, 120.4, 30.8. Melting point >230°C. HR-MS ($(M–H)^{-1}$ m/z): Calculated for $C_{28}H_{21}N_2O_8^-$, 513.13034, Found 513.12995. FT-IR (cm^{-1}): 3091 (br), 1515 (m), 1340 (s), 764 (m).



sCx4-1,3-CHO

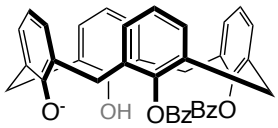
1,3-diformyl-calix[4]arene (0.47 g, 0.97 mmol) was added to CH_2Cl_2 (20 mL). While stirring, sulfuric acid (0.764 g, 7.79 mmol) was added. The solution was heated to 40°C overnight. The CH_2Cl_2 was decanted, and a minimum amount of ethyl acetate was added to the precipitate, after which a slurry was formed by sonication. The crude product was suspended by pouring the slurry into a 50 mL Falcon tube containing ice-cold ether (35 mL). The suspension was settled by centrifugation and the supernatant was discarded. The pellet was air dried and purified by high pressure liquid chromatography on a Shimadzu Prominence HPLC system over a 9.4 mm x 250 mm semi-preparative Agilent Eclipse XDB-C18 5 μM with UV detection at 280 nm. A gradient ran from 90% H_2O (+0.1% TFA)/10% CH_3CN (+0.1% TFA) to 85% H_2O (+0.1% TFA)/15% CH_3CN (+0.1% TFA) over 7 minutes, to 50% H_2O (+0.1% TFA)/50% CH_3CN (+0.1% TFA) over 13 minutes. The fractions were collected and lyophilized, yielding **sCx4-1,3-CHO** as a light brown solid (0.168 g, 27%). ^{1}H-NMR (300 MHz, D_2O): δ 9.12 (s, 2H), 7.58 (s, 4H), 7.41 (s, 4H), 3.78 (s, 8H). ^{13}C-NMR (300 MHz, D_2O): 193.8, 155.2, 151.1, 136.2, 131.5, 129.7, 128.2,

128.0, 126.7, 30.2. Melting point >230°C. HR-MS ((M–H)$^{-1}$ *m/z*): Calculated for $C_{30}H_{21}O_{12}S_2^-$, 639.06364, Found 639.06473. FT-IR (cm^{-1}): 3180 (br), 1670 (m), 1593 (m), 1133 (m), 1036 (m).



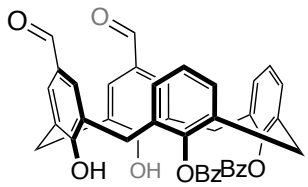
sCx4-1,3-NO_2

1,3-dinitro-calix[4]arene (0.53 g, 1.02 mmol) was added to CH_2Cl_2 (20 mL). While stirring, sulfuric acid (0.804 g, 8.19 mmol) was added, and the solution was heated to 40°C overnight. The CH_2Cl_2 was decanted, and a minimum amount of ethyl acetate was added to the precipitate, after which a slurry was formed by sonication. The crude product was suspended by pouring the slurry into a 50 mL Falcon tube containing ice-cold ether (35 mL). The suspension was settled by centrifugation and the supernatant was discarded. The pellet was air dried and purified by high pressure liquid chromatography using the same method as reported for **sCx4-1,3-CHO**. The fractions were collected and lyophilized, yielding **sCx4-1,3-NO_2** as a light brown solid (0.145 g, 21%). ^{1}H-NMR (300 MHz, D_2O): δ 7.87 (s, 4H), 7.55 (s, 4H), 3.85 (s, 8H). ^{13}C-NMR (300 MHz, D_2O): 156.1, 151.3, 140.5, 136.0, 128.2, 127.5, 126.8, 124.9, 30.3. Melting point >230°C. HR-MS ((M–2H)$^{-2}$ *m/z*): Calculated for $C_{28}H_{20}N_2O_{14}S_2^{-2}$, 336.01834, Found 336.01858. FT-IR (cm^{-1}): 3186 (br), 1518 (w), 1339 (m), 1110 (m), 1036 (m).



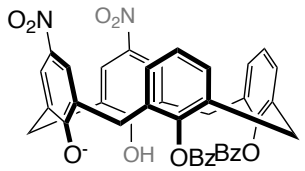
3,4-dibenzoylcalix[4]arene

2,4-dibenzoylcalix[4]arene (4.0 g, 6.32 mmol) and cesium carbonate (4.12 g, 12.6 mmol) were added to a dry round-bottom flask. The system was evacuated and flushed three times with nitrogen. Dry acetonitrile (150 mL) was added via cannula under a nitrogen atmosphere. The reaction was stirred at 55°C for 1.5 hours. The solvent was removed in vacuo and residue was re-dissolved in a minimum volume of CH_2Cl_2. The organic layer was washed three times with 1 M HCl and once with brine after which it was dried over magnesium sulfate. The CH_2Cl_2 was concentrated under reduced pressure and the product was precipitated in ice-cold ether. The precipitate was collected by vacuum filtration, yielding **3,4-dibenzoylcalix[4]arene** as a yellow solid (3.6 g, 90%) that was carried forward without further purification.



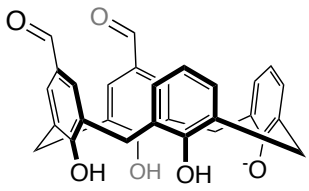
1,2-diformyl-3,4-dibezoylcalix[4]arene

3,4-dibenzoylcalix[4]arene (1.0 g, 1.58 mmol) and hexamethyltetramine (HMTA) (1.99 g, 14.2 mmol) were added to TFA (20 mL) and heated to 70°C overnight. The reaction was diluted with water and the precipitate collected by vacuum filtration, yielding **1,2-diformyl-3,4-dibezoylcalix[4]arene** as a pale-yellow solid (0.936 g, 86%) that was carried forward without further purification.



1,2-dinitro-3,4-dibezoylcalix[4]arene

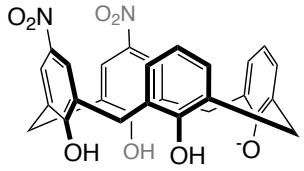
3,4-dibenzoylcalix[4]arene (1.0 g, 1.58 mmol) and potassium nitrate (0.399 g, 3.95 mmol) were added to a dry round bottom flask. The system was evacuated and flushed three times with nitrogen. Dry acetonitrile (100 mL) was added via cannula. Boron trifluoride diethyl etherate (0.561 g, 3.95 mmol) was added via syringe and the reaction was stirred under nitrogen atmosphere at room temperature for 48 h. The reaction was quenched with hydrochloric acid (1 M) and the precipitate was collected by vacuum filtration, yielding **1,2-dinitro-3,4-dibezoylcalix[4]arene** as a yellow solid (0.742 g, 65%) that was carried forward without further purification.



1,2-diformylcalix[4]arene

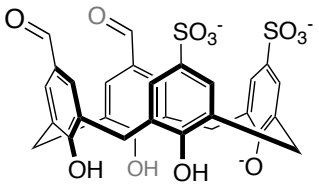
1,2-diformyl-3,4-dibezoylcalix[4]arene (0.94 g, 1.36 mmol) was added to methanol (50 mL). Sodium hydroxide (0.815, 20.4 mmol) was added to the suspension, which was heated at reflux at 70°C for 4 hours. The reaction was quenched with hydrochloric acid (1 M) and the precipitate was collected by vacuum filtration. The crude product was triturated in hot hexanes to remove the benzoic acid, yielding **1,2-diformylcalix[4]arene** as a dark brown/purple solid (0.411 g, 63%). ^{1}H-NMR (500 MHz, $CDCl_3$): δ 10.16 (s, 4H), 9.77 (s, 2H), 7.64 (s, 4H), 7.10 (d, J = 7.5 Hz, 2H), 7.08 (d, J = 7.7 Hz, 2H), 6.76 (t,

J = 7.7 Hz, 2H), 4.28 (s, 4H), 3.66 (s, 4H). ^{13}C-NMR (500 MHz, $CDCl_3$): δ 190.4, 154.4, 148.3, 131.4, 131.3, 131.0, 130.0, 129.5, 129.4, 129.2, 128.2, 128.1, 127.0, 122.7, 31.5. Melting point >230°C. HR-MS ($(M–H)^{-1}$ *m/z*): Calculated for $C_{30}H_{23}O_6^-$, 479.15001, Found 479.15009. FT-IR (cm^{-1}): 3189 (br), 1678 (m), 1596 (m), 1285 (m), 1128 (m), 756 (m).



1,2-dinitrocalix[4]arene

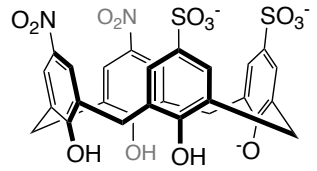
1,2-dinitro-3,4-dibezoylcalix[4]are (0.74 g, 1.03 mmol) was added to methanol (50 mL). Sodium hydroxide (0.616 g, 15.4 mmol) was added to the suspension. The solution was heated at reflux at 70°C for 4 hours. The reaction was quenched with hydrochloric acid (1 M) and the precipitate was collected by vacuum filtration. The crude product was triturated in hot hexanes to remove the benzoic acid, yielding **1,2-dinitrocalix[4]arene** as a dark brown/purple solid (0.465 g, 88%) that was carried forward without further purification.



sCx4-1,2-CHO

1,2-diformylcalix[4]arene (0.30 g, 0.63 mmol) was added to CH_2Cl_2 (20 mL). While stirring, sulfuric acid (0.491 g, 5.00 mmol) was added. The solution was heated to 40°C overnight. The CH_2Cl_2 was decanted, and a minimum amount of ethyl acetate was added to the precipitate after which a slurry was formed by sonication. The crude product was

suspended by pouring the slurry into a 50 mL Falcon tube with ice-cold ether (35 mL). The suspension was settled by centrifugation and the supernatant was discarded. The pellet was air dried and purified by high pressure liquid chromatography on a Shimadzu Prominence HPLC system over a 9.4 mm x 250 mm semi-preparative Agilent Eclipse XDB-C18 5 μM with UV detection at 280 nm. A gradient ran from 90% H_2O (+0.1% TFA)/10% CH_3CN (+0.1% TFA) to 75% H_2O (+0.1% TFA)/25% CH_3CN (+0.1% TFA) over 7 minutes, to 50% H_2O (+0.1% TFA)/50% CH_3CN (+0.1% TFA) over 13 minutes. The fractions were collected and lyophilized, yielding **sCx4-1,2-CHO** as a light brown solid (0.107 g, 17%). ^{1}H-NMR (500 MHz, D_2O): δ 9.38 (s, 2H), 7.62 (s, 2H), 7.58 (s, 2H), 7.56 (s, 4H), 4.02 (s, 2H), 3.97 (s, 4H), 3.78 (s, 2H). ^{13}C-NMR (500 MHz, D_2O): δ 193.74, 154.96, 151.11, 136.24, 132.19, 131.10, 129.95, 128.30, 128.10, 128.05, 127.90, 126.64, 126.59, 30.65, 30.42, 29.93. Melting point >230°C. HR-MS ($(M–H)^{-1}$ *m/z*): Calculated for $C_{30}H_{21}O_{12}S_2^-$, 639.06364, Found 639.06443. FT-IR (cm^{-1}): 3185 (br), 1674 (m), 1234 (s), 1036 (s), 623 (s).



sCx4-1,2-NO_2

1,2-dinitrocalix[4]arene (0.47 g, 0.90 mmol) was added to CH_2Cl_2 (20 mL). While stirring, sulfuric acid (0.709 g, 7.23 mmol) was added, and the solution was heated to 40°C overnight. The CH_2Cl_2 was decanted, and a minimum amount of ethyl acetate was added to the precipitate after which a slurry was formed by sonication. The crude product was suspended by pouring the slurry into a 50 mL Falcon tube containing ice-cold ether

(35 mL). The suspension was settled by centrifugation and the supernatant was discarded. The pellet was air dried and purified by high pressure liquid chromatography as was reported for **sCx4-1,2-CHO**. The fractions were collected and lyophilized, yielding **sCx4-1,2-NO_2** as a light brown solid (0.093 g, 17%). ^{1}H-NMR (300 MHz, D_2O): δ 8.00 (s, 2H), 7.91 (s, 2H), 7.53 (s, 4H), 3.97 (s, 6H), 3.73 (s, 2H). ^{13}C-NMR (500 MHz, D_2O) δ 162.71, 153.82, 138.39, 134.68, 130.66, 130.46, 129.72, 129.64, 126.08, 125.99, 125.11, 125.05, 31.89, 31.29, 30.90. Melting point >230°C. HR-MS ($(M–2H)^{-2}$ *m/z*): Calculated for $C_{28}H_{20}N_2O_{14}S_2^{-2}$, 336.01834. Found 336.01862. FT-IR (cm^{-1}): 3211 (br), 1679 (m), 1446 (w), 1340 (w) 1187 (s), 1135 (s) 1048 (m), 724 (m).

2.5.3 ^{1}H-, ^{13}C-NMR of novel calix[4]arenes

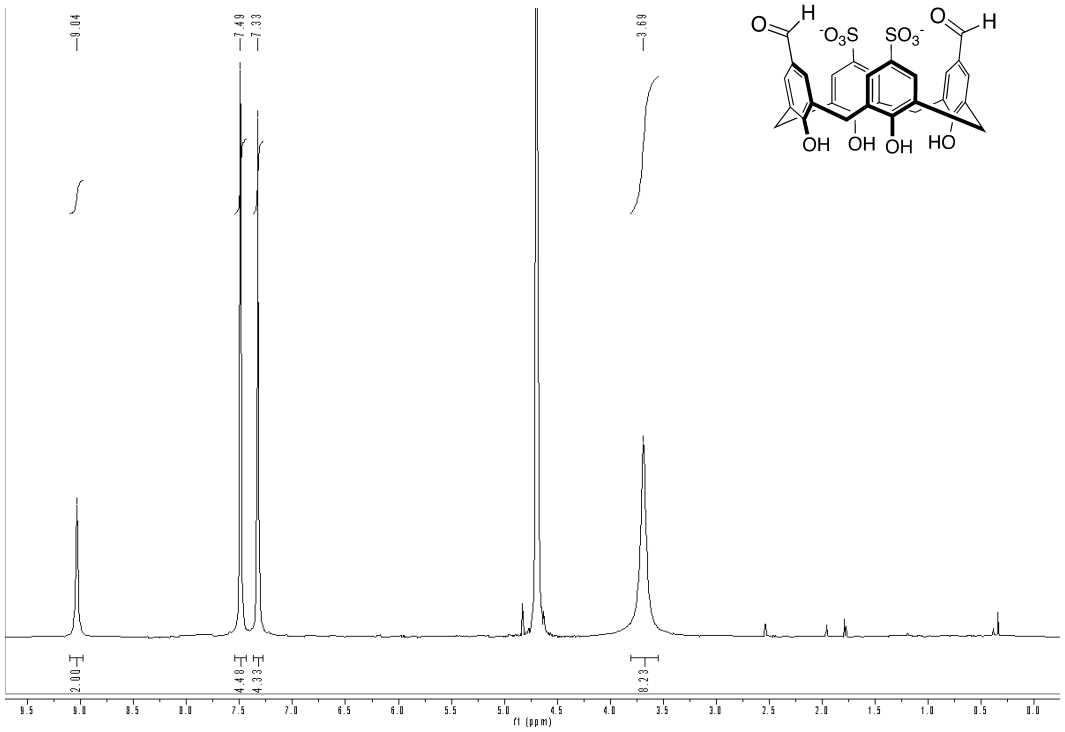


Figure S2.1. ^{1}H-NMR spectrum of **sCx4-1,3-CHO** (300 MHz) in D_2O

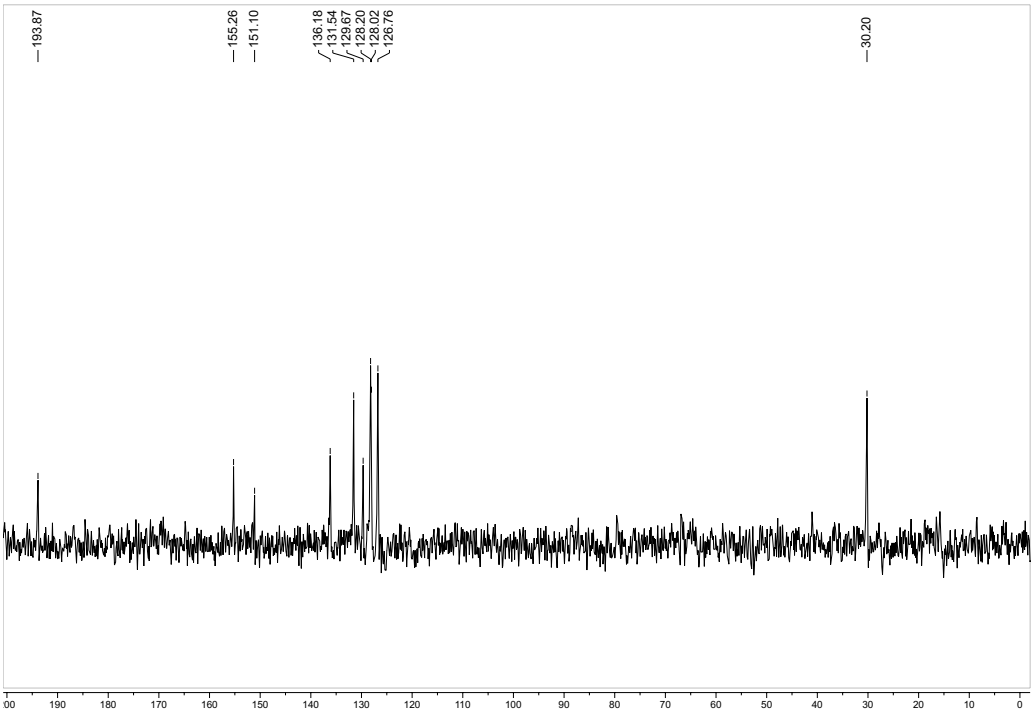


Figure S2.2. ^{13}C-NMR spectrum of **sCx4-1,3-CHO** (75 MHz) in D_2O

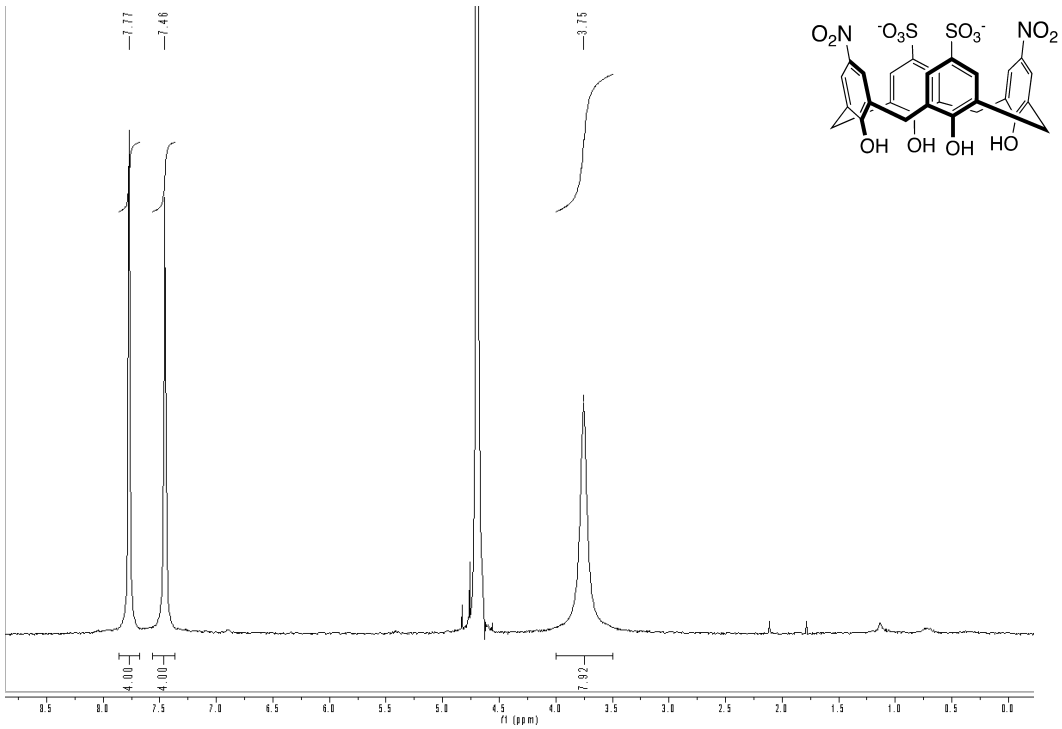


Figure S2.3. ^{1}H-NMR spectrum of **sCx4-1,3-NO_2** (300 MHz) in D_2O

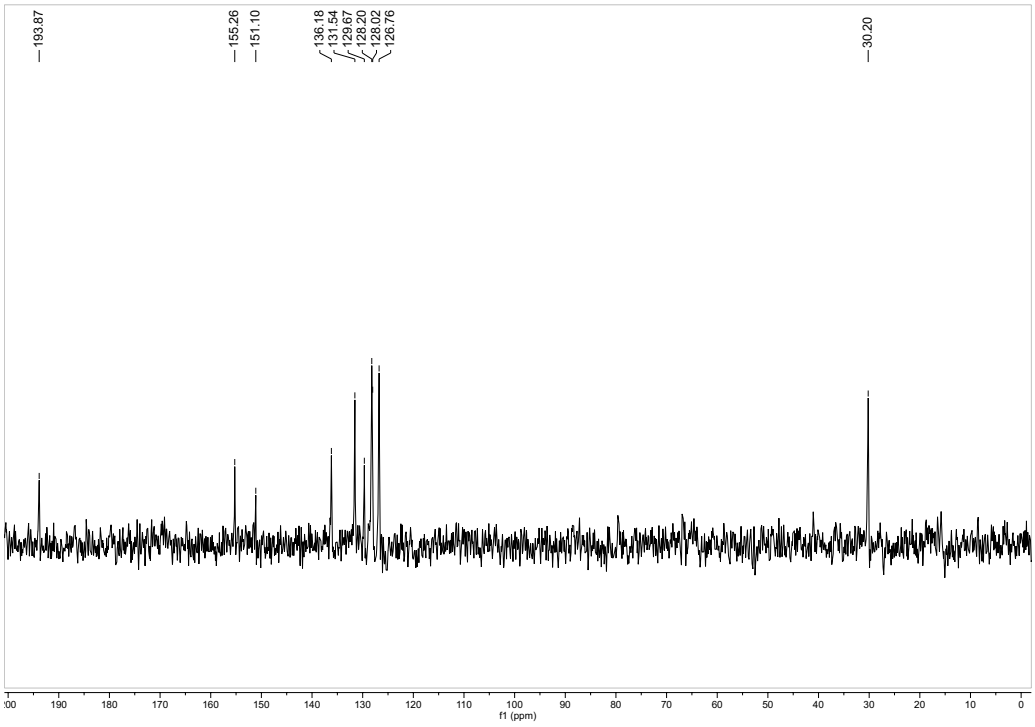


Figure S2.4. ^{13}C-NMR spectrum of **sCx4-1,3-NO_2** (75 MHz) in D_2O

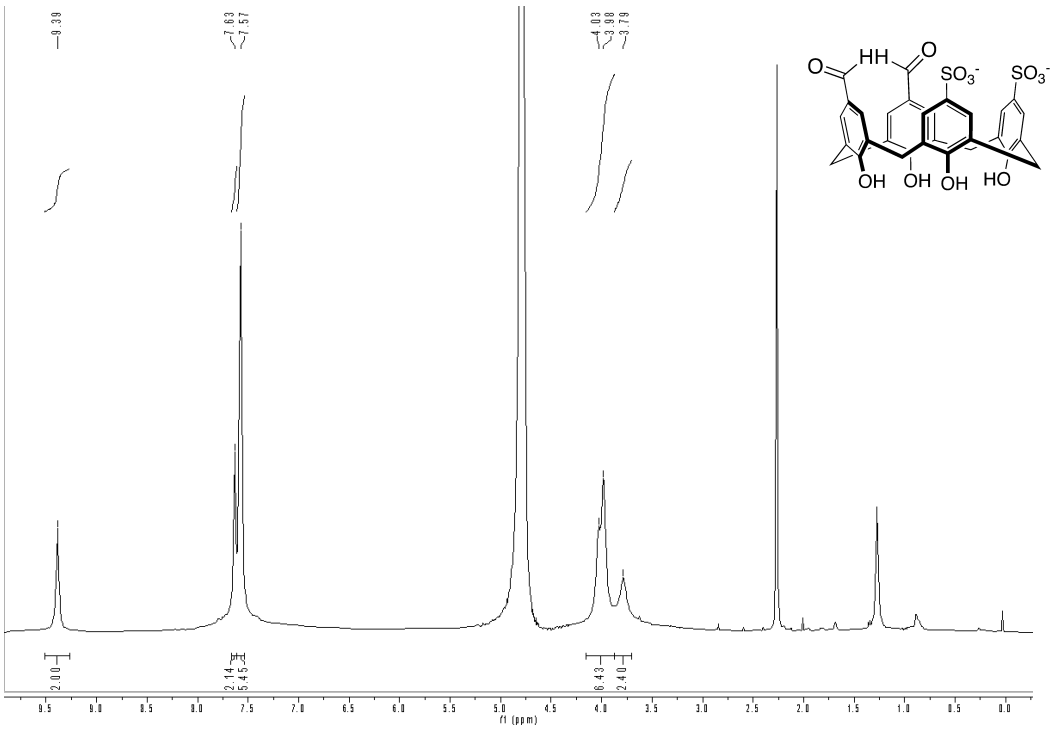


Figure S2.5. ^{1}H-NMR spectrum of **sCx4-1,2-CHO** (500 MHz) in D_2O

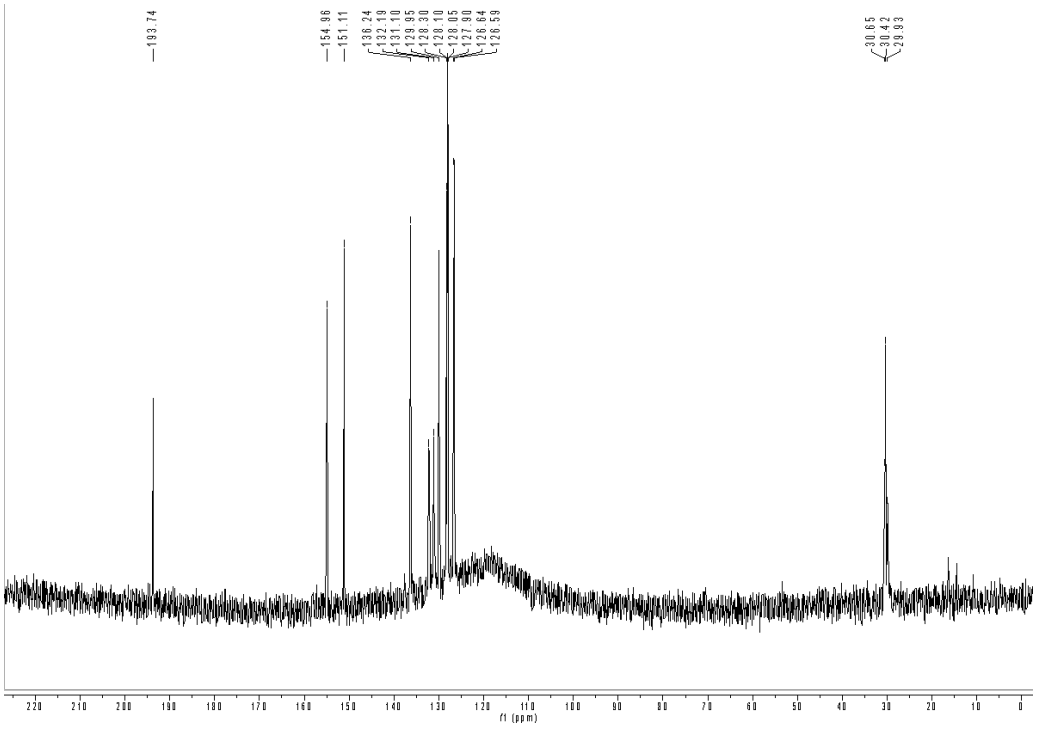


Figure S2.6. ^{13}C-NMR spectrum of **sCx4-1,2-CHO** (125 MHz) in D_2O

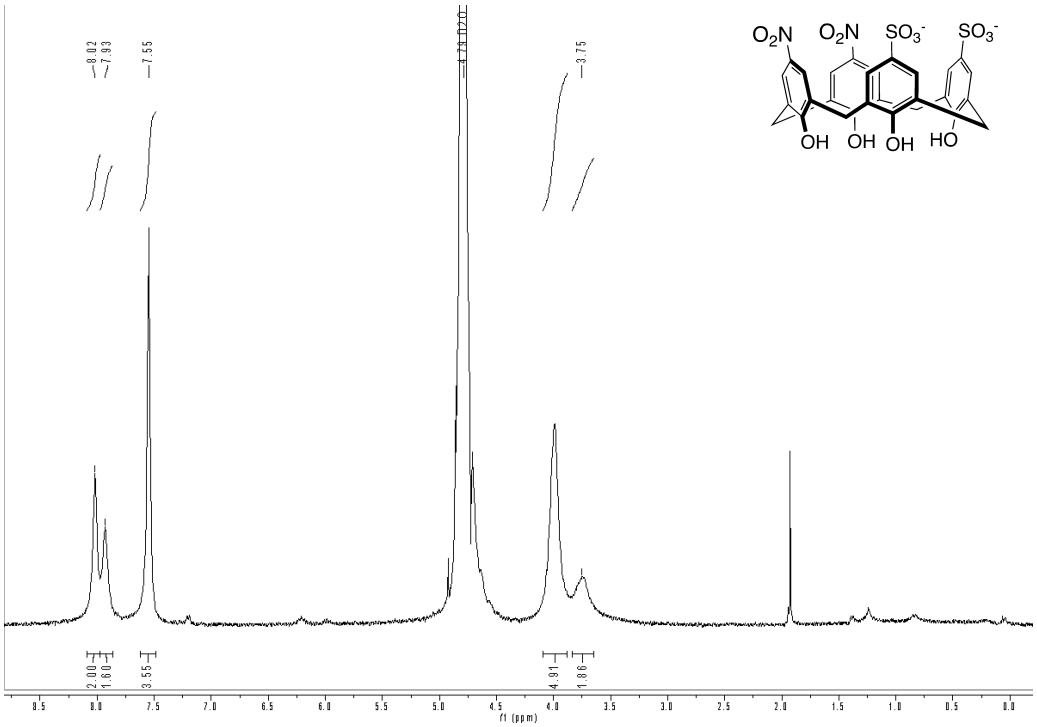


Figure S2.7. ^{1}H-NMR spectrum of **sCx4-1,2-NO_2** (300 MHz) in D_2O

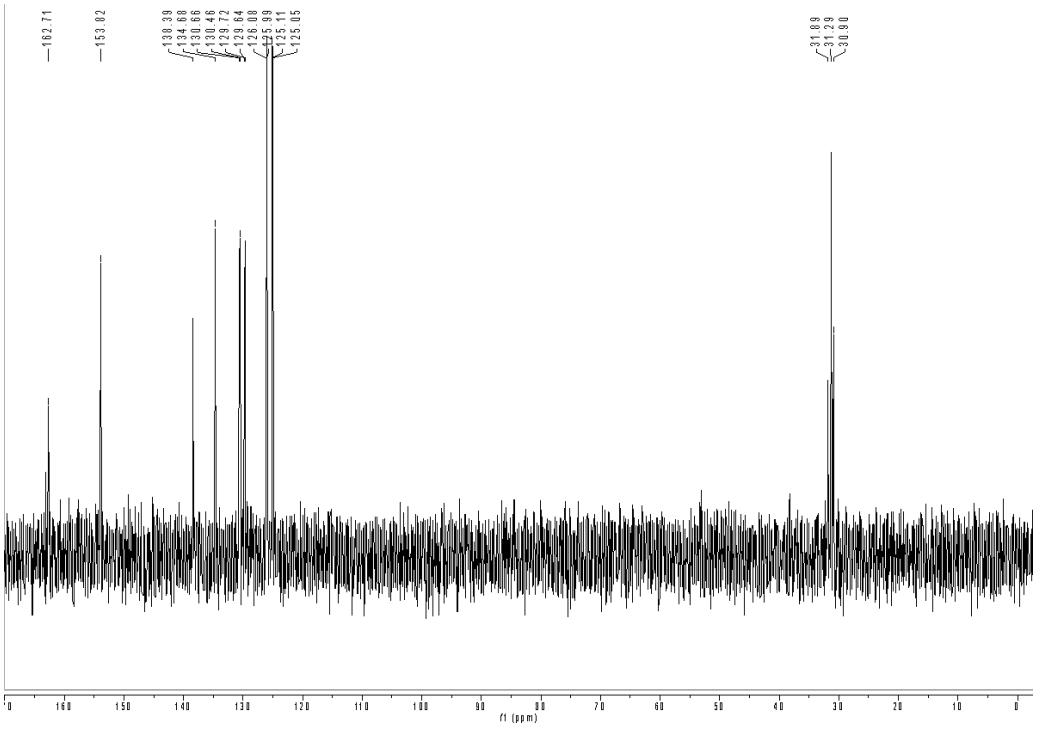


Figure S2.8. ^{13}C-NMR spectrum of **sCx4-1,2-NO_2** (125 MHz) in D_2O

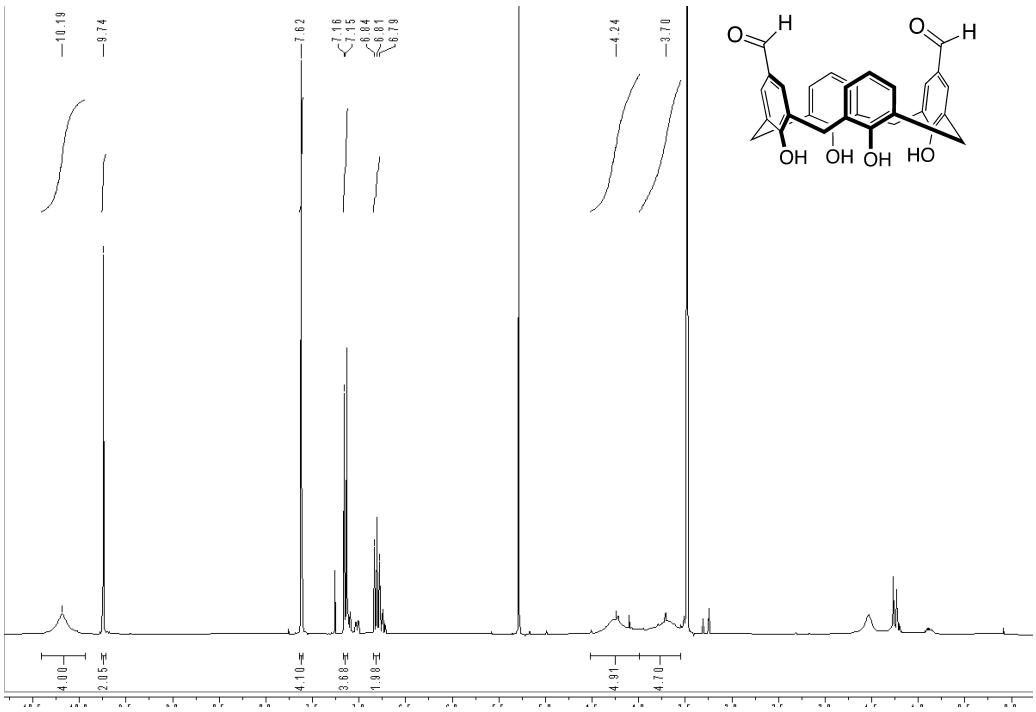


Figure S2.9. ^{1}H-NMR spectrum of **1,3-diformyl-calix[4]arene** (300 MHz) in D_2O

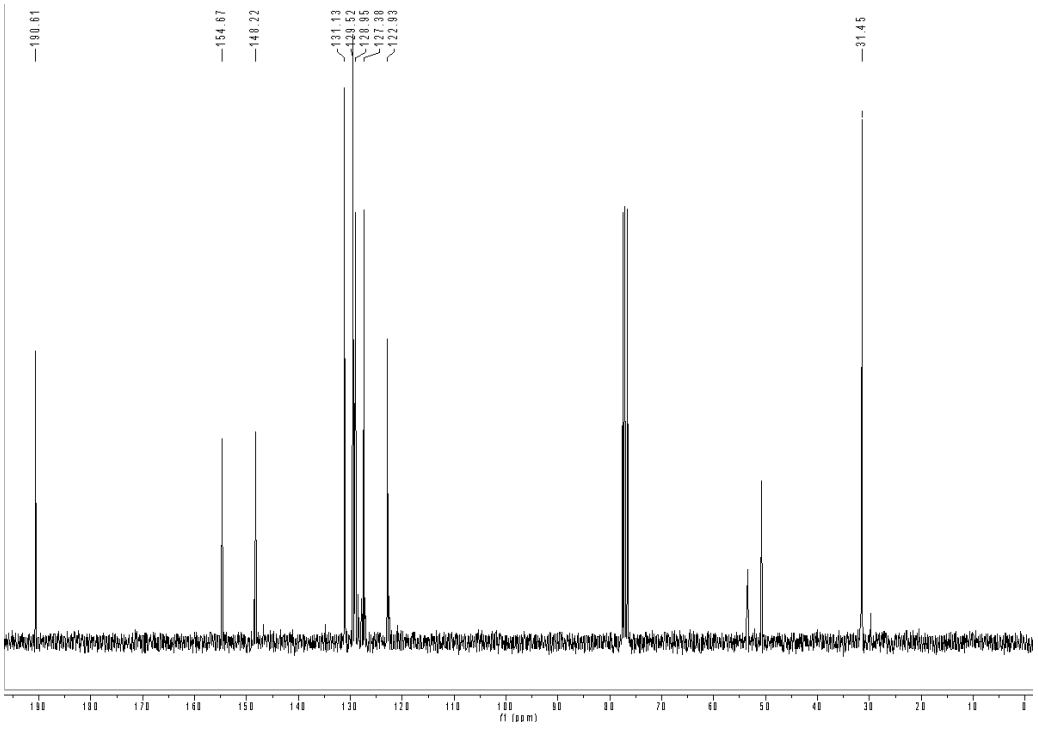


Figure S2.10. ^{13}C-NMR spectrum of **1,3-diformyl-calix[4]arene** (75 MHz) in D_2O

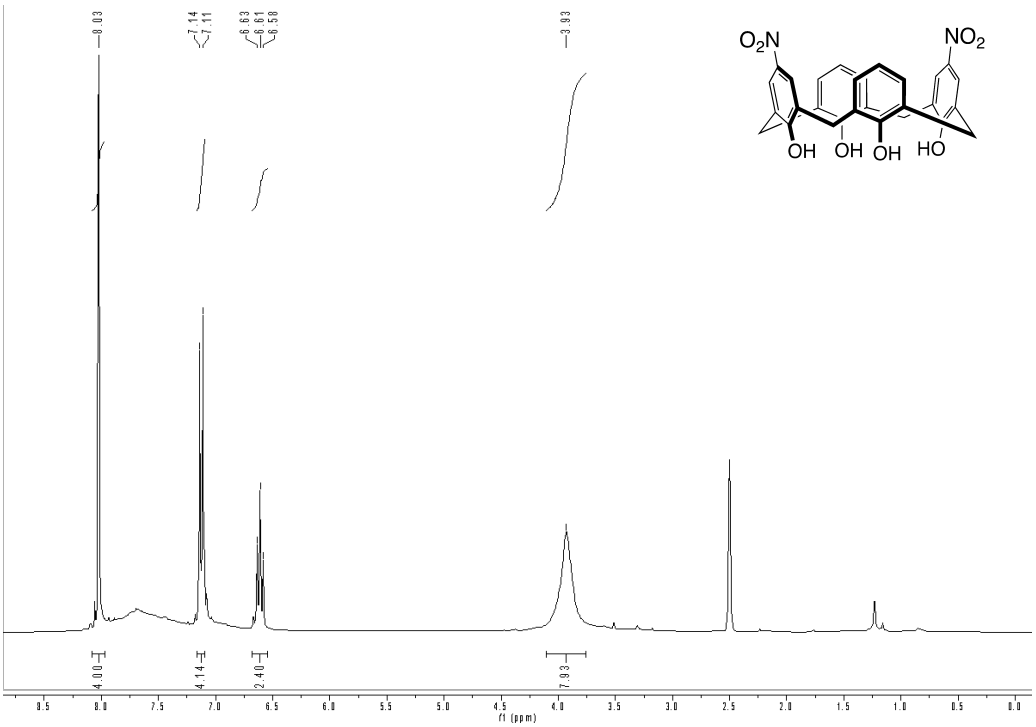


Figure S2.11. ^{1}H-NMR spectrum of **1,3-dinitro-calix[4]arene** (300 MHz) in D_2O

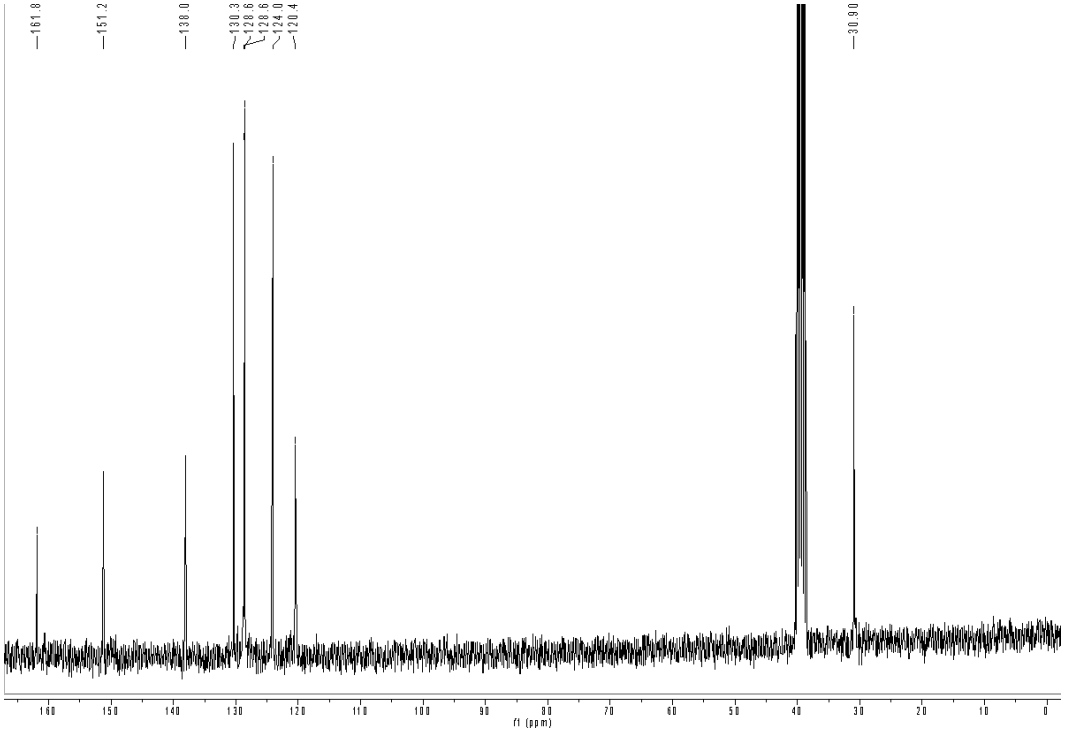


Figure S2.12. ^{13}C-NMR spectrum of **1,3-dinitro-calix[4]arene** (75 MHz) in D_2O

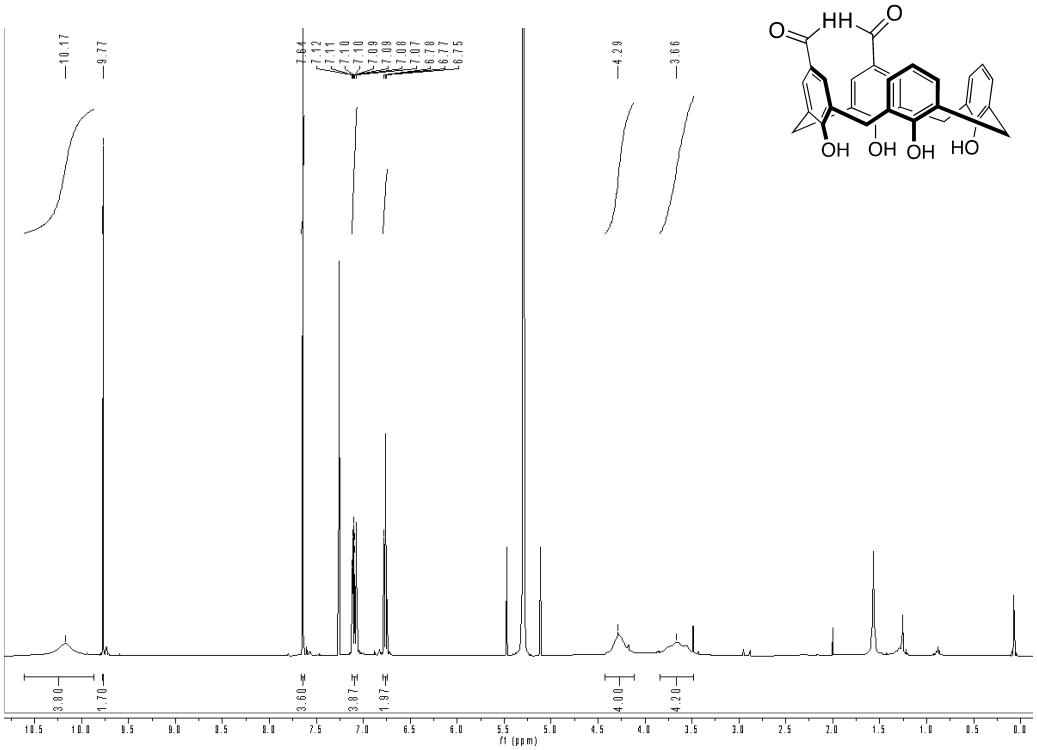


Figure S2.13. ^{1}H-NMR spectrum of **1,2-diformylcalix[4]arene** (500 MHz) in D_2O

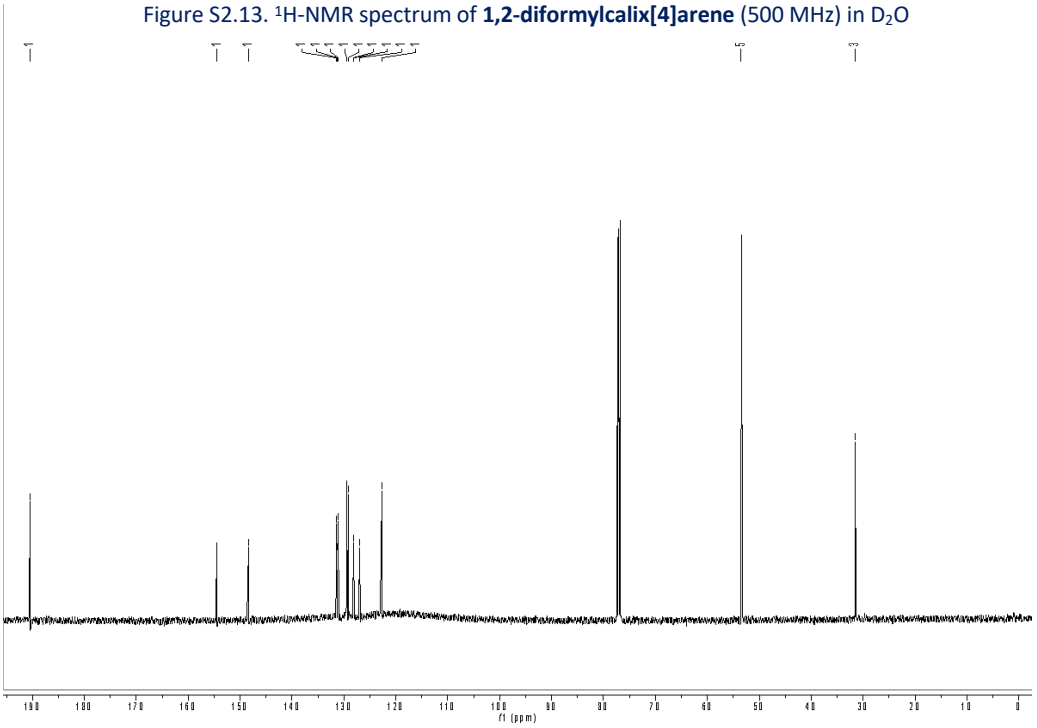


Figure S2.14. ^{13}C-NMR spectrum of **1,2-diformylcalix[4]arene** (125 MHz) in D_2O

2.5.4 IDA titrations with nicotine, nornicotine and cotinine

The IDAs were conducted in 384 well plates (Nunc™ 384-Well, Non-Treated, Flat-Bottom Microplate). The fluorescent signal was read on a Biotek Cytation 5 multimode plate reader (Software Version 3.05.11) at room temperature as a fluorescent endpoint measurement. 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. All wells had a final working volume of 50 μL. To measure the dye controls, three wells were prepared with a concentration of 0.25 μM LCG in 10 mM phosphate buffer, pH 7.4. To measure the host controls, three wells were prepared with a concentration of 5 μM host in 10 mM buffer. For the direct titrations of host into dye, a serial dilution ranging from 200 μM to 0.01 μM host into 0.25 μM LCG in 10 mM phosphate buffer was performed. The dissociation constants of the indicator (K_{ind}) that arrises from the direct titrations is used to fit the competitive titrations. For the competitive titrations, a serial dilution was ranging from 5 mM to 0.01 mM guest into 5 μM host, 0.25 μM LCG in 10 mM phosphate buffer was performed. All the IDA data was analyzed in GraphPad Prism Version 8.3.0 (328). The experiments for nicotine and nornicotine were performed as duplicates of triplicates and the experiments for cotinine were performed as duplicates of duplicates.

2.5.4.1 Calculations

Outliers

Outlier data points were 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 90% confidence, $Q > 0.941 = Q_{90\%,n=3}$, we conclude the data point is an outlier.

Equation 1. Dixon's Q-test

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

The standard error

The standard error was calculated for each triplicate, Equation 2.

Equation 2. Standard error

$$SD_{K_d} = \left(ln\ (10)\ * 10^{logK_d}\right) *\ SD_{logK_d}$$

SDK_d = Standard error of K_d

logK_d = Value of the log

SD$_{log}K_d$ = Standard error of the log K_d

The total standard error

The total standard error of the mean value determined from two triplicate measurements was 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

Curve fit for the direct titration

To curve fit for the direct titration Equation 4 and Equation 5 were used.

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

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

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_{ind} = Dissociation constant

Equation 5. Curve fit for the competitive titration

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

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

log_{EC50} = log of the concentration of the competitor binding half-way between F_{min} and F_{max}

K_d = Equilibrium dissociation constant in Molar

[D] = Concentration of dye in nM

K_{ind} = Equilibrium dissociation constant of the direct titration

F = Fitted data point

F_{max} = Maximum signal

F_{min} = Minimum signal

2.5.4.2 Fluorescence studies

Direct titrations of LCG with calix[4]arene hosts

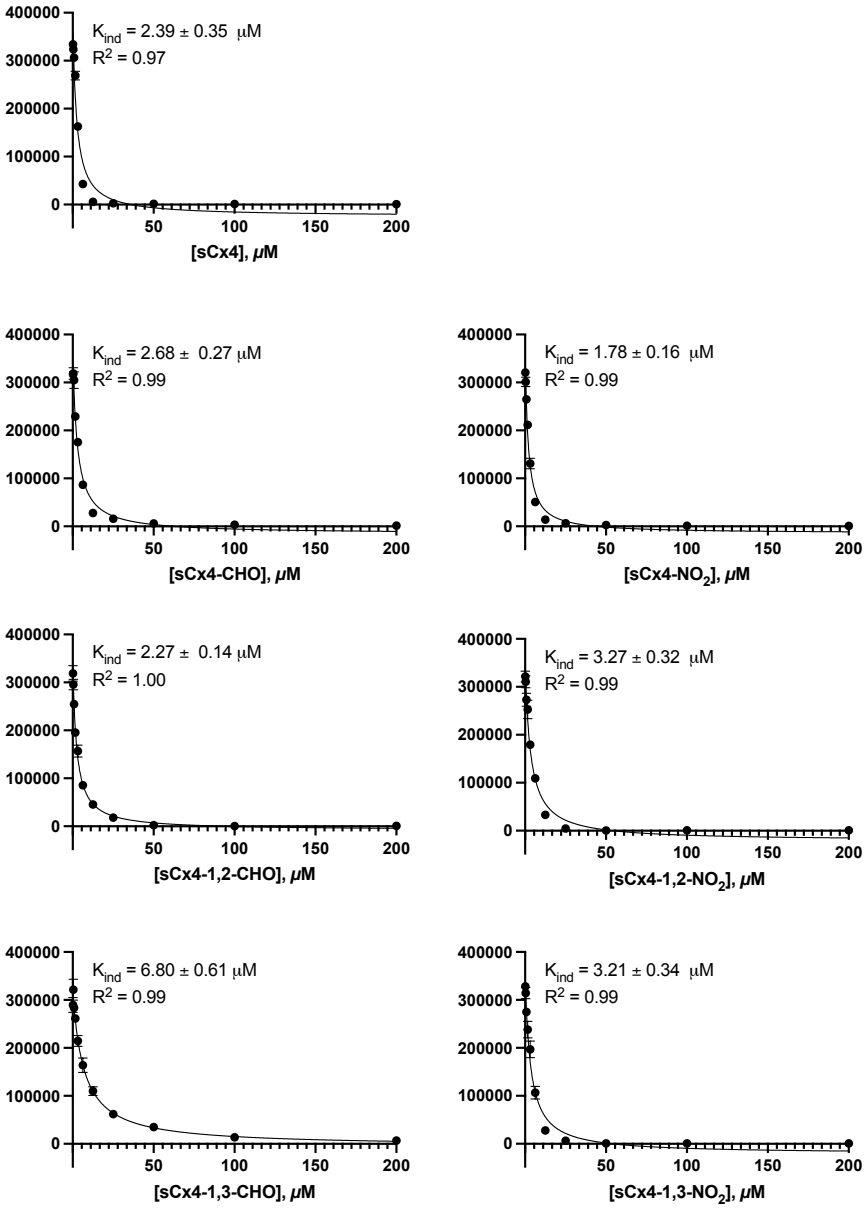


Figure S2.15. The direct titration of LCG (0.25 μM) with calix[4]arene hosts (0 – 200 μM)

Fluorescence studies of sCx4

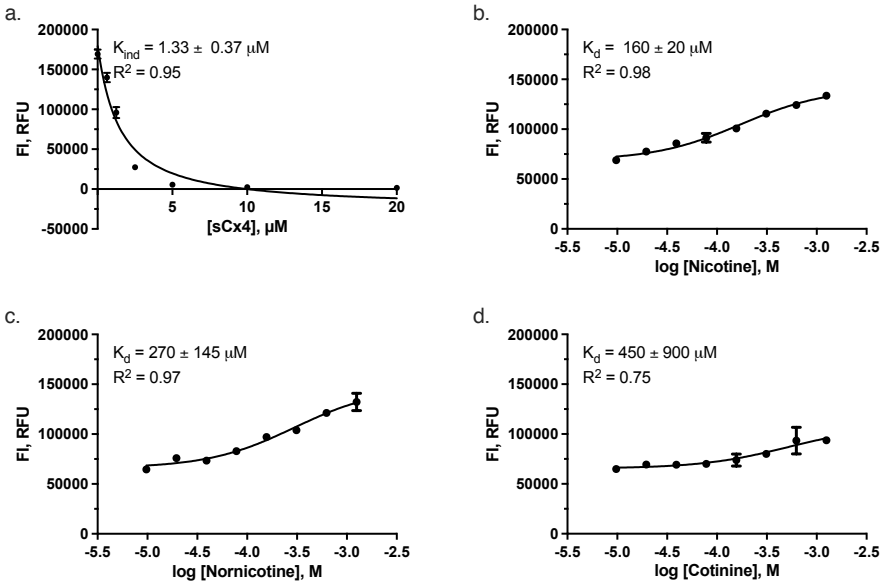


Figure S2.16. **sCx4** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

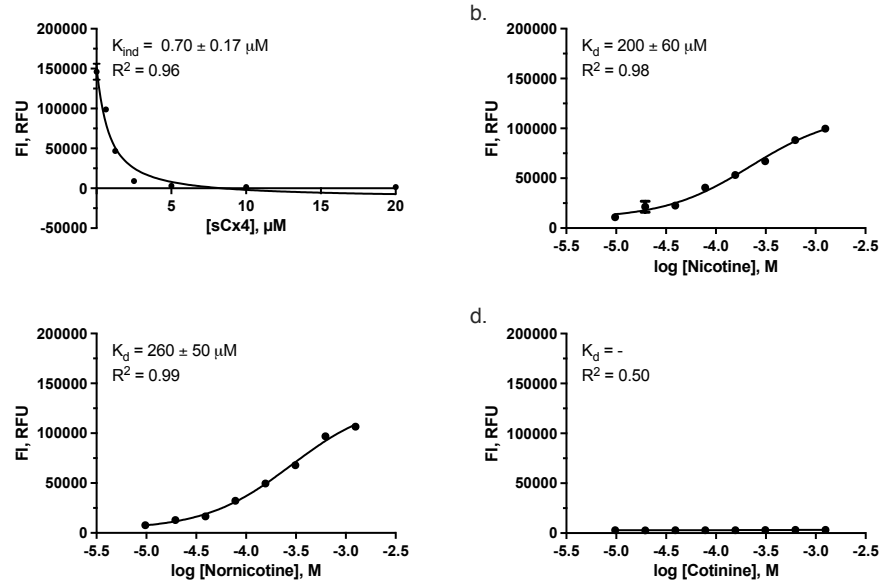


Figure S2.17. **sCx4** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

Fluorescence studies of sCx4-CHO

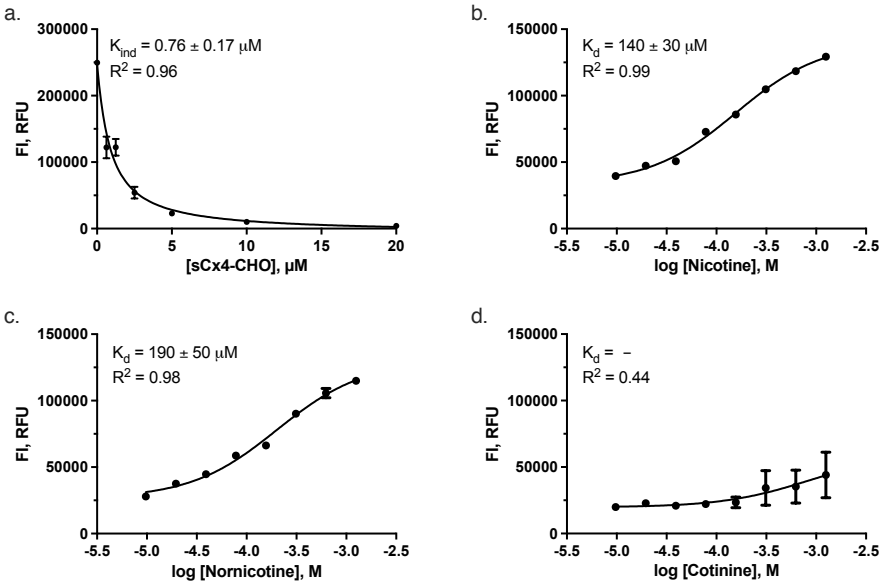


Figure S2.18. **sCx4-CHO** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-CHO** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

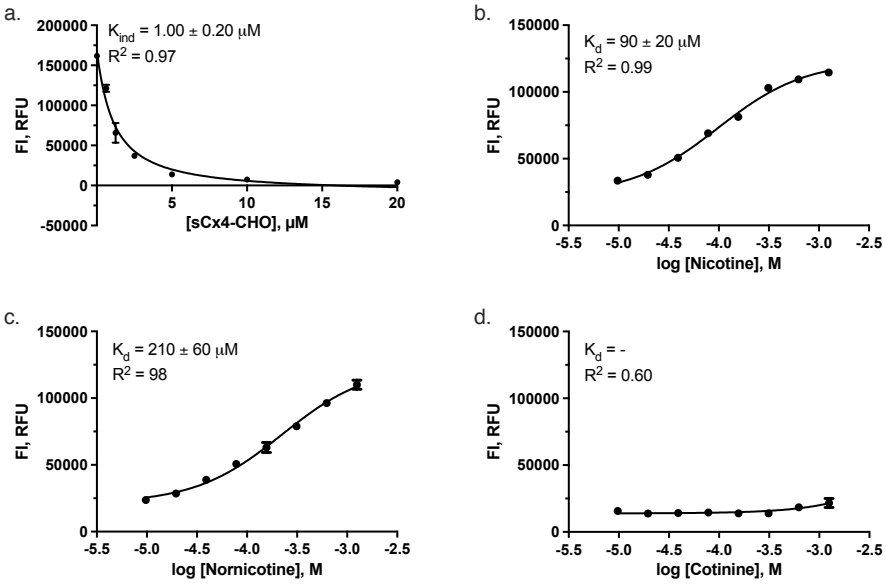


Figure S2.19. **sCx4-CHO** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-CHO** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

Fluorescence studies of sCx4-1,2-CHO

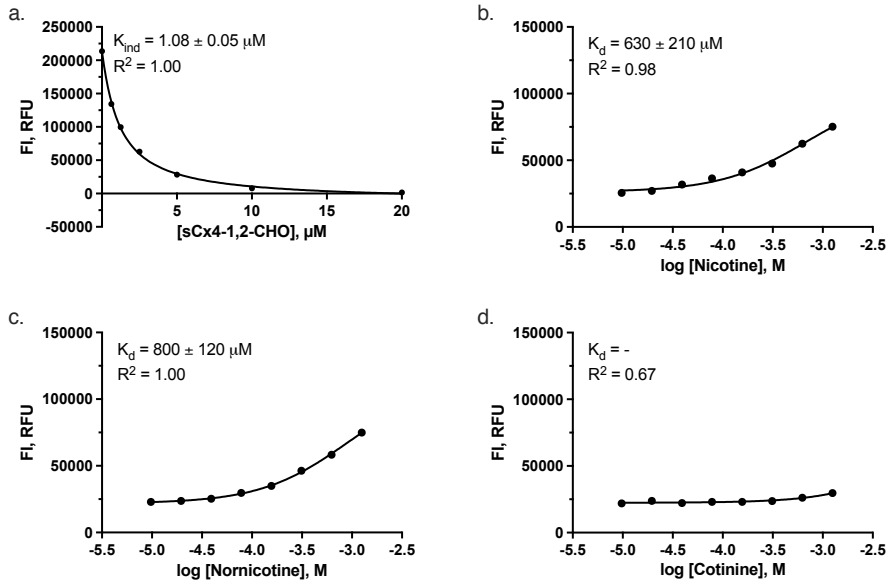


Figure S2.20. **sCx4-1,2-CHO** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-1,2-CHO** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

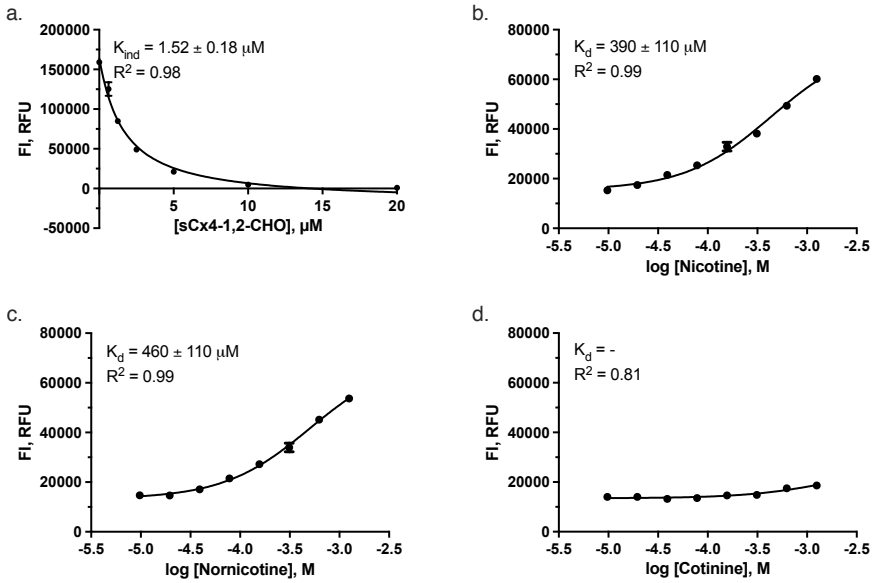


Figure S2.21. **sCx4-1,2-CHO** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-1,2-CHO** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

Fluorescence studies of sCx4-1,3-CHO

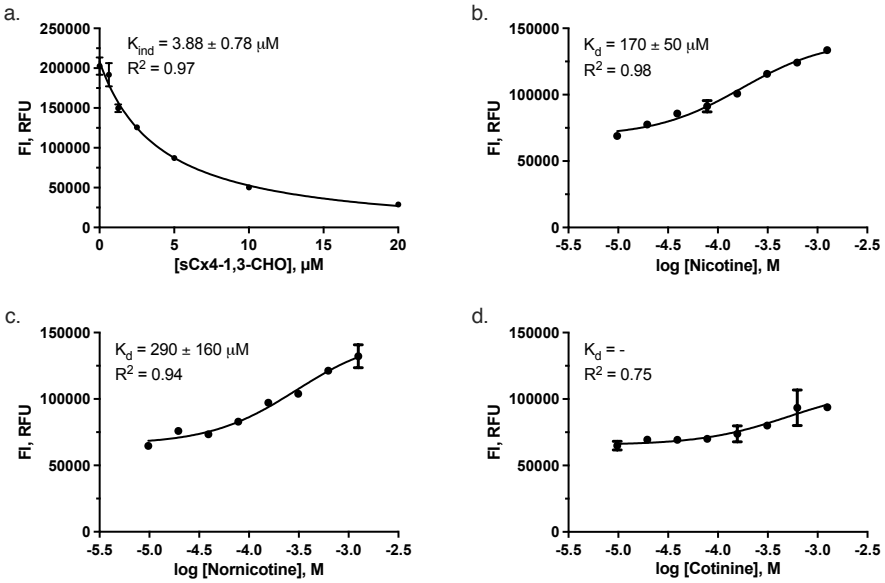


Figure S2.22. **sCx4-1,3-CHO** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-1,3-CHO** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

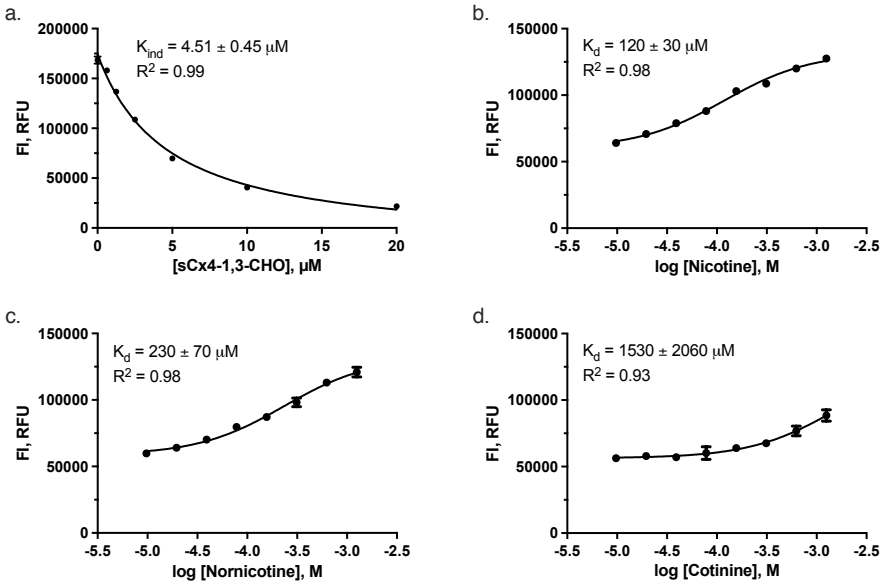


Figure S2.23. **sCx4-1,3-CHO** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-1,3-CHO** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

Fluorescence studies of sCx4-NO_2

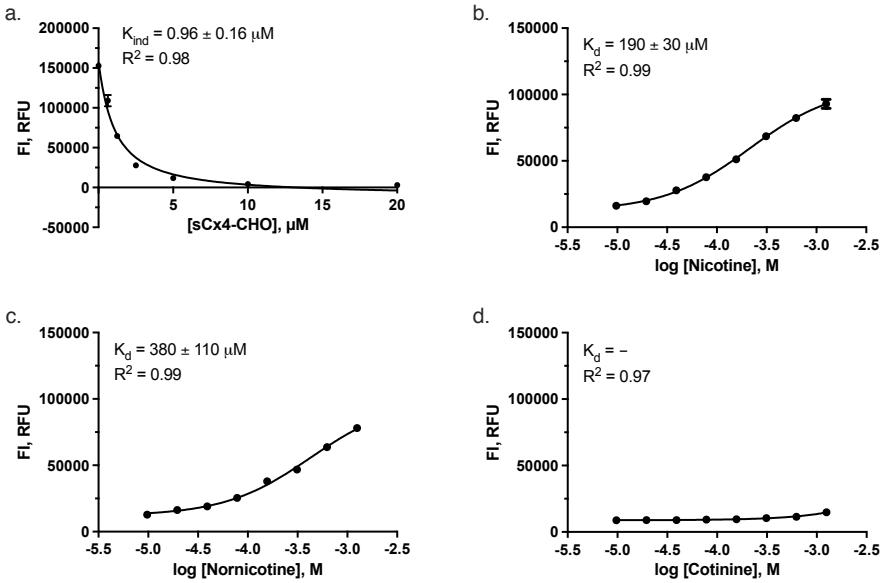


Figure S2.24. **sCx4-NO_2** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-NO_2** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

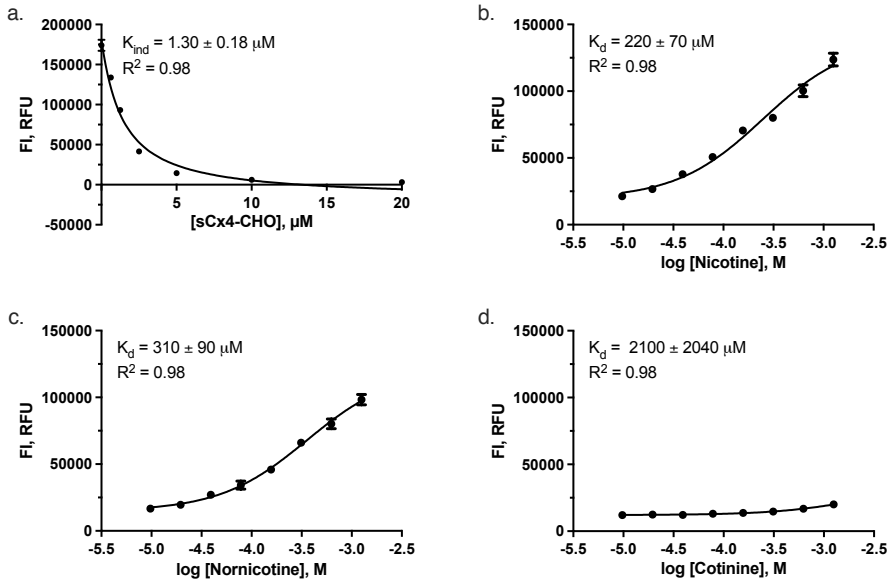


Figure S2.25. **sCx4-NO_2** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-NO_2** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

Fluorescence studies of sCx4-1,2-NO_2

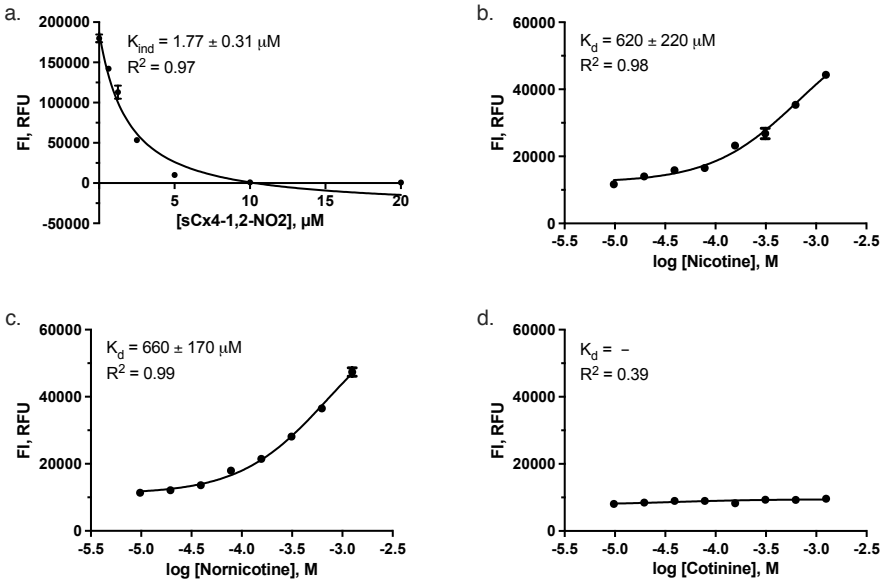


Figure S2.26. **sCx4-1,2-NO_2** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-1,2-NO_2** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

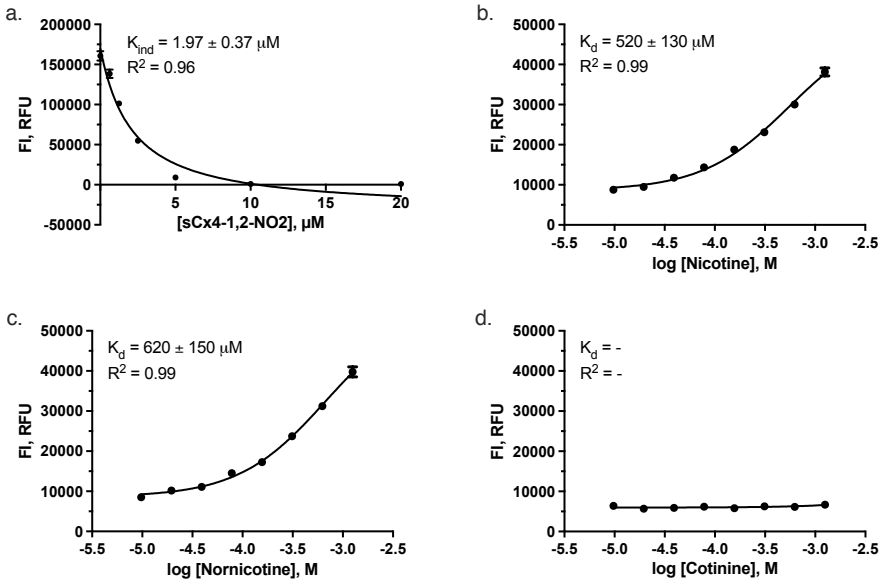


Figure S2.27. **sCx4-1,2-NO_2** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-1,2-NO_2** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

Fluorescence studies of sCx4-1,3-NO_2

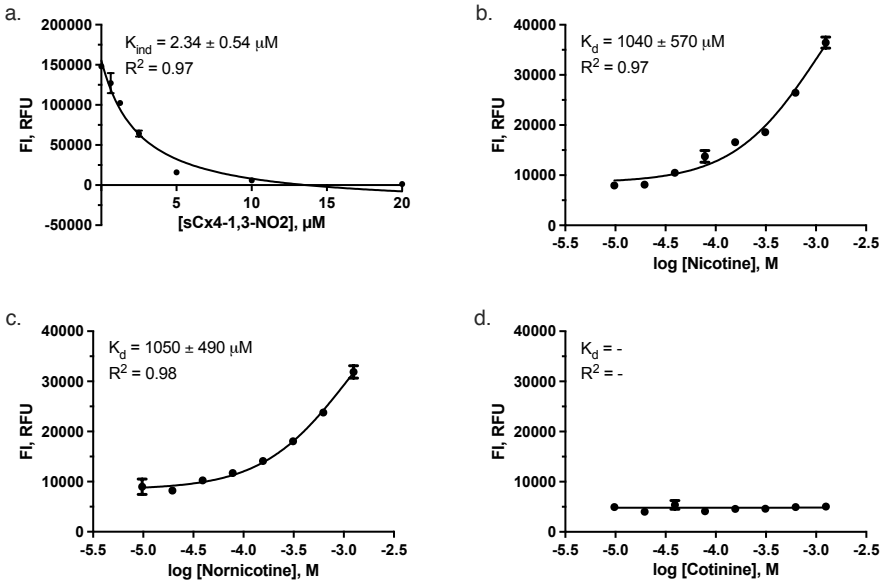


Figure S2.28. **sCx4-1,3-NO_2** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-1,3-NO_2** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

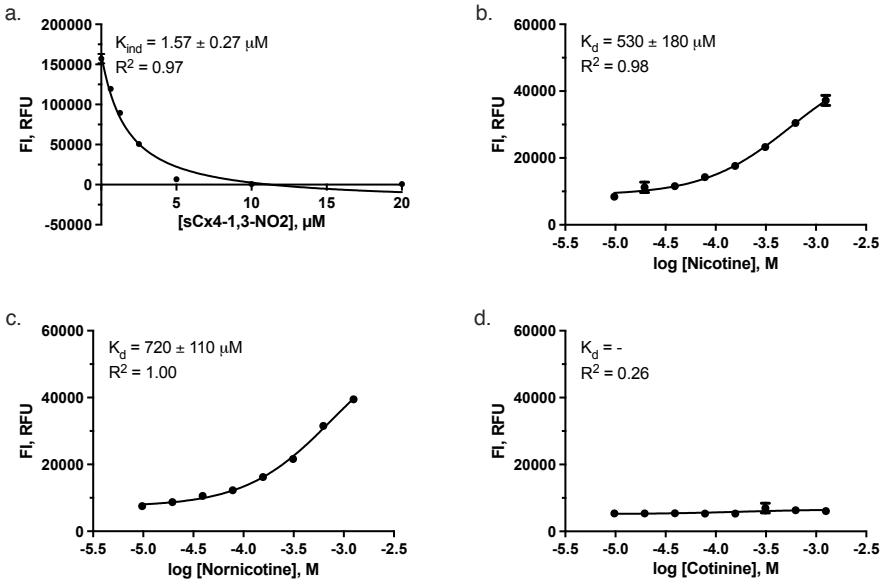


Figure S2.29. **sCx4-1,3-NO_2** with nicotine, nornicotine and cotinine. a) The direct titration of LCG (0.25 μM) with **sCx4-1,3-NO_2** (0 – 20 μM), and the competitive titrations of b) nicotine, c) nornicotine, and d) cotinine titrated into the host-LCG complex (0 – 1.25 mM guest, 5 μM host, 0.25 μM LCG).

2.5.5 ^{1}H-NMR titration between calix[4]arene hosts and nicotine

The NMR samples were prepared in a deuterated 10 mM phosphate buffer, pH 7.4. The NMR titrations were analyzed using the Bindfit method as implemented at http://supramolecular.org. The data was fitted using the NMR 1:1 model using the Nelder-Mead method, with the initial values subtracted. The links to the fitted data are provided, in addition to showing printouts of every replicate titration and its respective fits. When one of the protons overlapped with another NMR signal, that proton was not used to calculate the binding constant. Values reported in the main text are the mean of duplicate measurements. The percent error of the mean binding constant was propagated by equation 1 (below), where 'x' is the errors of the individual binding constant, and 'n' is the number or measurements.

$$error = \frac{\sqrt{(\delta x)^2 + (\delta x)^2}}{\sqrt{(n)}}$$

Equation 2 shows an exemplary error propagation for **sCx4**.

$$error = \frac{\sqrt{(7.004)^2 + (6.226)^2}}{\sqrt{(2)}}$$

^{1}H-NMR titration between sCx4 and nicotine

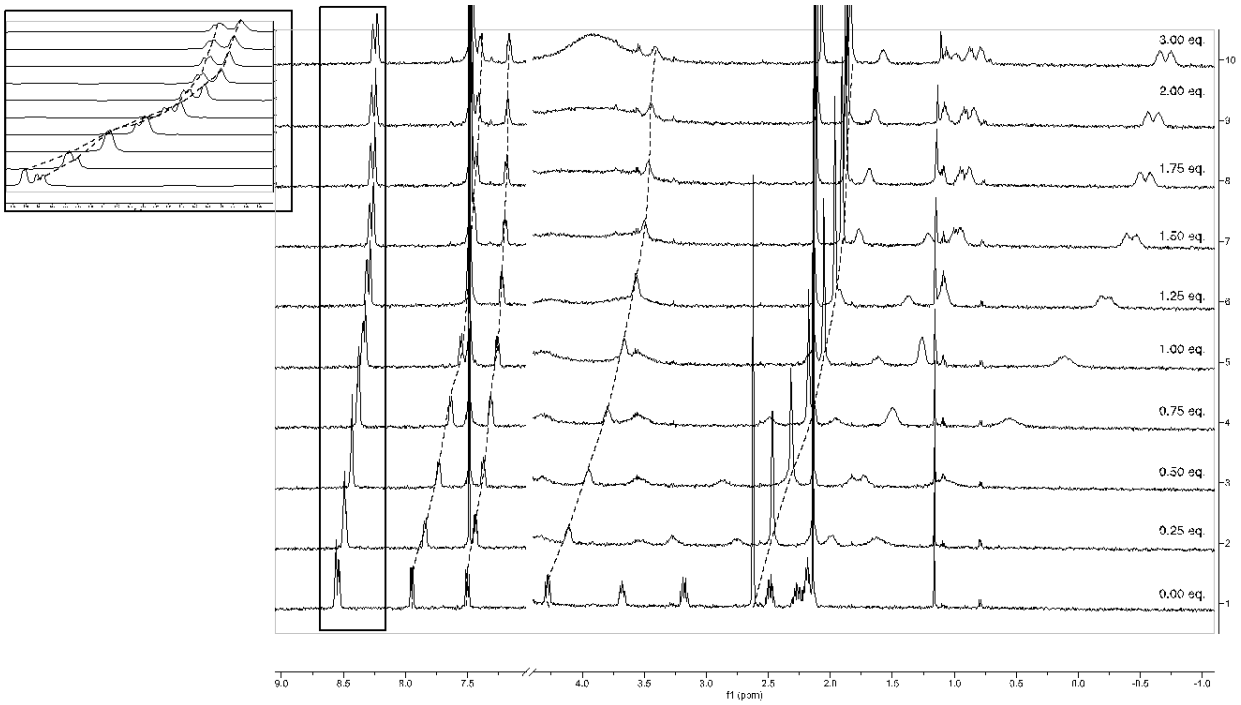


Figure S2.30. ^{1}H-NMR spectrum of nicotine (0.0005 M) titrated with **sCx4** (0–3 equivalents) in deuterated phosphate buffer (0.05 M, pH 7.4), 500 MHz.

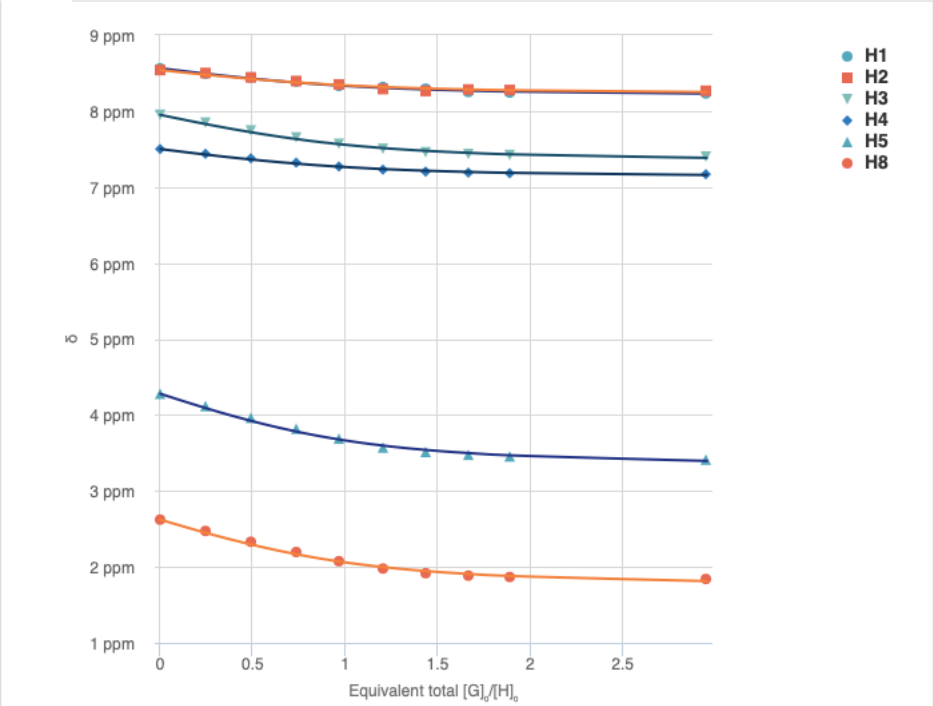


Figure S2.31. Fitted NMR data for **sCx4** by Bindfit. K_a= 9061 M^{-1} ± 7%

http://app.supramolecular.org/bindfit/view/b47b6dc6-8763-4f29-9408-7ad8cdeb9e6a

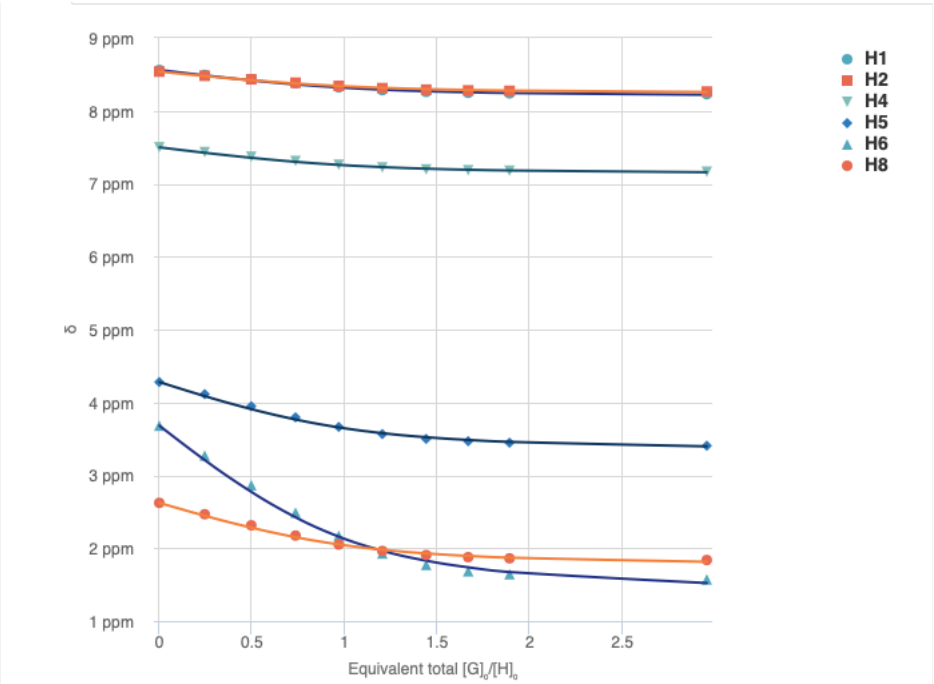


Figure S2.32. Fitted NMR data for **sCx4** by Bindfit. K_a= 11695 M^{-1} ± 6%

http://app.supramolecular.org/bindfit/view/b2479a00-2f46-4823-8fe1-1d693add9ddd

^{1}H-NMR titration between sCx4-CHO and nicotine

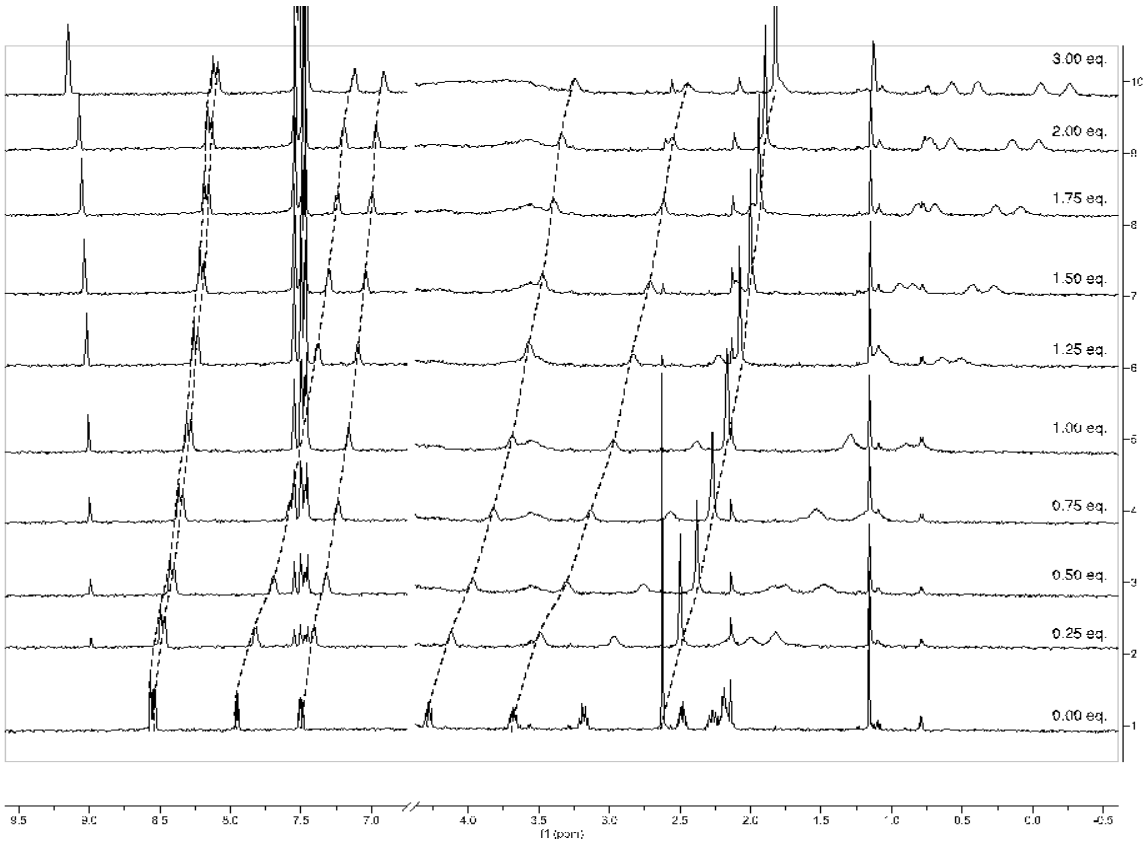


Figure S2.32. ^{1}H-NMR spectrum of nicotine (0.0005 M) titrated with **sCx4-CHO** (0–3 equivalents) in deuterated phosphate buffer (0.05 M, pH 7.4), 500 MHz.

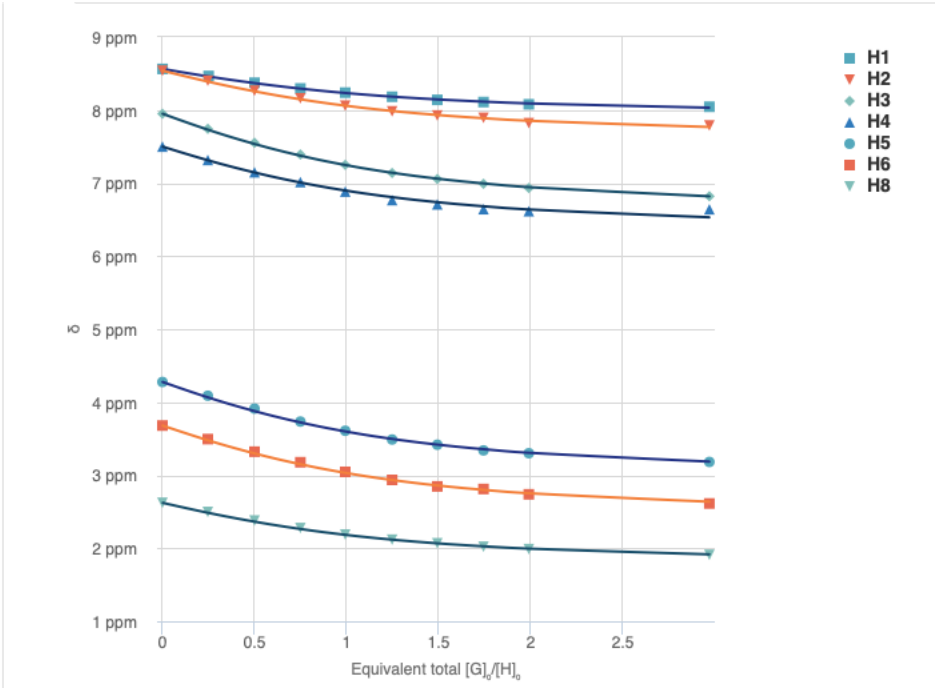


Figure S2.33. Fitted NMR data for **sCx4-CHO** by Bindfit. K_a= 3689 M^{-1} ± 2%

http://app.supramolecular.org/bindfit/view/9e41cbcd-9b5d-41ed-b936-21b22499c297

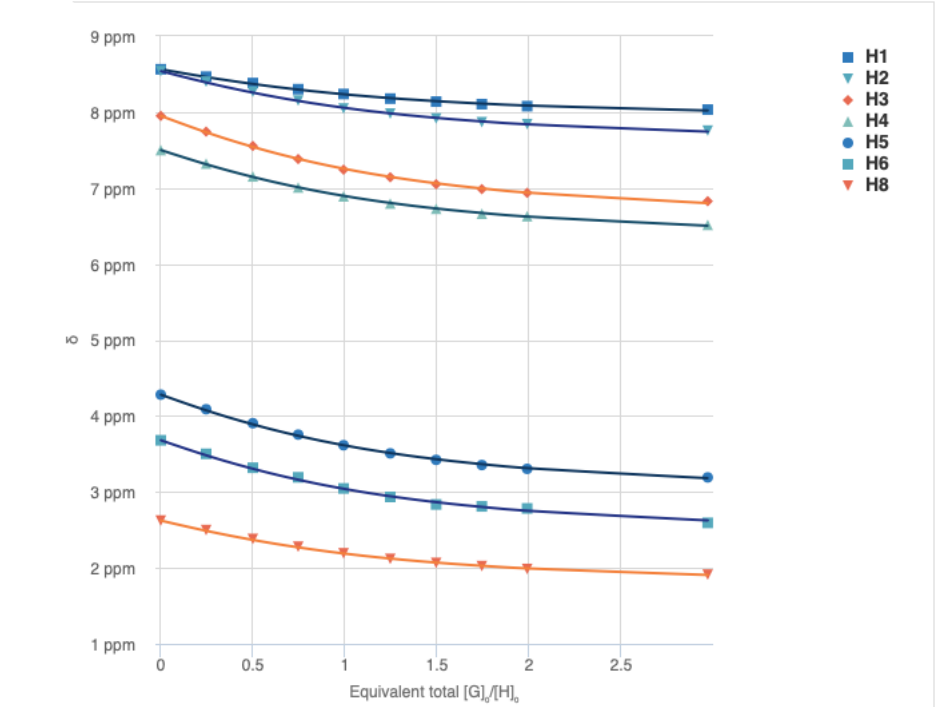


Figure S2.34. Fitted NMR data for **sCx4-CHO** by Bindfit. K_a= 4357 M^{-1} ± 3%

http://app.supramolecular.org/bindfit/view/f0fe1bb5-a6cf-4e1f-a585-b370fcd3fb03

^{1}H-NMR titration between sCx4-NO_2 and nicotine

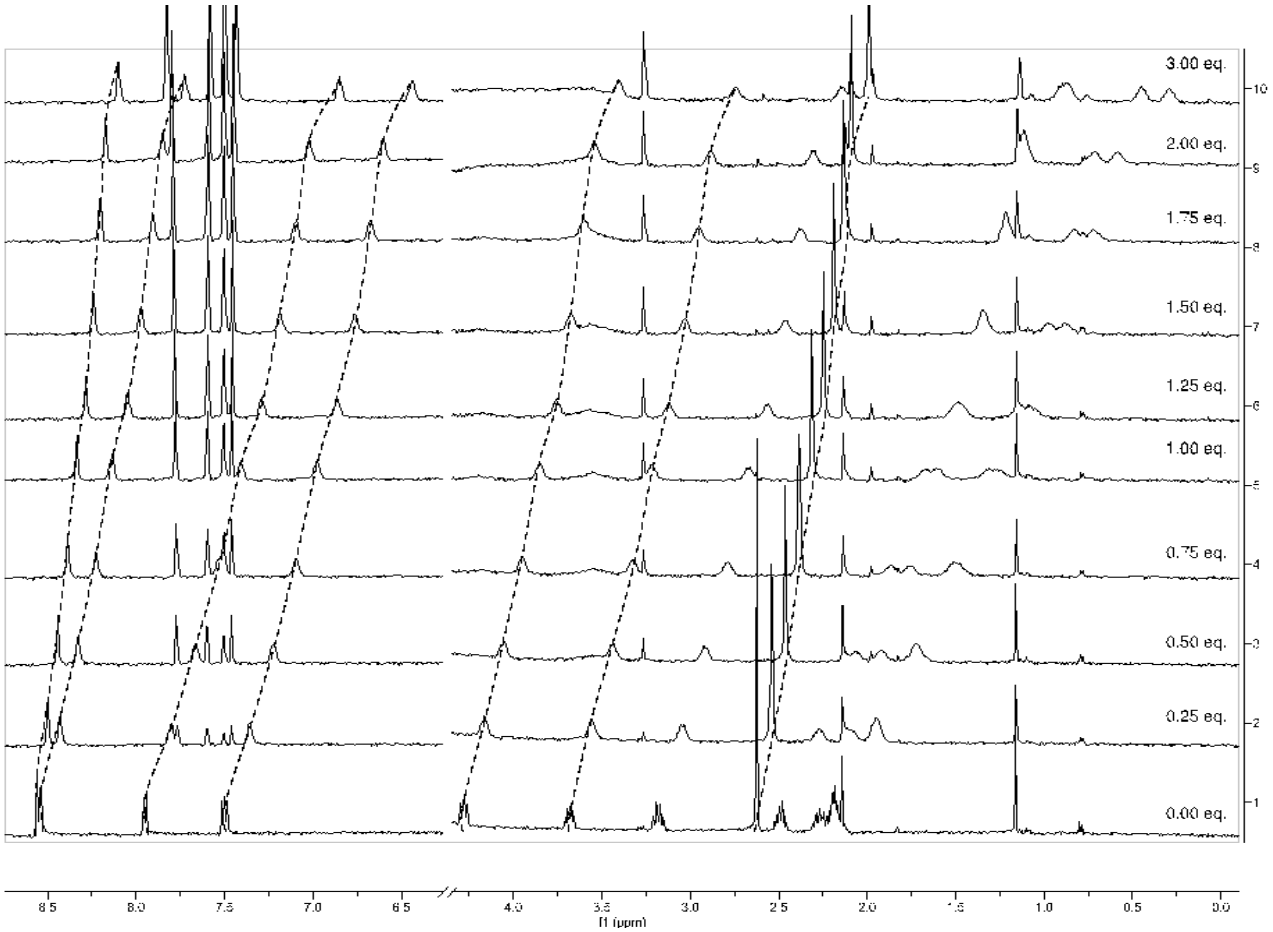


Figure S2.35. ^{1}H-NMR spectrum of nicotine (0.0005 M) titrated with **sCx4-NO_2** (0–3 equivalents) in deuterated phosphate buffer (0.05 M, pH 7.4), 500 MHz.

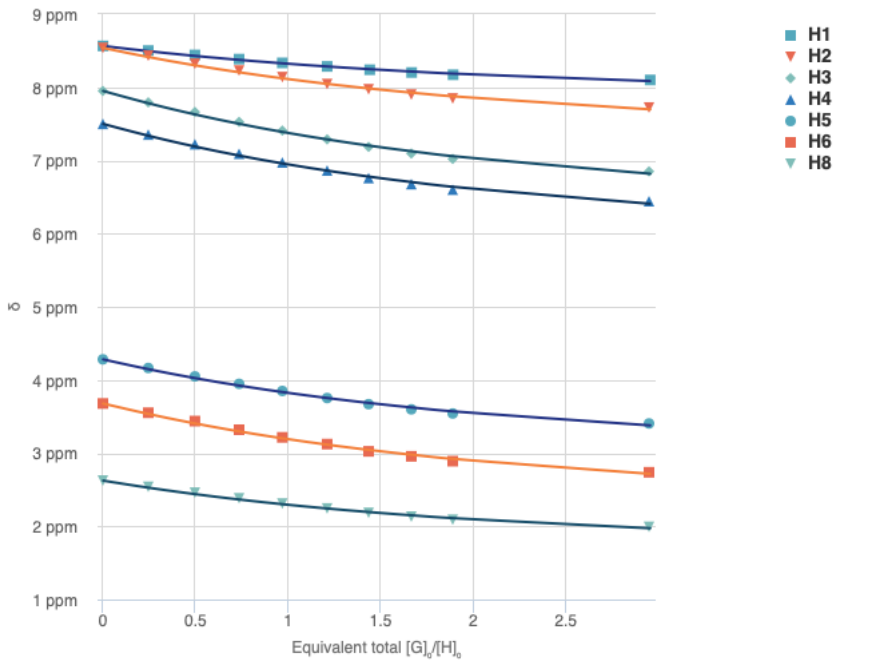


Figure S2.36. Fitted NMR data for **sCx4-NO$_2$** by Bindfit. K_a = 1223 M^{-1} ± 2%

http://app.supramolecular.org/bindfit/view/cc7d02cc-beef-49a7-91f8-9de6901c751d

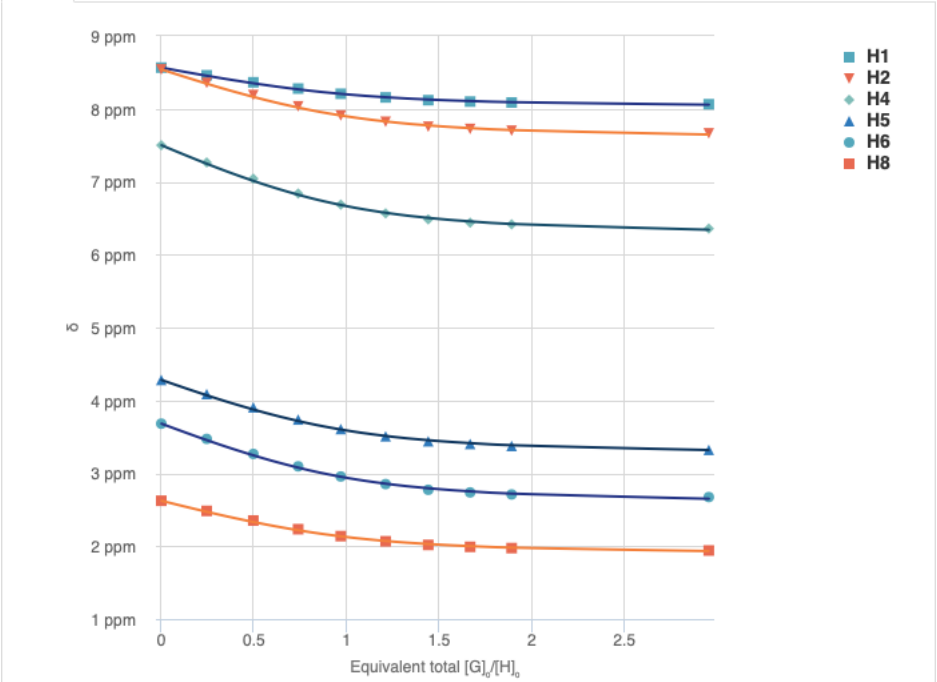


Figure S2.37. Fitted NMR data **for sCx4-NO$_2$** by Bindfit. K_a = 11933 M^{-1} ± 4 %

http://app.supramolecular.org/bindfit/view/15b96050-e269-4b93-a443-0e5e091f75d6

^{1}H-NMR **titration between sCx4-1,3-CHO and nicotine**

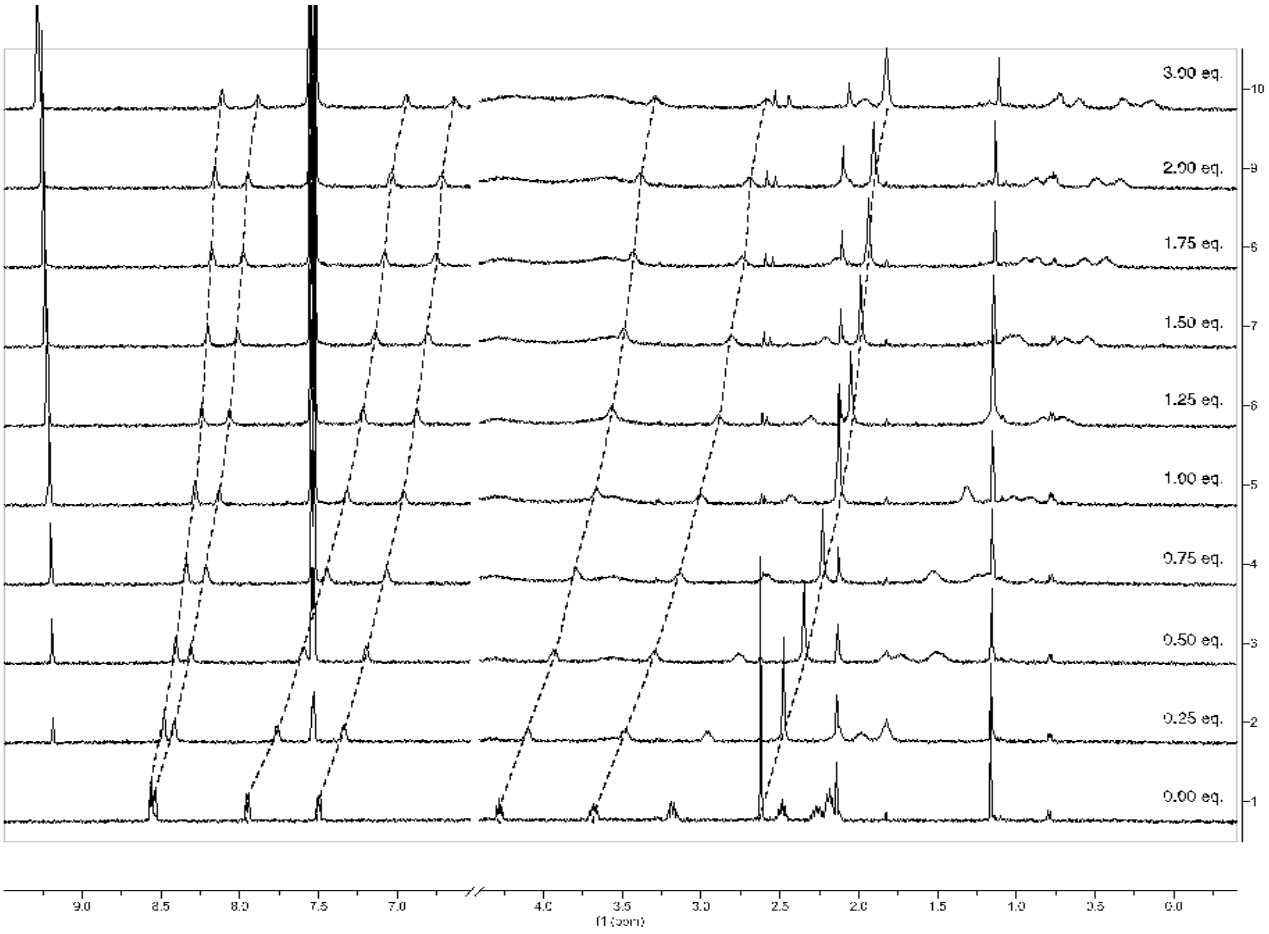


Figure S2.38. ^{1}H-NMR spectrum of nicotine (0.0005 M) titrated with **sCx4-1,3-CHO** (0–3 equivalents) in deuterated phosphate buffer (0.05 M, pH 7.4), 500 MHz.

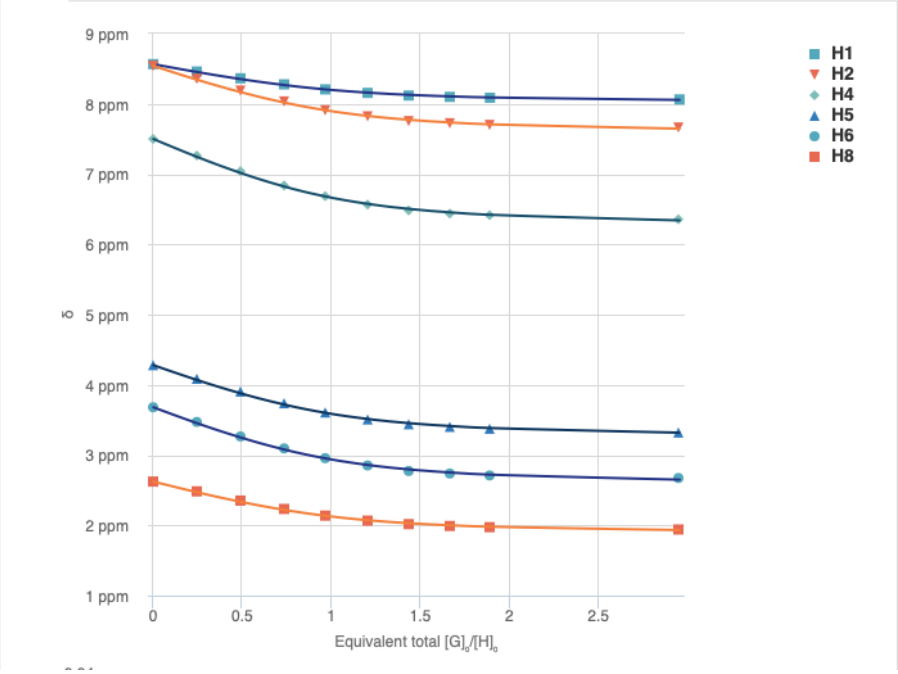


Figure S2.39. Fitted NMR data for **sCx4-1,3-CHO** by Bindfit. K_a = 1223 M^{-1} ± 2%

http://app.supramolecular.org/bindfit/view/e3afe697-a414-4a44-944a-3669ed909ad7

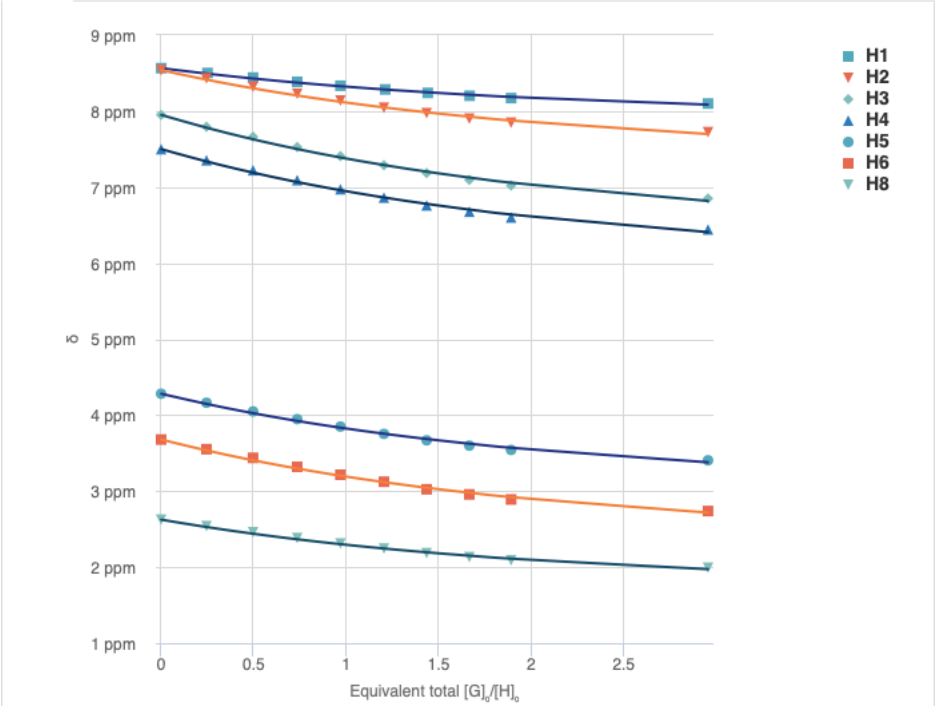


Figure S2.40. Fitted NMR data for **sCx4-1,3-CHO** by Bindfit. K_a= 11933 M^{-1} ± 4%

http://app.supramolecular.org/bindfit/view/b881418b-7039-43be-9693-68dd80ba740b

^{1}H-NMR titration between sCx4-1,3-NO_2 and nicotine

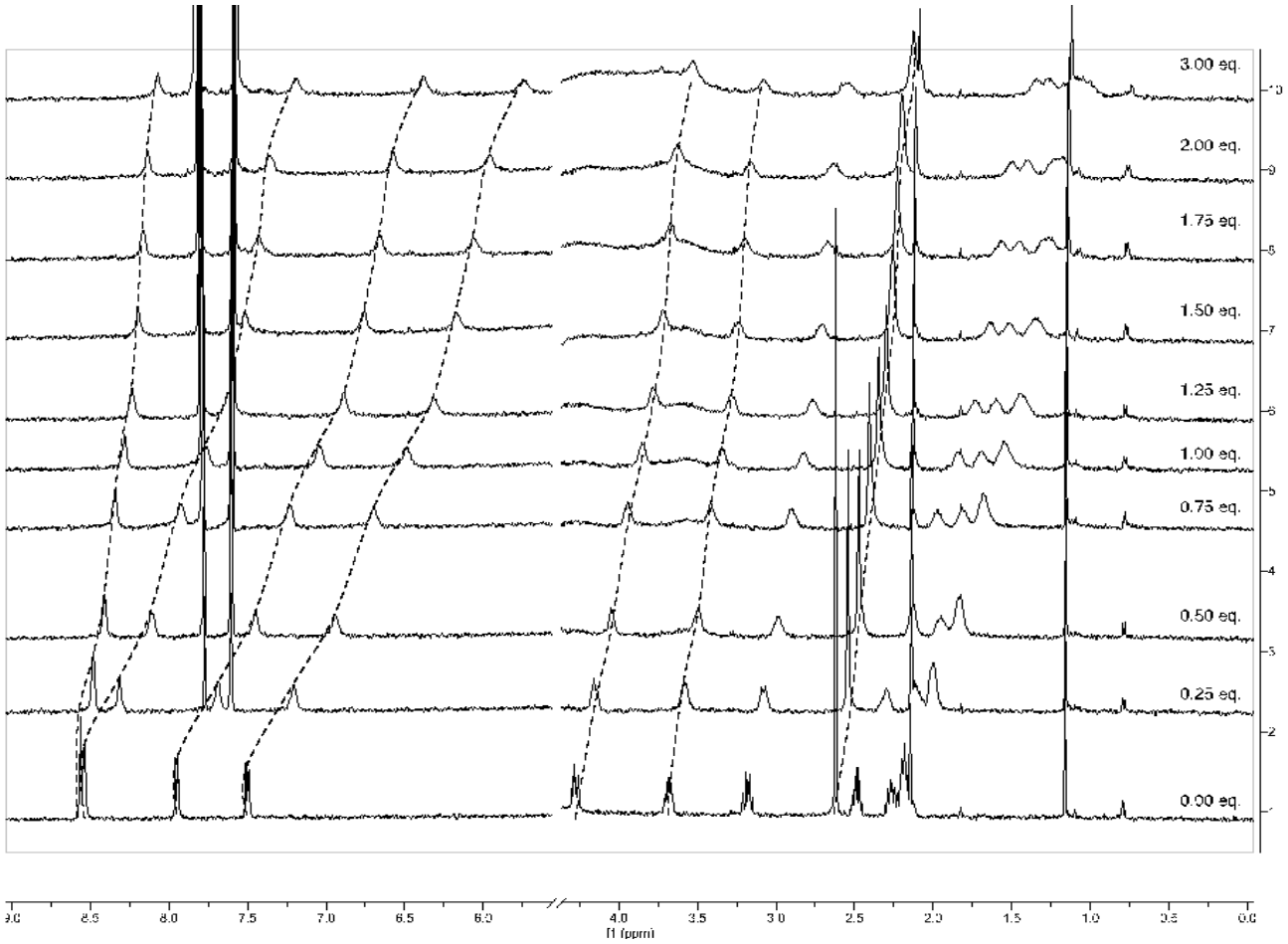


Figure S2.41. ^{1}H-NMR spectrum of nicotine (0.0005 M) titrated with **sCx4-1,3-NO_2** (0–3 equivalents) in deuterated phosphate buffer (0.05 M, pH 7.4), 500 MHz.

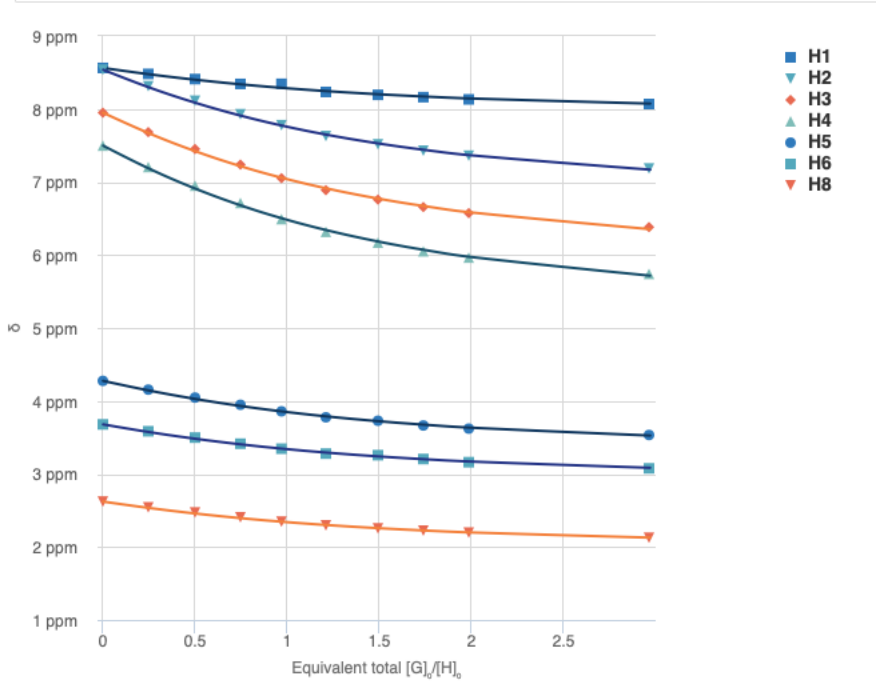


Figure S2.43. Fitted NMR data for **sCx4-1,3-NO_2** by Bindfit. K_a= 2627 M^{-1} ± 1%

http://app.supramolecular.org/bindfit/view/01575fe5-4fd6-4303-8783-6cb41e31e619

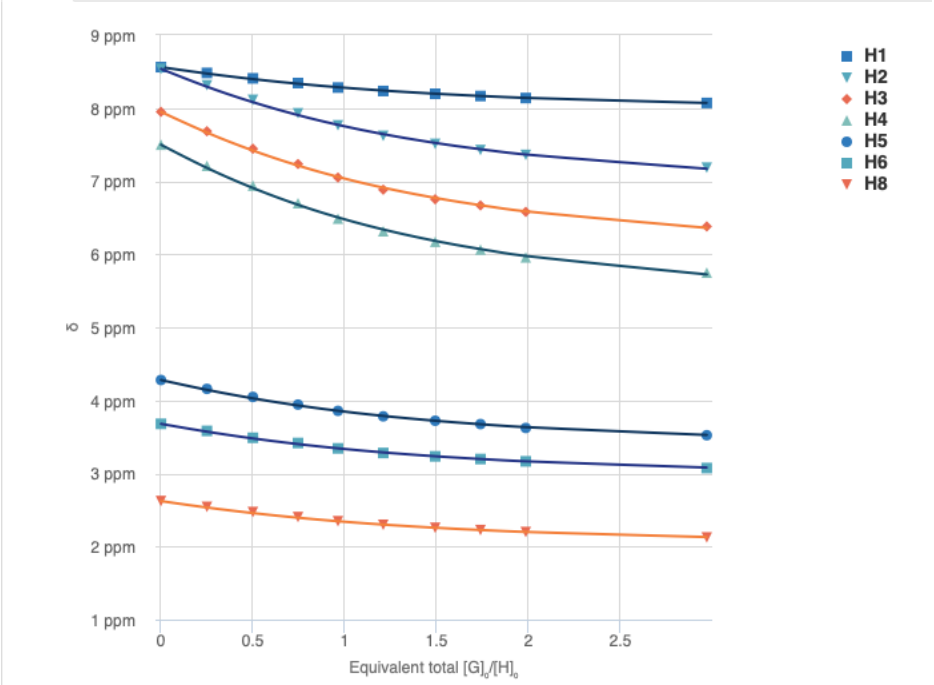


Figure S2.44. Fitted NMR data for **sCx4-1,3-NO_2** by Bindfit. K_a= 2504 M^{-1} ± 2 %

http://app.supramolecular.org/bindfit/view/88df5a0b-c365-4c67-89a2-5ae7e024e813

^{1}H-NMR **titration between sCx4-1,2-CHO and nicotine**

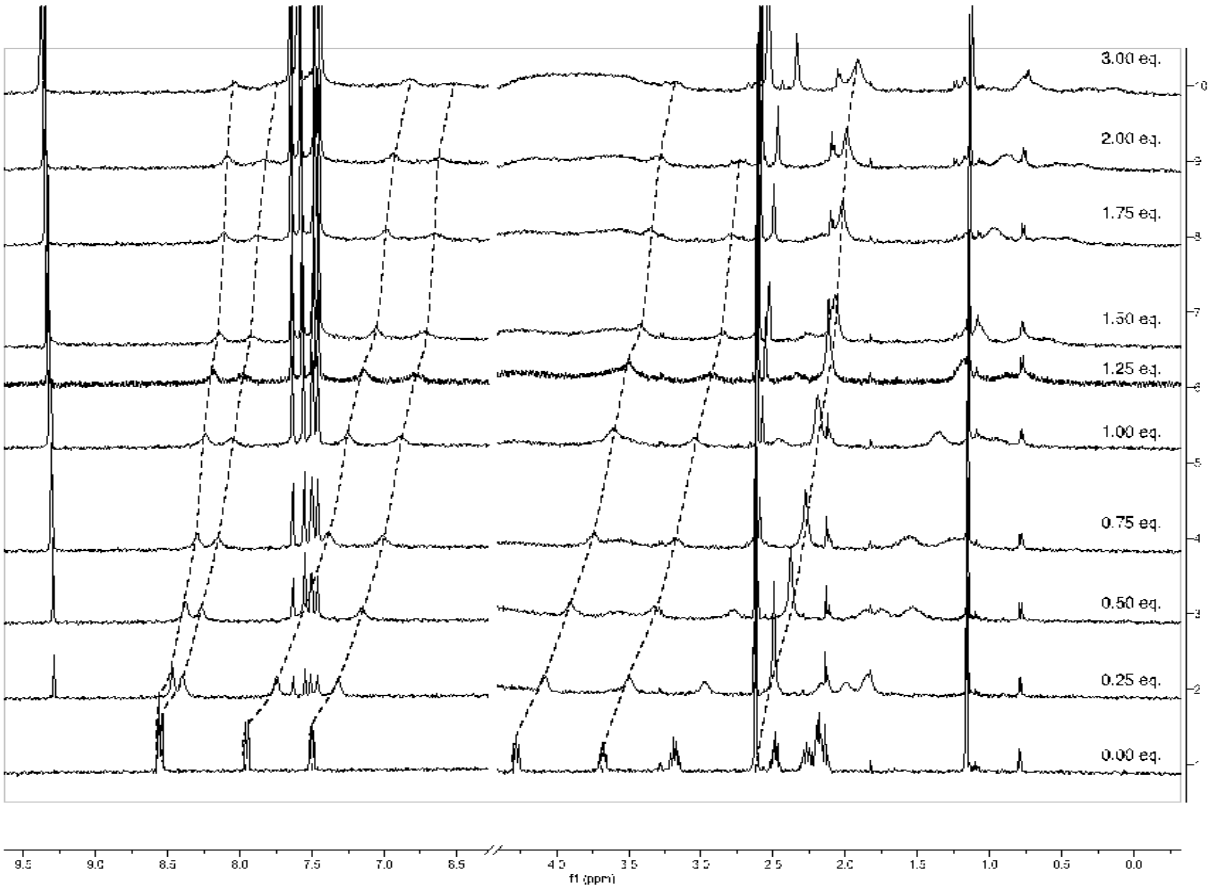


Figure S2.44. ^{1}H-NMR spectrum of nicotine (0.0005 M) titrated with **sCx4-1,2-CHO** (0–3 equivalents) in deuterated phosphate buffer (0.05 M, pH 7.4), 500 MHz.

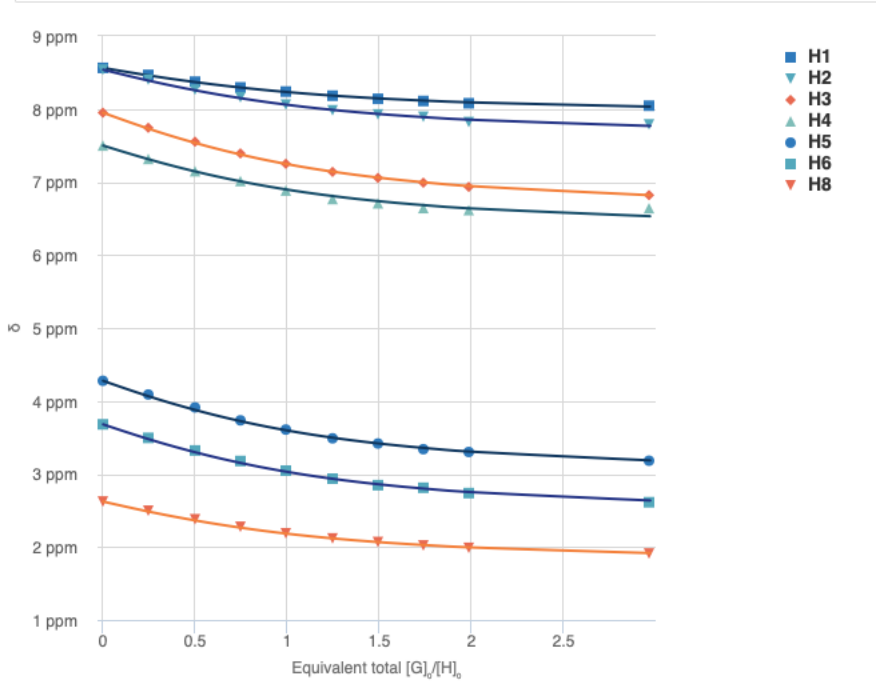


Figure S2.45. Fitted NMR data for **sCx4-1,2-CHO** by Bindfit. K_a= 3689 M^{-1} ± 2%

http://app.supramolecular.org/bindfit/view/9e41cbcd-9b5d-41ed-b936-21b22499c297

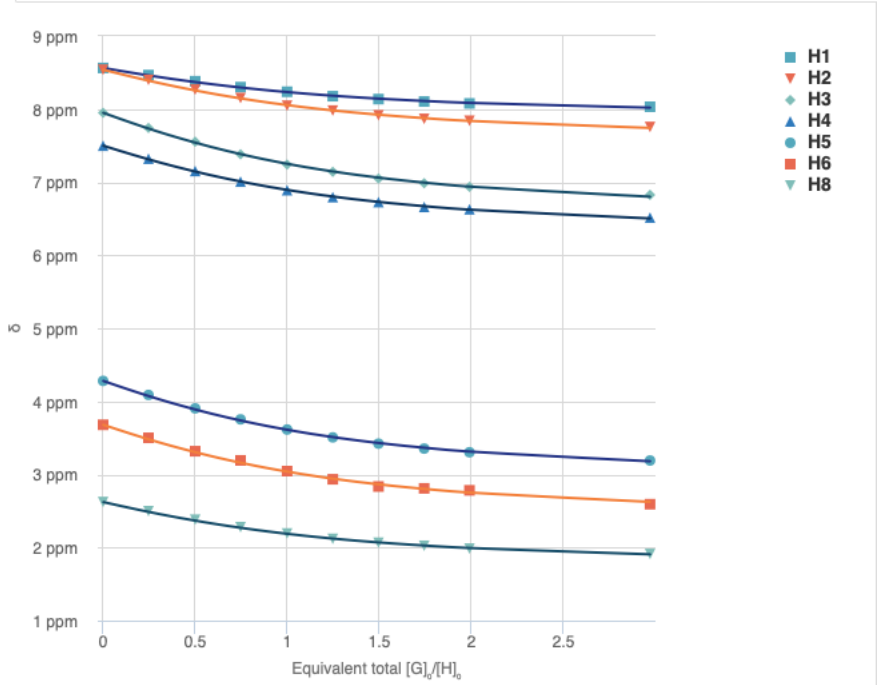


Figure S2.46. Fitted NMR data for **sCx4-1,2-CHO** by Bindfit. K_a= 4357 M^{-1} ± 3%

http://app.supramolecular.org/bindfit/view/f0fe1bb5-a6cf-4e1f-a585-b370fcd3fb03

^{1}H-NMR **titration between sCx4-1,2-NO$_2$ and nicotine**

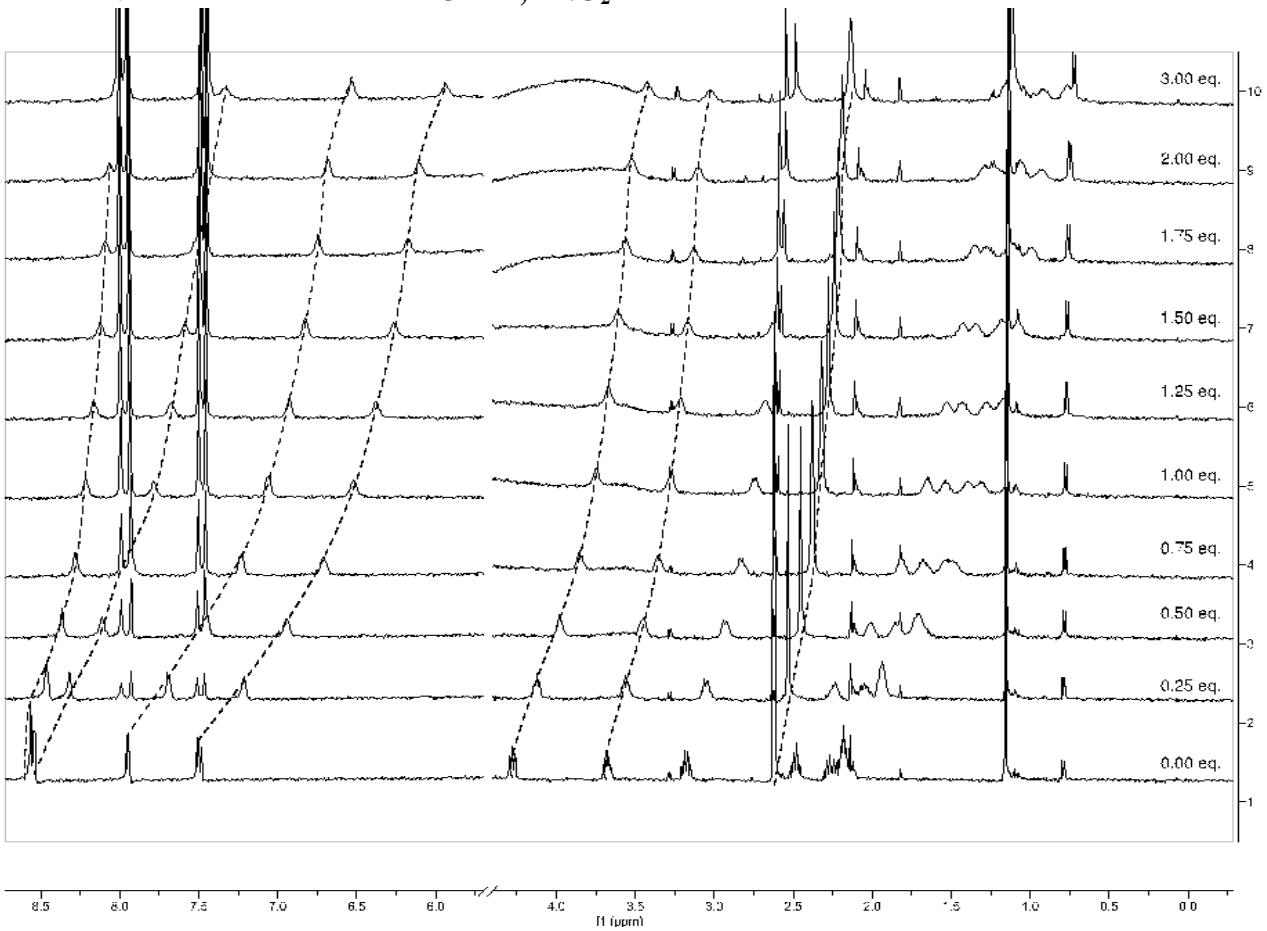


Figure S2.47. ^{1}H-NMR spectrum of nicotine (0.0005 M) titrated with **sCx4-1,2-NO$_2$** (0–3 equivalents) in deuterated phosphate buffer (0.05 M, pH 7.4), 500 MHz

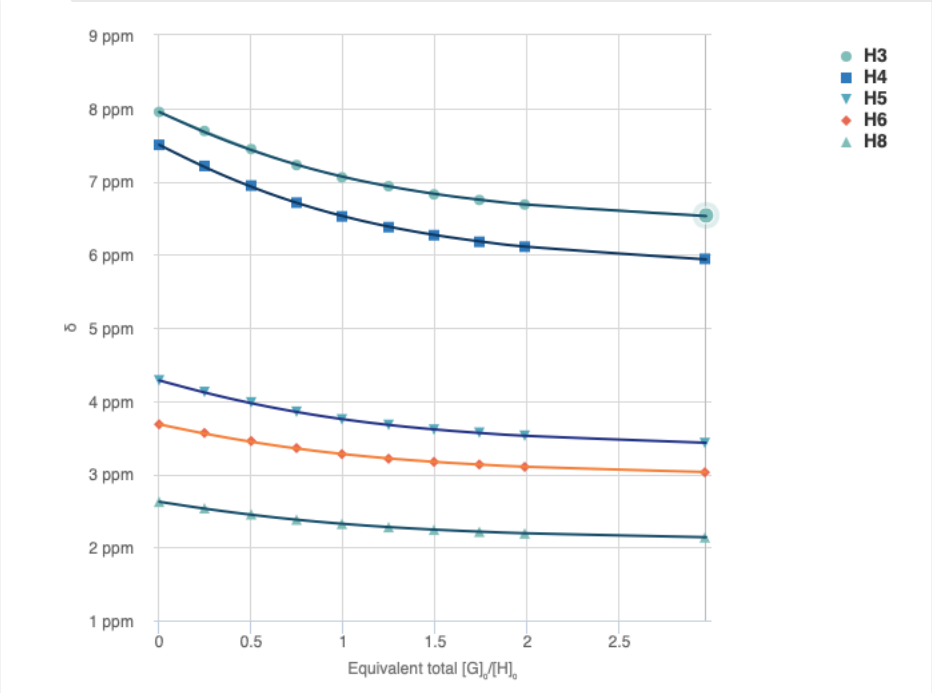


Figure S2.48. Fitted NMR data for **sCx4-1,2-NO_2** by Bindfit. K_a= 3855 M^{-1} ± 1%

http://app.supramolecular.org/bindfit/view/b78dfc30-c67f-41f9-8bc2-bd7dc71ff9a9

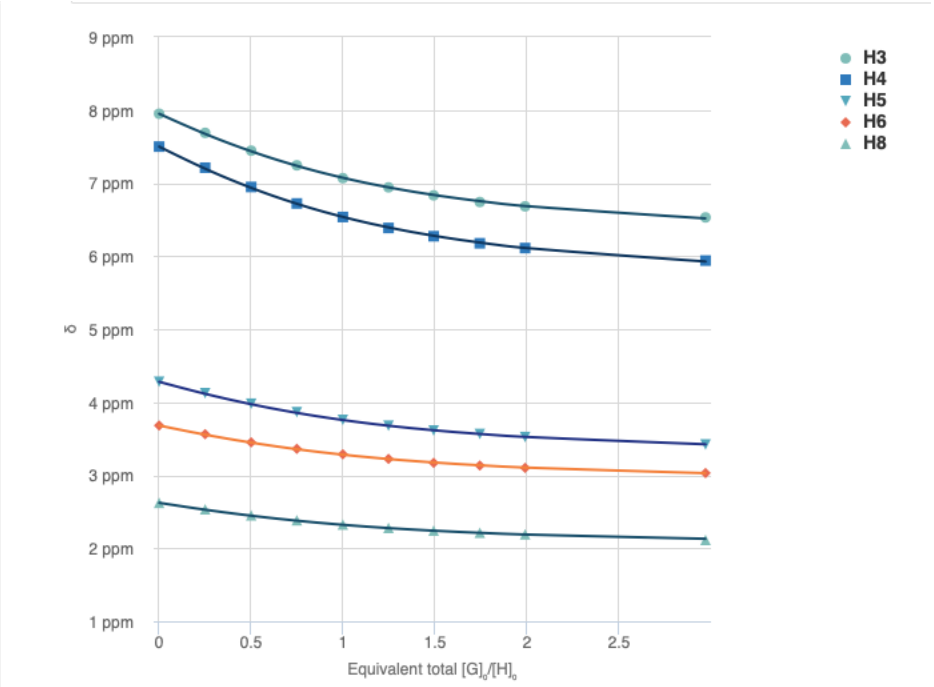


Figure S2.49.Fitted NMR data for **sCx4-1,2-NO_2** by Bindfit. K_a= 4310 M^{-1} ± 1 %

http://app.supramolecular.org/bindfit/view/ed8fa055-2c54-4fba-b9e2-891e82930b40

Chapter 3: Binding methylarginines and methyllysines as free amino acids: a comparative study of multiple host classes

This work was adapted from a previously published paper.

Z. Warmerdam, B. E. Kamba, M. H. Le, T. Schrader, L. Isaacs, P. Bayer, F. Hof, *ChemBioChem*, **2022**, 23, e202100502

This work was accomplished in collaboration between three research groups. Dr. Lyle Isaacs (University of Maryland, United States) supplied compounds **M1** and **M2**. Drs. Thomas Schrader and Peter Bayer (Universität Duisburg Essen, Germany) and their students Bianca E. Kamba and My-Hue Le supplied compounds **PC** and **CLR01**. Bianca E. Kamba developed the IDA assays for **PC** and **CLR01**, IDAs for the other hosts were previously established. Zoey Warmerdam and Bianca E. Kamba both performed IDA studies on the all the host-guest combinations. In this chapter the data for **M1**, **M2**, **PC** and **CLR01** was collected by Zoey Warmerdam, the data for the calix[4]arene hosts was collected by Bianca E. Kamba. Zoey Warmerdam analyzed all the data sets and drew conclusions from them. My-Hue Le did the computational modelling for host-guest complexes. Zoey Warmerdam wrote the paper.

3.1 Foreword

Chapter 3 progresses the work from Chapter 2, this time to study a different group of small molecule guests — methylated amino acids. We used the same calix[4]arene host library, but also collaborated with two other research groups who both provided their own functionalized host molecules. During this study we able to compare the different host

classes and draw conclusions that go beyond the host functionalization and towards a more holistic view of the impact of host geometries.

3.2 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 compared to other amino acids. 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.[13]

There are dozens of diseases caused by disordered amino acid metabolism.[114,115] 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.[116,117] Antibodies are known that bind amino acids but have some inherent shortcomings[118-121] 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 Basílio et al.[122] Examples of hosts binding to free

amino acids include calixarenes[111,123,124], pillararenes[125,126], cucurbiturils[127-129], and cyclodextrins.[130]

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.[131] We chose a parent anionic calixarene, **sCx4**, and the functionalized analogues **sCx4-NO_2** and **sCx4-CHO** (Figure 3.1a-c) to include in this study after a preliminary screen ruled out some weaker-binding calixarenes (**sCx4-1,2-CHO, sCx4-1,3-CHO, sCx4-1,2-NO_2** and **sCx4-1,3-NO_2** see supporting information). 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 3.1d, e) can be functionalized along the edge of their binding surfaces, and are increasingly being used in biomedical applications.[132] Another more rigid cleft-like host family includes the 'tweezers' and 'clips' introduced by Klärner and Schrader.[133,134] 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 3.1f, g).

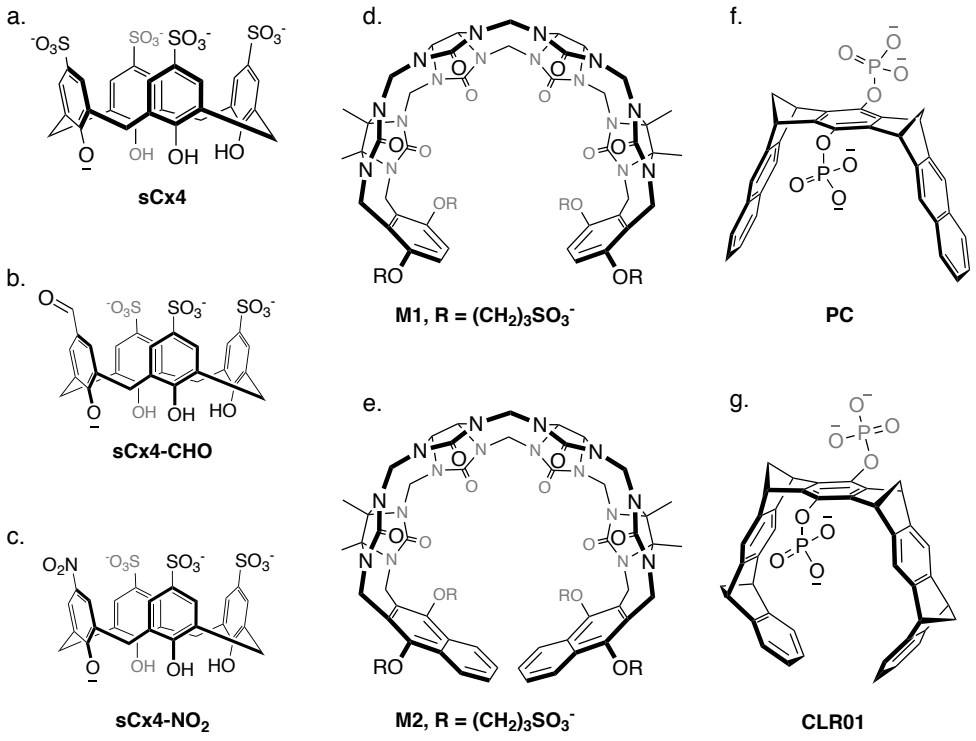


Figure 3.1. Hosts studied in this report. a) **sCx4**. b) **sCx4-CHO**. c) **sCx4-NO_2**. d) **M1**. e) **M2**. f) **PC**. g) **CLR01**.

For this fundamental, comparative study we have selected methylated amino acids as binding targets. Our test set (Figure 3.2) 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)). The amino acids have subtle structural differences among them,[135] 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,[136,137] renal disease,[138-140] and

others.[141] 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 ADMA plasma concentration for disease conditions as high as 5-7 μM).[142,143] SDMA plasma levels are lower (<1 μM).[144] 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.

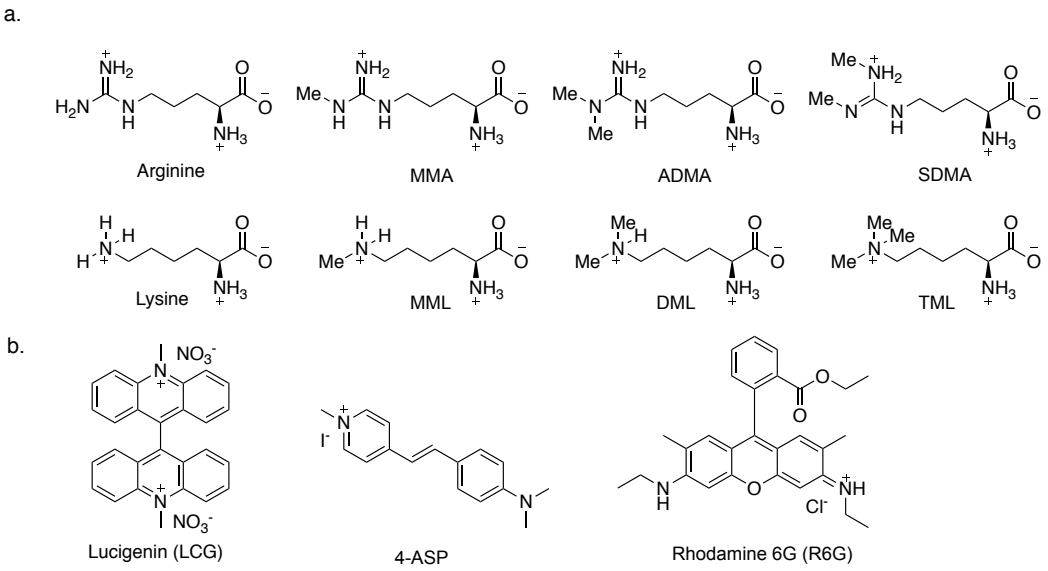


Figure 3.2. Guests and indicators studied in this report. a) Guests' arginine, MMA, ADMA, SDMA, Lysine, MML, DML, and TML, b) Indicators LCG, 4-ASP, and R6G

A considerable amount of research has been focused on binding methylated lysines,[59,111,145,146] while the selective binding of methylated arginine has received relatively less attention.[60,63] 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[136,147-150] demonstrating that some of the free methylated

amino acids are metabolites that play critical biological roles in pathways that are distinct from those involving proteins and peptides with post-translationally methylated residues.

3.3 Results and Discussion

Indicator displacement assays (IDAs) were used to determine the affinities between the different host-guest combinations. We adapted previously reported IDAs[151] for **sCx4** (using lucigenin, Figure 3.2b)[152] and **M1** and **M2** (using Rhodamine 6G, Figure 3.2b).[153,154] 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 3.2b) 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 equilibrium dissociation constants (K_d) 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 phosphate 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 3.1), which we categorize as follows for convenient presentation: strong binding (K_d < 200 μM), medium to strong binding (K_d 200–1500 μM) and weak to no binding (K_d > 1500 μM).

We also presented the data as a 3D bar graph (Figure 3.3), to help visualise the binding trends.

Table 3.1. K_d values determined by IDA for each host-guest complex in 10 mM phosphate buffer.[a]

	sCx4[b]	**sCx4-CHO** [b]	**sCx4-NO_2** [b]	**M1**[c]	**M2**[c]	**CLR01** [d]	**PC** [d]
	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)
Arginine	>1500	>1500	>1500	>1500	>1500	75 ± 20	1100 ± 630
MMA	>1500	>1500	>1500	45 ± 10	>1500	135 ± 30	1230 ± 540
ADMA	500 ± 50	230 ± 50	290 ± 60	10 ± 5	60 ± 20[e]	>1500	650 ± 130
SDMA	1090 ± 300	370 ± 150	>1500	35 ± 10	20 ± 5	>1500	810 ± 270
Lysine	>1500	>1500	>1500	>1500	>1500	30 ± 5	>1500
MML	930 ± 120	470 ± 110	610 ± 130	160 ± 60	>1500	15 ± 5	>1500
DML	300 ± 40	240 ± 50	410 ± 60	70 ± 10	130 ± 90[e,f]	30 ± 5	980 ± 710
TML	120 ± 20	100 ± 30	150 ± 30	15 ± 5	12 ± 3	65 ±20	1040 ± 360

[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] $R^2 \geq 0.92$. [f] Value arises from one triplicate.

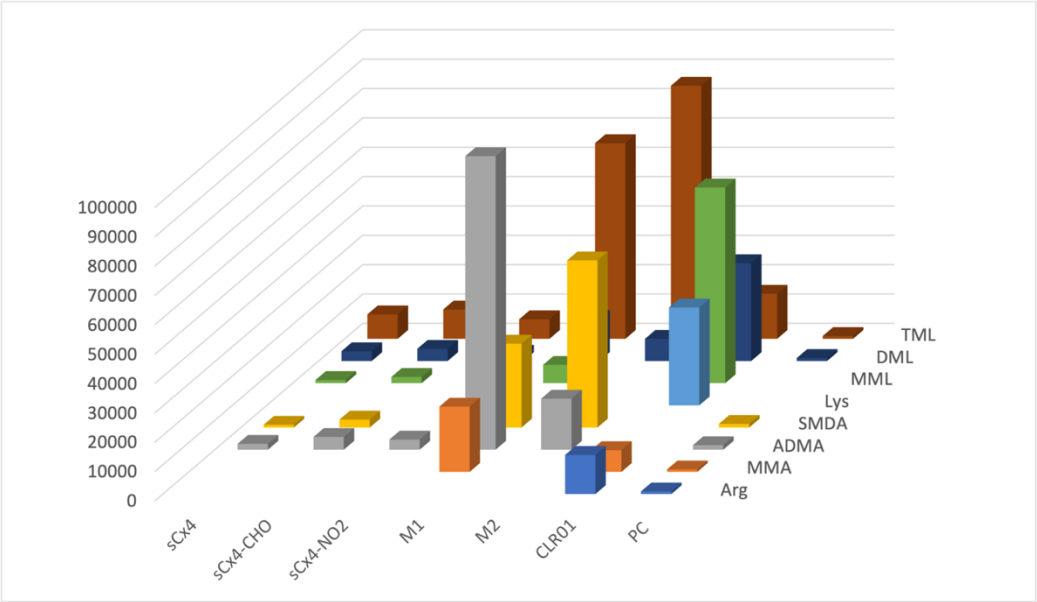


Figure 3.3. 3D bar graph to visualize the binding trends, represented as the Ka values in units M^{-1} ($K_a = 1/K_d$). See Table 3.1 for experimental details.

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, 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's conditions. Comparing the symmetric and asymmetric isomers of dimethylarginine, ADMA displays a ~2-fold stronger binding. Interestingly, **sCx4-NO$_2$** 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.

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 ($K_d = 10\ \mu M$) over MMA ($K_d = 45\ \mu M$) and SDMA ($K_d = 35\ \mu$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 μM) over DML (K_d = 70 μM) and MML (K_d = 160 μM). **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.[155] Among the stronger binding guests, **M2** has a >3-fold selectivity for SDMA (K_d = 20 μM) over ADMA (K_d = 60 μM), and a >10-fold selectivity for TML (K_d = 12 μM) over DML (K_d = 130 μM).

Despite their functional group similarity, **PC** and **CLR01** have different behaviours. **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. 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 μM binding of each unmethylated amino acid.[156] Lysine binds with a K_d = 30 μM, and the slightly more hydrophobic MML binds more strongly at K_d = 15 μM. Besides this one exception, each other amino acid binds progressively weaker with increasing methylation. Unlike all other hosts, dimethylarginines are not measurably bound by **CLR01** under the conditions of the experiment.

The salt concentration of the IDA has an influence on the weaker binding calix[4]arenes, where a higher salt concentration interrupts the complexation with LCG.[105,157-159] To determine the effect of a higher salt concentration in the stronger binding host, IDAs on the amino acids were also performed in a more concentrated 50 mM

phosphate buffer with: **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 remains in the "strong" range of K_d values, see Table 3.2.

Table 3.2. 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 (μM)	K_d (μM)	K_d (μM)
Arginine	>1500	>1500	90 ± 25
MMA	180 ± 60	>1500	155 ± 30
ADMA	30 ± 10	125 ± 45	>1500
SDMA	55 ± 15	40 ± 10	>1500
Lysine	>1500	>1500	50 ± 15
MML	300 ± 150	>1500	35 ± 5
DML	210 ± 50	190 ± 110[d]	55 ± 10
TML	30 ± 5	25 ± 5	40 ± 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] $R^2 \geq 0.93$.

Studies with a small set of simpler guests show how the impact of hydrophobicity manifests differently in the different host classes. We performed IDA studies with dimethylammonium chloride, tetramethylammonium chloride, and benzyl trimethylammonium chloride as controls that display increasing hydrophobicity without

neighbouring polar functionality and with relatively simple changes in size and shape (see supporting information). Dimethylammonium is the smallest, least hydrophobic, resulting in weak binding ($K_d \geq 1$ mM) across the set of hosts. In shallow hosts like the calixarenes and **PC**, tetramethylammonium generally binds in the range of K_d = 300–900 μM, with approximately the same affinity as benzyl trimethylammonium in each case. The deeper hosts **M1**, **M2**, and **CLR01** show clear preference for benzyl trimethylammonium over tetramethylammonium, as would be expected from the guest's increased size and hydrophobicity. We rationalise that this difference in trends for benzyl trimethylammonium arises because the shallow hosts do not have any additional hydrophobic surface area available to interact with the larger guest.

The hosts studied in this chapter (except **sCx4-CHO**, **sCx4-NO$_2$** and **PC**) have previously been studied via ITC with lysine and arginine.[111,160,161] From those studies, we see that enthalpy is the dominant driving force across all hosts (see supporting information). ITC studies for **sCx4** with methyl lysine 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 performed in salty aqueous solution, because salt effects are often large and unpredictable contributors to the heats of binding.[162]

We performed molecular modelling for six of the hosts to provide a general view of the differences and similarities between the hosts for certain key complexes. Molecular modelling was completed using minimization in explicit water for all the indicated

complexes using Maestro (Schrödinger, Inc). Host **sCx4-CHO**, **sCx4-NO$_2$**, **M1**, **M2**, **PC**, **CLR01** were modelled in complex with ADMA (Figure 3.4) and hosts **M1**, **M2**, **PC**, **CLR01** were also modelled each in complex with MML and TML (see 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**, and **CLR01**). They also reveal differences in host-guest interactions among the different host classes. The connections between these structural features and guest-binding selectivity 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,[111] where the tweezer-type host do have a precedent as strong binding host.[156,161] 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.[163] 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_d 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 interactions in the complexes of **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 3.4c-d). The calixarenes also have multiple sulfonate groups, and in these complexes host-guest salt bridges are prominent features of the complexes (Figure 3.4a-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 (Figure 3.4e-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.[163] 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. 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.

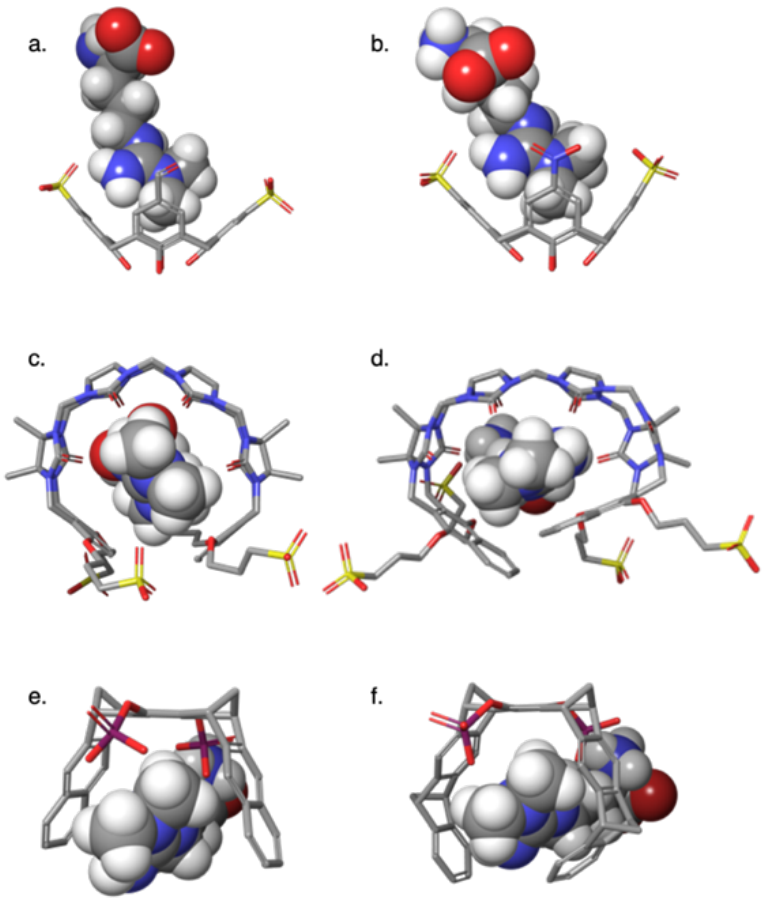


Figure 3.4. Molecular modelling of each host in complex with ADMA. a) **sCx4-CHO**, front view, b) **sCx4-NO_2**, front view/ c) **M1**, top view. d) **M2**, top view. e) **PC**, front view. f) **CLR01**, front view. Molecules were energy-minimized in explicit water (not shown) using OPL_2005 as implemented in Maestro (Schödinger, Inc). See supporting information for more views and for other host-guest complexes.

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.[164] 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 selectivity 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 (Figure 3.4). 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.

3.4 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_d 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_d 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 selectivity, 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 geometry 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[165] would lead to more such insights.

3.5 Supporting information

3.5.1. General information and materials

The host used in these studies have all been previously published or are commercially available. **sCx4** (4-sulfocalix[4]arene) was purchased from Tokyo Chemical Industry (TCI) (CAS 112269-92-8, >94.0%). **sCx4-CHO,**[102] **sCx4-1,2-CHO,**[152] **sCx4-1,3-CHO,**[152]**sCx4-NO_2,**[71] **sCx4-1,2-NO_2,**[152] **sCx4-1,3-NO_2,**[152] **M1,**[166] **M2,**[166] **CLR01**[133] and

PC[134] 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%), TML (55528-53-5, ≥97%), dimethylamine hydrochloride (506-59-2, 99%) and tetramethylammonium chloride (75-57-0, >98%) were purchased from Sigma-Aldrich. Benzyl trimethylammonium chloride (56-93-9, >99%) was purchased from Tokyo Chemical Industry Co., Ltd. (TCI).

3.5.2. IDA titrations

The IDAs were conducted in 384 well plates (Nunc™ 384-Well, Non-Treated, Flat-Bottom Microplate). All wells had a final working volume of 50 μL. All the experiments are performed in a 10 mM phosphate buffer unless otherwise indicated. The fluorescent signal was read on a Biotek Cytation 5 multimode plate reader (Software Version 3.05.11) at room temperature.

The fluorescent signal for the calix[4]arene hosts titrations were 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 were 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 were 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.

For the dye controls, three wells were prepared with dye (LCG 0.25 μM, 4-ASP (0.20 μM), R6G (0.10 μM). For the host controls, three wells were prepared with host (5 μM calix[4]arene, 70 μM **CLR01**, 30 μM **PC**, 10 μM **M1** and **M2**).

For the direct titrations, a serial dilution was performed titrating the hosts into their respective dyes (0-50 μM calix[4]arene into LCG, 0-90 μM **CLR01** into 4-ASP, 0-200 μM **PC** into 4-ASP, 0-80 μM **M1** and **M2** into R6G). The dissociation constants of the indicator (K_{ind}) that arises from the direct titrations is used to fit the competitive titrations. For the competitive titrations, a serial dilution was performed, 0.01-2.50 mM, guest into the respective host and dye concentrations. All the IDA data was corrected, by subtracting the host control wells, and the analyzed in GraphPad Prism Version 8.3.0 (328).

One exemplary replicate is shown for each host. All the experiments for the 10 mM phosphate buffer were performed as duplicates of triplicates, except for the control compounds (dimethylamine, tetramethylammonium, and benzyl trimethylammonium) which were performed as a single set of triplicates. All the experiments in 50 mM phosphate buffer were performed in triplicate.

3.5.2.1 Equations

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.

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

The standard error

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

Equation 2. Standard error

$$SD_{K_d} = (ln\ (10)\ * 10^{logK_i}) *\ SD_{logK_d}$$

SDK_d = Standard error of K_d

logK_d = Value of the log

SD$_{log}K_d$= Standard error of the logK_d

The total standard error

The total standard error between 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

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.

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

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

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

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_{ind} = Dissociation constant

Curve fit for the competitive titration

$$log_{EC50} = log\ (10^{\mathrm{d}} * \left(1 + \frac{[D]}{K_{ind}}\right))$$

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

log_{EC50} = log of the concentration of the competitor binding half-way between F_{min} and F_{max}

K_d = Equilibrium dissociation constant in Molar

[D] = Concentration of dye in nM

K_{ind} = Equilibrium dissociation constant of the direct titration

F = Fitted data point

F_{max} = Maximum signal

F_{min} = Minimum signal

3.5.2.2. IDA studies in 10 mM buffer

Table S3.1. K_d values determined by IDA for each of the calix[4]arene amino acids complexes [a]

	sCx4[b]	**sCx4-CHO**	**sCx4-1,2-CHO**	**sCx4-1,3-CHO**	**sCx4-NO_2**	**sCx4-1,2-NO_2**	**sCx4-1,3-NO_2**
	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)
Arginine	>1500	>1500	>1500	>1500	>1500	>1500	>1500
MMA	>1500	>1500	>1500	>1500	>1500	>1500	>1500
ADMA	500 ± 50	230 ± 50	>1500	>1500	290 ± 60	610 ± 190	560 ± 140
SDMA	1090 ± 300	370 ± 150[b]	>1500	>1500	>1500	700 ± 260[b]	690 ± 210
Lysine	>1500	>1500	>1500	>1500	>1500	>1500	>1500
MML	930 ± 120	470 ± 110	>1500	>1500	610 ± 130	>1500	>1500
DML	300 ± 40	240 ± 50	170 ± 90[b, c]	>1500	400 ± 60	190 ± 60[c]	440 ± 130[d]
TML	120 ± 20	100 ± 30	160 ± 70	>1500	150 ± 30	160 ± 40[b]	190 ± 60

[a] All titrations were carried out in 10 mM NaH_2PO_4 buffer at pH 7.4. Lucigenin (LCG) was used as indicator. See below for titration curves. All K_d values arise from fits with $R^2 \geq 0.95$ except where indicated. [b] value arise from one triplicate. [c] $R^2 \geq 0.93$. [c] $R^2 \geq 0.94$.

IDA study of sCx4

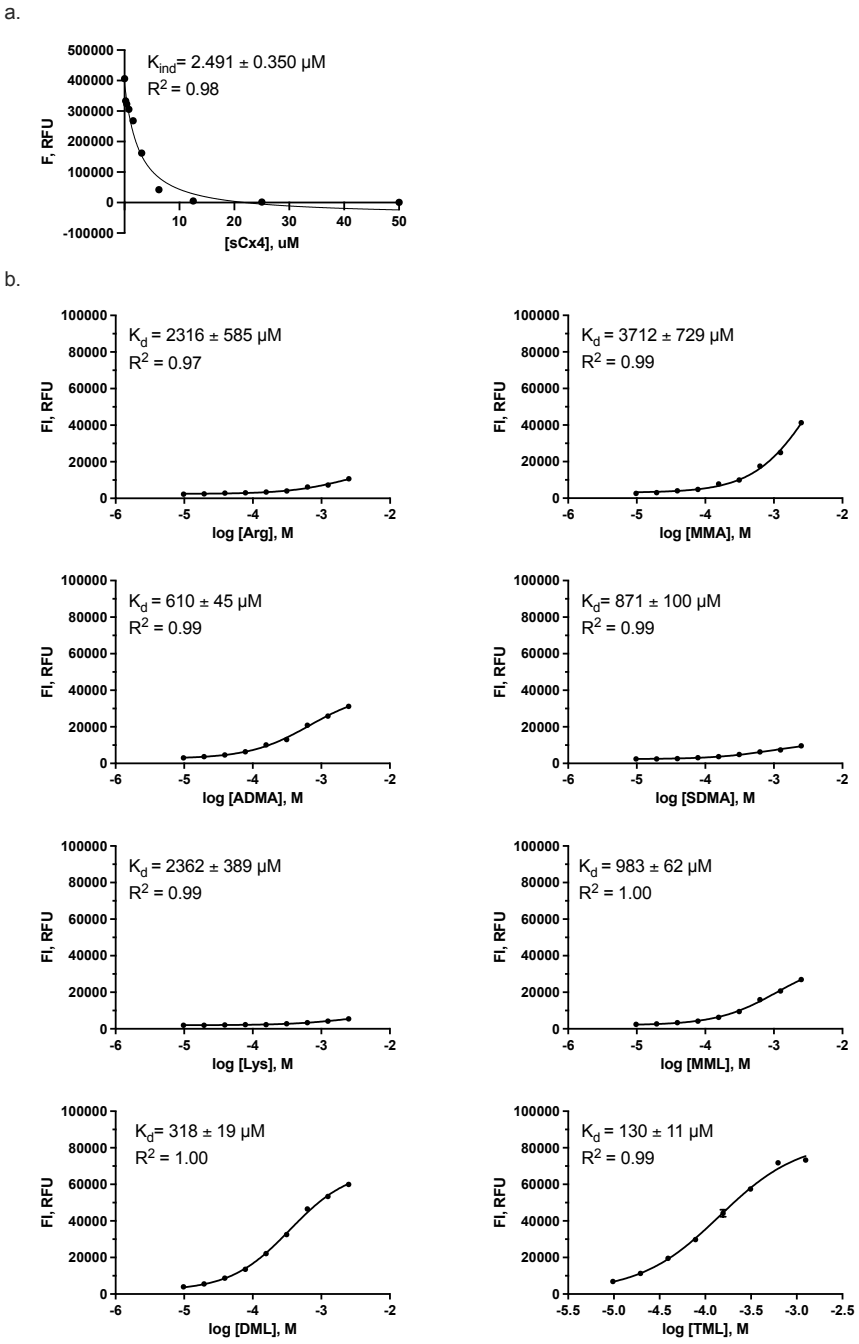


Figure S3.1. Fluorescence based studies of **sCx4**. a) Direct titration of LCG (0.25 μM) with **sCx4** (0 – 50 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **sCx4**-LCG (5 μM **sCx4**, 0.25 μM LCG) complex.

IDA study of sCx4-CHO

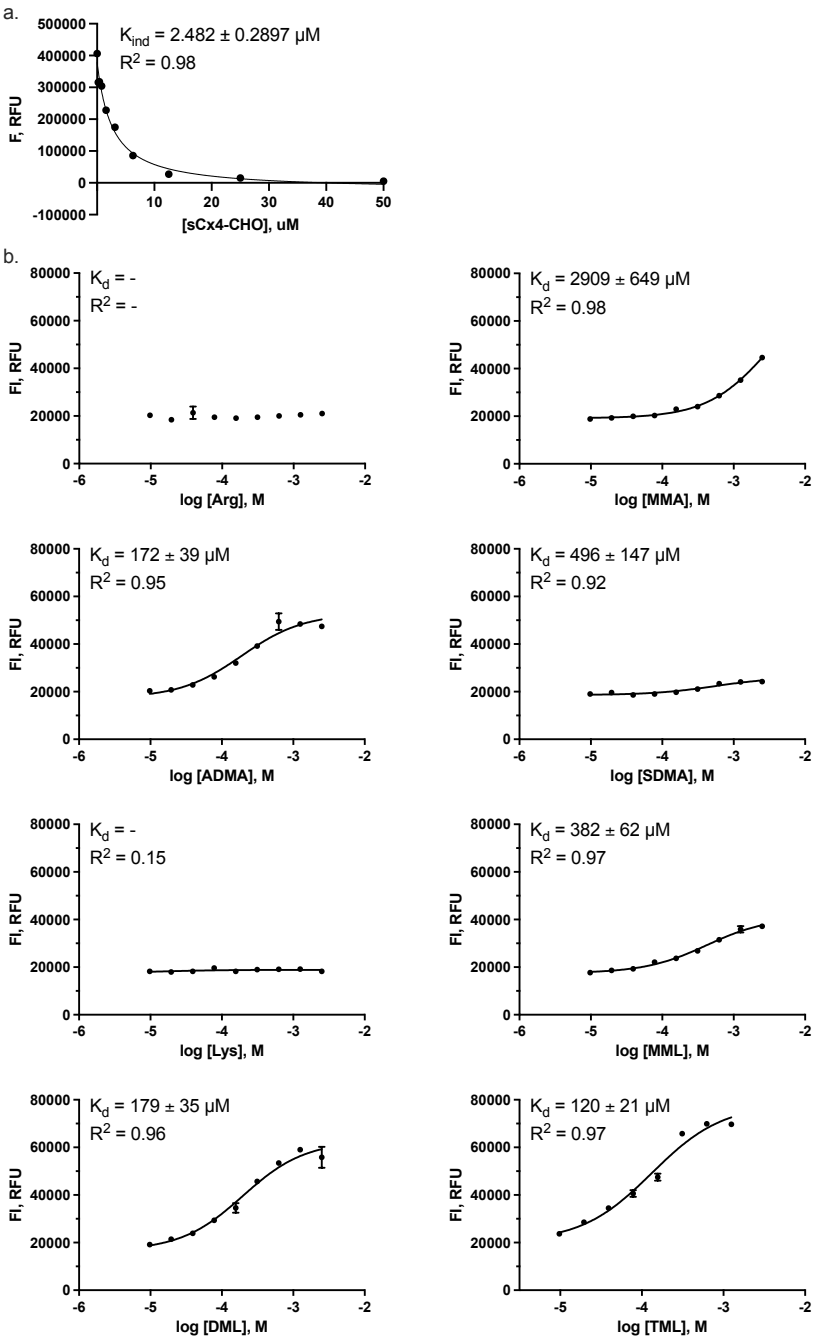


Figure S3.2. Fluorescence based studies of **sCx4-CHO**. a) Direct titration of LCG (0.25 μM) with **sCx4-CHO** (0 – 50 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **sCx4-CHO**-LCG (5 μM **sCx4-CHO**, 0.25 μM LCG) complex.

IDA study of sCx4-1,2-CHO

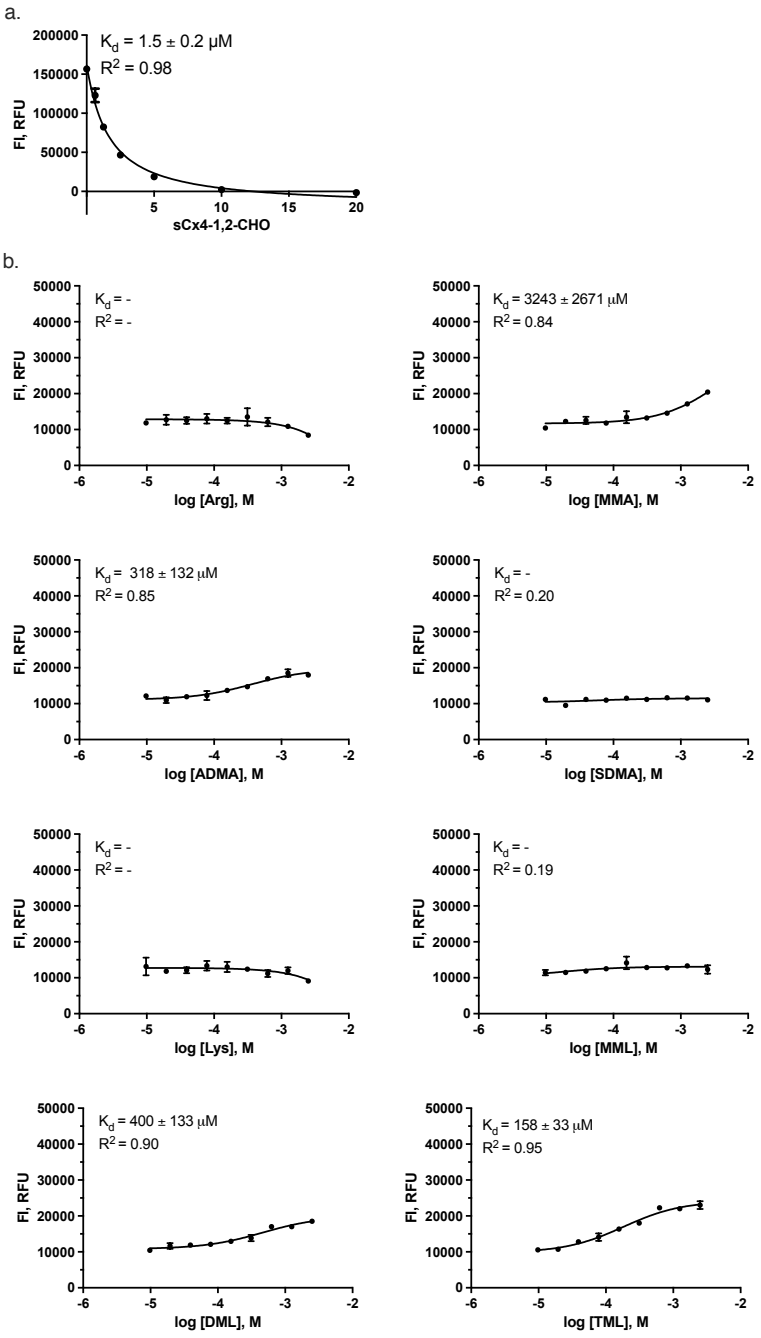


Figure S3.3. Fluorescence based studies of **sCx4-1,2-CHO**. a) Direct titration of LCG (0.25 μM) with **sCx4-1,2-CHO** (0 – 20 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **sCx4-1,2-CHO**-LCG (5 μM **sCx4-1,2-CHO**, 0.25 μM LCG) complex.

IDA study of sCx4-1,3-CHO

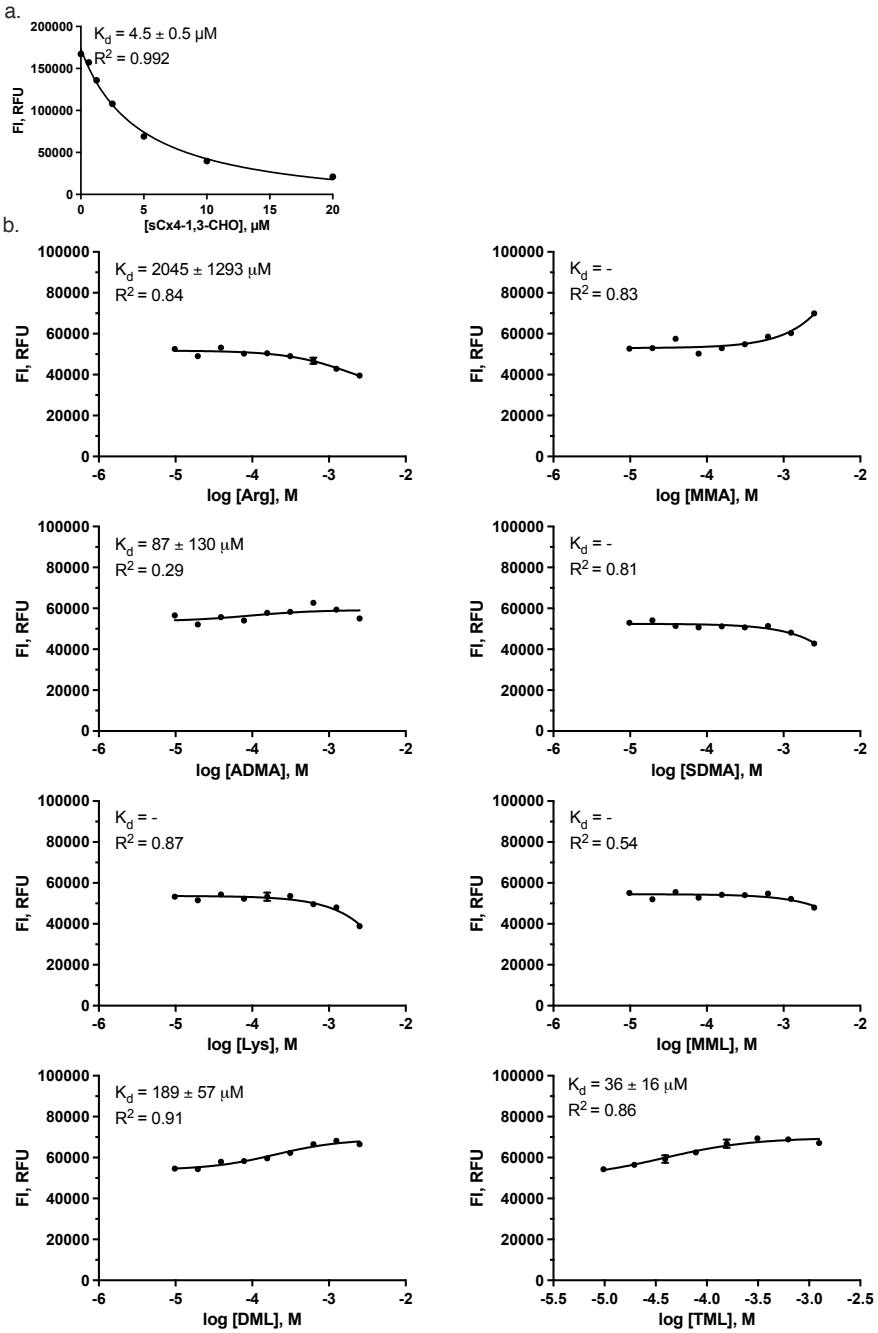


Figure S3.4. Fluorescence based studies of **sCx4-1,3-CHO**. a) Direct titration of LCG (0.25 μM) with **sCx4-1,3-CHO** (0 – 20 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **sCx4-1,3-CHO**-LCG (5 μM **sCx4-1,3-CHO**, 0.25 μM LCG) complex.

IDA study of sCx4-NO_2

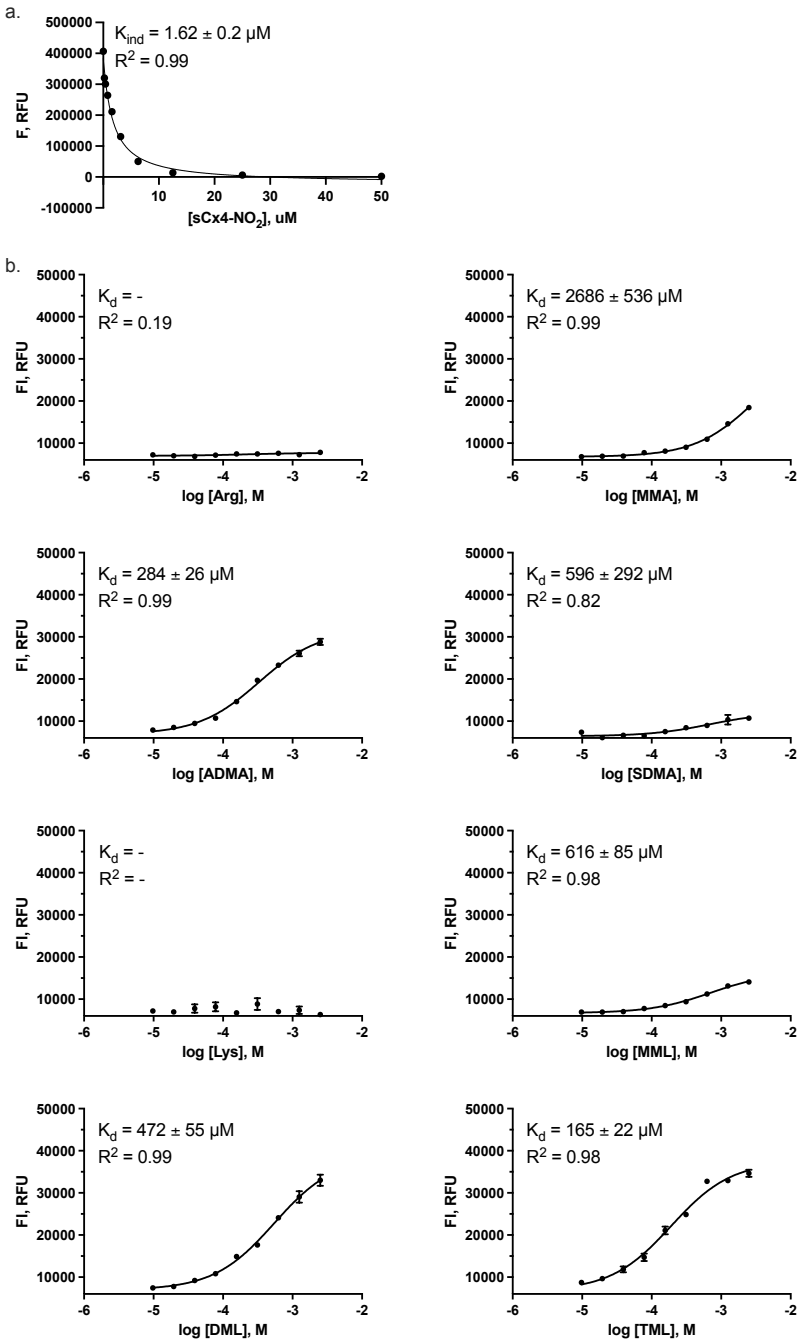


Figure S3.5. Fluorescence based studies of **sCx4-NO_2**. a) Direct titration of LCG (0.25 μM) with **sCx4-NO_2** (0 – 50 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **sCx4-NO_2**-LCG (5 μM **sCx4-NO_2**, 0.25 μM LCG) complex.

IDA study of sCx4-1,2-NO_2

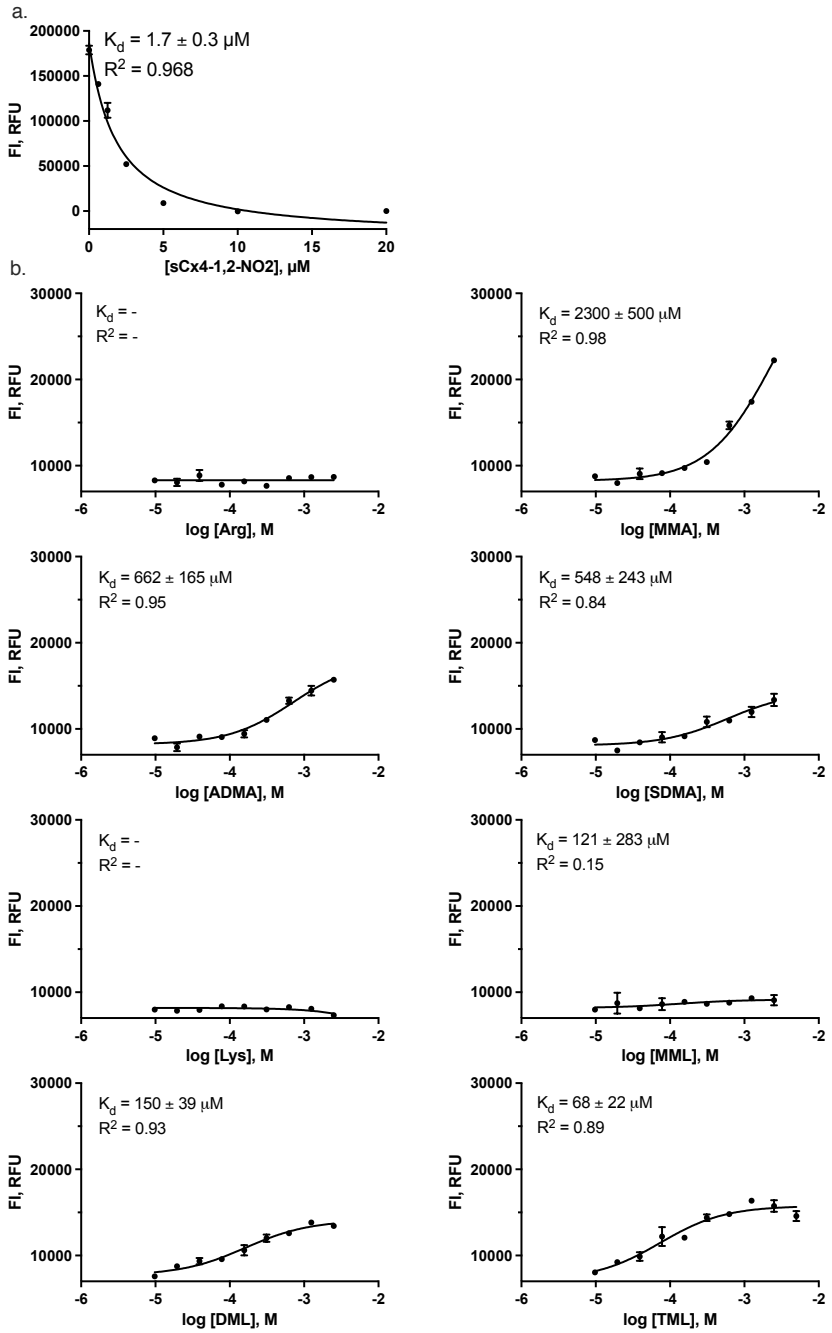


Figure S3.6. Fluorescence based studies of **sCx4-1,2-NO_2**. a) Direct titration of LCG (0.25 μM) with **sCx4-1,2-NO_2** (0 – 20 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **sCx4-1,2-NO_2**-LCG (5 μM **sCx4-1,2-NO_2**, 0.25 μM LCG) complex.

IDA study of sCx4-1,3-NO_2

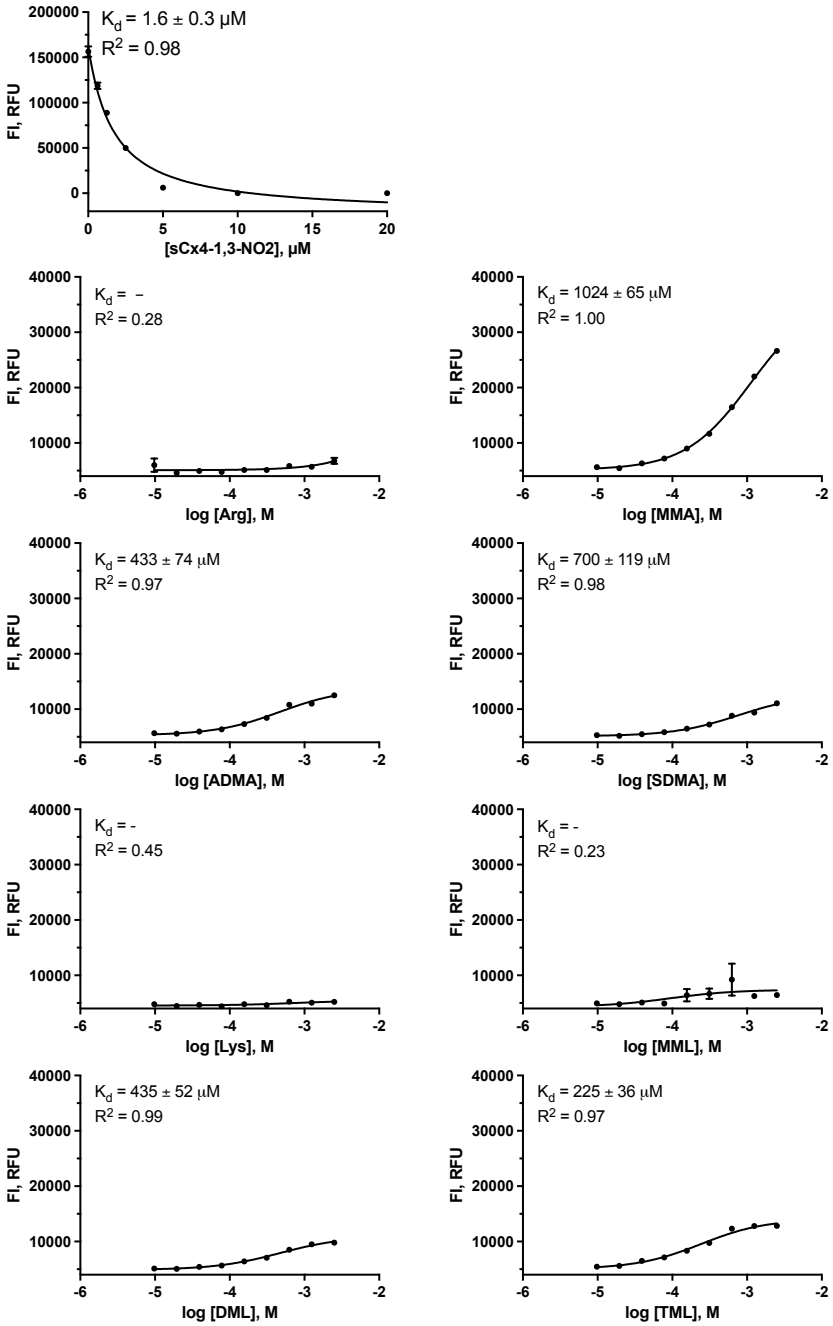


Figure S3.7. Fluorescence based studies of **sCx4-1,3-NO_2**. a) Direct titration of LCG (0.25 μM) with **sCx4-1,3-NO_2** (0 – 20 μM). b) Competitive titrations of Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **sCx4-1,3-NO_2**-LCG (5 μM **sCx4-1,3-NO_2**, 0.25 μM LCG) complex.

IDA study of CLR01

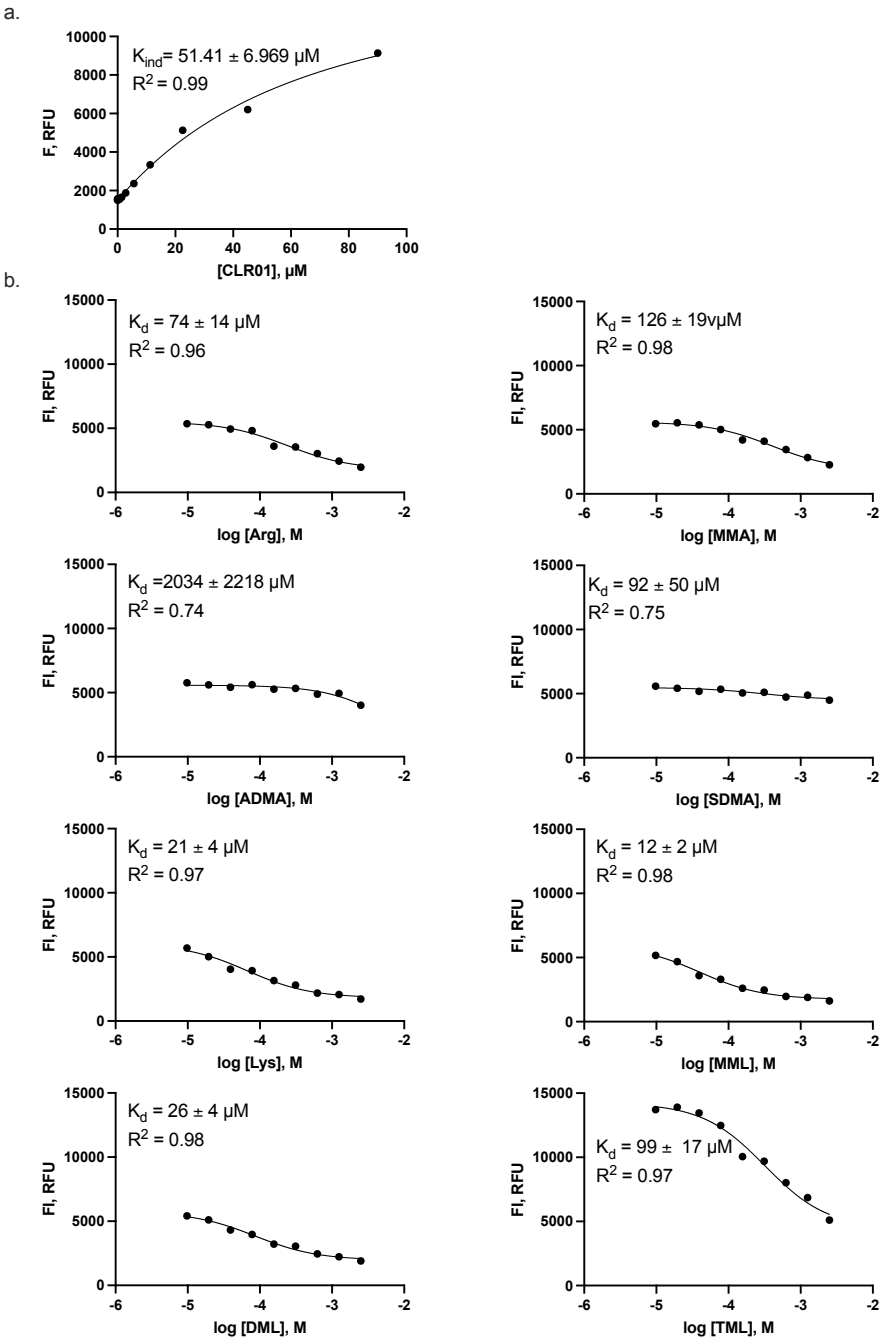


Figure S3.8. Fluorescence based studies of **CLR01**. a) Direct titration of 4-ASP(20 μM) with **CLR01** (0 – 50 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **CLR01**-4-ASP (70 μM **CLR01**, 20 μM 4-ASP) complex.

IDA study of PC

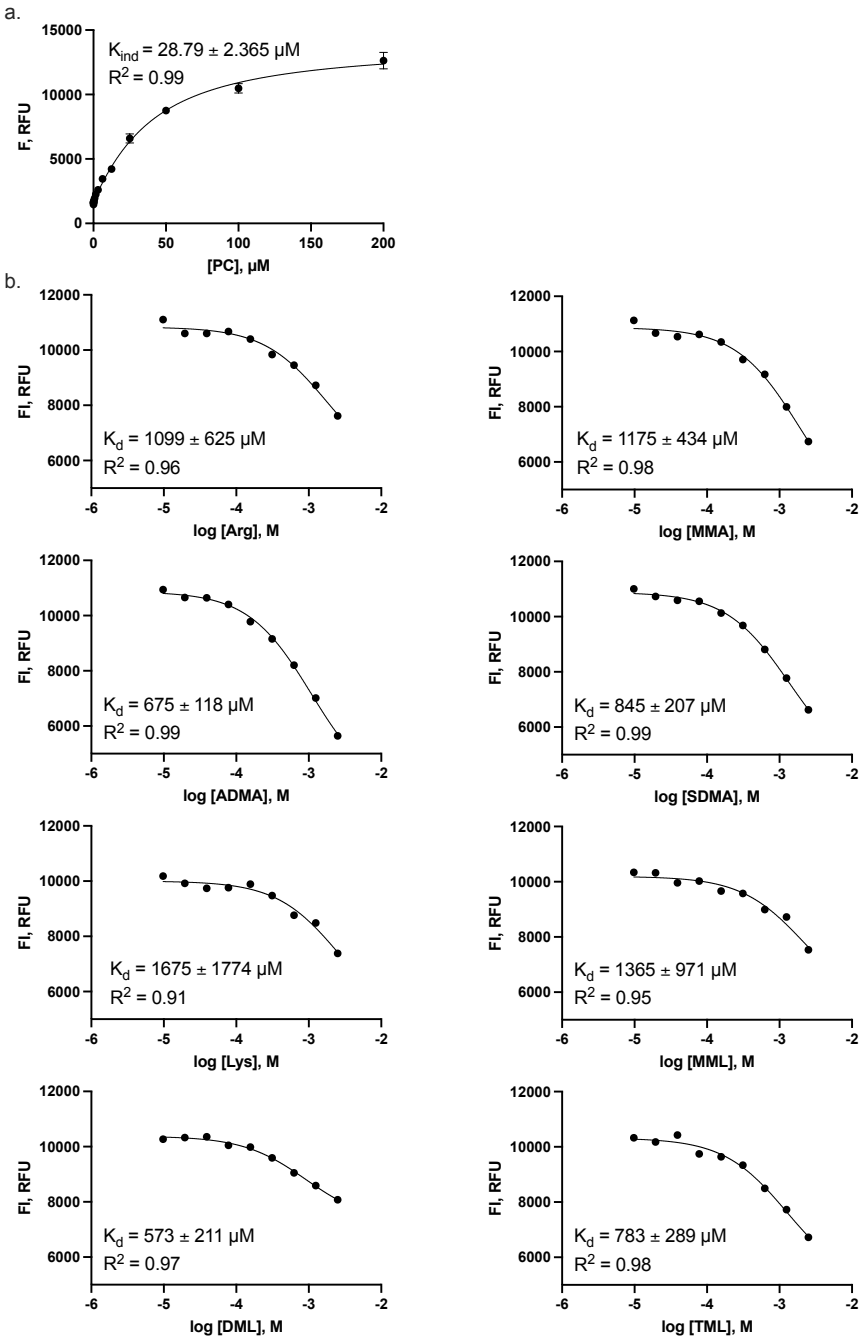


Figure S3.9. Fluorescence based studies of **PC**. a) Direct titration of 4-ASP (20 μM) with **PC** (0 – 200 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **PC**-4-ASP (30 μM **PC**, 20 μM 4-ASP) complex.

IDA study of M1

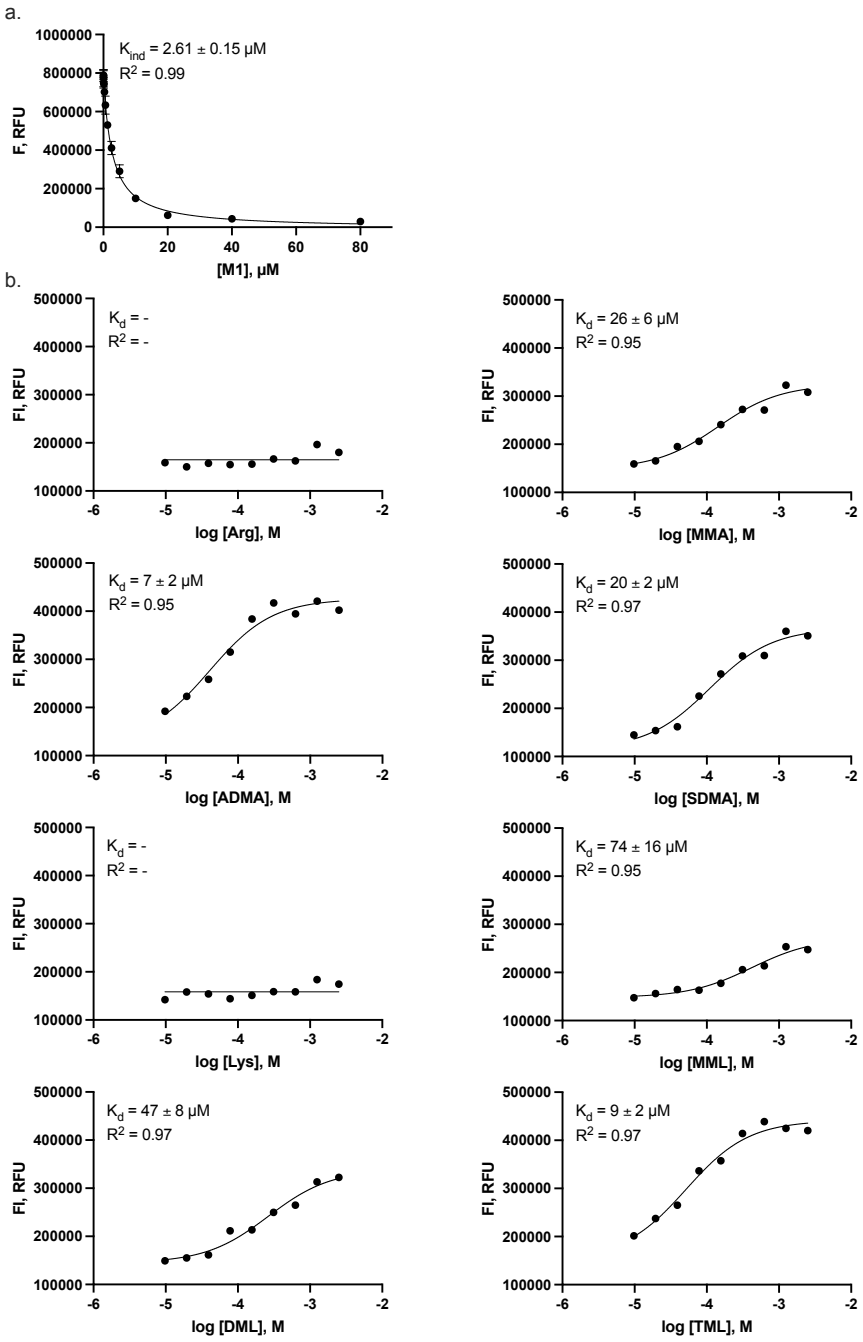


Figure S3.10. Fluorescence based studies of **M1**. a) Direct titration of R6G (10 μM) with **M1** (0 – 80 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **M1**-R6G (10 μM **M1**, 10 μM R6G) complex.

IDA study of M2

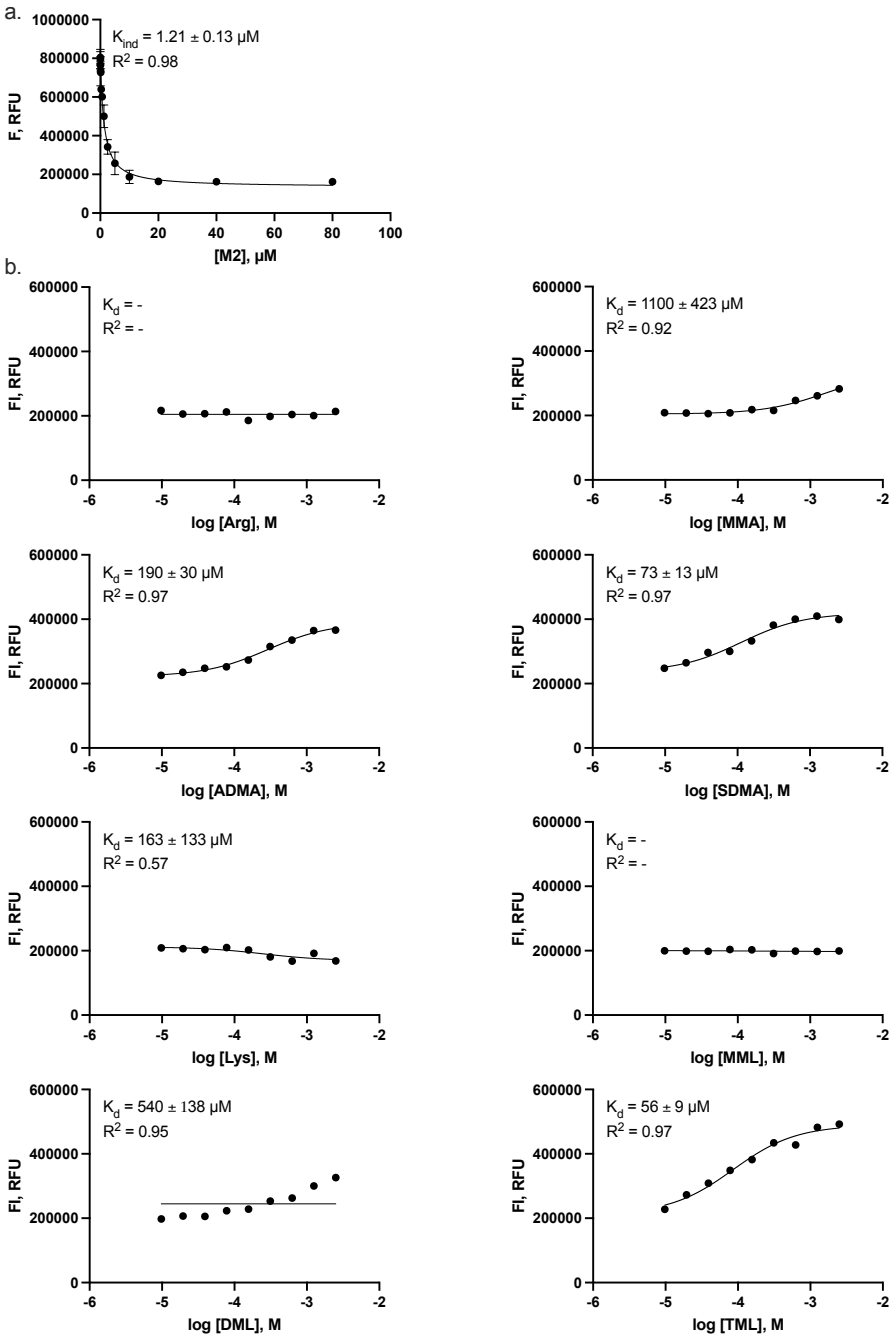


Figure S3.11. Fluorescence based studies of **M2**. a) Direct titration of R6G (10 μM) with **M2** (0 – 80 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **M2**-R6G (10 μM **M2**, 10 μM R6G) complex.

3.5.2.3. Fluorescence based studies with control compounds

We performed IDA studies dimethylamine, tetramethylammonium, and benzyl trimethylammonium (Figure S3.12) as controls for evaluating hydrophobicity and size/shape-matching. **M1** and **M2** displayed such strong binding to benzyl trimethylammonium that we performed a second IDA with a lower concentration range (0 to 0.31 mM) to observe the binding curve.

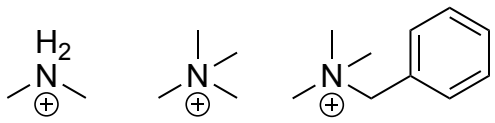


Figure S3.12. Control compounds for hydrophobicity and shape-matching. Left to right dimethylammonium, tetramethylammonium and benzyltrimethylammonium. Each guest was used as its chloride salt.

Table S3.2. K_d values determined by IDA for complexes of control guests in 10 mM phosphate buffer. [a]

	sCx4[b]	**sCx4-CHO**[b]	**sCx4-NO_2**[b]	**M1**[c]	**M2**[c]	**CLR01**[d]	**PC**[d]
	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)	K_d (μM)
Dimethylamine	>1500	>1500	>1500	>1500	>1500	1100 ± 1100	>1500
Tetramethylammonium	370 ± 70	390 ± 110	660 ± 90	12 ± 5	40 ± 40[e]	460 ± 360[f]	370 ± 440[g]
Benzyl trimethylammonium	940 ± 200	340 ± 60	660 ± 180	3 ± 1	1.5 ± 0.4	170 ± 30	960 ± 560

[a] All titrations were carried out in 10 mM NaH_2PO_4 buffer at pH 7.4. See below 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] $R^2 \geq 0.83$. [f] $R^2 \geq 0.92$. [g] $R^2 \geq 0.79$.

IDA study of sCx4 and control compounds

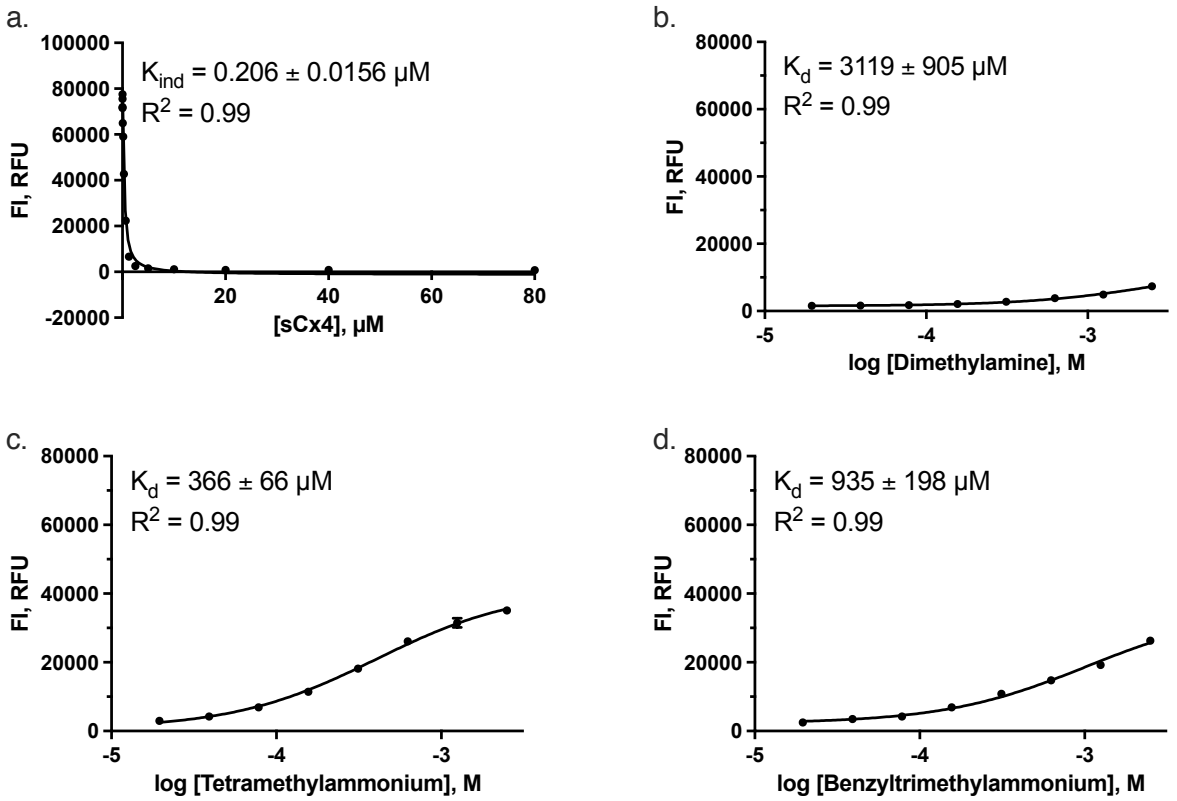


Figure S3.13. Fluorescence based studies of **sCx4**. a) Direct titration of LCG (0.25 μM) with **sCx4** (0 – 80 μM). b) Competitive titration with dimethylamine (0 – 2.5 mM) titrated into the **sCx4**-LCG (5 μM **sCx4**, 0.25 μM LCG) complex. c) Competitive titration with tetramethylammonium (0 – 2.5 mM) titrated into the **sCx4**-LCG (5 μM **sCx4**, 0.25 μM LCG) complex. d) Competitive titration with benzyl trimethylammonium (0 – 2.5 mM) titrated into the **sCx4**-LCG (5 μM **sCx4**, 0.25 μM LCG) complex. See Table S3.1 for buffer conditions.

IDA study of sCx4-CHO and control compounds

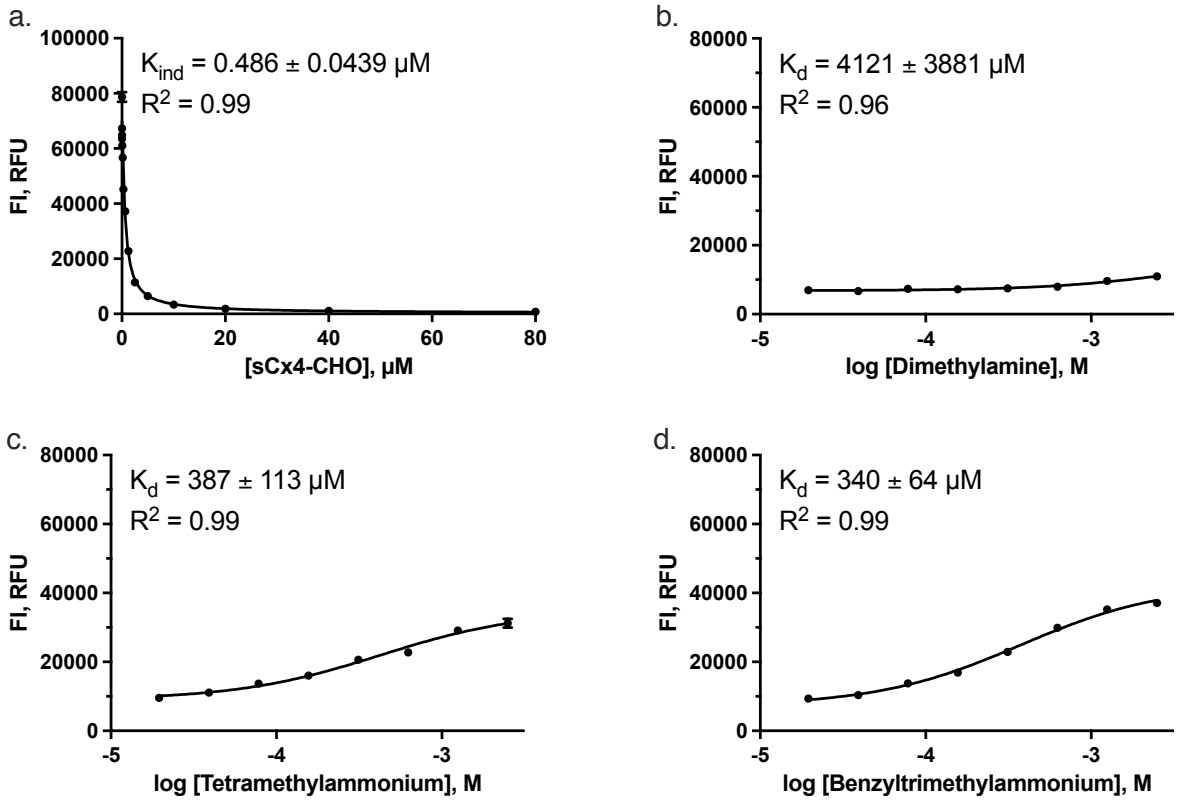


Figure S3.14. Fluorescence based studies of **sCx4-CHO**. a) Direct titration of LCG (0.25 μM) with **sCx4-CHO** (0 – 80 μM). b) Competitive titration with dimethylamine (0 – 2.5 mM) titrated into the **sCx4-CHO**-LCG (5 μM **sCx4-CHO**, 0.25 μM LCG) complex. c) Competitive titration with tetramethylammonium (0 – 2.5 mM) titrated into the **sCx4-CHO**-LCG (5 μM **sCx4-CHO**, 0.25 μM LCG) complex. d) Competitive titration with benzyl trimethylammonium (0 – 2.5 mM) titrated into the **sCx4-CHO**-LCG (5 μM **sCx4-CHO**, 0.25 μM LCG) complex. See Table S3.1 for buffer conditions.

IDA study of sCx4-NO_2 and control compounds

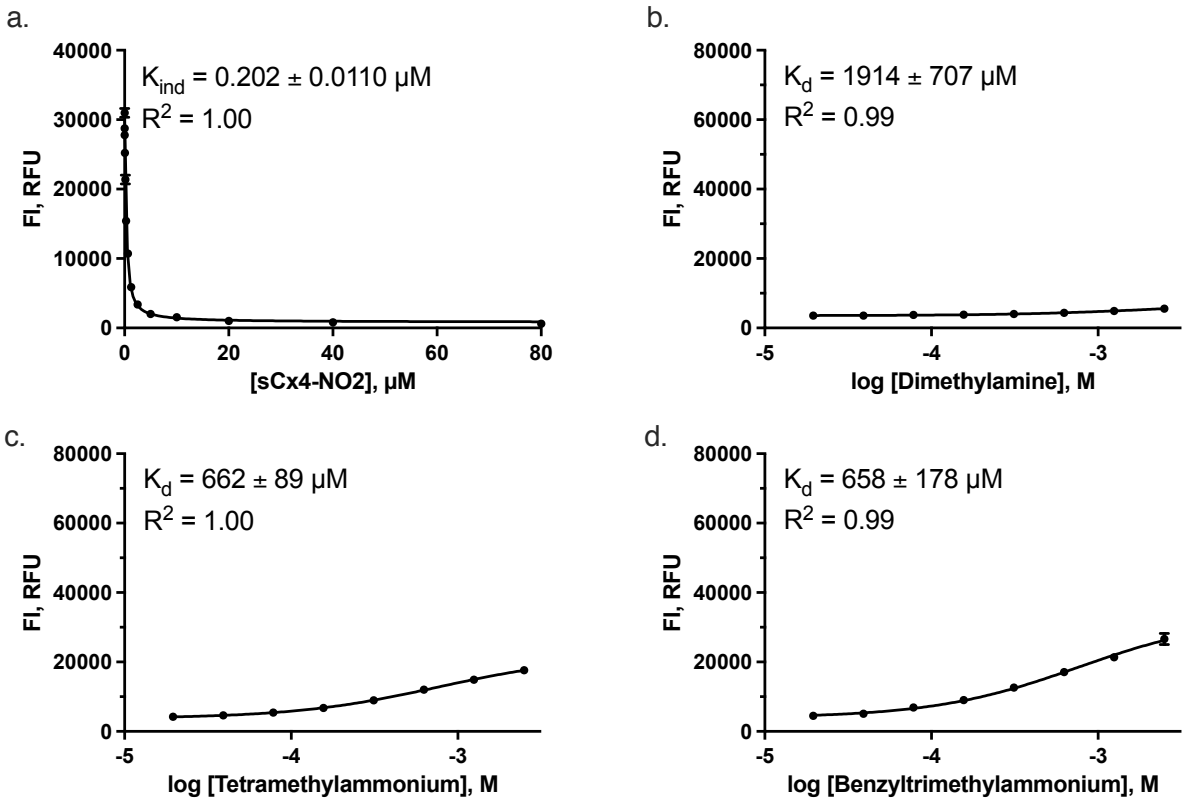


Figure S3.15. Fluorescence based studies of **sCx4-NO_2**. a) Direct titration of LCG (0.25 μM) with **sCx4-NO_2** (0 – 80 μM). b) Competitive titration with dimethylamine (0 – 2.5 mM) titrated into the **sCx4-NO_2**-LCG (5 μM **sCx4-NO_2**, 0.25 μM LCG) complex. c) Competitive titration with tetramethylammonium (0 – 2.5 mM) titrated into the **sCx4-NO_2**-LCG (5 μM **sCx4-NO_2**, 0.25 μM LCG) complex. d) Competitive titration with benzyl trimethylammonium (0 – 2.5 mM) titrated into the **sCx4-NO_2**-LCG (5 μM **sCx4-NO_2**, 0.25 μM LCG) complex. See Table S3.1 for buffer conditions.

IDA study of M1 and control compounds

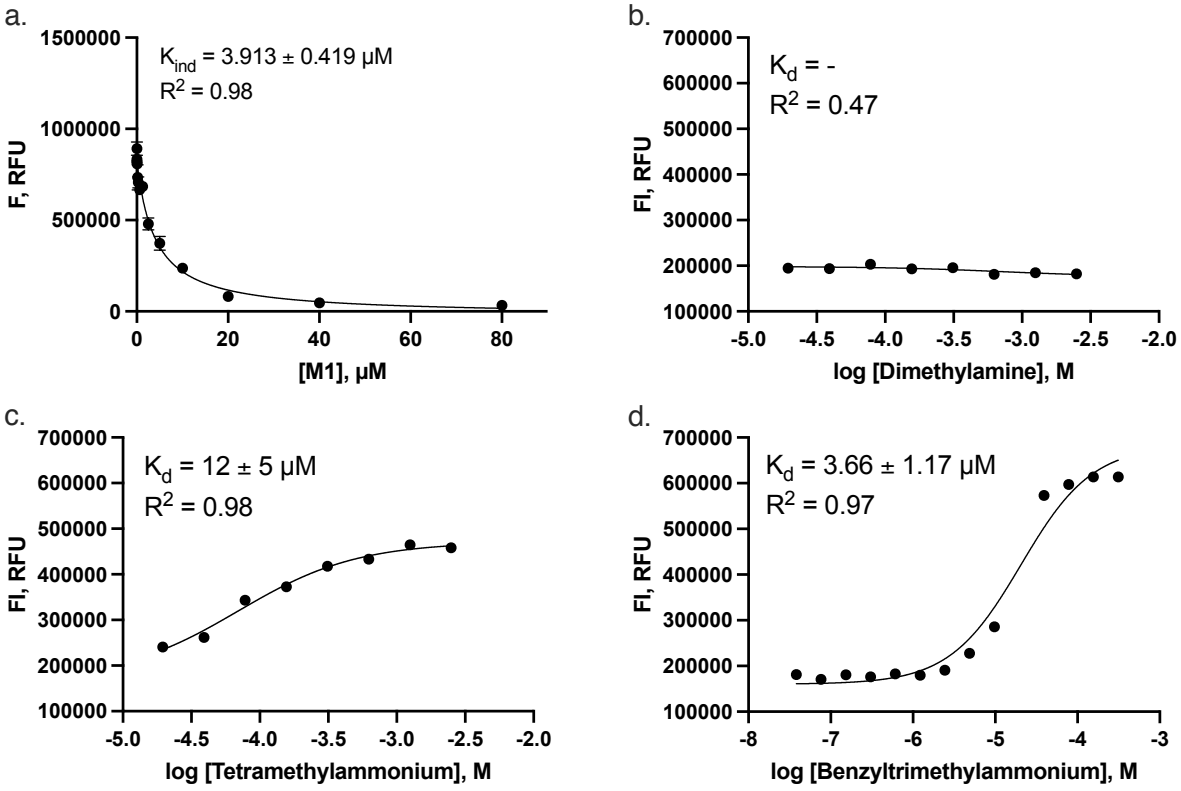


Figure S3.16. Fluorescence based studies of **M1**. a) Direct titration of R6G (10 μM) with **M1** (0 – 80 μM). b) Competitive titration with dimethylamine (0 – 2.5 mM) titrated into the **M1**-R6G (10 μM **M1**, 10 μM R6G) complex. c) Competitive titration with tetramethylammonium (0 – 2.5 mM) titrated into the **M1**-R6G (10 μM **M1**, 10 μM R6G) complex. d) Competitive titration with benzyl trimethylammonium (0 – 0.31 mM) titrated into the **M1**-R6G (10 μM **M1**, 10 μM R6G) complex. See Table S3.1 for buffer conditions.

IDA study of M2 and control compounds

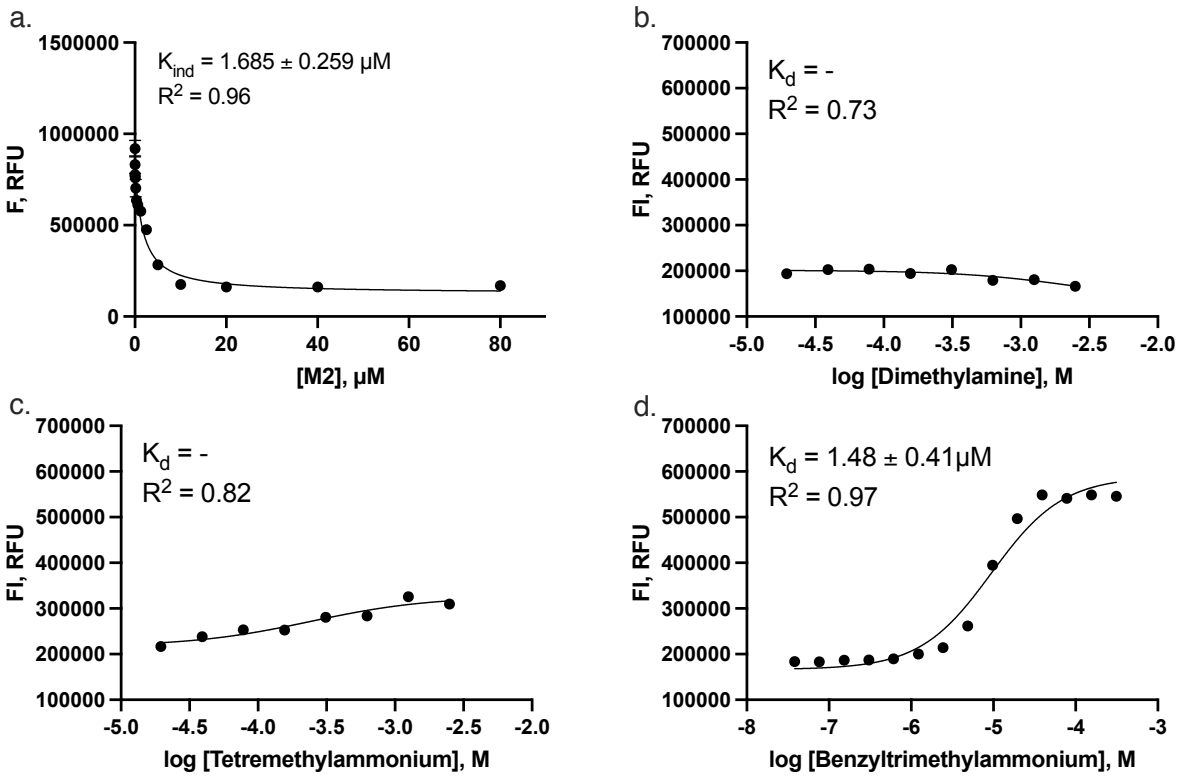


Figure S3.17. Fluorescence based studies of **M2**. a) Direct titration of R6G (10 μM) with **M2** (0 – 80 μM). b) Competitive titration with dimethylamine (0 – 2.5 mM) titrated into the **M2**-R6G (10 μM **M2**, 10 μM R6G) complex. c) Competitive titration with tetramethylammonium (0 – 2.5 mM) titrated into the **M2**-R6G (10 μM **M2**, 10 μM R6G) complex. d) Competitive titration with benzyl trimethylammonium (0 – 0.31 mM) titrated into the **M2**-R6G (10 μM **M2**, 10 μM R6G) complex. See Table S3.1 for buffer conditions.

IDA study of PC and control compounds

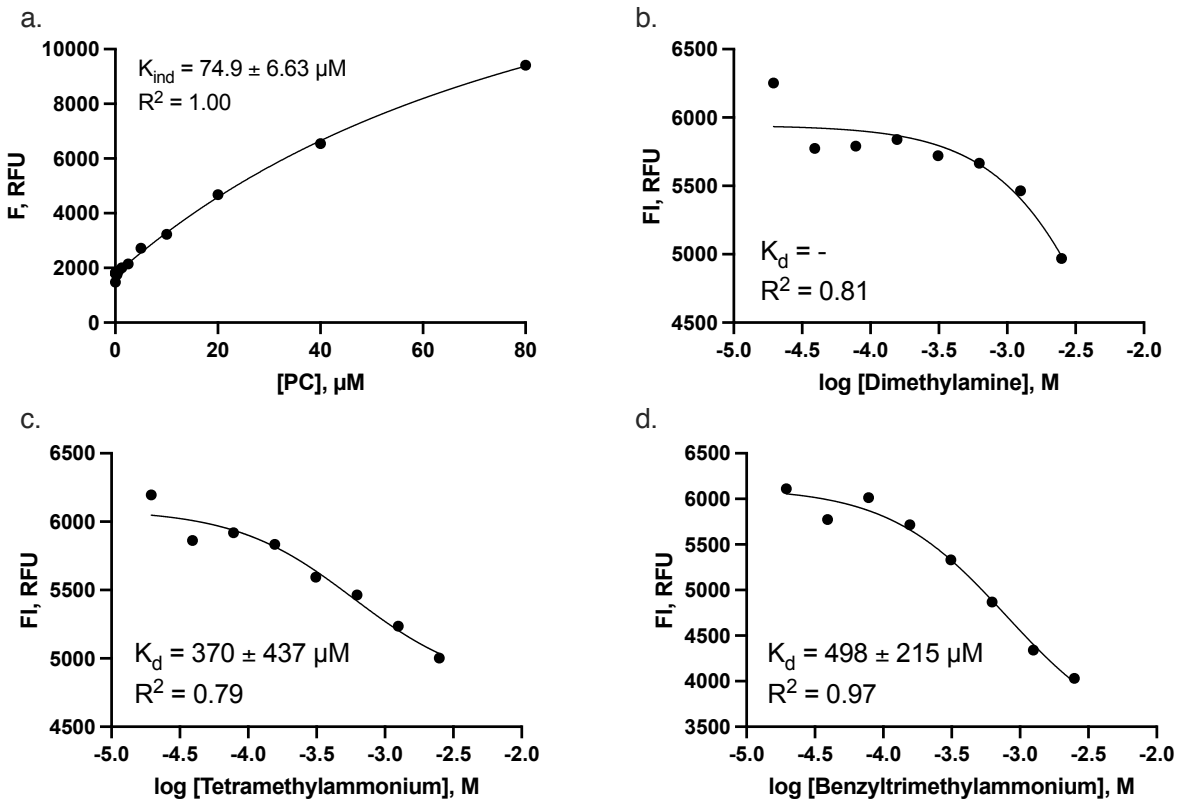


Figure S3.18. Fluorescence based studies of **PC**. a) Direct titration of 4-ASP (20 μM) with **PC** (0 – 50 μM). b) Competitive titrations with dimethylamine (0 – 2.5 mM) titrated into the **PC**-4-ASP (30 μM **PC**, 20 μM 4-ASP) complex. c) Competitive titrations with tetramethylammonium (0 – 2.5 mM) titrated into the **PC**-4-ASP (30 μM **PC**, 20 μM 4-ASP) complex.d) Competitive titrations with benzyl trimethylammonium (0 – 2.5 mM) titrated into the **PC**-4-ASP (30 μM **PC**, 20 μM 4-ASP) complex. See Table S3.1 for buffer conditions.

IDA study of CLR01 and control compounds

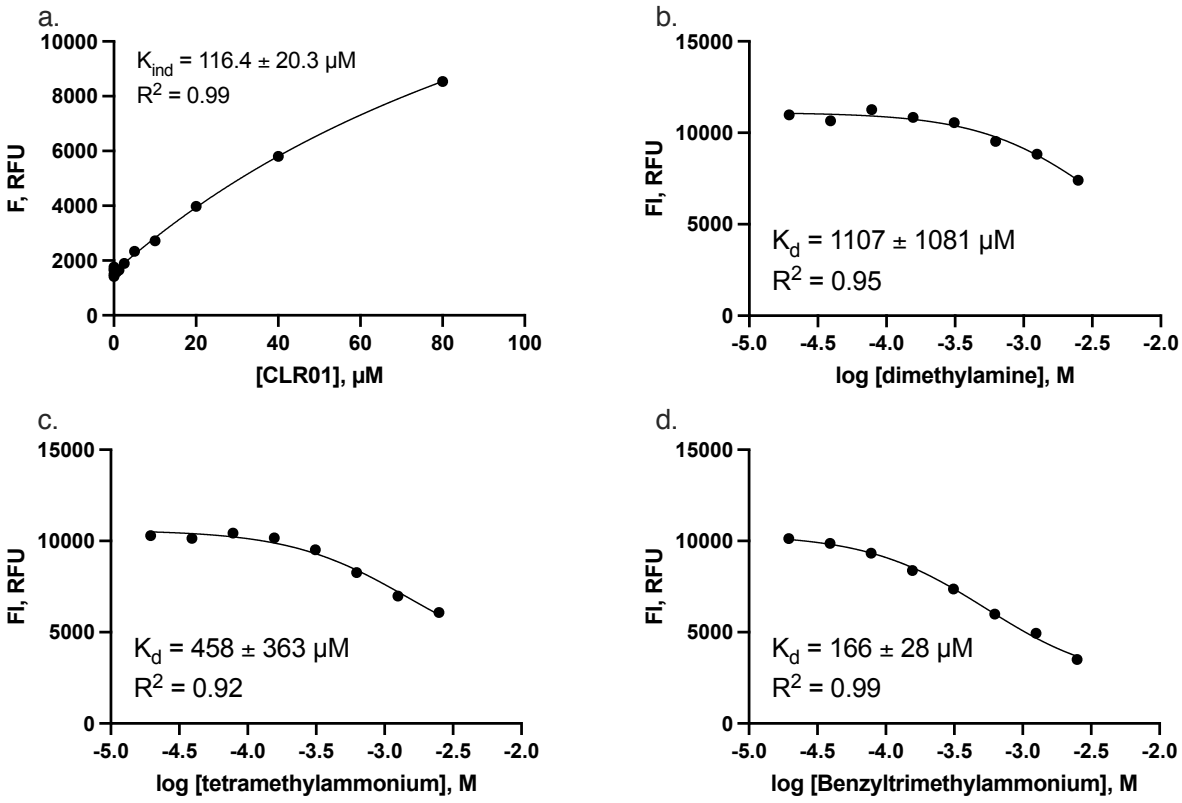


Figure S3.19. Fluorescence based studies of **CLR01**. a) Direct titration of 4-ASP (20 μM) with **CLR01** (0 – 50 μM). b) Competitive titrations with dimethylamine (0 – 2.5 mM) titrated into the **CLR01**-4-ASP (70 μM **CLR01**, 20 μM 4-ASP) complex. c) Competitive titrations with tetramethylammonium (0 – 2.5 mM) titrated into the **CLR01**-4-ASP (70 μM **CLR01**, 20 μM 4-ASP) complex.d) Competitive titrations with benzyl trimethylammonium (0 – 2.5 mM) titrated into the **CLR01**-4-ASP (70 μM **CLR01**, 20 μM 4-ASP) complex. See Table S3.1 for buffer conditions.

3D bar graph visualizing binding trends between hosts and dimethylamine, tetramethylammonium and benzyl trimethylammonium

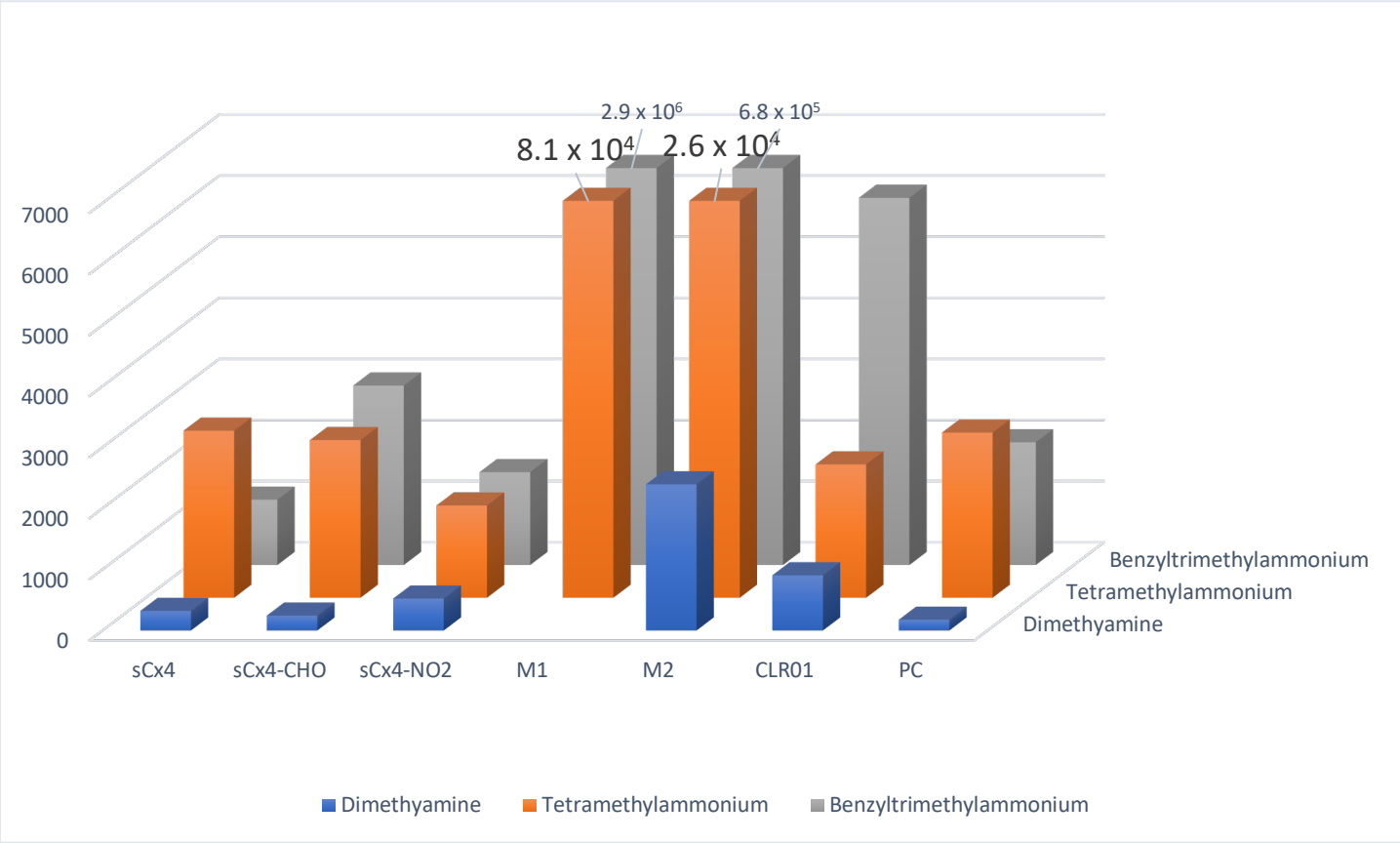


Figure S3.20. 3D bar graph visualizing binding trends between the hosts and dimethylamine, tetramethylammonium and benzyltrimethylammonium. The binding strengths are presented as the *K*a values in units M^{-1} ($K_a=1/K_d$). The bar for **M1** and **M2** with tetramethylammonium and benzyltrimethylammonium are truncated, with the actual K_a values presented above the bars.

3.5.2.4. IDA studies in 50 mM buffer

IDA study of M1

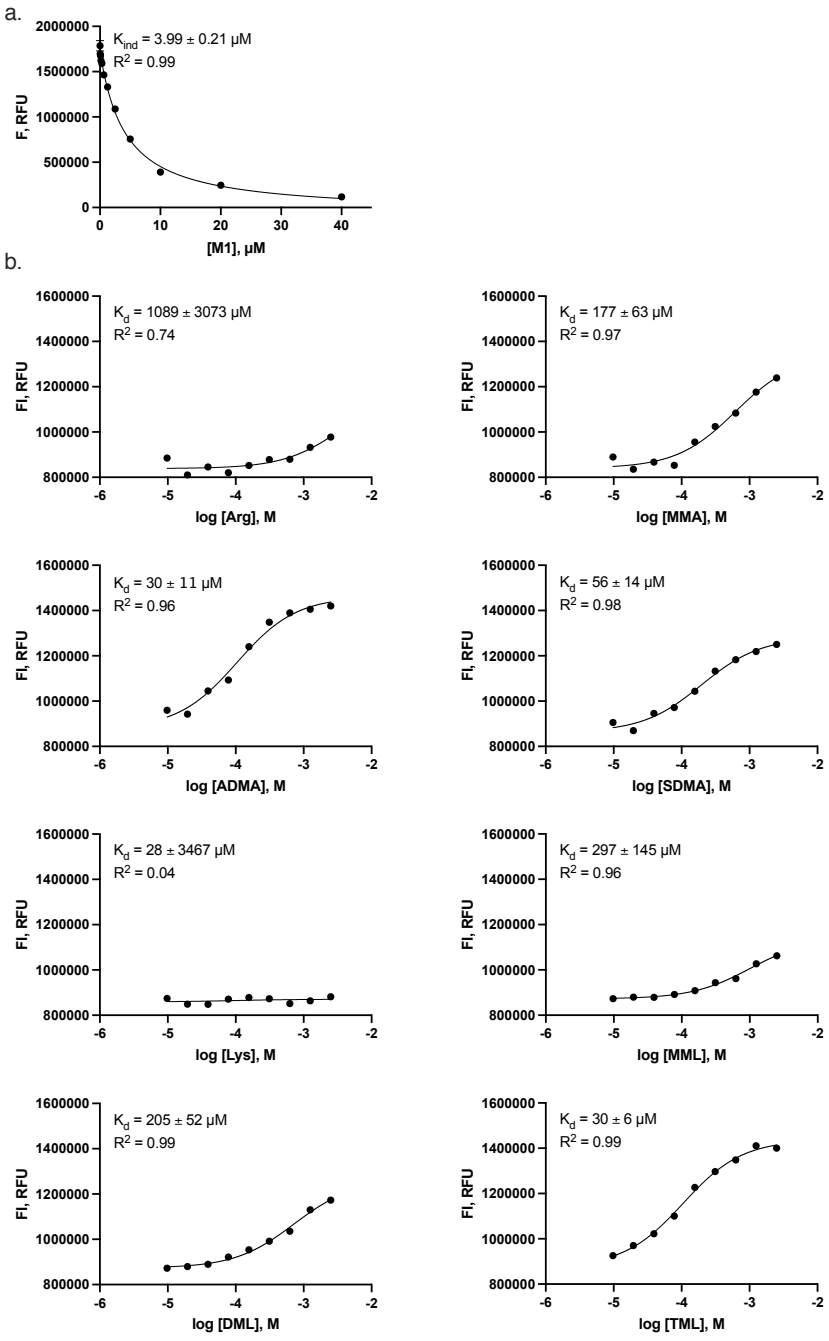


Figure S3.21. Fluorescence based studies of **M1**. a) Direct titration of R6G (10 μM) with **M1** (0 – 40 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **M1**-R6G (10 μM **M1**, 10 μM R6G) complex.

IDA study of M2

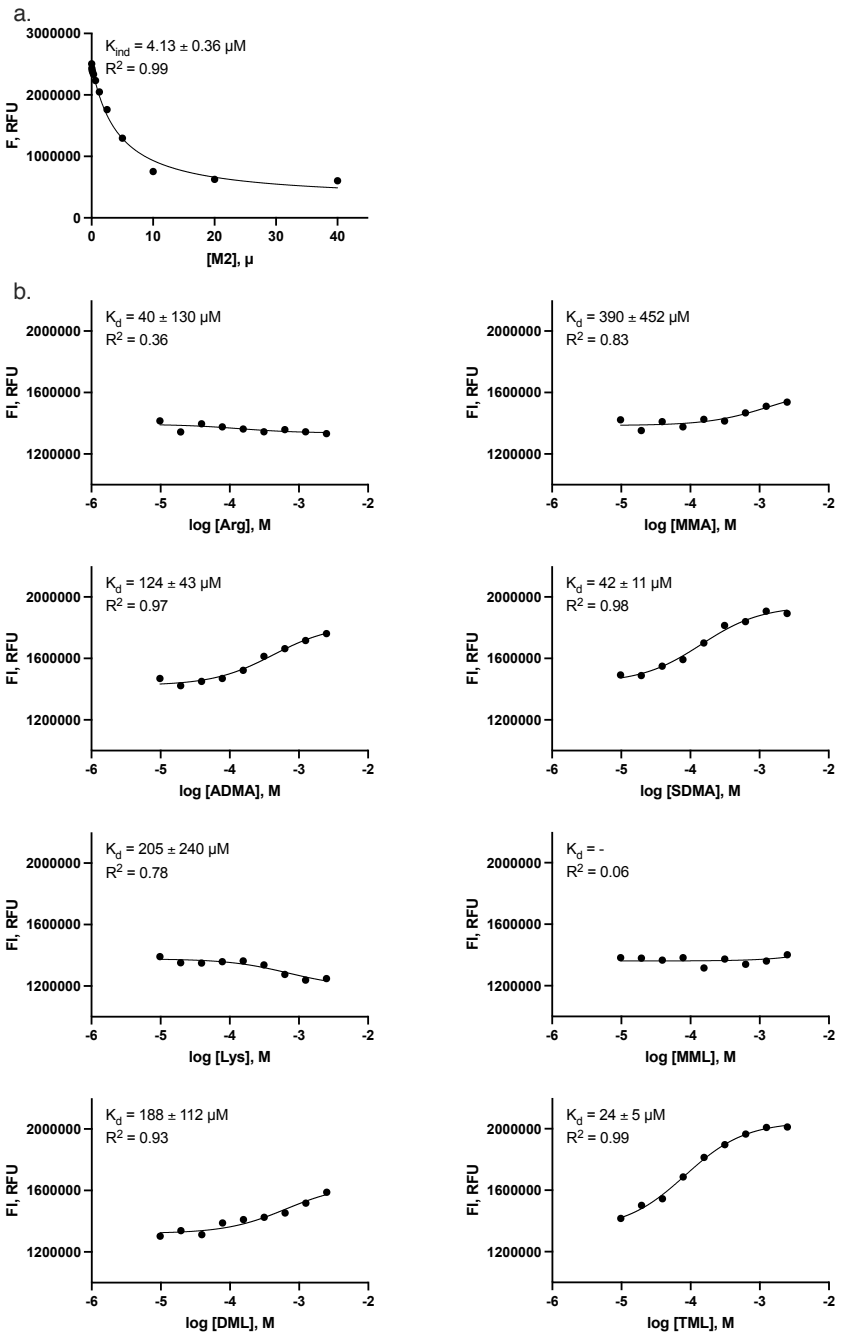


Figure S3.22. Fluorescence based studies of **M2**. a) Direct titration of R6G (10 μM) with **M2** (0 – 40 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **M2**-R6G (10 μM **M2**, 10 μM R6G) complex.

IDA study of CLR01

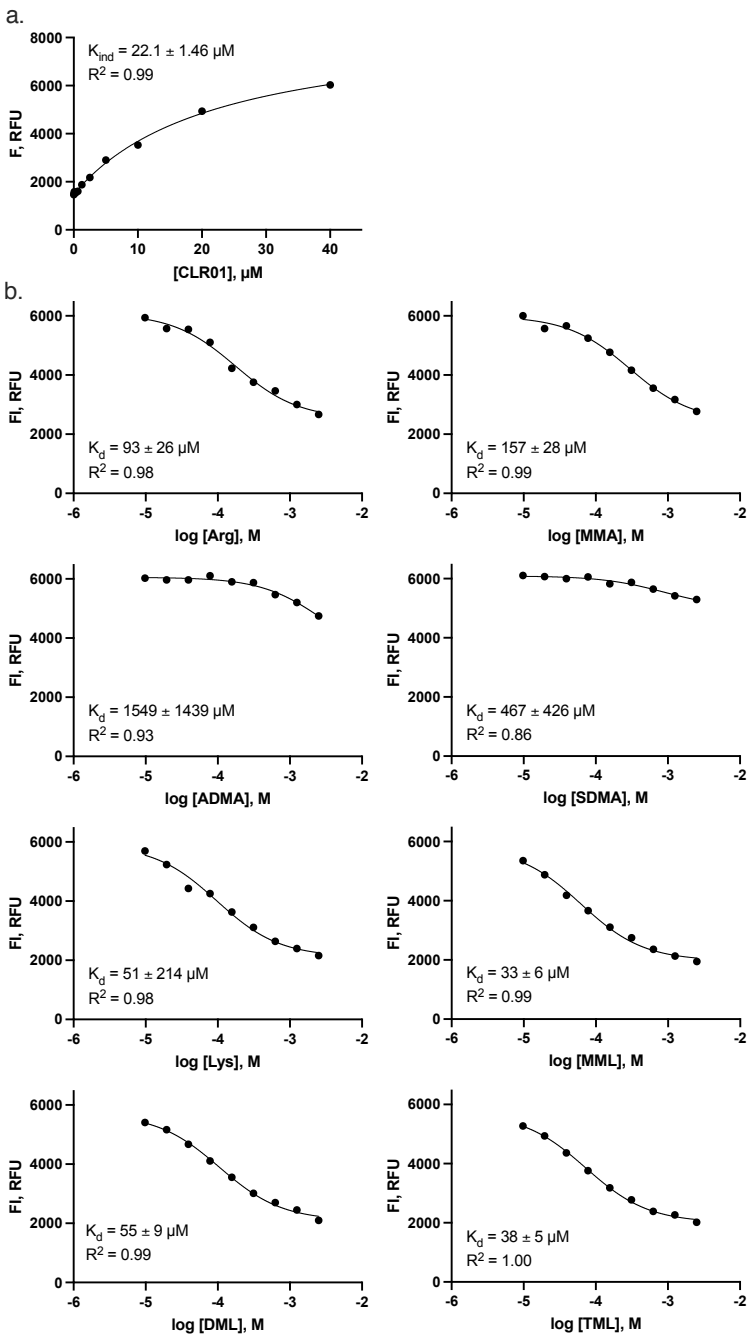


Figure S3.23. Fluorescence based studies of **CLR01**. a) Direct titration of 4-ASP(20 μM) with **CLR01** (0 – 40 μM). b) Competitive titrations between Arginine, MMA, ADMA, SDMA, Lysine, MML, DML and TML (0 – 2.5 mM) individually titrated into the **CLR01**-4-ASP (70 μM **CLR01**, 20 μM 4-ASP) complex.

3.5.3. Isothermal Titration Calorimetry — literature data

Table S3.3. Summary of previously published ITC data on relevant complexes.

		K_d	ΔH	TΔS
		μM	kJ/mol	kJ/mol
sCx4[a111]	Lysine	n.d.	n.d.	n.d.
	MML	335	–16	3.5
	DML	95	–20	3.4
	TML	30	–22	4.3
M1[b161]	H-Lys-COO^-	3	–49	34
	H-Arg-COO^-	6.6	–16	–2
M2[b161]	H-Lys-COO^-	2010	–9	–6
	H-Arg-COO^-	420	–41	22
CLR01[c160]	Ac-Lys-OMe	0.15	–23	–4
	Ac-Arg-OMe	0.34	–29	4

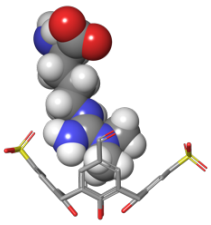
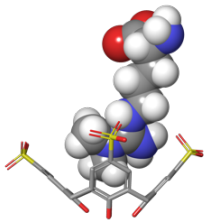
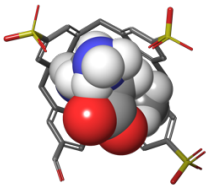
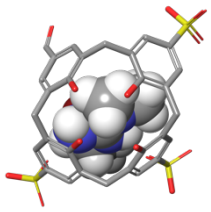
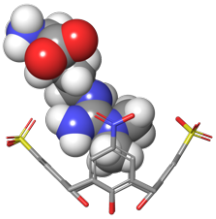
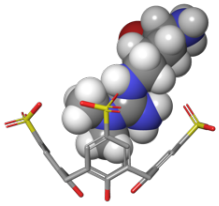
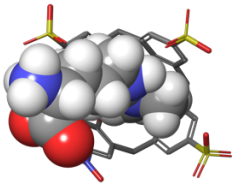
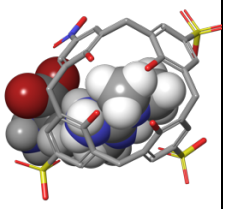
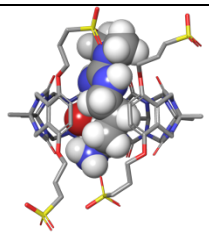
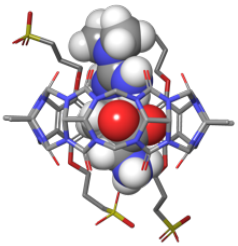
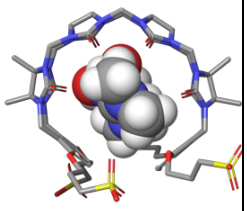
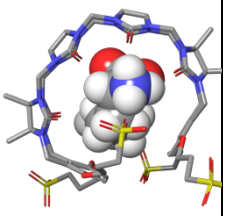
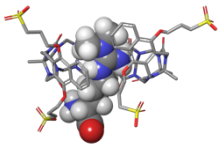
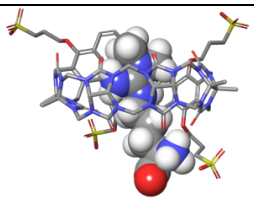
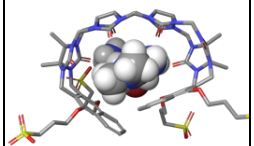
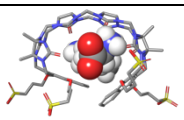
[a] ITC were carried out in 40 mM sodium phosphate buffer (pH 7.4). [b] ITC were carried out in 20 mM sodium phosphate buffer (pH 7.4). [c] ITC were carried out in 10 mM sodium phosphate buffer (pH 7.6).

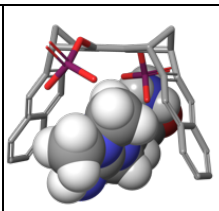
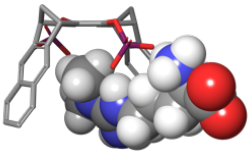
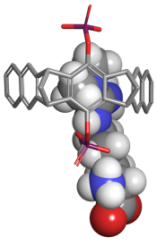
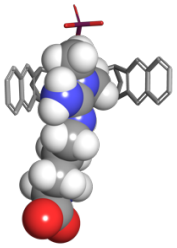
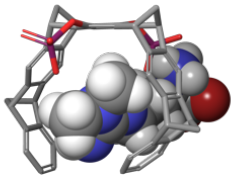
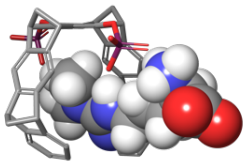
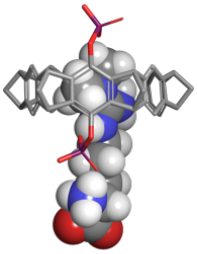
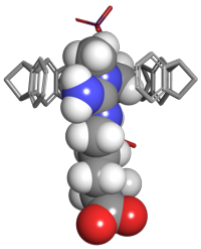
3.5.4. Molecular Modelling

Modelling was performed 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. Colour Code for atoms: C = gray, N = blue, H = white, O = red, P = purple, S = yellow. Each host is modelled with ADMA (**sCx4-CHO**, **sCx4-NO_2**, **M1**, **M2**, **PC**, **CLR01**). Hosts **M1**, **M2**, **PC**, **CLR01** are also modelled with MML and with TML.

3.5.4.1. Each host modelled with ADMA. Front, back, top, and bottom views for all complexes.

Table S3.4.Energy-minimized complexes of hosts with ADMA (OPLS_2005, explicit water)

	Front	Back	Top	Bottom
sCx4-CHO				
sCx4-NO_2				
M1				
M2				

PC				
CLR01				

3.5.4.2. Hosts M1, M2, PC, CLR01 each modelled with MML and with TML.

Table S3.5.Energy-minimized complexes of host **M1** with MML and TML (OPLS_2005, explicit water)

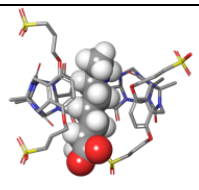
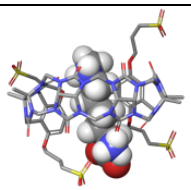
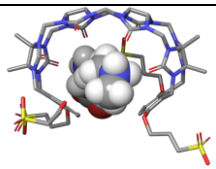
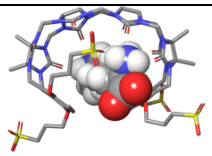
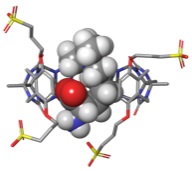
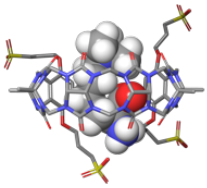
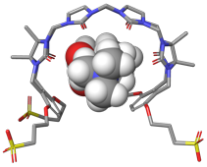
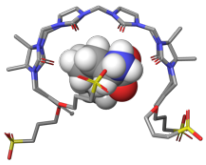
	Front	Back	Top	Bottom
MML				
TML				

Table S3.6. Energy-minimized complexes of host **M2** with MML and TML (OPLS_2005, explicit water)

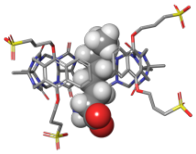
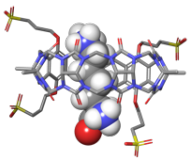
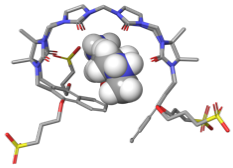
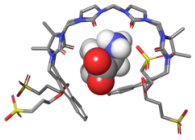
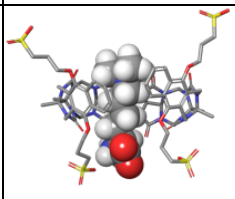
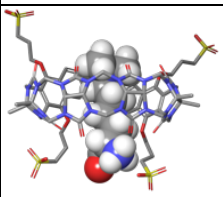
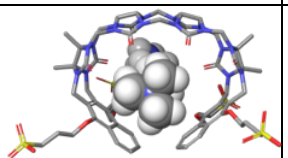
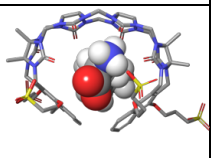
	Front	Back	Top	Bottom
MML				
TML				

Table S3.7. Energy-minimized complexes of host **PC** with MML and TML (OPLS_2005, explicit water)

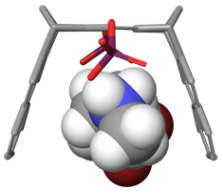
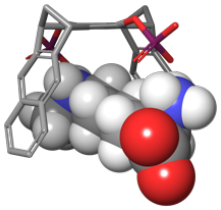
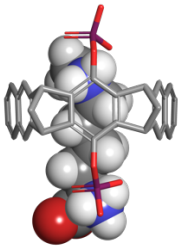
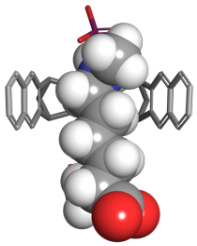
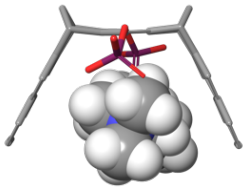
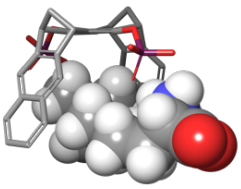
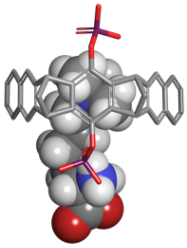
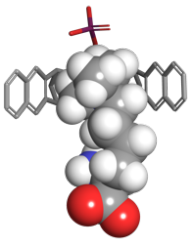
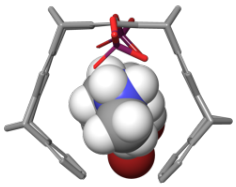
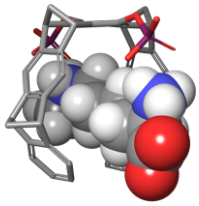
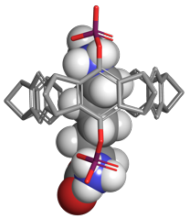
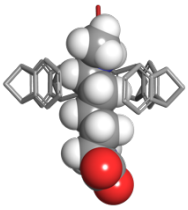
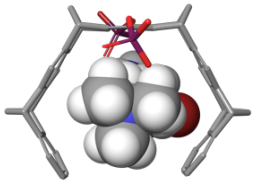
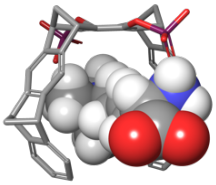
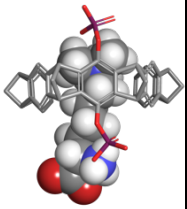
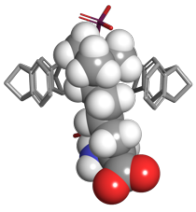
	Front	Back	Top	Bottom
MML				
TML				

Table S3.8. Energy-minimized complexes of host **CLR01** with MML and TML (OPLS_2005, explicit water)

	Front	Back	Top	Bottom
MML				
TML				

Chapter 4: Constructing large libraries of macrobicyclic calix[4]arene hosts using phage display

The work in Chapter 4 is unpublished. Zoey Warmerdam established the synthesis of compounds **sCx4-1,3-NH_2** and **sCx4-1,3-$NHCOCH_2Cl$**, as well as the phage display assays, IDA- and NMR studies. Claire Hodson assisted the phage display experiments by doing routine upkeep of stocks and experimental needs. Zoey Warmerdam performed routine upkeep of stocks and experimental needs, all phage manipulations, development of protocols, PCR, and sequencing data analysis. Arunika Ekanayake (University of Alberta) assisted in the phage sequencing process. Zoey Warmerdam wrote the chapter.

4.1 Foreword

Chapter 4 reports ambitious efforts to combine the key lessons covered in the previous chapters: high-throughput synthesis of host libraries is desired, and a more encapsulating host geometry leads to stronger binding interactions. In this chapter I developed a new method to make a library of potentially > 1 billion macrobicyclic hosts by combining phage displayed peptides with a calixarene structural element. This library is then put through an affinity pull-down with two targeted guests in an effort to identify hosts that have high affinities and selectivities for each of the guests. However, detailed studies showed that the selected hosts do not display the desired high guest affinities. A discussion on the potential reasons why I was unable to select a strong binding host from this library follows.

4.2 Introduction

The ability to design the perfect host for any given guest is a major challenge in supramolecular chemistry. Computationally predicting the geometries and energies of host-guest interactions in aqueous conditions is challenging. Unconscious bias and over-fitting of known data sets both contribute to disappointing experimental results when using established computational methods to predictively design host-guest binding energies.[167] This is widely experienced by supramolecular chemists, and is demonstrated concretely by the Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) challenges. The SAMPL challenges are a series of blind predictive challenges to test computational models on predicting properties related to protein-ligand and host-guest binding. A series of host-guest complexes are studied experimentally to determine their binding energies, and then laid out as a challenge for computational chemists to predict the as-yet undisclosed binding energies. This sort of true blind host-guest binding prediction exercises takes away the prejudice.

Prior to SAMPL6, no computational method had been developed that had acceptable predictive accuracy across a range of different host-guest interactions.[168,169] However the challenges SAMPL7[170] and SAMPL8[171] showed promising results. In SAMPL7 host-guest binding free energies for three host families (cucurbituril-derivatives, Gibb deep cavity cavitands (GDCCs) and modified cyclodextrins) and a range of guests were predicted.[170] The benchmark for a good performing method is for the root mean squared error (RMSE) to be under 3 kcal/mol across multiple host classes. During SAMPL7 this was achieved for the cucurbituril-derivative hosts. One method was found that could predict 7 of 8 free binding energies within a RMSE of 2 kcal/mol. For the

cyclodextrin derivatives, many methods got to a relatively low RMSE. From the 30 total submissions, 10 had an RMSE < 3 kcal/mol. In SAMPL8 absolute binding affinities for two host families (cucurbituril-derivatives and GDCCs) and a range of guests were predicted. During this challenge RMSEs of less than 1 kcal/mol and 1.75 kcal/mol were achieved for the GDCCs, but the cucurbituril-derivatives proofed more challenging all with RMSE greater than 4 kcal/mol. From the 51 total submissions, 14 had an RMSE < 3 kcal/mol.

Although progress is being made toward accurate predictions of host-guest interactions, there are still no accurate methods for computationally designing a new host that binds a targeted guest in water. The challenging problems with computational predictions lead to wasted synthetic effort when the designed molecular hosts do not perform as expected or predicted. Therefore, I stepped away from design, creating a host library using a genetically encoded peptide library to introduce structural diversity in the host binding pocket.

Peptides are an excellent source of diverse molecular structures. Peptides are defined as a chain of 2-50 amino acids, linked through peptide bonds. The diversity of each peptide comes from its constituent amino acids. In nature there are 20 commonly occurring (canonical) amino acids which could be classified into four classes; polar neutral, non-polar neutral, positively charged, and negatively charged.[172] Peptides can be derived from many sources, including plants, animals, humans, and peptides discovered from chemical libraries. Chemical peptide libraries can be classified into two groups: synthetic peptide libraries and biological peptide libraries.

Synthetic peptide libraries are achieved through combinatorial solid-phase peptide synthesis (SPPS).[173] In SPPS, a peptide is synthesized by attaching the first (C-terminal) amino acid of the peptide to a polymeric solid support (beads) and adding more amino acids sequentially building toward the N-terminus of the desired peptide sequence. The advantage of this technique is that, because the peptide is attached to the solid support, excess reagents can simply be washed off after each reaction. One way of creating libraries of peptides using SPPS is the split and pool method.[174] The split and pool method involves setting up multiple reaction flasks in parallel. The beads within each flask are coupled with a different amino acid before being combined, mixed, and then redivided (split) into different flasks again. This process is repeated until the desired peptide length is reached, resulting in a different peptide sequence on each bead. When a binding assay or panning study can be performed on the peptide-linked beads, then the selection of a given bead can be followed by identification of the specific peptide sequence on that bead, allowing for the discovery of high potency peptide binders of a given target. This method has been used to synthesize synthetic receptors,[175] drugs[176] and chemosensors,[177] among others, and these kinds of libraries are more generally called one-bead-one-compound (OBOC) libraries.[178]

Biological peptide libraries are achieved by genetically encoding the peptide libraries. The key to these libraries is the direct link between the genotype (genetic information) and phenotype (expression of the peptide). These libraries can be either acellular (systems based on in vitro transcription and translation) or cellular (libraries propagated in living cells). Examples of acellular approaches are ribosome display,[179] mRNA display[180], *cis*-activity based (CIS) display,[181] and covalent antibody display

(CAD).[182] Cellular libraries have been created in which the encoded peptides are displayed on the surfaces of bacteria,[183] yeast,[184] and filamentous bacteriophage (or simply "phage"). Of all these approaches to making biological peptide libraries, phage display is the most widely used. These phage display libraries are commercially available in forms containing millions-to-billions of different peptide sequences of unprotected natural amino acids that do not require chemical synthesis.[185]

Bacteriophage viruses were first described in 1917 by Felix d'Herelle, when he published a paper describing the lysing of bacteria by viruses that he named bacteriophage.[186] Bacteriophage ("phage") got their name after the Greek "*phageîn* – "to eat, devour". Phage are viruses that can infect and replicate in bacteria. The first therapeutic use of phage dates to 1919, where they were used for the treatment of dysentery. With the rise of antibiotics, the use of phage as therapeutic agents became disfavored by western scientists, although the upsurge of antibiotic-resistant bacterial strains has led to a rediscovery of this idea.[187,188]

The main modern use for phage is the generation of peptide libraries through phage display. Phage display is a method that provides a physical link between the genotype (inside) and phenotype (outside) of the phage virus, by encoding a foreign peptide gene into the phage genome, after which the peptide is displayed on the outside of the phage coat protein (Figure 4.1). In most common applications of phage display, filamentous phages are used since they are easy to grow in bacterial culture and, unlike many other phages, do not lyse the bacterial cells that they infect. In 2018, George P. Smith was awarded the Nobel prize in Chemistry for his work on phage display.[189] Smith

achieved the first example of phage display, by the installment of a fragment of the EcoRI protein on to the phage surface, after which an EcoRI antibody was used to show the efficiency of the expression.[190]

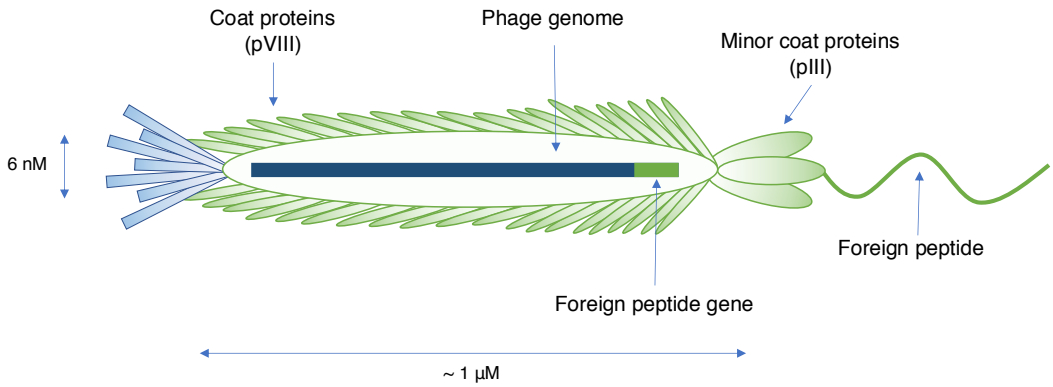
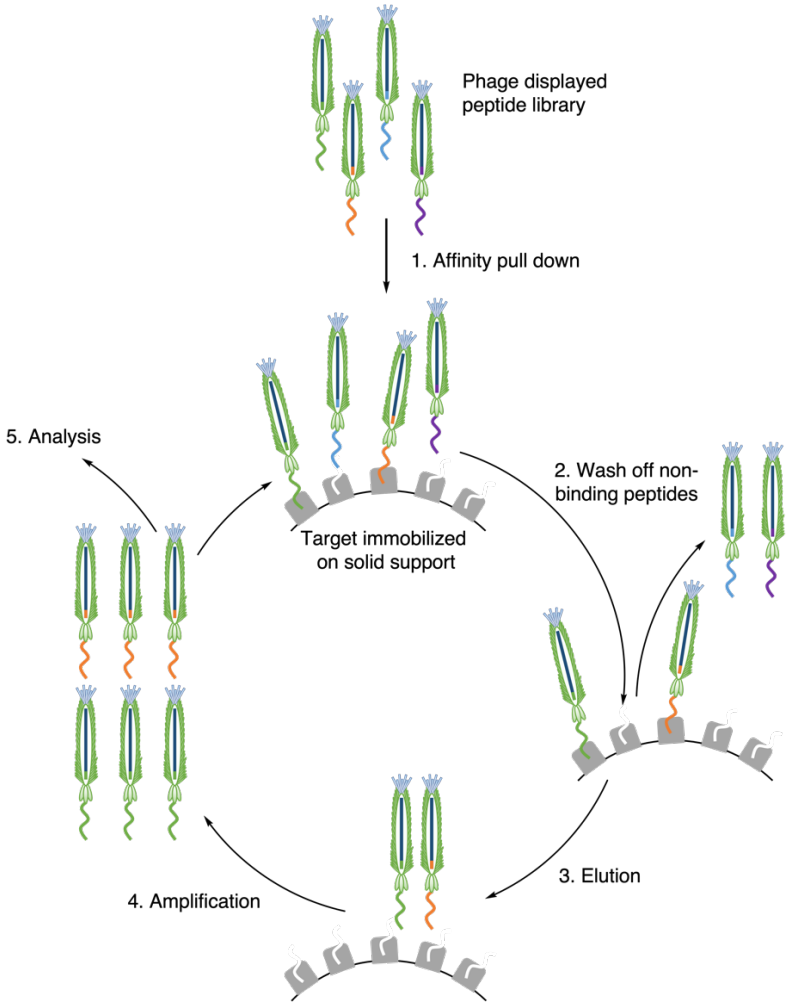


Figure 4.1. Schematic representation of a peptide fused to the minor coat protein (pIII) of M13 phage. Different viral vapsid proteins are depicted using different schematics. For clarity, only one copy of the foreign peptide is shown, but in reality there are five copies of the coat protein on the intact phage capsid that each carry a copy of the foreign peptide.[191]

The peptides present in a phage displayed peptide library can be selected for active binding to a target in an affinity pull-down assay. This happens in a sequence of actions called the phage display cycle (Figure 4.2).[192] The first step of the phage display cycle is the immobilization of the target molecule on solid support followed by the affinity pull-down of the phage displayed peptide library. Here, any peptides with an affinity for the target form a complex with the target and are immobilized. This is a dynamic exchange, where peptides (still tethered to their respective phage capsids) equilibrate on and off the target in a competitive matter. Phage displayed peptides that have weak or no affinity for the target are out competed by peptides that have a stronger affinity. The peptides that have weak or no affinity are washed off in step 2. The next steps in the phage display

cycle are eluting the higher affinity phage peptides off the target with a stronger elution condition and amplifying those phage viruses in *E. coli* to create larger amounts of the phage library members that display the most strongly binding peptides. After this bio-



panning the DNA from the selected and amplified phage is isolated and sequenced, revealing the sequences of the peptides that were selected.

Figure 4.2. Schematic diagram of the phage display cycle.

Unmodified peptides have little chance of forming strong complexes with small "guest" type molecules on their own. Peptides are potentially good ligands for larger targets like protein surfaces. But the flexible nature of a peptide means that it cannot form

a concave binding pocket, and therefore cannot envelope and form strong complexes with a small molecule guest. Phage display libraries on their own have not been useful for identifying binders of small organic molecules.

Chemical modifications on phage-displayed peptide libraries can be used to overcome shortcomings that are intrinsic to unmodified peptides. Genetically encoded peptide libraries are limited in their application due to their intrinsic properties; the peptides can only consist of the 20 natural amino acids, which limits their coverage of the chemical space. Additionally, the high rotational flexibility of the peptide backbone, leads to a high entropic cost upon binding.[193] These properties can be improved by chemically modifying the peptide libraries to generate more diverse libraries and introduce novel functionality. Chemical modifications of phage displayed peptides have been achieved in numerous ways[194], for example by installing pharmacophores,[195,196] photoswitches,[197] glycans,[198] unnatural amino acids,[199,200] and cross-linkers.[201-203] Heinis and co-workers reported on bicyclic peptides ligands synthesized by using a set of chemical linkers to cyclize phage peptide.[204] The phage displayed peptides used here have a sequence that contains three cysteines which can be reacted with a linker (Figure 4.3). With the phage introducing diversity and the linker introducing rigidity (compared to cyclic peptides), these bicyclic peptide ligands have been developed against a range of diverse target classes including enzymes, receptors, and cytokines with nano to picomolar affinities.

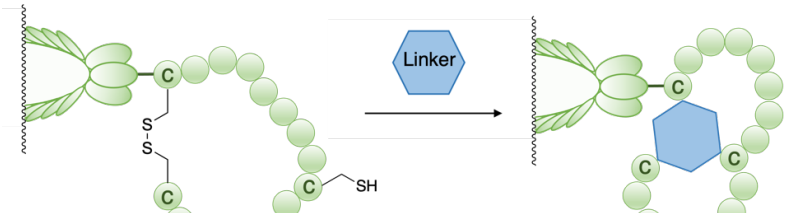


Figure 4.3. Bicyclic phage peptide

In Chapter 4 I will demonstrate a new approach in which libraries of calixarene-peptide hybrids are synthesized and screened on the surface of phage. I chose two small molecule targets: asymmetric dimethyl arginine (ADMA) as an important biomarker, and benzoylecgonine as a structural proxy for cocaine (which is also itself the main metabolite of cocaine that is present in the urine after cocaine use; Figure 4.4). While the short-term motivations here are about establishing a fundamentally new technique, each small molecule binding partner has a biological activity that is worth targeting in the longer term. A host with the ability to encapsulate ADMA could be developed into a therapeutic and/or diagnostic tools for heart disease. ADMA is a biomarker in cardiovascular disease.[137,205,206] When the ADMA and MMA levels are upregulated they inhibit nitric oxide synthase (NOS). The inhibition of NOS reduces NO-production causing vascular distress. ADMA plasma levels are 10-fold higher than the MMA plasma levels[136], therefore the focus in this chapter will be on ADMA.

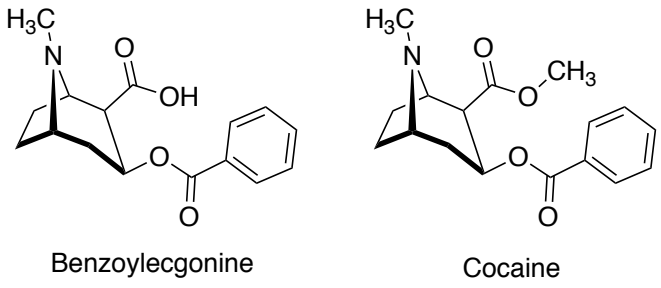


Figure 4.4. Structures of benzoylecgonine (left) and cocaine (right)

Supramolecular tools that can encapsulate illicit drugs could be instrumental in the ever-growing overdose epidemic. Supramolecular chemists have been able to design elaborate hosts that can detect or reverse toxins, drugs, biomarkers, and metabolites.[9,10] However, designing, synthesizing, and testing one unique host at a time is a major hindrance. With the ever-growing drug crisis, time is of the essence and phage display gives us access to hundreds of thousands of unique hosts at once. As a proof of concept, I chose to look at cocaine as a representation of illicit drugs. Cocaine is a recreational drug commonly found in overdoses.[207] Relative to other overdose agents like fentanyl and its analogs, cocaine and its proxy benzoyl ecgonine have the advantage of being more readily available for purchase in the significant quantities needed for scientific research.

The scaffold of calix[4]arene has been shown to bind small molecules like amino acids and drugs.[71,79,91,208,209] Modifications to the parent calix[4]arenes directly influence the binding pocket. However, small modifications, such as functional group changes and position do not adequately discriminate between structurally similar compounds.[152] From Chapter 3 we have learned that a more diverse and enclosed topology is desired in the pursuit of selective hosts.[210] In Chapter 4, libraries of macrobicyclic calix[4]arene hosts are synthesized by grafting a calixarene onto a commercially available Ph.D. C7C Phage Display Peptide Library. This library expresses peptides that consist of seven randomized amino acids flanked by a cysteine on each side (CX7C). The cysteines are soft nucleophiles that can react with soft electrophiles in biorthogonal reaction conditions.[211] I designed a calix[4]arene host containing chloroacetamide on two upper rim positions (**sCx4-1,3-NHCOCH$_2$Cl**, Figure 4.5), with the idea that the reaction of CX7C-peptides would generate a unique macrobicyclic calixarene host in which the peptide enclosing the

binding pocket is made up of diverse peptide sequences. These macrobicyclic calix[4]arene hosts have an increased binding surface area, compared to the parent host (**sCx4-1,3-NHCOCH$_2$Cl**), creating a host that surrounds the target. After the library synthesis is completed, an affinity pull-down assay with the targets is used to select strong binding hosts, as in the phage panning procedure shown in Figure 4.2.

4.3 Results and Discussion

Synthesis

sCx4-1,3-NHCOCH$_2$Cl was synthesized by reducing the previously published calixarene **sCx4-1,3-NO$_2$**[152] to the diamino calixarene **sCx4-1,3-NH$_2$** using Raney Nickel (Figure 4.5). The reduction was carried forward without purification and characterized in crude form with ^{1}H-NMR, and high-resolution mass spectrometry. **sCx4-1,3-NH$_2$** was reacted with chloroacetyl chloride in the presence of tertiary amine base to give **sCx4-1,3-NHCOCH$_2$Cl**. This product was purified via HPLC and characterized with ^{1}H-NMR, ^{13}C-NMR and high-resolution mass spectrometry see supporting information for experimental methods and characterization data.

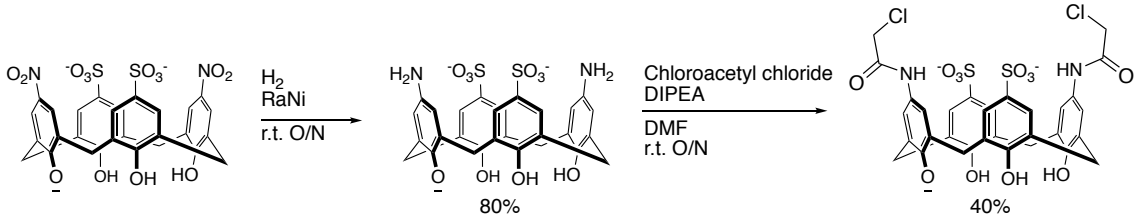


Figure 4.5. Synthesis scheme of **sCx4-NH$_2$** and **sCx4-1,3-NHCOCH$_2$Cl**

The two small molecule targets ADMA and benzoylecgonine were attached to a biotinylated bait molecule in order to set the stage for later attachment to a solid support.

The bait molecules are comprised of the general structure biotin-PEG3-target (Figure 4.6). For the ADMA bait molecule, ADMA and biotin-PEG3-NHS were dissolved in a HEPES buffer (pH 7.4) to allow the free amine of ADMA to react with the NHS ester. The reaction was diluted with water and purified via HPLC. For the benzoylecgonine bait molecule, benzoylecgonine, HBTU and DIPEA were premixed in dimethylformamide to activate the carboxylic acid on benzoylecgonine. Biotin-PEG3-NH_2 was added, the reaction was stirred at 60 °C overnight. The reactions purified via HPLC. See supporting information for experimental methods and characterization data.

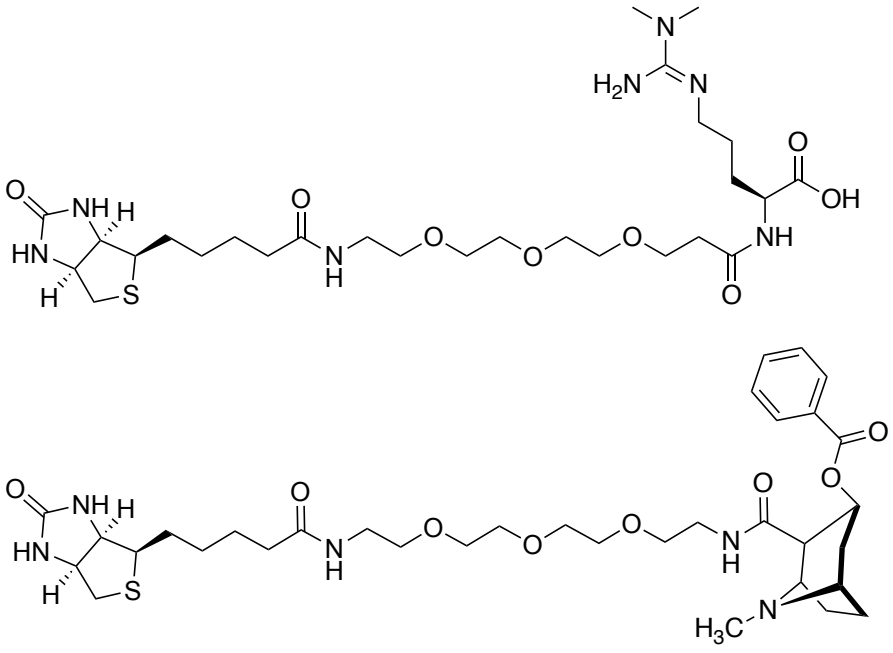


Figure 4.6. Bait molecules, **Biotin-PEG3-ADMA** (top), **Biotin-PEG3-benzoylecgonine** (bottom)

Phage display

The first step towards doing an affinity pull-down is alkylating the phage displayed peptide library with **sCx4-1,3-NHCOCH$_2$Cl**. The modification (alkylation) of

the phage displayed peptide library cannot be measured directly by any method that analyzes product formation (NMR, TLC, MS, etc), therefore a control experiment that measure the overall efficiency of the alkylation efficiency was performed. Two more control experiments were performed, simultaneously to the alkylation, to guarantee that each part of the alkylation experiment worked properly (Figure 4.7).

For the alkylation experiments the disulfide bridges formed between the cysteine residues on the phage peptide are first reduced with tris(2-carboxyethyl)phosphine (TCEP). After the reduction, the phage is incubated with **sCx4-1,3-NHCOCH$_2$Cl**, to react the now free cysteines on the phage peptides with the chloroacetamide on **sCx4-1,3-NHCOCH$_2$Cl**. After the reaction period has been completed, biotin iodoacetamide (BIA) is added, to react with any unreacted cysteines while simultaneously tagging them with biotin in order to allow monitoring of the amount of unmodified cysteine residues. This is followed by the incubation of the modified phage library on streptavidin beads. The streptavidin on the beads forms a strong complex with biotin, pulling all the biotinylated phages out of the solution, these are the phage peptides that had not or incompletely been alkylated with calixarene in the first phase of the reaction. This process is called "the capture". The phage mixtures before the capture and after the capture are titered in *E. coli* K12 and grown on agar plates overnight at 37 °C. The next day the plaque forming units (PFUs) are counted to determine the percent of alkylation by calixarene. For a successful alkylation, a high number of PFU is expected before and after the capture, as the cysteines should not be available to react with BIA since they should have already reacted with **sCx4-1,3-NHCOCH$_2$Cl**.

The first control is the BIA pull-down assay, to check the pull-down efficiency of the streptavidin beads. In this experiment the cysteines in the phage library are reduced with TCEP, reacted with BIA, and immediately incubated with the streptavidin beads. The phage mixtures before and after the capture are titered in *E. coli* K12 and grown on agar plates overnight at 37 °C. The next day the PFUs are counted to determine the maximum capture of the experiment that day. Here, a high number of PFU is expected before the capture and a low number of PFU after the capture. The percent capture from this control is the maximum capture for that day and is used to correct the following experiments.

The second control is for the viability of the phage on the day of the experiments and is also established by a BIA pull-down assay. The cysteines on the displayed peptides are reduced using TCEP. Then the reduced library is incubated with water instead of **sCx4-1,3-NHCOCH$_2$Cl**, as the blank control. After this, BIA is added to react with the free reduced cysteines. This mixture is then incubated on streptavidin beads. The phage mixtures before and after the capture are titered in *E. coli* K12 and grown on agar plates overnight at 37 °C. The next day the PFU are counted to calculate the maximum capture of the experiment that day. Here a high number of PFU is expected before the capture and a low number of PFU after the capture. This blank condition is used to correct for the loss in phage viability from going through the incubation periods during the alkylation. Full experimental details and calculations can be found in the supporting information.

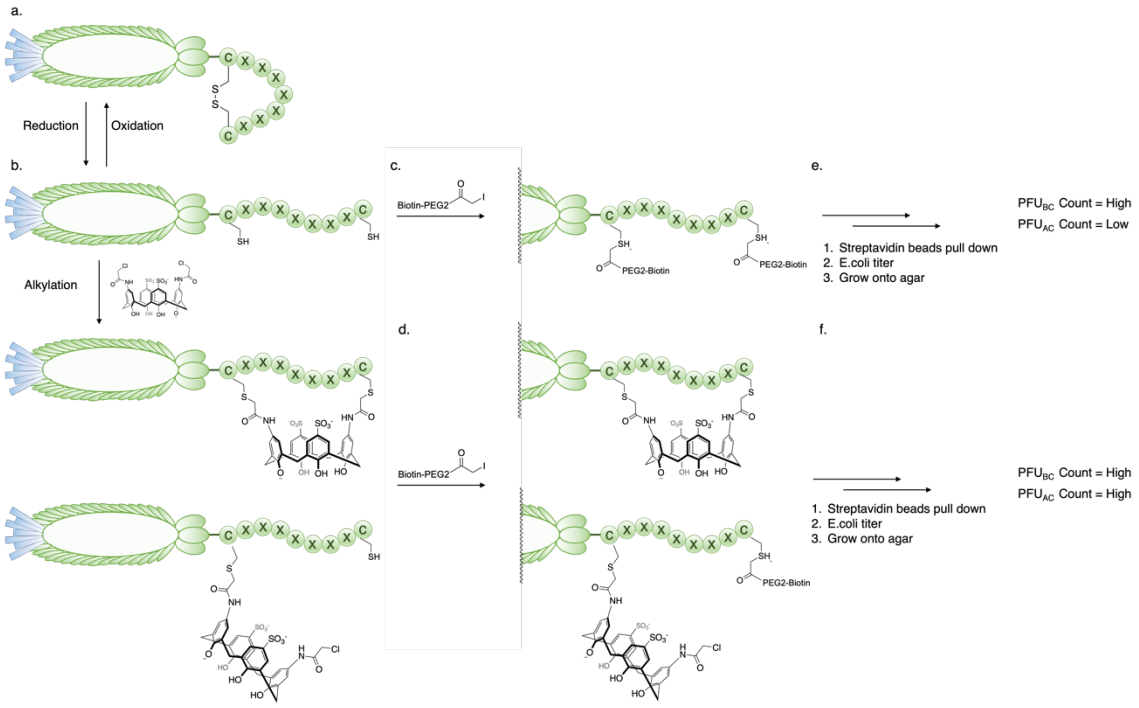


Figure 4.7. Experimental lay-out for the phage pull-down experiments, including controls. a) Reduction of the cysteines to create reactive thiol handles. b) Alkylation of the phage peptides with **sCx4-1,3-NHCOCH$_2$Cl**. c/d) Reacting the unoccupied cysteines with BIA. e/f) pull-down with streptavidin beads and E. coli growth.

Using the general set of experiments outlined above, I adjusted the reaction conditions to optimize the percent alkylation of the phage peptide library while also maximizing the viability of the phage themselves. Viability represents the ability of the phage to infect *E. coli*, and it can decrease dramatically under different chemical treatment conditions. Viability is important because the *E. coli* plaque count, PFUs, is used to determine the success of the alkylation experiments. The percent alkylation represents the amount of phage peptide present that were modified by **sCx4-1,3-NHCOCH$_2$Cl**. The starting conditions for our phage modification experiments were taken from literature that

used different phage alkylating agents.[197,212] For the optimization of the experimental conditions, a high viability and high alkylation percentage are desired.

The phage modifications and viability were optimized by altering the alkylation and reduction conditions. To improve the phage modification (percent alkylation), the alkylation reaction was pulsed with the reducing agent TCEP halfway through the reaction time with **sCx4-1,3-NHCOCH$_2$Cl**, to re-reduce any thiols that had spontaneously oxidized back to disulfides, and therefore increase the free cysteines available for alkylation by the calixarene. The addition of the extra TCEP pulse caused the percent alkylation to improve from 48% to 64%, while the viability decreased from 82% to 72% (Figure 4.8a). By using the TCEP pulse, the overall amount or alkylated phage was increased, so the pulse was implemented into the experimental procedure moving forward.

To further optimize the degree of alkylation the reduced phage was incubated using different concentrations (0.125, 0.25, 0.5, 1, 2 mmol) of **sCx4-1,3-NHCOCH$_2$Cl**, to determine the best compromise between improved phage alkylation and reduced phage viability. I observed that higher calixarene concentrations results in better alkylation efficiency but lower phage viability. This trend continues logically, at lower calixarene concentrations there is a lower percentage of alkylation and a higher phage viability. During the concentration studies I determined that 0.5 mM was the best compromise and had a 64% alkylation efficiency and 95% phage viability (Figure 4.8b). It must be noted that the alkylation experiment at a concentration of 0.5 mM with the TCEP pulse, over the course of optimizing all experimentation in this chapter, were repeatedly performed

in triplicate. An average of ten of these experiments showed that there was a 74% alkylation efficiency alongside 77% phage viability — a level of variability from the original concentration study that is not unexpected for these kinds of biological experiments.

The final condition that was monitored is the stability of the alkylated phage library, to see how long it could be stored and if it is possible to use the modified library over multiple days for the pull-down assay with the target molecules. For this experiment the alkylated phage library was stored at 4 °C and titered into *E. coli* over four days, day zero being the day of the alkylation experiment. Day zero was set to have a 100% count where after 24 hours the count drastically decreased to 54% on day one. The PFU on day two was 33% and on day three it had decreased to 12% (Figure 4.8c). Beside the count PFU going down, the plaques themselves also lost their intense blue color, indicating that the phage is not able to infect the *E. coli* anymore. I concluded that I would only attempt the affinity pull-down assay with the target molecules with libraries alkylated on the same day.

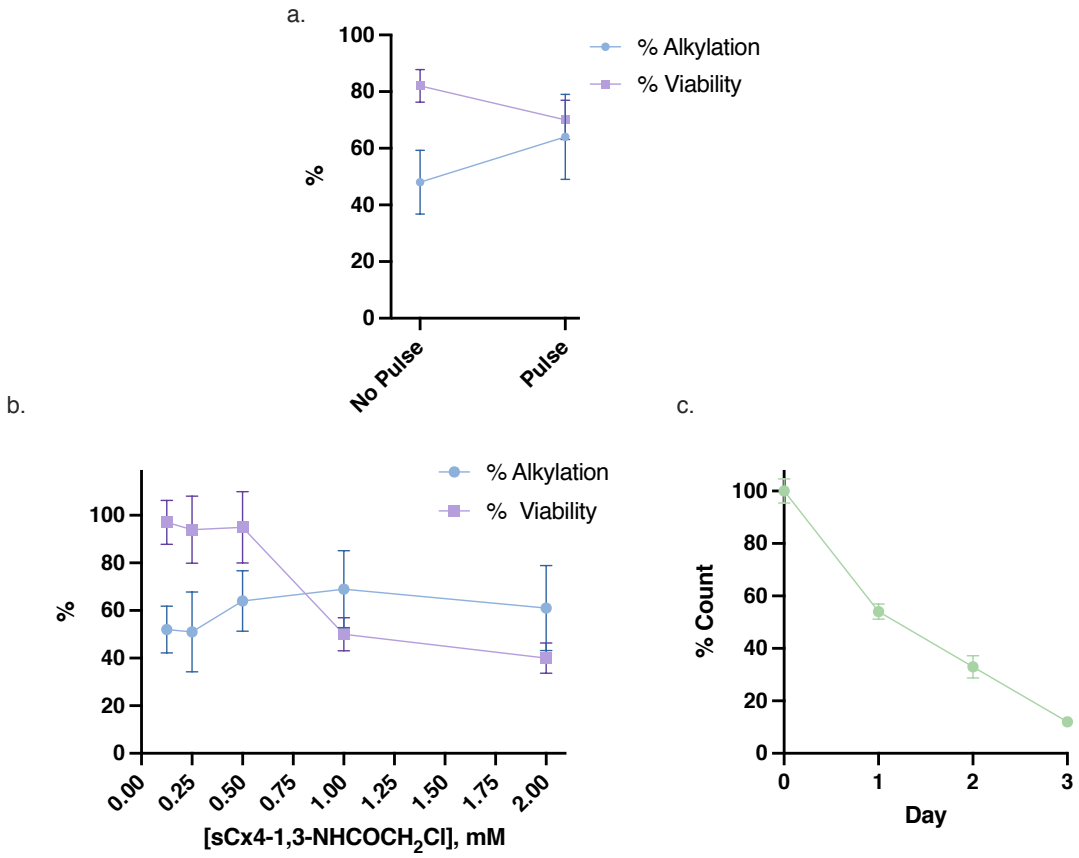


Figure 4.8. Optimization of the alkylation experiments. a) Percent alkylation and viability without and with a TCEP pulse b) Percent alkylation and viability over a concentration range of **sCx4-1,3-NHCOCH$_2$Cl** (0.125-2.00 mM) c) Stability of phage library over four days

Conditions for pull-down assays were individually established with ADMA and benzoylecgonine. After alkylating the phage peptide library, it is assumed that there are still many possible peptide sequences that could have formed hosts that bind the targets. To identify the strong binding sequences, an affinity pull-down assay was established. In this assay streptavidin magnetic beads are covered with bait molecules (for experimental details see supporting information). The modified beads are incubated with the alkylated library, a reversible exchange between the hosts and the target takes place until an

equilibrium is reached where the stronger binding hosts out-compete the weaker. The bound hosts stay on the beads, while the unbound hosts are washed off and discarded. The bound hosts are eluted by using heat, this also denatures the protein phage coating, thus releasing the DNA (Figure 4.9). The DNA was isolated and sent for sequencing.

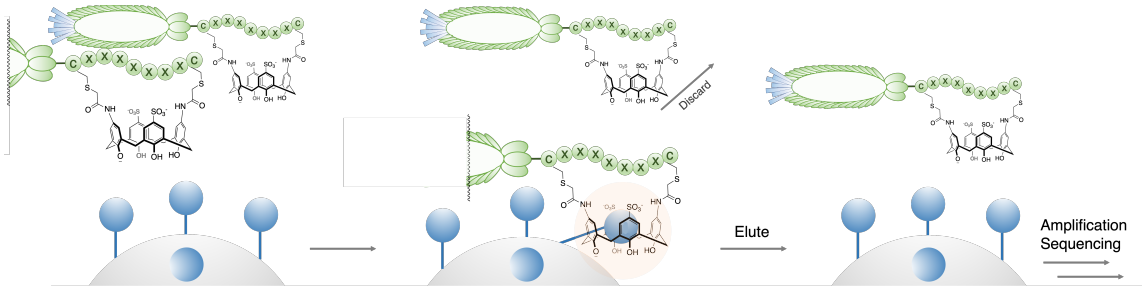


Figure 4.9. Affinity pull-down assay with the target on a solid support

The DNA was sequenced using next-generation sequencing.[213] To correct for unselective binding during the affinity pull-down assay, the DNA sequence counts were corrected by subtracting the peptide sequences collected from two control experiments. The first control experiment is an affinity pull-down assay performed on unmodified phage. Any peptide sequences resulting from this experiment are from the unmodified peptide and the binding was not induced by alkylation with **sCx4-1,3-NHCOCH$_2$Cl**. The second control experiment was an affinity pull-down assay performed on modified phages, with unmodified streptavidin beads. Any peptide sequences resulting from this experiment show binding to the bead rather than the target. After the corrections a list of

~3000 peptides sequences for each target were amplified in some measurable amount. For experimental details and data see supporting information.

The sequencing was performed by 48Hour Discovery (University of Alberta startup, Prof. Ratmir Derda) and the data was analyzed in collaboration with Wenrui Huang, the 48Hour Discovery technician. Each target (ADMA and benzoylecgonine) had approximately 3000 sequences amplified (see supporting information). From these sequences, those that were amplified were considered for resynthesis and follow-up confirmatory testing. When selecting the peptides, the degree of amplification and their likely water solubility were considered. If the degree of amplification was similar, peptides with more polar amino acids were chosen over peptides containing more non-polar amino acids. For the ADMA bait pull-downs, the peptide sequences **ACNGWPGASC**, **ACDRSTTKIC**, **ACEPRSLANC**, **ACNTGSPYEC**, **ACTLTTFRSC**, were selected. For the cocaine bait pull-downs, the peptide sequences **ACASPNYTVC**, **ACSNYSGPFC**, **ACNSSKLHMC** were selected. After the peptide selection, to help with the solubility, the peptides were modified to include one serine and two aspartic acid amino acids on the C-terminus, outside of the CX7C motif.

To study the binding strength of the selected peptides the calixarene-peptide-hybrids were synthesized. The peptides were ordered from GenScript and cyclized using **sCx4-1,3-NHCOCH$_2$Cl**, see supporting information. The cyclization conditions were based on the previously published cryptand-peptide cyclization by Taki et al.[53] Our macrocyclizations were conducted in a phosphate buffer (10 mM, pH 7.4), where the calixarene (500 μM, 5 eq) and neutralized TCEP (500 μM, 5 eq) were available in excess

relative to the peptide (100 μM, 1 eq)(Figure 4.10). This is important to stimulate the formation of 1:1 calixarene-peptide complexes and prevent larger aggregates. This was also promoted by preforming the reaction in dilute conditions. The reactions were stirred overnight, and their progress was monitored using UPLC-MS. From the eight selected sequences, six were successfully cyclized. Under the experimental conditions used **ACSNYSGPFCSDD** was not soluble and **ACNTGSPYECSDD** did not show any sign of cyclization with **sCx4-1,3-NHCOCH$_2$Cl**. After the initial cyclisation, the reactions were scaled up to a scale more suitable for purification. After synthesis and purification, four calixarene-peptide hosts were synthesized on a scale suitable for binding studies, **sCx4-P1** (**sCx4-ACNGWPGASCSDD**), **sCx4-P2** (**sCx4-ACDRSTTKICSDD**), **sCx4-P3** (**sCx4-ACEPRSLANCSDD**) and **sCx4-P4** (**sCx4-ACASPNYTVCSDD**). Figure 4.10 show the final structure of the hosts, **sCx4-P1** is shown as an example.

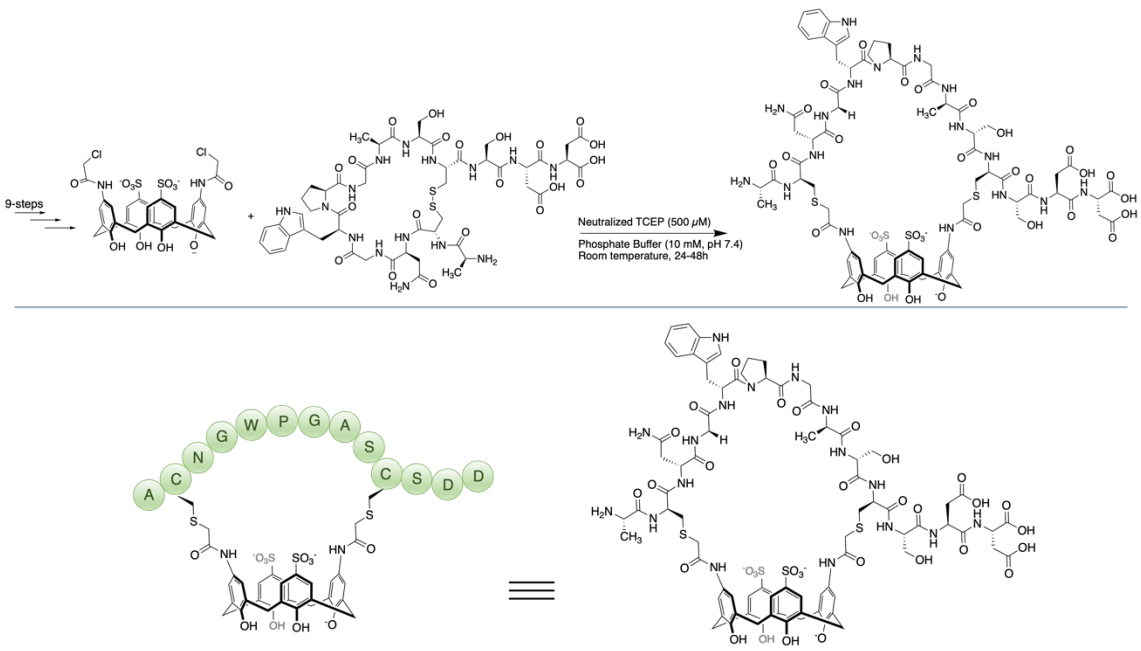


Figure 4.10. Synthetic scheme **sCx4-P1** (top), cartoon style and chemical drawing of **sCx4-P1** (bottom)

The binding constants of the macrobicyclic hosts were determined *via* indicator displacement assays (IDAs). We adapted previously reported IDAs for **sCx4**. [151,152] Each host was first titrated directly into lucigenin (LCG) to determine the host-dye dissociation constant (K_{ind}) values, which range from, 13.9–35.2 μM (see supporting information). This was followed by competitive titrations, where the guest competes with LCG, displacing it from the binding pocket, resulting in an increased fluorescence emission.[105,151] The fluorescence was measured and plotted against the concentration of the guest, and the resulting curve was fitted to determine the binding constant for each host-guest pair, see supportive information.

Unfortunately, only weak or no binding was observed between the new macrobicyclic hosts and the guests (choline, arginine (Arg), asymmetric dimethylarginine (ADMA), benzoylecgonine, and cocaine). The only host-guest pair that showed any measurable binding under these conditions was **sCx4-P1** to cocaine ($K_d = 236 \pm 115$ μM) (Figure 4.11 and supporting information). This host was selected in an affinity pull-down assay with ADMA as the target, therefore it is presumed to be just a coincidence that it shows affinity for cocaine and not a genuine example of a successful selection from the library.

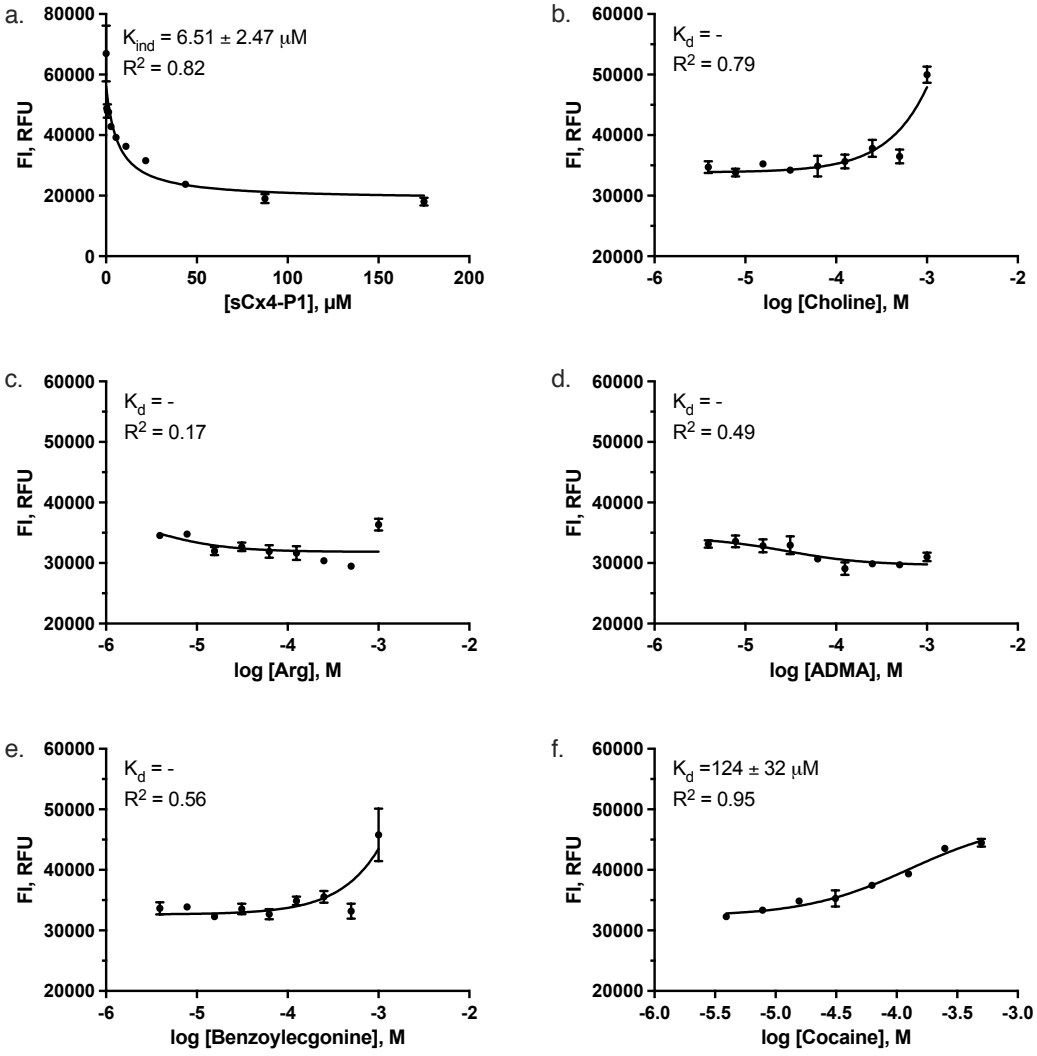


Figure 4.11.Exemplary dataset of fluorescence-based studies of **sCx4-P1**. a) Direct titration of LCG (0.25 μM) with **sCx4-P1** (0 – 175 μM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into **sCx4-P1**

Based on the IDA results, combined with the large number of amplified peptides, it is evident that the affinity pull-down needs to be improved. The selected peptide-calixarene hosts do not display a selectivity or binding to the targets they were selected with. To improve the affinity pull-down assay in future experiments, multiple rounds of

bio-panning should be performed. The repeating of the bio-panning steps (e.g. 3-6 times) in a phage display cycle, will lead to a better selection of strong-binding peptides.[214]

The binding structure between **sCx4-P1** and cocaine was studied by NMR to acquire insight into the reasons for the weak complexation. In this study we compare the 1D-^{1}H-NMR spectra of **sCx4-P1** on its own and **sCx4-P1** with cocaine. Looking at **sCx4-P1**, the peptide structure creates a complex NMR spectrum with many overlapping resonances, but the well-known 1D chemical shifts of amino acid resonances (supplemented with COSY data) allowed for the confident assignment of some signals. The spectra are very complex but based on the ^{1}H-NMR and COSY correlations we were able to assign the methyl protons from two alanine (A) residues, as well as the proline (P) protons, see supporting information. NOESY NMR shows through space interactions, instead of through bonds, helping with determining if molecules are in proximity to each other (generally <5 Å). The NOESY spectra provides further information on the conformation of the **sCx4-P1** (Figure 4.12). We see extensive correlations of the aliphatic peptide protons (4.7 - 2.5 ppm) correlating with resonances in the aromatic region, suggesting that the peptide is undergoing a hydrophobic collapse into the calix[4]are binding pocket. Looking at the protons that are assigned, we see that only the protons of one alanine show NOE correlations in the aromatic region as well as all three methylene's from the proline side chain. When looking at just the aromatic region (8.0- 6.4 ppm), there are additional NOE correlations present indicating that the tryptophan side chain is also in close contact with the calixarene's binding pocket. The collective evidence for close contact between these protons with the binding pocket gives support for an overall

hydrophobic collapse, which would explain the weak binding from our targeted small molecule guests. When we compare our macrobicyclic host to the host seen in earlier chapters the guests did not have to compete with a peptide chain, therefore there was no competition for a place in the binding pocket.

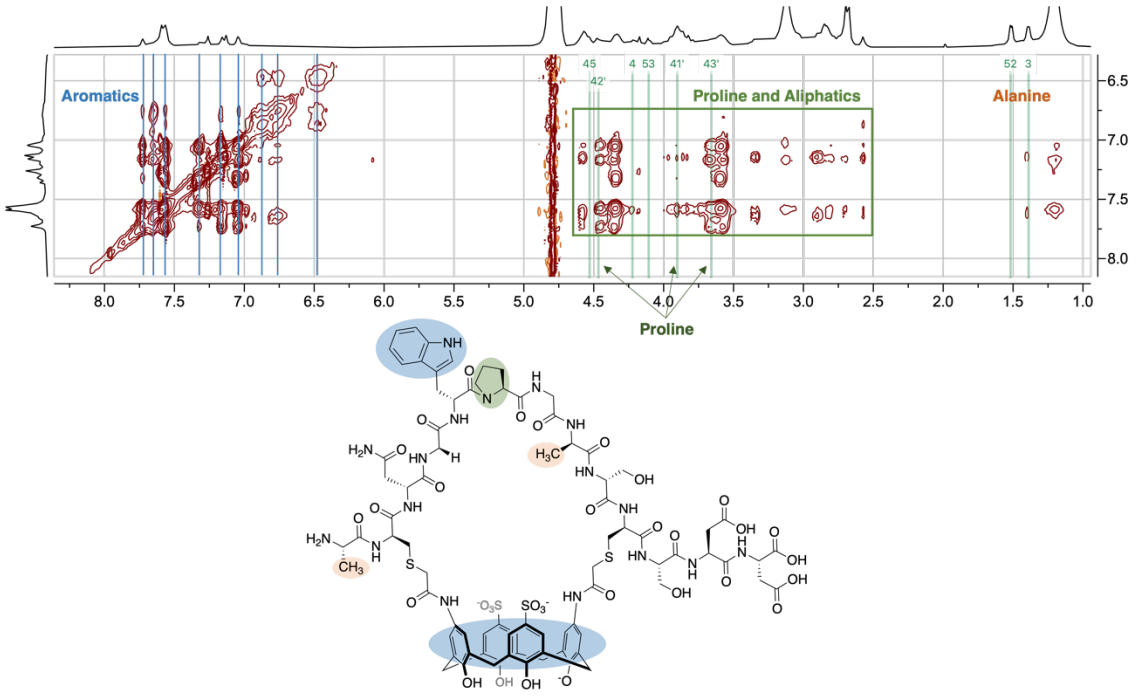


Figure 4.12. NOESY **sCx4-P1**, zoomed in on aromatic region. The green vertical line are the assigned protons of **sCx4-P1** (proline and alanine protons), the arrows are highlighting the proline protons and the box is highlighting the aliphatic peptide protons (4.6-2.5 ppm). The alanine protons are downfield around 1.5 ppm only one alanine has a weak NOE signal with the aromatic region. The blue lines are showing the aromatic protons (8.0-6.5 ppm), not assigned in the structure.

Next, we investigated **sCx4-P1** and cocaine together, expecting to see a shift in the ^{1}H-NMR peaks of the cocaine protons. This shift is expected to go upfield, as the protons are shielded by the electron cloud of the arene units of the calix[4]arene as the cocaine is complexed.[110,215] We also expect to see the peaks broaden as the cocaine protons

are now complexed with **sCx4-P1**, larger complexes rotate slower that small molecules, changing their exchange on the NMR time scale, broadening the peaks.[216] For the ^{1}H-NMR we focused on cocaine since it was present in excess and all off its corresponding peaks were assigned, making it easier to assign the changes observed.

Figure 4.13 shows zoomed-in stacked plots of **sCx4-P1** (bottom) and **sCx4-P1** with cocaine (top), see supporting information for the full spectra. As expected, we see the protons shift upfield to varying degrees. For the aromatic protons we only observe a small chemical shift difference(~ 0.01 ppm) (Figure 4.13a), indicating that these protons are not experiencing shielding to a large degree, the peaks do however broaden, showing that cocaine complexed with **sCx4-P1**. The protons on the cycloheptane moiety are experiencing the biggest chemical shift differences (~0.04-0.14 ppm) indicating that these protons are shielded the most. It should also be noted that the protons of the methyl group on the nitrogen are shifted, ~0.02 ppm (Figure 4.13e) more than the protons of the methyl group on the ester moiety, ~-0.01 ppm (Figure 4.13d), confirming this geometry. The protons of the methyl group on the ester are shifting downfield, and this deshielding indicates that the methyl ester is oriented away from the calixarene's electron-rich binding pocket. Extensive NOESY and ROESY experiments with **sCx4-P1** complexed with cocaine failed to reveal useful information.

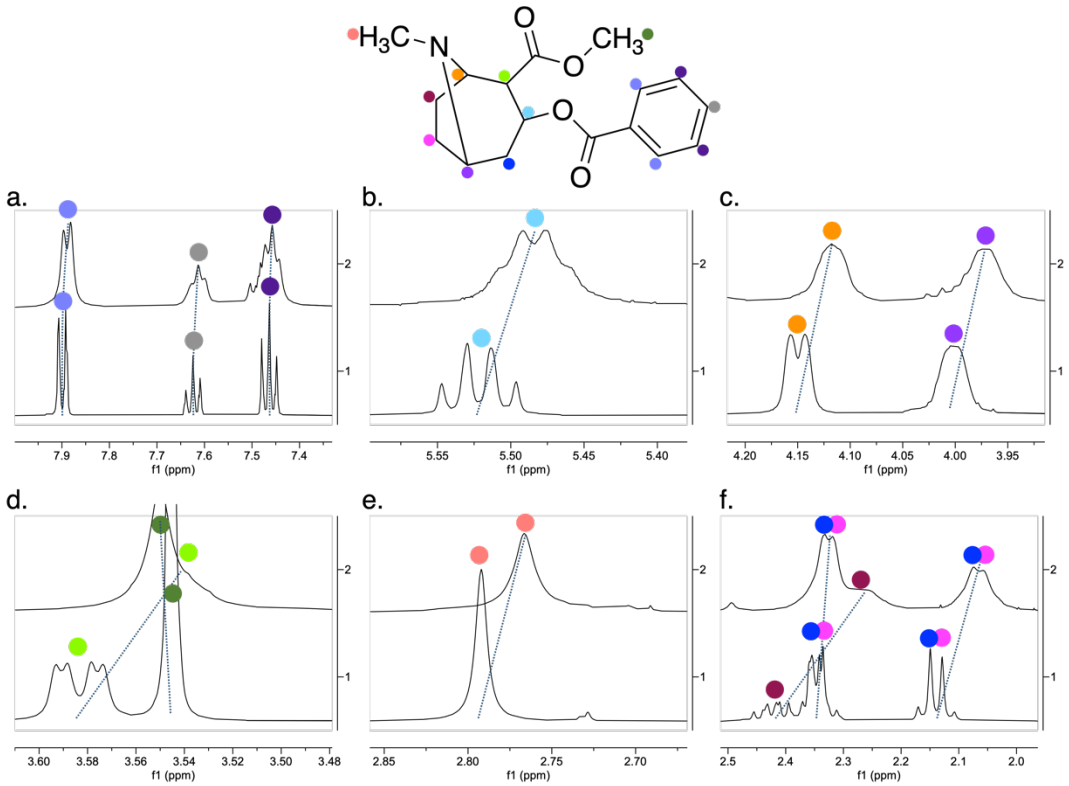


Figure 4.13. Zoomed in stacked plots of **sCx4-P1** (bottom) and **sCx4-P1** with cocaine (top) with the cocaine protons labelled

4.4 Conclusion

In this chapter we showed our efforts towards the synthesis of a diverse host library. This was achieved by the synthesis of a novel calixarene host **sCx4-1,3-NHCOCH$_2$Cl** combined with phage display. We successfully established the methods to modify phage displayed peptide libraries. This was achieved through the alkylation of the cysteine-containing peptides with **sCx4-1,3-NHCOCH$_2$Cl** creating a library of macrobicyclic calixarene hosts attached to phage. The hosts from this library then underwent an affinity pull-down assay with the targets of our interest: ADMA and benzoylecgonine, narrowing the library from over a billion possible hosts to thousands of

hosts. From these hosts we selected the peptide sequences that were amplified the most, compared to our library before the affinity pull-down assay, and resynthesized them for further binding studies. However, IDA experiments revealed that these hosts did not have the expected binding properties for the targeted guests, showing that we were unsuccessful in the selection of a host for the desired targets. This shows that we need to improve our affinity pull-down to achieve the eventual goal of selecting a strong binding host from the library for a large array of targets. We did by chance find a host, **sCx4-P1**, that bound cocaine with a K_d of 236 μM. To gain a further understanding on the geometry of this host-guest complex we carried out detailed NMR studies. The NMR studies showed that **sCx4-P1** experiences a hydrophobic collapse of the peptide into the calix[4]arene binding pocket in the absence of guest. This limits the accessibility for other guests as they must outcompete the intramolecular peptide chain for binding. During the studies of **sCx4-P1** with cocaine, we confirmed that complexation does take place (albeit weakly), in addition to gaining further insides insights on the binding geometry.

4.5 Supporting information

4.5.1 General information

The ^{1}H-NMR and ^{13}C-NMR spectra for **sCx4-1,3-NH_2** and **sCx4-1,3-$NHCOCH_2Cl$** were recorded on a Bruker AV300, 300 MHz spectrometer. The ^{1}H-NMR, ^{13}C-NMR, COSY, NOESY and ROESY spectra were recorded on a Bruker NEO500, 500 MHz spectrometer. All NMR data is processed with MestReNova by Mestrelab Research S.L. All reported chemical shifts were reported in ppm. Deuterated solvents were purchased from Sigma Aldrich. The UV traces were recorded on a Waters UPLC-DAD-MS over a 50 mm x 2.1 mm Waters, C18 1.7 u with UV detection at 280 nm, flowrate 0.5 mL/min. The gradient ran at 90% H_2O (+0.1% FA)/10% CH_3CN (+0.1% FA) for 1 min then went to 10% H_2O (+0.1% FA)/90% CH_3CN (+0.1% FA) over 5 minutes. Then to 90% H_2O (+0.1% FA)/90% CH_3CN (+0.1% FA) in 1 min hold for 1 min.

4.5.2 Experimental

4.5.2.1. sCx4-1,3-NH_2

sCx4-1,3-NO_2 (0.0589 g, 0.088 mmol) was dissolved in water (10 mL). 1 M NaOH solution was added till pH 8. The system was bubbled with a nitrogen balloon, Raney-nickel was added (cat.) the system was bubbled with a nitrogen balloon. The solution was bubbled with hydrogen gas overnight at room temperature. The solution was filtered over celite to remove the catalyst and lyophilized, yielding **sCx4-1,3NH_2** as a green/brown solid (0.0392 g, 73% by NMR) that was carried forward without further purification. ^{1}H-NMR (300 MHz, D_2O): δ 7.70 (s, 4H), 6.35 (s, 4H), 3.85 (d, 8H). Melting point >230°C. HR-MS (MH^+(-1) *m/z*): Calculated for $C_{28}H_{26}N_2O_{10}S_2$, 614.09561, Found 614.10289. FT-IR (cm^{-1}): 3385 (br), 1615 (w), 1472 (m), 1436 (m), 1172 (s), 1102 (m), 1040 (s).

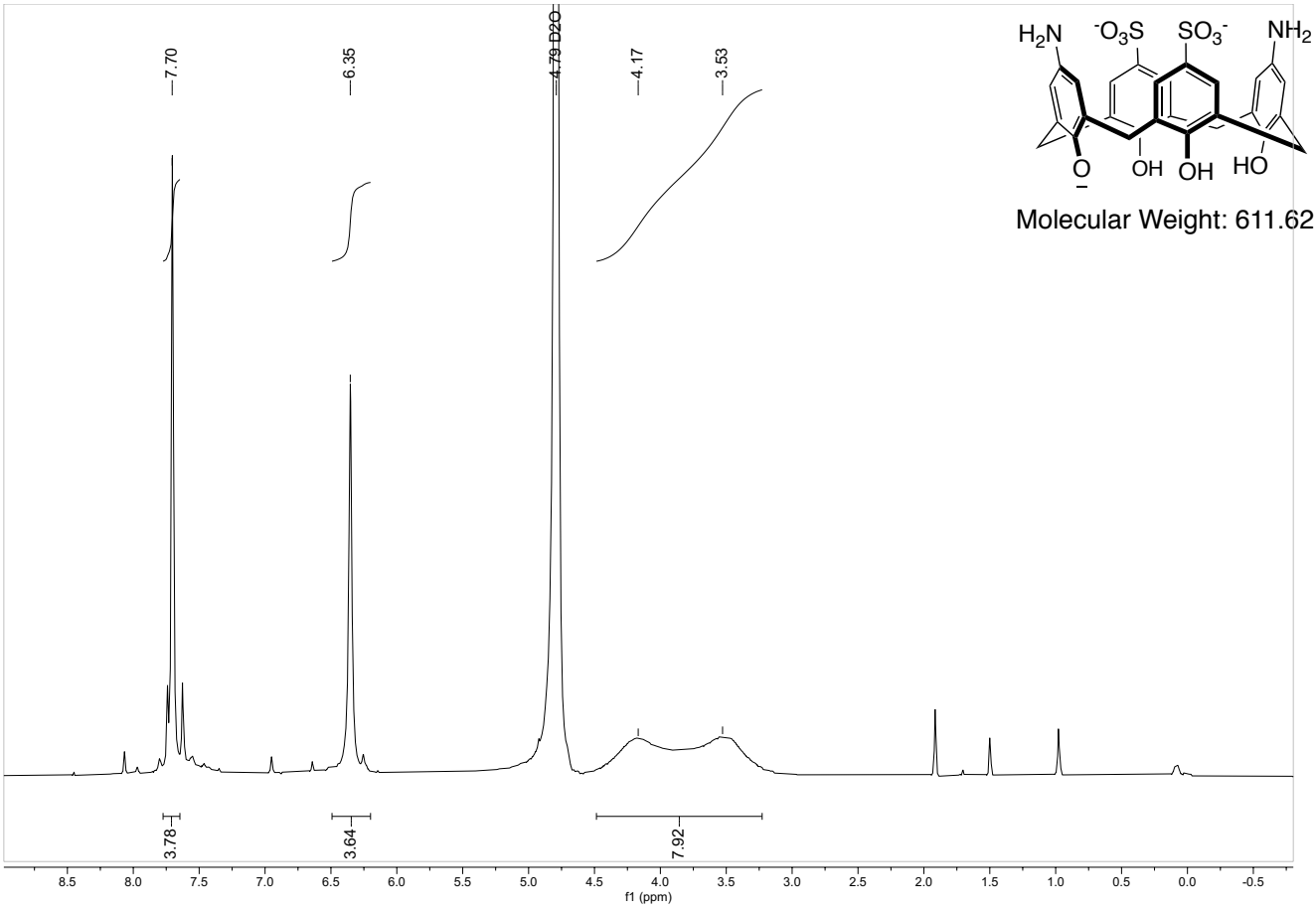


Figure S4.1. ^{1}H-NMR spectrum of **sCx4-1.3-NH_2** (300 MHz) in D_2O

4.5.2.2. sCx4-1,3-NHCOCH$_2$Cl

sCx4-1,3-NH$_2$ (0.0347 g, 56 mmol) was dissolved in DMF (minimal). DIPEA was added till the solution became clear. Chloroacetylchloride (0.063 g, 10 eq) was added. The reaction was stirred overnight at room temperature. The reaction was diluted with water and purified by high pressure liquid chromatography on an ACCQPrep Waters HPLC system over a 9.4 mm x 250 mm semi-preparative Agilent Eclipse XDB-C18 5 μm with UV detection at 280 nm, flowrate 7 mL/min. The gradient ran at 90% H_2O (+0.1% TFA)/10% CH_3CN (+0.1% TFA) for 1 min then went to 65% H_2O (+0.1% TFA)/35% CH_3CN (+0.1% TFA) over 13 minutes. Then to 10% H_2O (+0.1% TFA)/90% CH_3CN (+0.1% TFA) in 1 min hold for 1 min. Then to 90% H_2O (+0.1% TFA)/10% CH_3CN (+0.1% TFA) over 0.5 minute and hold for 1 minute. The fractions were collected and lyophilized, yielding **sCx4-1,3-NHCOCH$_2$Cl** as a light-yellow solid (0.0134 g, 31%). ^{1}H-NMR (300 MHz, D_2O): 7.64 (s, 4H), 7.11 (s, 4H), 4.04 (s, 4H), 3.95 (s, 8H). HR-MS (MH^+(-1) *m/z*): Calculated for $C_{32}H_{28}C_{l2}N_2O_{12}S_2$, 766.03879, Found 766.04607.

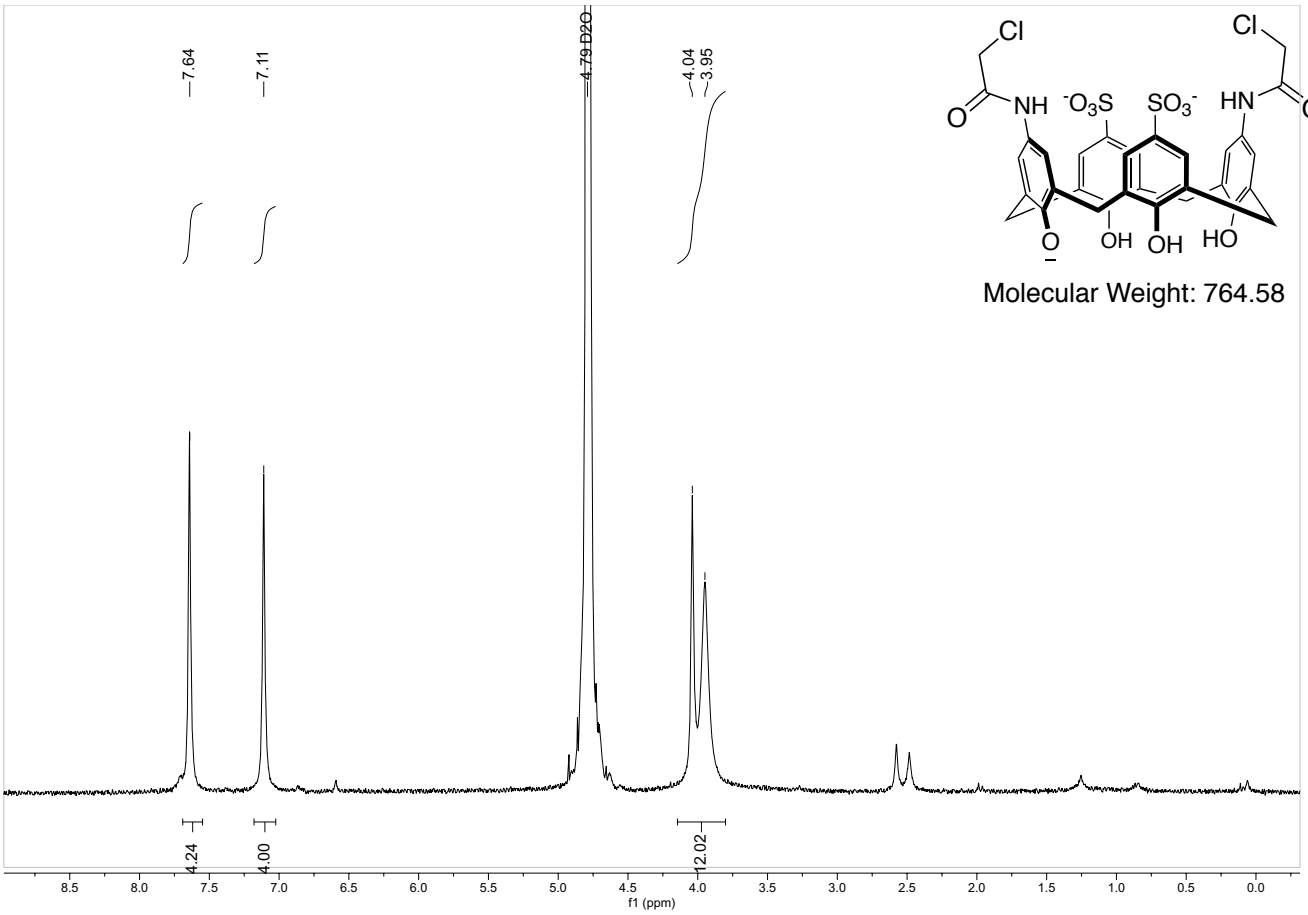


Figure S4.2. ^{1}H-NMR spectrum of **sCx4-1.3-NHCOCH$_2$Cl** (300 MHz) in D_2O

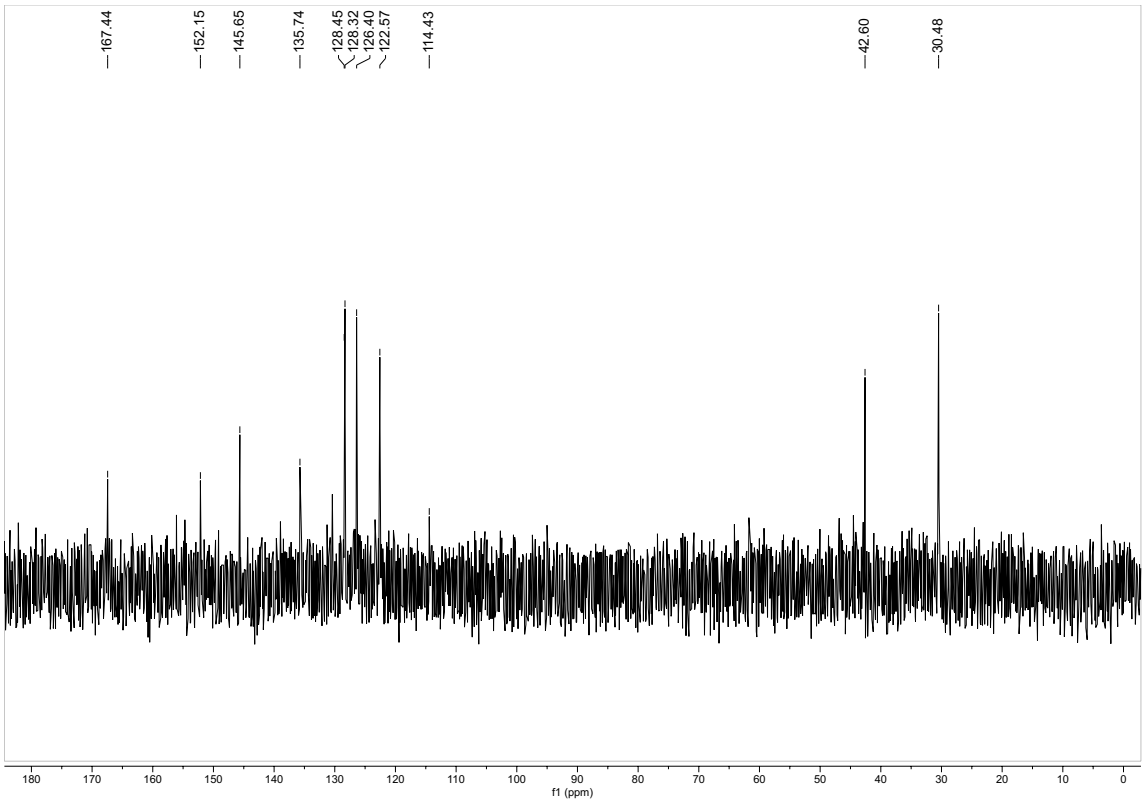
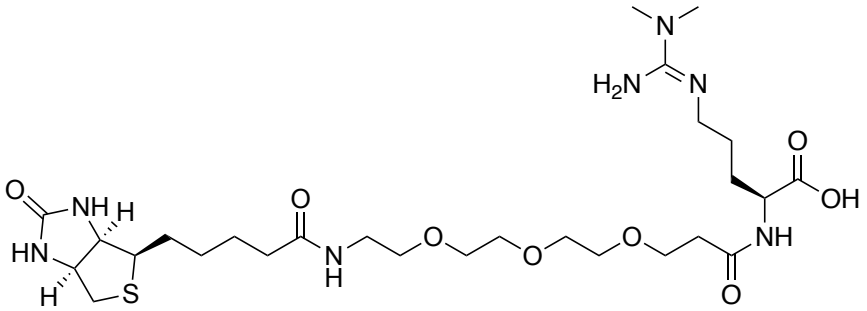


Figure S4.3. ^{13}C-NMR spectrum of **sCx4-1.3-NHCOCH$_2$Cl** (500 MHz) in D_2O

4.5.2.3. Biotin-PEG3-ADMA

ADMA (0.0253 g, mmol) and biotin-PEG3-NHS (0.0518 g, 0.0951mmol) were dissolved in HEPES buffer (pH 8.4) and stirred for 4 hours. The reaction was diluted with water and purified by high pressure liquid chromatography on an ACCQPrep Waters HPLC system over a 9.4 mm x 250 mm semi-preparative Agilent Eclipse XDB-C18 5 μm with UV detection at 280 nm, flowrate 7 mL/min. The gradient ran at 95% H_2O (+0.1% TFA)/5% CH_3CN (+0.1% TFA) for 8 min then to 87.5% H_2O (+0.1% TFA)/12.5% CH_3CN (+0.1% TFA) over 8 minutes, then to 10% H_2O (+0.1% TFA)/90% CH_3CN (+0.1% TFA) over 1 minutes. Hold at 10% H_2O (+0.1% TFA)/90% CH_3CN (+0.1% TFA) for 1 min, then to 50% H_2O (+0.1% TFA)/50% CH_3CN (+0.1% TFA) over 1 minute and hold for 1 minute. The fractions were collected and lyophilized, yielding **Biotin-PEG3-ADMA** as a white solid (0.0212 g, 33%).



Molecular Weight: 631.79

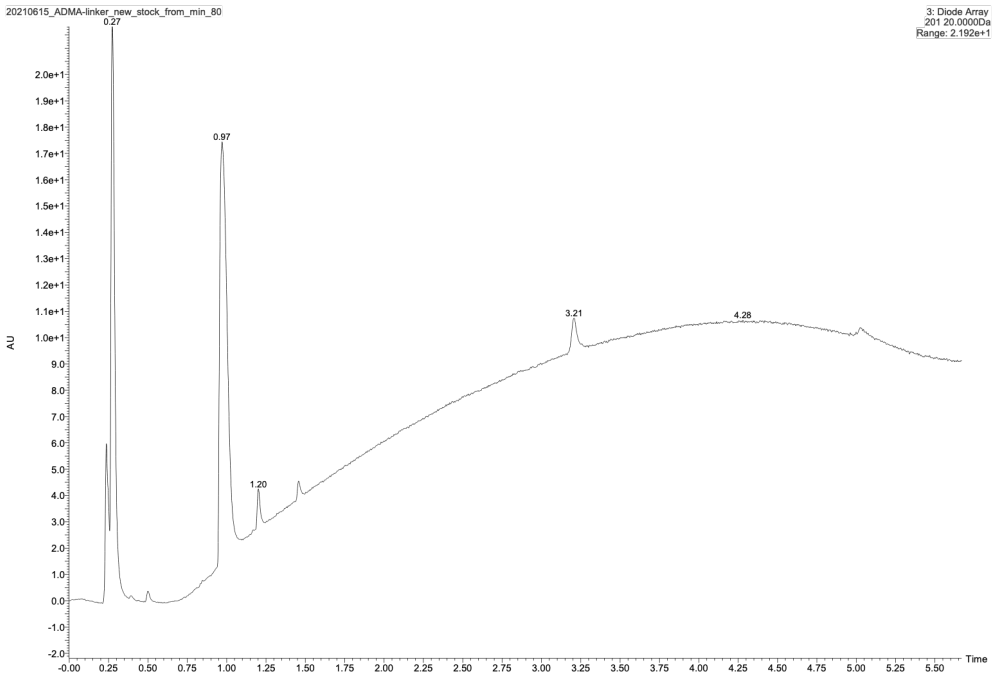


Figure S4.4. Diode array 201 nm for **Biotin-PEG3-ADMA**. Injection peak at 0.27. Product peak at 0.97 min, by-products 1.30, 1.52 and 3.21 min.

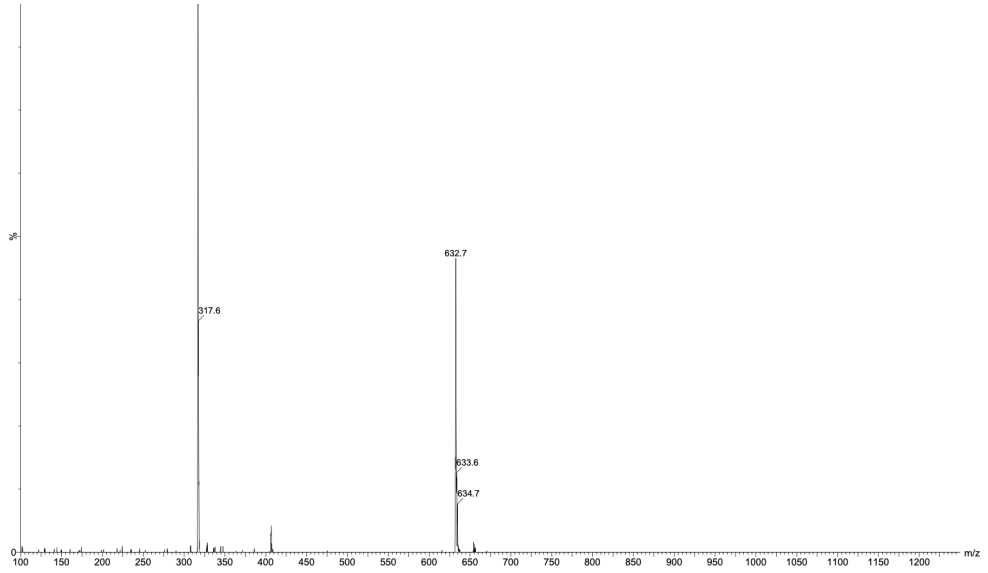
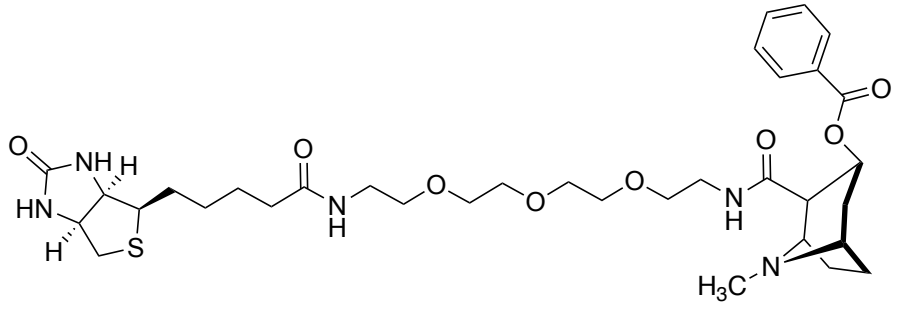


Figure S4.5. Mass spectrum for **Biotin-PEG3-ADMA**, positive mode. MS (MH^+(-1) m/z): Product mass 631.79 calculated $C_{27}H_{49}N_7O_8S$, Found 632.7.6 and half mass 317.0

4.5.2.4. Biotin-PEG3-Benzoylecoginine

Benzoylecgonine (0.0100 g, 0.0346 mmol), HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) (0.0273 g, 0.0720 mmol and DIPEA (N, N-diisopropylethylamine) (0.0178 g, 0.1378 mmol) were dissolved in dimethylformamide. Stir at 60 °C for 10 minutes. Add biotin-PEG3-NH_2 (0.0318 g, 0.0760 mmol), stir at 60 °C overnight. The reaction was diluted with water and purified by high pressure liquid chromatography on an ACCQPrep Waters HPLC system over a 9.4 mm x 250 mm semi-preparative Agilent Eclipse XDB-C18 5 μm with UV detection at 280 nm, flowrate 7 mL/min. The gradient ran at 90% H_2O (+0.1% TFA)/10% CH_3CN (+0.1% TFA) for 2 min then 10% H_2O (+0.1% TFA)/90% CH_3CN (+0.1% TFA) over 25 minutes. Hold at 10% H_2O (+0.1% TFA)/90% CH_3CN (+0.1% TFA) for 3 min, then to 10% H_2O (+0.1% TFA)/90% CH_3CN (+0.1% TFA) to 90% H_2O (+0.1% TFA)/10% CH_3CN (+0.1% TFA) over 1 minute and hold for 2 minutes. The fractions were collected and lyophilized, yielding biotin-PEG3-Benzoylecoginine as a white solid (0.017 g, 71%).



Molecular Weight: 689.87

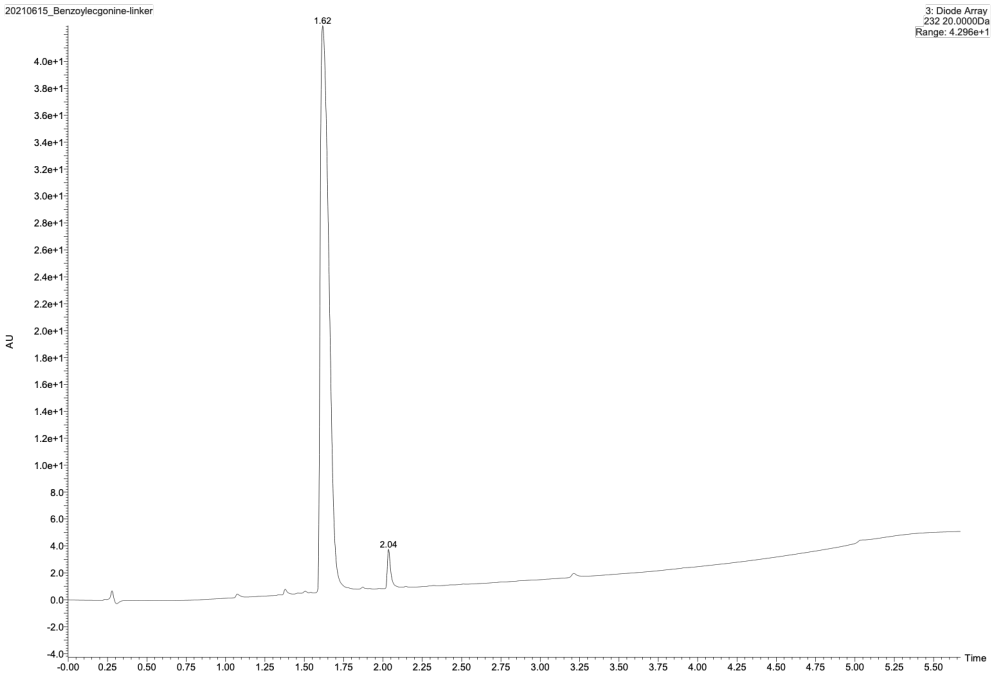


Figure S4.6. Diode array 323 nm for **Biotin-PEG3-Benzoylecoginine**. Product peak at 1.62 min, by-product 2.04 min.

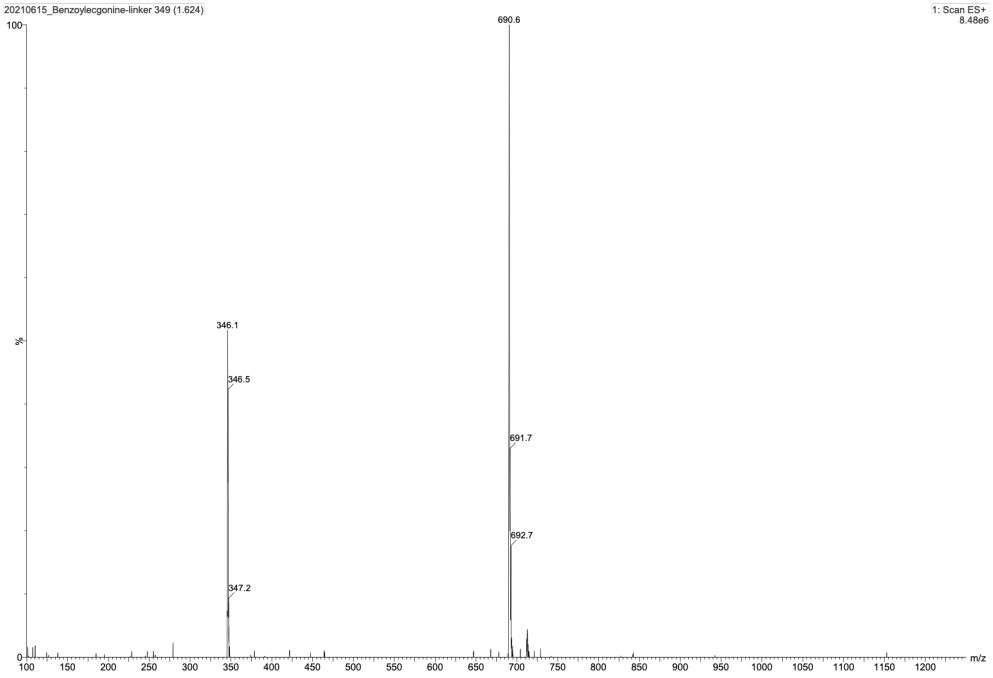


Figure S4.7. Mass spectrum for **Biotin-PEG3-Benzoylecoginine**, positive mode. MS (MH^+(-1) m/z): Product mass 689.87 calculated for $C_{34}H_{51}N_5O_8S$, Found 690.6 and half mass 346.1.

4.5.3. Phage display

General information

All the reactions are performed in a sterile environment in a volume of 100 μL at room temperature in a 1.5 mL Eppendorf tube unless otherwise indicated.

Tetracycline LB plates (1L)

Tryptone (10 g), sodium chloride (5 g), yeast extract (5 g) and agar (15 g) were added to 1 L of water and autoclaved with a stir bar. After cooling down, 1 mL of 20 mg/mL tetracycline solution (filtered) was added. Pour agar in Petri dishes (15 cm) let completely cool down before storing in a dark fridge (4 °C).

LB-broth (1L)

Tryptone (10 g), sodium chloride (5 g), yeast extract (5 g) and agar (15 g) were added to 1 L of water and autoclaved. After cooling down.

IPTG/X-gal plates (1L)

Tryptone (10 g), sodium chloride (5 g), yeast extract (5 g) and agar (15 g) were added to 1 L of water and autoclaved with a stir bar. After cooling down, 40 mg IPTG and 50 mg X-gal in 1 mL DMF were added. Pour agar in Petri dishes (15 cm) let completely cool down before storing in the fridge (4 °C).

Top Agar (1L)

Tryptone (10 g), sodium chloride (5 g), yeast extract (5 g), Magnesium Chloride (1 g) and agar (15 g) were added to 1 L of water and autoclaved. Divide in smaller bottles that fit a microwave and autoclave.

Labeling C7C phage library with sCx4-1,3-NHCOCH$_2$Cl, with TCEP pulse

Disulfide bridge reduction

67.5 μL of sterile H_2O, 10 μL of $1 \bullet 10^{13}$ pfu/mL of the phage (PhD-CX7C library, (NEB E8120S). 20 μL of Tris-HCl buffer (500 mM in deionized H2O, pH 8.5, filter 0.22 um) and 2.5 μL of TCEP solution (20mM) were combined. Let it stand at room temperature, 30 min.

Phage alkylation

Blank control

67.5 μL of sterile H_2O, 10 μL of the reduced phage and 20 μL of Tris-HCl buffer (500 mM in deionized H_2O, pH 8.5, filter 0.22 um) were combined, after 30 min 2.5 μL of TCEP solution (10 mM) was added, incubate the solution for an additional 30 min. Add 1 μL BIA (EZ-Link Iodoacetyl-PEG2-Biotin, 100 mM in DMF). Let it stand at room temperature for 30 min in a dark place.

Unmodified phage (-)-Phage

67.5 μL of sterile H_2O, 10 μL of the reduced phage and 20 μL of Tris-HCl buffer (500 mM in deionized H2O, pH 8.5, filter 0.22 um) after 30 min add 2.5 μL of TCEP solution (10mM) were combined, incubate the solution for an additional 30 min. Let it stand at room temperature for 30 min in a dark place.

Modified phage with 500 μM **sCx4-1,3-NHCOCH$_2$Cl** (+)-Phage

57.5 μL of sterile H_2O, 10 μL of the reduced phage, 10 μL **sCx4-1,3-NHCOCH$_2$Cl** (5 mM in deionized H2O, filter 0.22 um) and 20 μL of Tris-HCl buffer (500 mM in deionized H2O, pH 8.5, filter 0.22 um) were combined, incubate for 30 min. Add 2.5 μL of TCEP solution (10mM) incubate the solution for an additional 30 min. Add 1 μL BIA (EZ-Link Iodoacetyl-PEG2-Biotin), 100 mM in DMF). Let it stand at room temperature for 30 min in a dark place.

Quality of the streptavidin beads

Prepare 5 min before adding the BIA to the alkylation conditions above. Add the BIA to all the reactions at the same time.

71.5 μL of sterile H_2O, 5 of $1 \bullet 10^{13}$ pfu/mL of the phage (PhD-CX7C library, (NEB E8120S) and 20 μL of Tris-HCl buffer (500 mM in deionized H2O, pH 8.5, filter 0.22 um), 2.5 μL of TCEP solution (20 mM) and 1 μL BIA(EZ-Link Iodoacetyl-PEG2-Biotin), 100 mM in DMF) were combined. Let it stand at room temperature for 30 min in a dark place.

Quenching the reactions

The alkylation reactions are quenched by dilution.

Dilute the blank control and the alkylation condition in four sterile 1.5 mL tubes each, where the first three tubes have 990, 900, and 900 μL of sterile 1X PBS, while tube four has 990 μL of 2% BSA in 1X PBS. Dilute the assay control in four sterile 1.5 mL tubes each, where the first three tubes have 990, 900, and 990 μL of sterile 1X PBS, while tube 4 has 990 μL of 2% BSA in 1X PBS.

Dilute the Blank ((-)-Phage) and the alkylation with **sCx4-1,3-NHCOCH$_2$Cl** ((+)-Phage) for the pull-down assay in 900 μL 1x PBS.

Capture of biotinylated phage by streptavidin magnetic beads

Start the capture protocol right after adding the BIA to the alkylation reaction.

Resuspend the beads by gently shaking the tube. Use 50 μL resuspended beads, for each control conditions, use 100 μL resuspended beads for the (-)-Phage and (+)-Phage. Place the magnetic beads in a 1.5 mL sterile Eppendorf tube, and on a magnetic stand for 5 minutes, retain the beads, discard supernatant. Wash the beads with sterile 1X PBS (500 μL for each 50 μL of beads), gently shake, place on magnetic stand to retain beads and dispose of supernatant. Repeat two more times. Add sterile 2% BSA in 1X PBS (500 μL for each 50 μL of beads) and gently agitate for at least 15 minutes. Place the beads on a magnetic stand for 5 minutes and retain the beads, discard supernatant, add 100 μL of the

final dilution from the reaction quench for the control conditions and use 200 μL for the (-)-Phage and (+)-Phage. Gently agitate for 20 minutes. Place the beads on a magnetic stand for 5 minutes and retain the supernatant, discard beads.

Titer of control conditions

Start the titer protocol two days before the labeling experiment.

Streak out an *E. coli* K12 plate on Tetracycline LB plate, incubate overnight at 37 °C. Inoculate 30 mL LB-broth (see recipe below) with a single colony from the plate, incubate overnight shaking 200 rpm at 37 °C. On the day of the labeling experiment pre-heat a water bath to 60 °C and place IPTG/X-Gal plates in a 37 °C incubator. Melt top agar in microwave and dispense 4 mL per plate into sterile 5 ml polystyrene round-bottom, push-cap tubes. Store tubes in the warmed-up water bath. Add 200 μL of overnight culture into a sterile 0.6 mL tube per plate. Plate the final dilutions before capture (final dilution from quenching reaction) and after capture (retained supernatant from the capture) by adding 10 μL of the solution to the 0.6 mL tube. Plate each condition in triplicate, for a total of six plates per condition (three before- and three after capture). Pipet the entire contents of the 0.6 mL tube into one of push cap tubes containing melted top agar and vortex briefly. Invert and pour the inoculated top agar onto a pre-warmed plate. 'Swirl' the plate around to make sure it is evenly and completely covered with top agar. Avoid bubbles if possible. Cover plate with lid and allow top agar to solidify (5-10 mins). Invert the plates and incubate at 37 °C overnight. Count plaques on the following day.

Pull-down assay of calixarene modified C7C phage library with biotinylated target

There are 6 conditions for the pull-down assay, Table S4.1, each performed in four replicates.

Table S4.1. Pull-down conditions

Conditions	Phage	Beads
1	-	
2	+	
3	-	-
4	-	+
5	+	-
6	+	+

Note: Each condition is performed in four replicates, (-) unmodified, (+) modified

Wash beads three times with 1x PBS. For the modified (+) beads incubate 20 μL beads with 5 μg biotinylated target in 100 μL 1x PBS per condition. For the modified (-) beads incubate 20 μL beads in 100 μL 1x PBS per condition, rotate overnight at 4 °C Per condition incubate 20 μL beads with 100 μL phage overnight at 4 °C. The next day wash the modified beads 6x with 1x PBS and incubate for one hour in 2%BSA in 1x PBS. Put the phage-beads in the magnetic stand and collect the supernatant. After the one-hour blocking has finished Incubate the beads and phage according to the conditions (Table S4.1) for 2 hours at room temperature. Capture the beads by putting them on a magnetic stand, 5 min, wash 4x with 0.1% Tween 1x PBS and 2x with 1x PBS to wash off any unbound phage, retain the beads. Add 35 μL DNA-free water and boil the beads for 10 min at 95 °C. Collect the supernatant, this is send-off sequencing.

Calculations

BIA pull-down

$$\frac{pfu}{mL} = \frac{average\ counted\ plaques}{0.1} \times dilution$$

$$\%\ Capture = \frac{\frac{pfu}{mL} before\ capture - \frac{pfu}{mL} after\ capture}{\frac{pfu}{mL} before\ capture} \times 100$$

Blank control

$$\frac{pfu}{mL} = \frac{average\ counted\ plaques}{0.1} \times dilution$$

$$\%\ Capture = \frac{\frac{pfu}{mL} before\ capture - \frac{pfu}{mL} after\ capture}{\frac{pfu}{mL} before\ capture} \times 100$$

$$phage\ viability = \frac{\frac{pfu}{mL} before\ capture\ from\ blank}{\frac{pfu}{mL} before\ capture\ from\ BIA} \times 100$$

Alkylation

$$\frac{pfu}{mL} = \frac{average\ counted\ plaques}{0.1} \times dilution$$

$$\% \, Capture = \frac{\frac{pfu}{mL} before \; capture - \frac{pfu}{mL} after \; capture}{\frac{pfu}{mL} before \; capture} \times 100$$

$$phage \; viability = \frac{\frac{pfu}{mL} before \; capture \; from \; alkylation}{\frac{pfu}{mL} before \; capture \; from \; BIA} \times \; 100$$

Sequencing data

The sequencing was performed by 48Hour Discovery (https://48hourdiscovery.com) and the data was analyzed in collaboration with, Wenrui Huang, the 48Hour Discovery technician. Each target (ADMA and benzoylecgonine) had around ~3000 sequences amplified.

From these the top 35 most amplified sequences are shown below. The sequences chosen for further studies are **bolded**.

Table S4.2. Benzoylecgonine sequencing data

	Target count replicate			
	1	2	3	4
ACHVSSQSLC	0	0	0	3327
ACADQRSAYC	0	0	0	1794
ACSNYSGPFC	**0**	**1729**	**0**	**0**
ACVSSSGSRC	0	0	0	1690
ACASPNYTVC	**1681**	**0**	**0**	**0**
ACKHTWRSEC	0	0	0	1632
ACNSSKLHMC	**2028**	**142**	**561**	**8**
ACWDAINSRC	1377	0	0	0
ACYMSSPTLC	1344	0	0	0
ACVGHSHHMC	1063	0	0	0
ACTVEHARSC	1022	0	0	0
ACSALKHHLC	1021	0	0	0
ACWMTRSQPC	1014	0	0	0
ACEQSFAAYC	1013	0	0	0

ACGNDRMRLC	988	0	0	0
ACKDTALEHC	987	0	0	0
ACMECTNRAC	973	0	0	0
ACMLFNTLHC	968	0	0	0
ACYTNAFMIC	968	0	0	0
ACRATSAGPC	966	0	0	0
ACLHDTRLIC	959	0	0	0
ACQQYTYPWC	955	0	0	0
ACPERYHNAC	1012	0	0	0
ACPVSSSKTC	957	0	0	0
ACFPLHTKQC	946	0	0	0
ACFTYTHPHC	950	0	0	0
ACTTPWSYPC	937	0	0	0
ACSQMQNRLC	932	0	0	0
ACMAMSPASC	940	0	0	0
ACWEKPRSMC	934	0	0	0
ACYDETRAKC	926	0	0	0
ACSLHLGSAC	917	0	0	0
ACDSIQNITC	910	0	0	0
ACDRASFRYC	915	0	0	0
ACNASRQTTC	923	0	0	0

Table S4.3. ADMA sequencing data

	Target count replicate			
	1	2	3	4
ACNTGSPYEC	**3434**	**0**	**0**	**0**
ACHVSSQSLC	0	3327	0	0
ACNGWPGASC	**92**	**479**	**145**	**1887**
ACITTPYSPC	0	0	1918	0
ACADQRSAYC	0	1794	0	0
ACVSSSGSRC	0	1690	0	0
ACKHTWRSEC	0	1632	0	0
ACPKDFDHRC	1623	0	0	0
ACNSMNSARC	1477	0	0	0
ACPGPKSLWC	1359	0	0	0
ACITRAVTSC	0	0	0	1220
ACEWLTNFVC	0	0	0	1140
ACTLTTFRSC	**1190**	**24**	**27**	**0**
ACPATDIMTC	958	0	0	0
ACRVSPSTWC	885	10	0	1
ACHHWTGVAC	887	0	0	0
ACSSKQPQLC	876	0	0	0
ACDRSTTKIC	**516**	**191**	**468**	**543**
ACVMADWARC	0	0	852	0
ACLPTVMVPC	849	0	0	0
ACDSPFKGEC	842	0	0	0
ACPQLAHIKC	835	0	0	0
ACGSLVKHGC	804	0	0	0

ACSAVEPGNC	791	0	0	0
ACKEMYEFAC	0	0	741	0
ACLPQSSTFC	722	0	0	0
ACGELGALDC	721	0	0	0
ACSPLLKKNC	0	710	0	0
ACTRLHNLNC	705	0	0	0
ACLETDVFSC	643	0	0	0
ACGLPPHESC	621	0	0	0
ACMPNHGPLC	0	0	2	962
ACEPRSLANC	**251**	**330**	**122**	**260**
ACGAVSSTHC	567	0	0	0
ACGHTLASKC	484	0	0	0

4.5.4. Host cyclizations

The hosts were cyclized following a modified protocol from Taki *et al*.[53] The selected peptides were modified with a hydrophobic tail (SDD) to improve water solubility. The peptides, P1: ACNGWPGASCSDD, P2: ACDRSTTKICSDD, P3: ACEPRSLANCSDD, P4: ACASPNYTVCSDD, were purchased from GenScript Inc. (NJ, USA). Each reaction takes place in a total volume of 50 mL. **sCx4-NHCOCHCl** (500 μM), peptide (500 μM, 20% DMSO), neutralized TCEP (500 μM), were dissolved in a phosphate buffer (10 mmol, pH 7.4) and stirred at room temperature in the dark till the reactions had gone to completion (48-96 h). The reaction progress was followed by UPLC-MS. The reactions were then lyophilized to reduce the volume before purification. The reactions were dissolved in 90% water and 10% acetonitrile with 0.1% TFA and purified by high pressure liquid chromatography on an ACCQPrep Waters HPLC system over a 20 mm x 150 mm preparative C18 5 μm with UV detection at 280 nm, flowrate 18.9 mL/min. The gradient ran at 90% H_2O (+0.1% TFA)/10% CH_3CN (+0.1% TFA) for 1.7 min then 79% H_2O (+0.1% TFA)/21% CH_3CN (+0.1% TFA) for 4 minutes. Then 63.3% H_2O (+0.1% TFA)/36.7% CH_3CN (+0.1% TFA) for 11.8 minutes. Then 10% H_2O (+0.1% TFA)/90% CH_3CN (+0.1% TFA) for 1 minutes. Hold at 10% H_2O (+0.1% TFA)/90% CH_3CN (+0.1% TFA) for 1 min, then to 10% H_2O (+0.1% TFA)/90% CH_3CN (+0.1% TFA) to 90% H_2O (+0.1% TFA)/10% CH_3CN (+0.1% TFA) hold for 1 minutes. The product was collected and lyophilized, yielding the cyclized hosts **sCx4-P1** as an off white solid (light yellow) (0.0032 g, 32%), HR-MS: *m/z* calculated for $(C_{82}H_{93}C_{l2}N_{17}O_{33}S_4)^{-4}$, 492.8757, $[M\text{-}4H]^{4-}$ Found 482.8757, **sCx4-P2** as a white solid (0.0043 g, 46%), **sCx4-P3** as an off white (light beige) (0.0040 g, 39%), **sCx4-P4** as an off white solid (light yellow) (0.0024 g, 23 The unreacted **sCx4-1,3-NHCOCHCl** was recovered. The hosts were characterized by UPLC-MS.

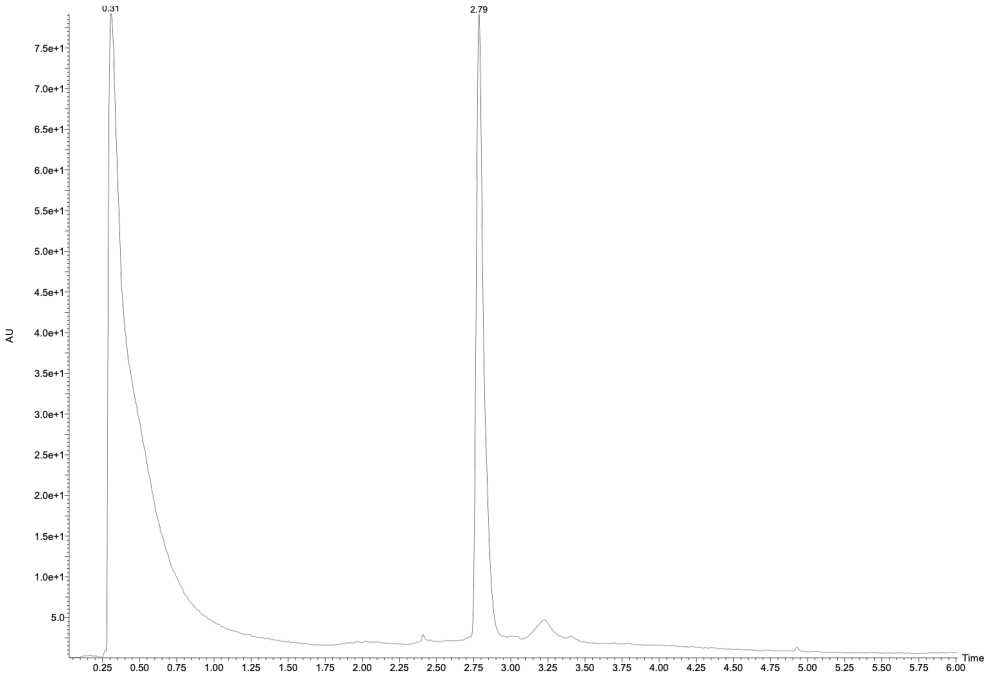


Figure S4.8. Diode array for **sCx4-P1**. Injection peak at 0.31 min. Product peak at 2.79 min

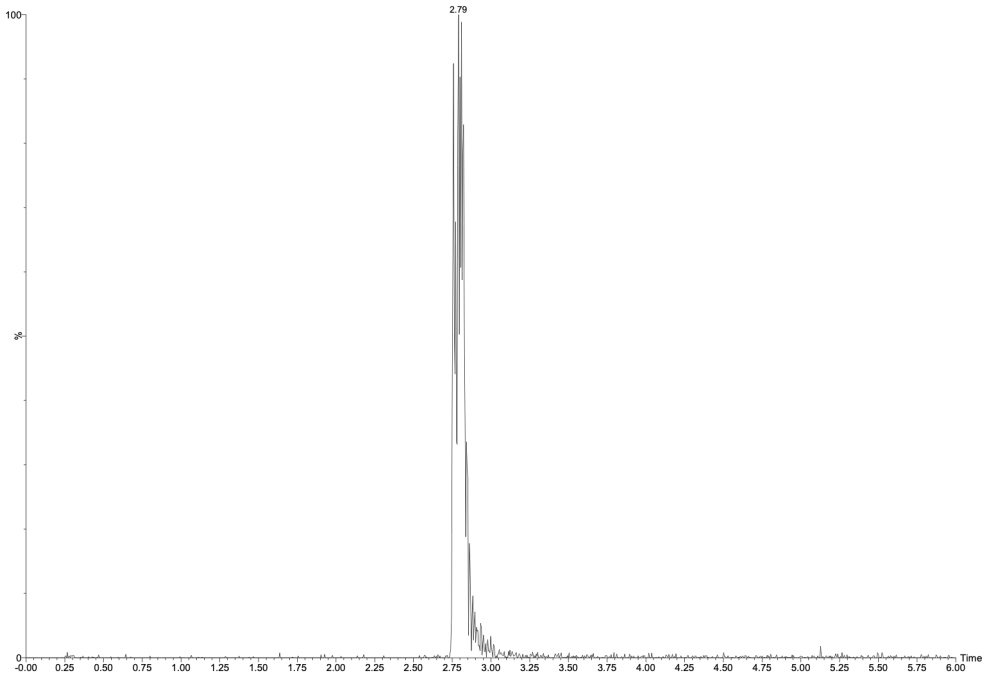


Figure S4.9. Mass spectrum for **sCx4-P1**, negative mode. MS: m/z calculated for $(C_{82}H_{95}C_{l2}N_{17}O_{33}S_4)^{-2}$, 986.76, $[M-2H]^{2-}$ found 987

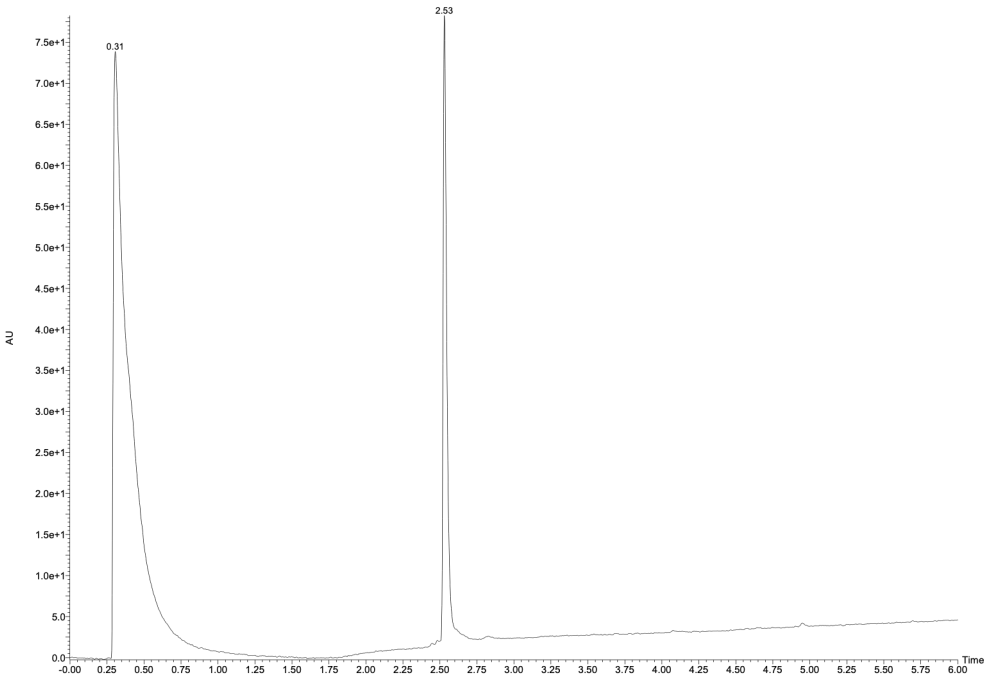


Figure S4.10. Diode array for **sCx4-P2**. Injection peak at 0.31 min. Product peak at 2.53 min

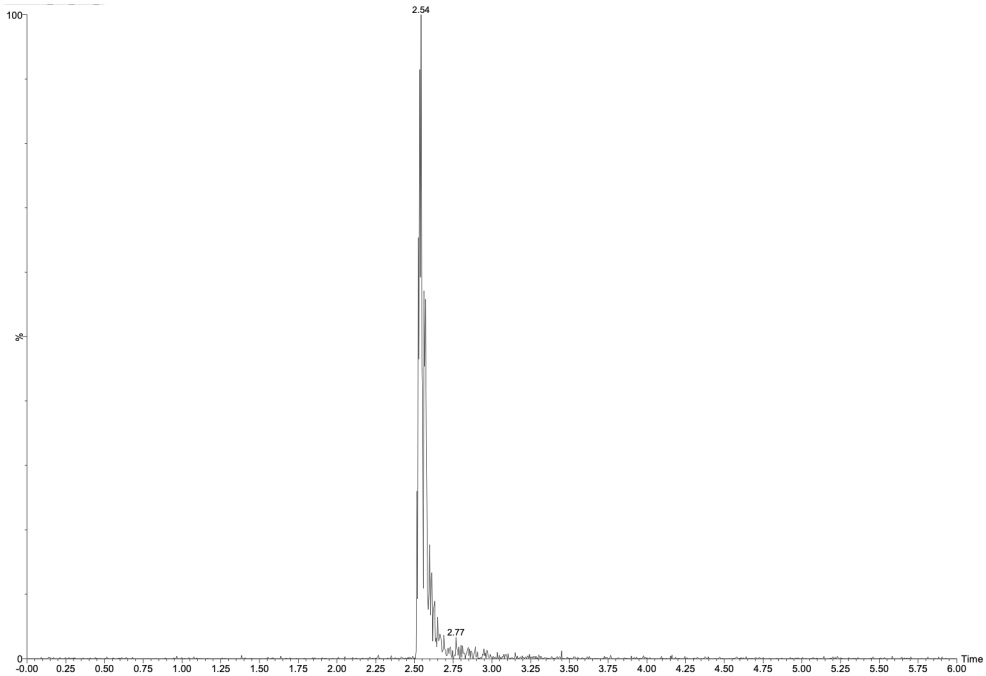


Figure S4.11. Mass spectrum for **sCx4-P2**, negative mode. MS (MH^-(-2) m/z): Product mass: 2207.19

Calculated for $C_{85}H_{115}N_{19}O_{36}S_4^{2-}$,half mass 1052.83, Found 1053.

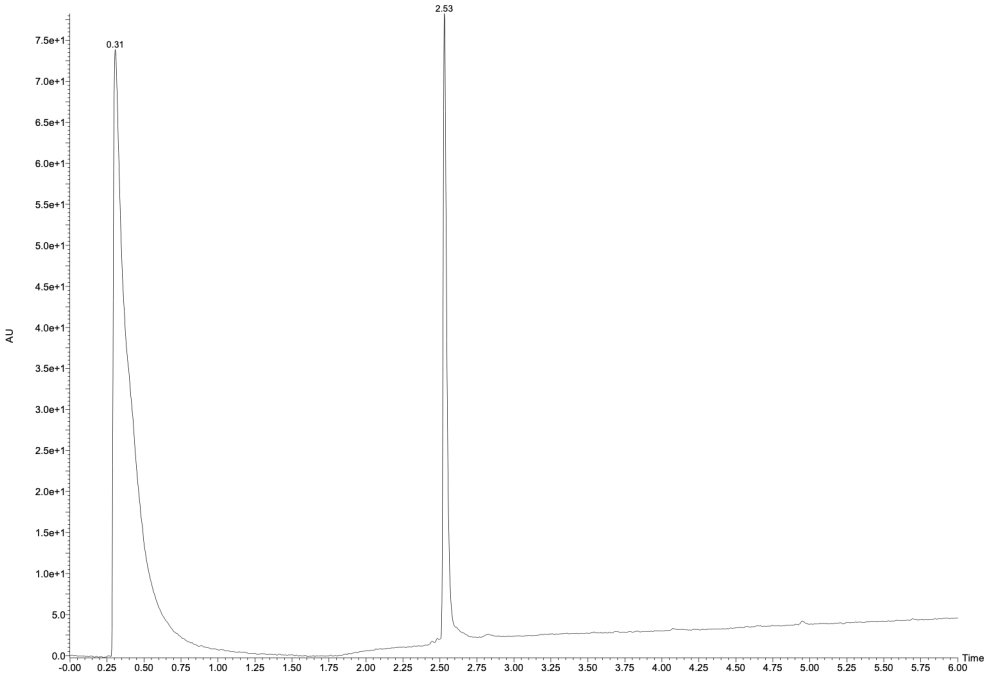


Figure S4.12. Diode array for **sCx4-P3**. Injection peak at 0.31 min. Product peak at 2.61 min, by-product 2.80 min.

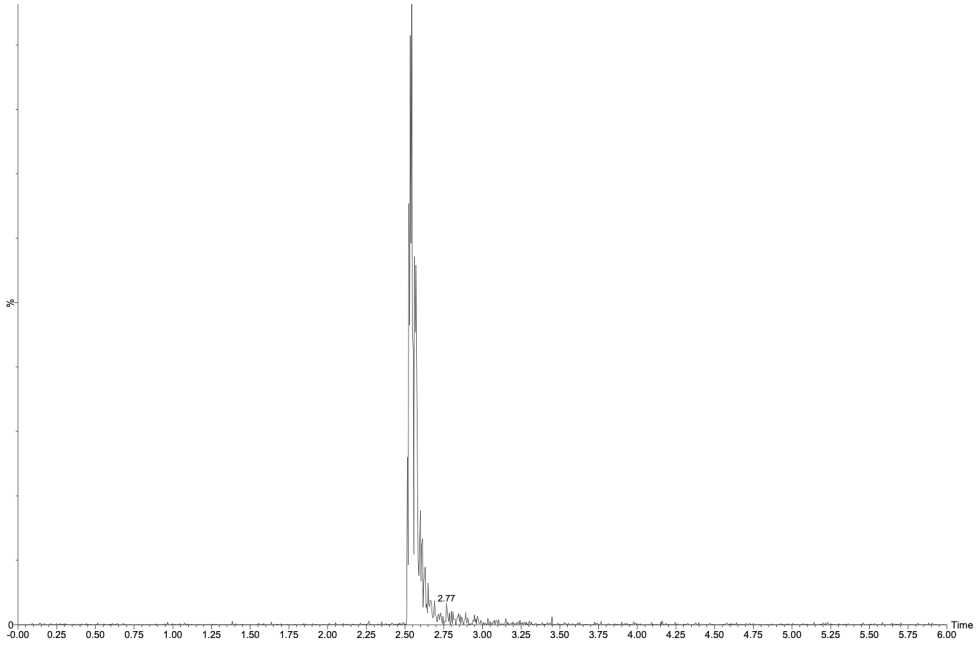


Figure S4.13. Mass spectrum for **sCx4-P3**, negative mode. MS: m/z calculated for $(C_{84}H_{109}N_{19}O_{35}S_4)^{-2}$, 1036.31, $[M-2H]^{2-}$ found 1036

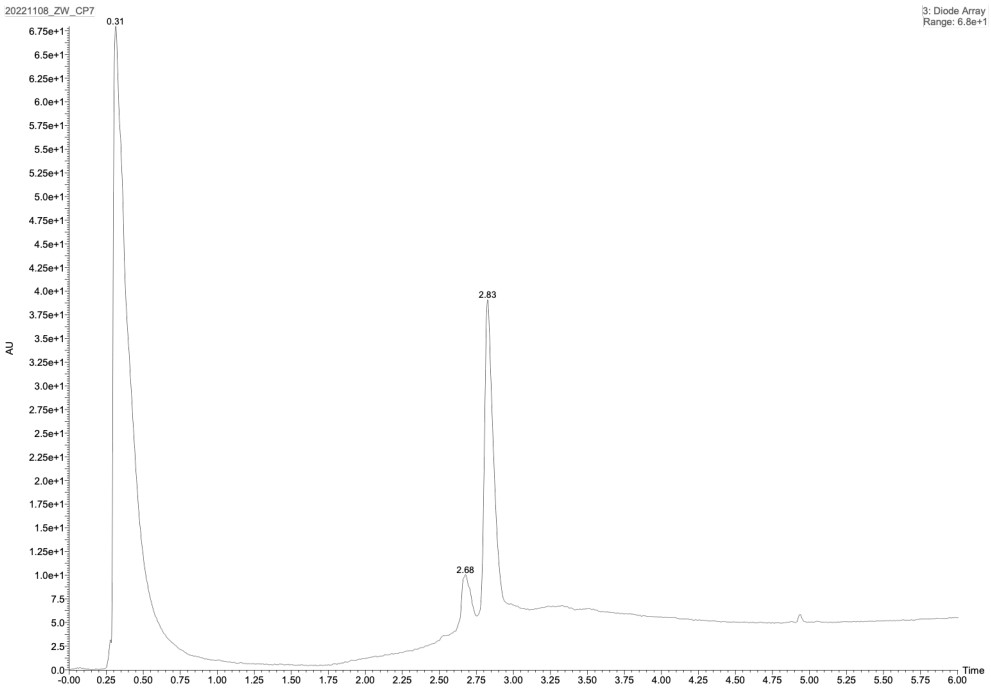


Figure S4.14.Diode array for **sCx4-P4**. Injection peak at 0.31 min. By-product 2.68 min, product peak at 2.83 min

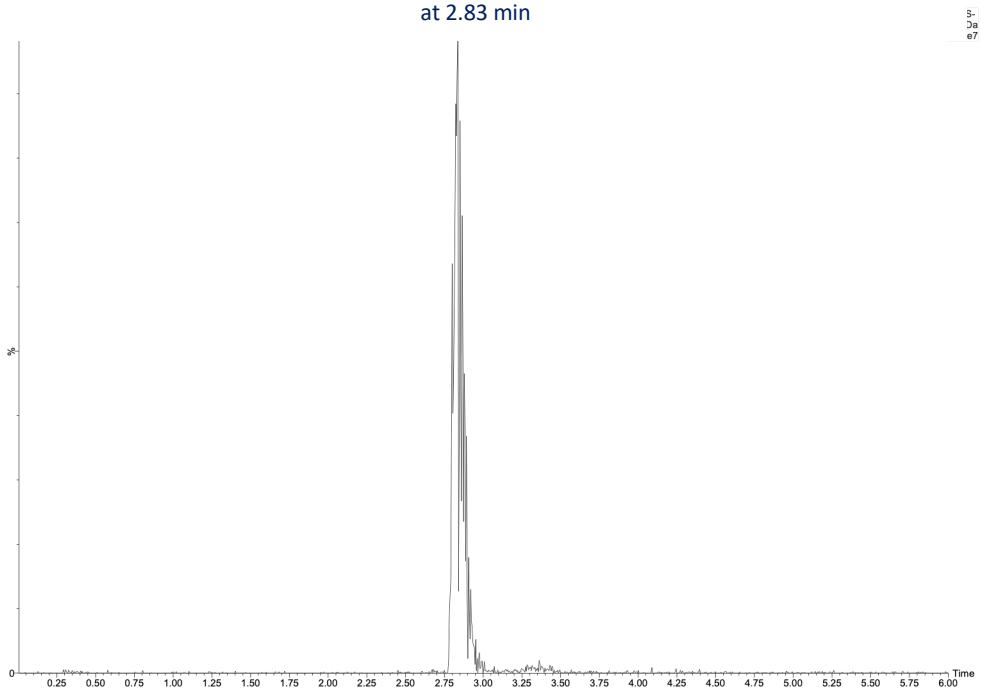


Figure S4.15.Figure S18. Mass spectrum for **sCx4-P4**, negative mode. MS: m/z calculated for $(C_{85}H_{104}N_{16}O_{35}S_4)^{-2}$, 1018.29, $[M-2H]^{2-}$ found 1018

4.5.5. Indicator displacement assays

The IDAs were conducted in 384 well plates (NUNC, optical black bottom). The fluorescent signal was read on a Biotek Cytation 5 multimode plate reader (Software Version 3.05.11) as a fluorescent endpoint measurement. 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. All wells had a final working volume of 50 μL. To measure the dye controls, three wells were prepared with a concentration of 0.25 μM LCG in 10 mM phosphate buffer, pH 7.4. To measure the host controls, three wells were prepared with a concentration of 50 μM host in 10 mM buffer. For the direct titrations of host into dye, a serial dilution ranging from 200 μM to 0.04 μM host into 0.25 μM LCG in 10 mM phosphate buffer was performed. For the competitive titrations, a serial dilution ranging from 1 mM to 0.04 mM guest into 50 μM host, 0.25 μM LCG in 10 mM phosphate buffer was performed. All the IDA data was analyzed in GraphPad Prism Version 8.3.0 (328). The experiments are performed as duplicates of triplicates and the experiment for **sCx4-P3** were performed as a single triplicate.

Calculations

Outliers

Outlier's data points were 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 90% confidence, $Q > 0.941 = Q_{90\%,n=3}$, we conclude the data point is an outlier.

Equation 1. Dixon's Q-test

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

The standard error

The standard error was calculated for each triplicate, Equation 2.

Equation 2. Standard error

$$SD_{K_d} = \left(ln\ (10)\ * 10^{logK_d}\right) *\ SD_{logK_d}$$

SDK_d = Standard error of K_d

logK_d = Value of the log

SD$_{log}K_d$ = Standard error of the log K_d

The total standard error

The total standard error of the mean value determined from two triplicate measurements was 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

Curve fit for the direct titration

To curve fit for the direct titration Equation 4 and Equation 5 were used.

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

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

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_{ind} = Dissociation constant

Equation 5. Curve fit for the competitive titration

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

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

log_{EC50} = log of the concentration of the competitor binding half-way between F_{min} and F_{max}

K_d = Equilibrium dissociation constant in Molar

[D] = Concentration of dye in nM

K_{ind} = Equilibrium dissociation constant of the direct titration

F = Fitted data point

F_{max} = Maximum signal

F_{min} = Minimum signal

IDA study of sCx4

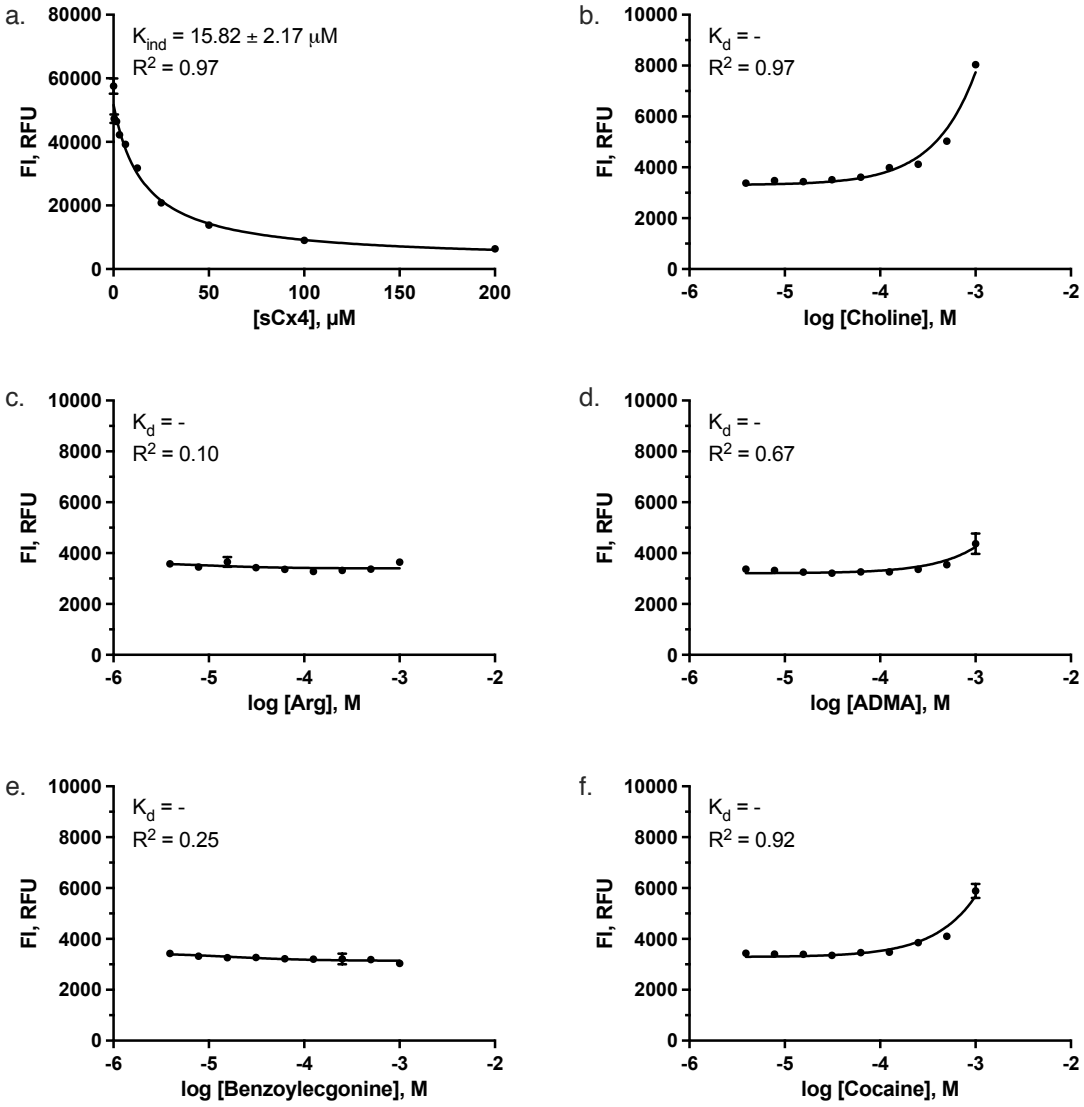


Figure S4.16. Fluorescence based studies of **sCx4**. a) Direct titration of LCG (0.25 μM) with **sCx4** (0 – 200 μM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into the **sCx4**

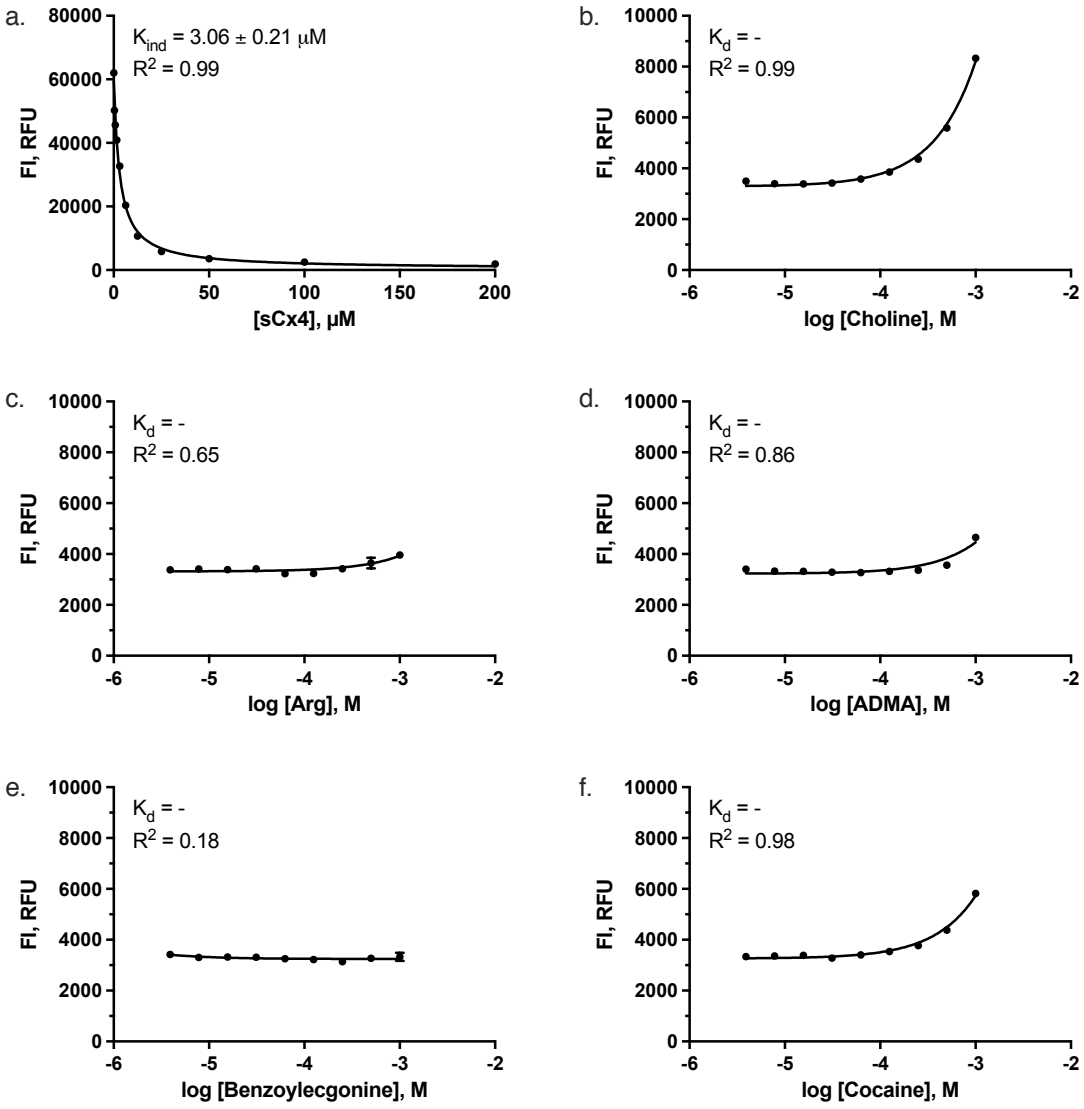


Figure S4.17. Fluorescence based studies of **sCx4**. a) Direct titration of LCG (0.25 μM) with **sCx4** (0 – 200 μM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into the **sCx4**

IDA study of sCx4-P1

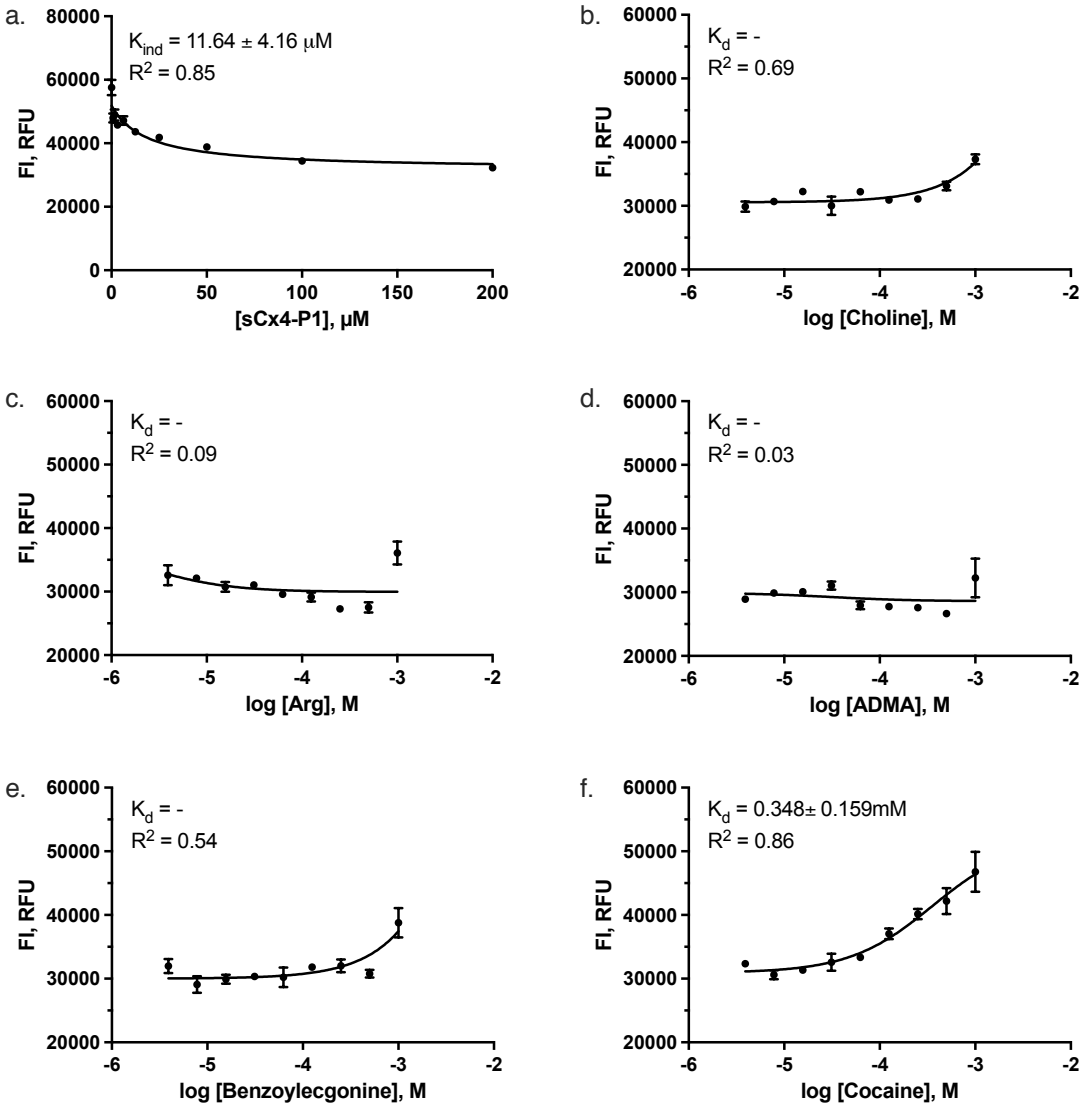


Figure S4.18. Fluorescence based studies of **sCx4-P1**. a) Direct titration of LCG (0.25 μM) with **sCx4-P1** (0 – 200 μM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into the **sCx4-P1**

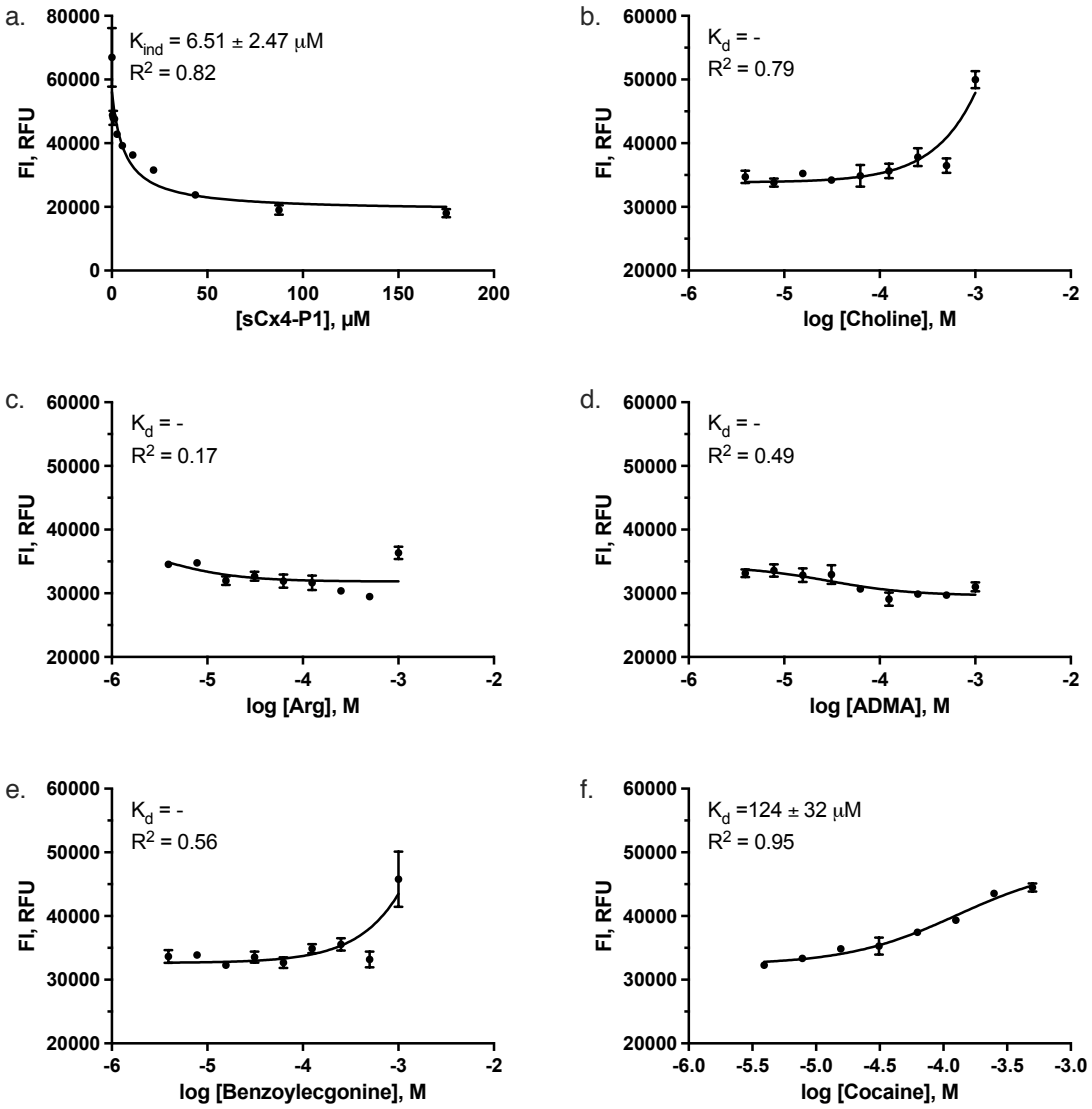


Figure S4.19. Fluorescence based studies of **sCx4-P1**. a) Direct titration of LCG (0.25 μM) with **sCx4-P1** (0 – 175 μM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into the **sCx4-P1**

IDA study of sCx4-P2

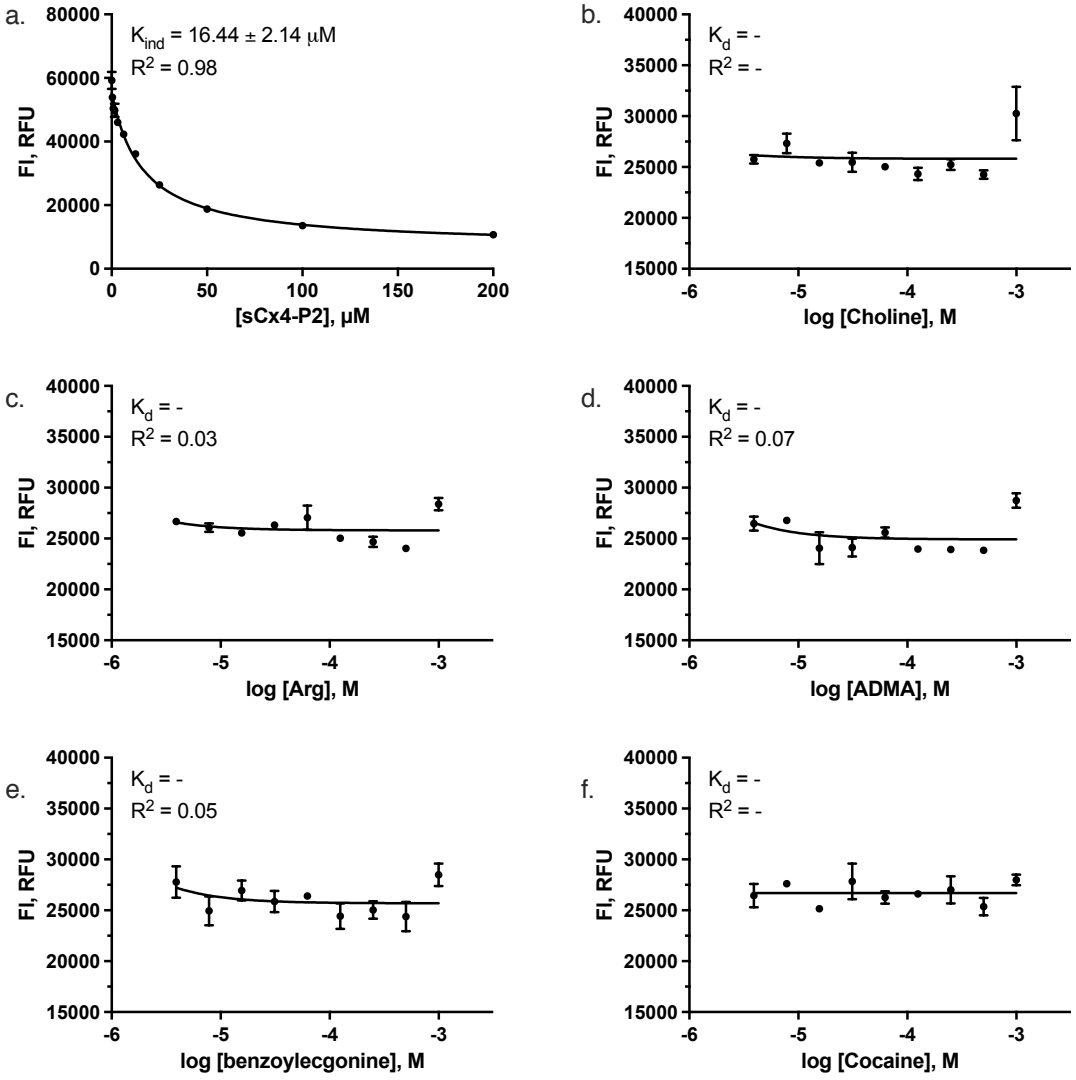


Figure S4.20.Fluorescence based studies of **sCx4-P2**. a) Direct titration of LCG (0.25 µM) with **sCx4-P2** (0 – 200 µM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into the **sCx4-P2**

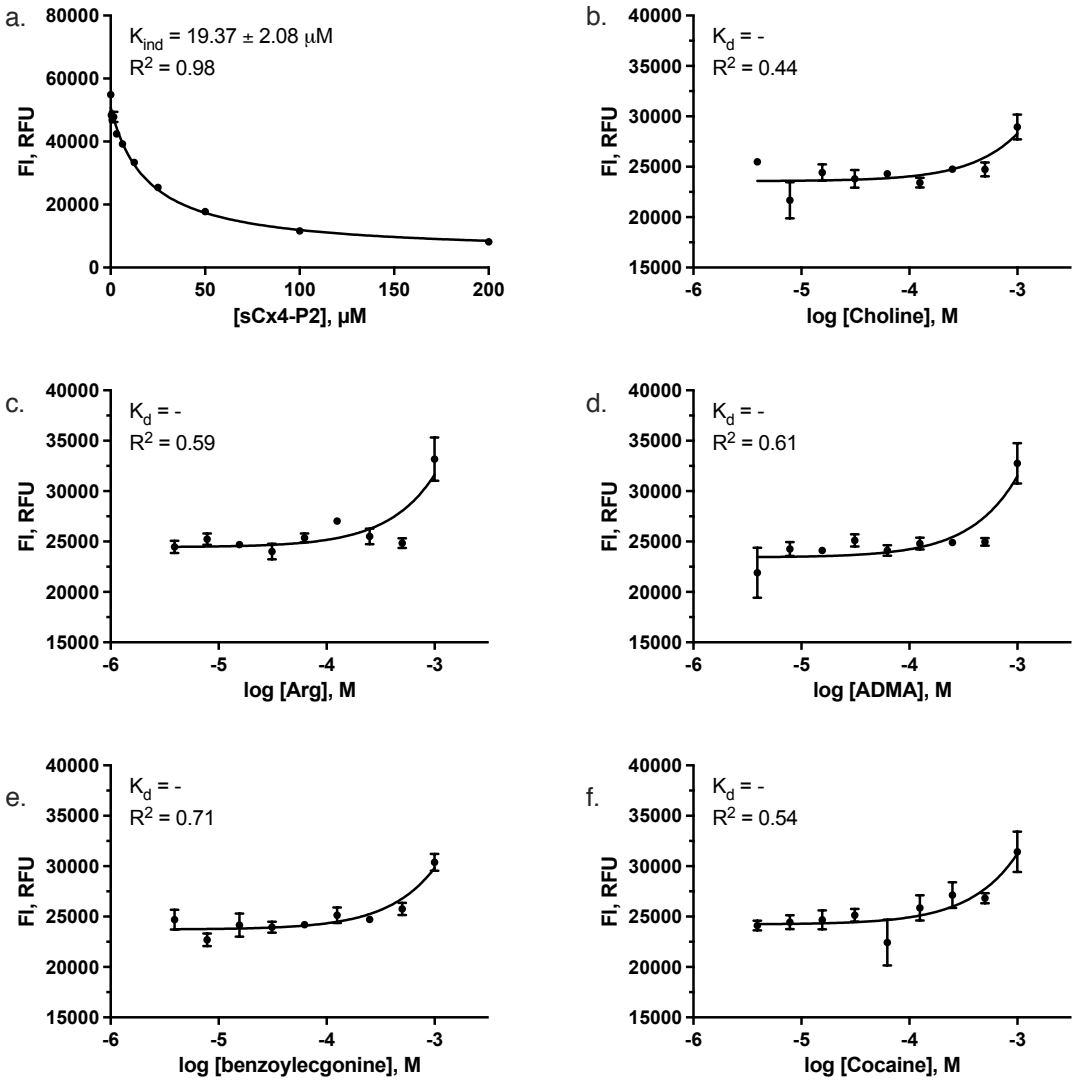


Figure S4.21. Fluorescence based studies of **sCx4-P2**. a) Direct titration of LCG (0.25 µM) with **sCx4-P2** (0 – 200 µM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into **sCx4-P2**

IDA study of sCx4-P3

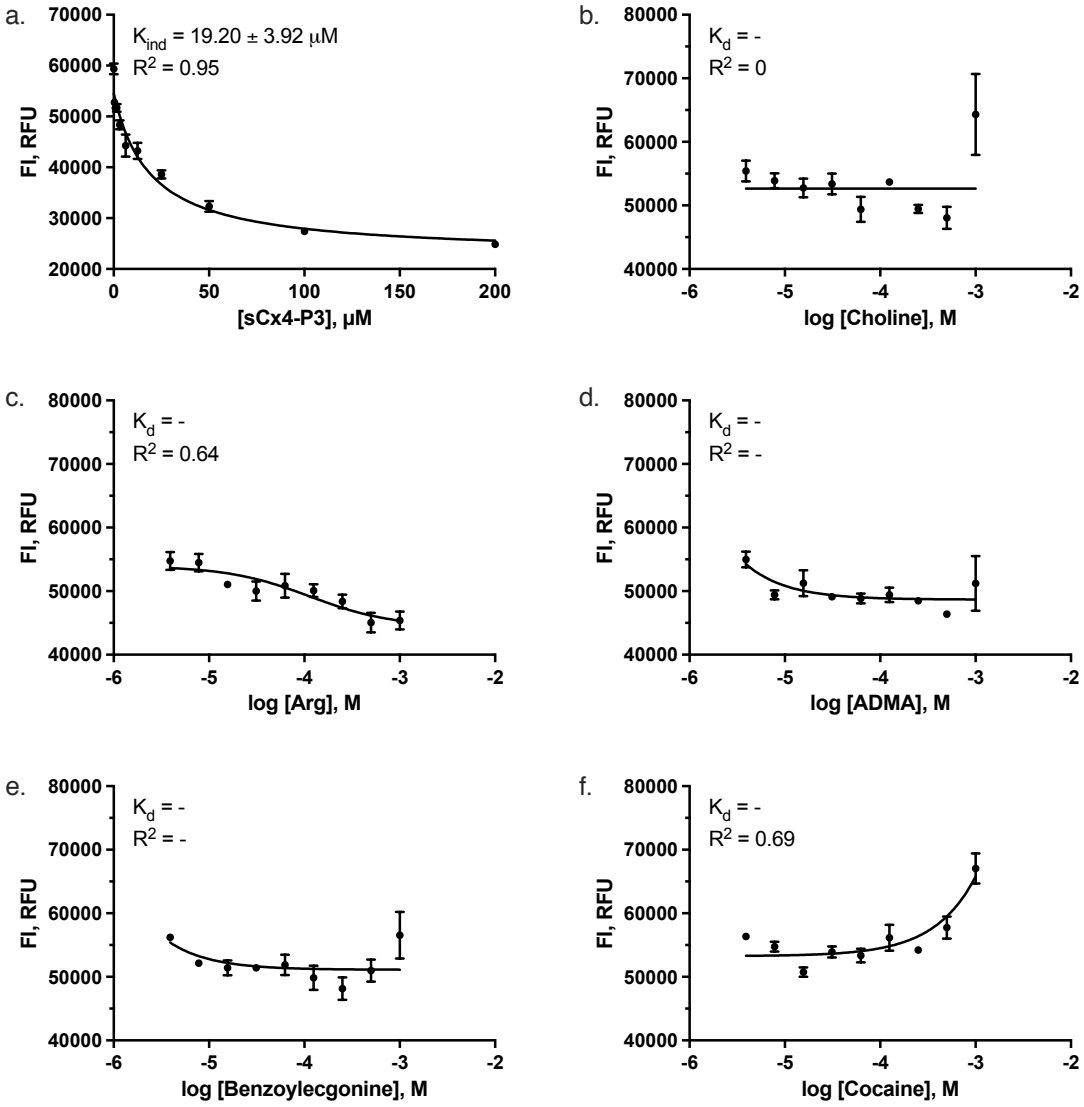


Figure S4.22. Fluorescence based studies of **sCx4-P3**. a) Direct titration of LCG (0.25 μM) with **sCx4-P3** (0 – 200 μM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into the **sCx4-P3**

IDA study of sCx4-P4

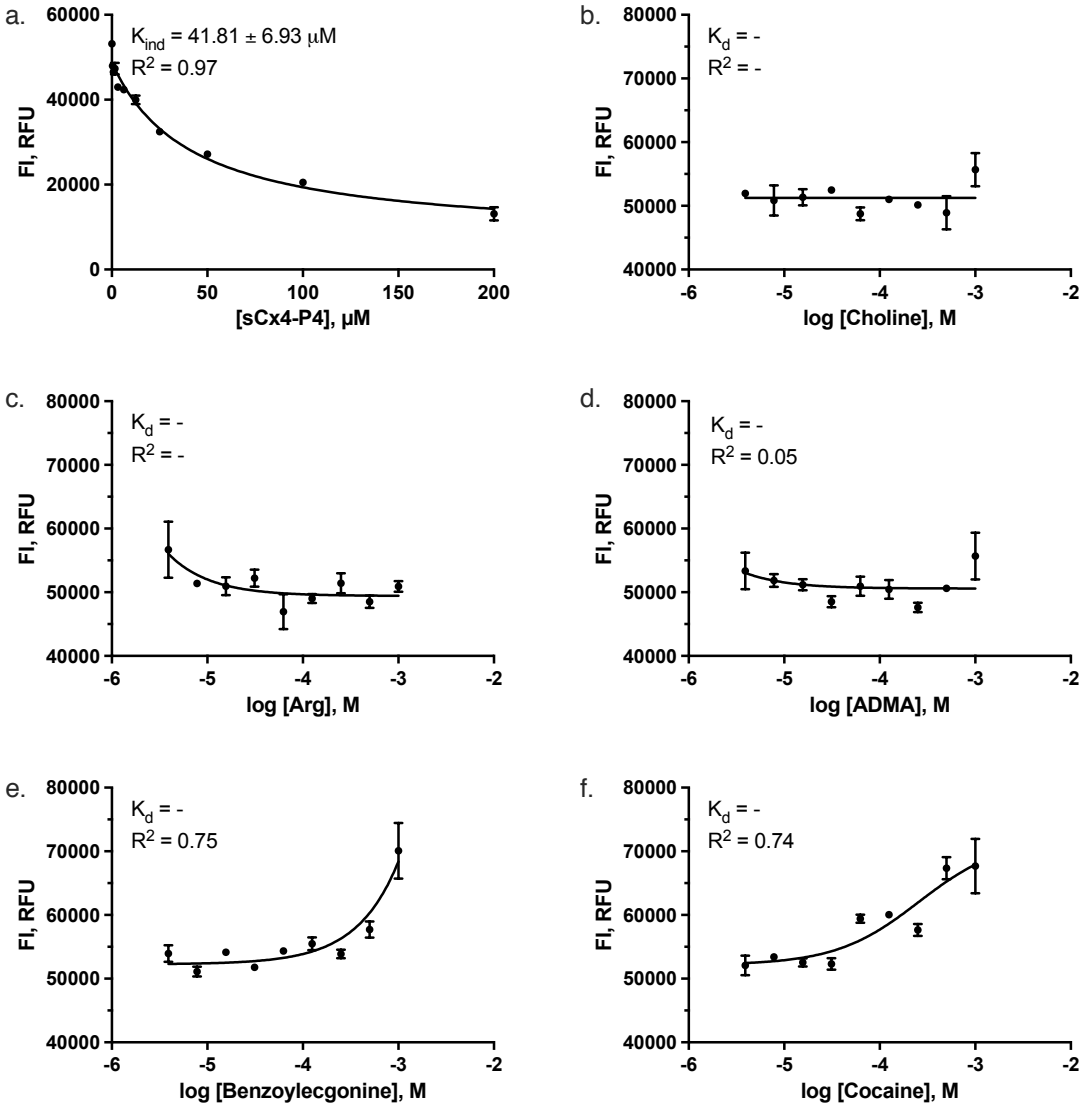


Figure S4.23. Fluorescence based studies of **sCx4-P4**. a) Direct titration of LCG (0.25 μM) with **sCx4-P4** (0 – 200 μM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into the **sCx4-P4**

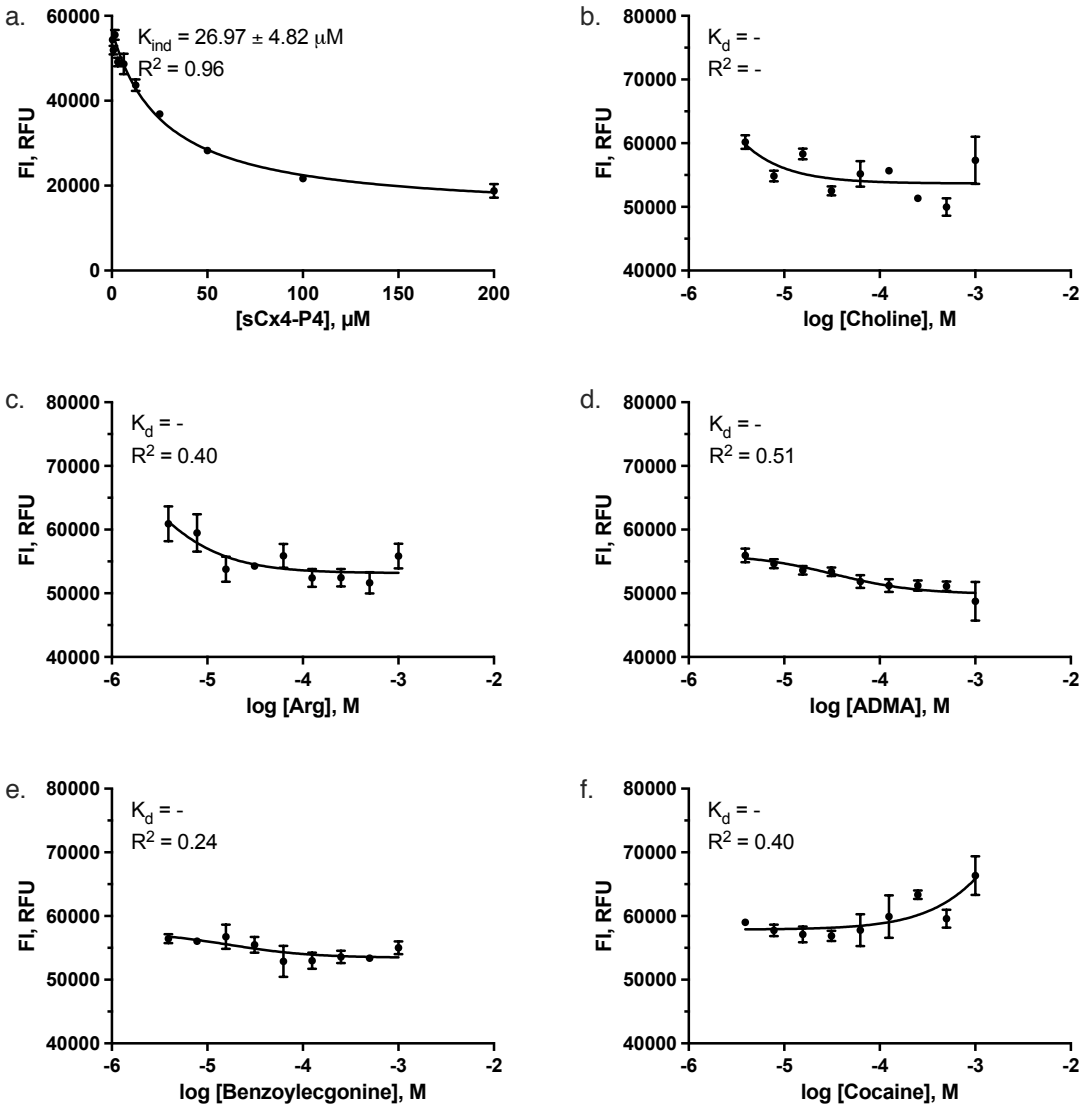


Figure S4.24.Fluorescence based studies of **sCx4-P4**. a) Direct titration of LCG (0.25 μM) with **sCx4-P4** (0 – 200 μM). b-f) Competitive titrations of choline, arginine, asymmetric dimethylarginine, benzoylecgonine and cocaine (0 – 1 mM) individually titrated into the **sCx4-P4**

4.5.6. NMR studies

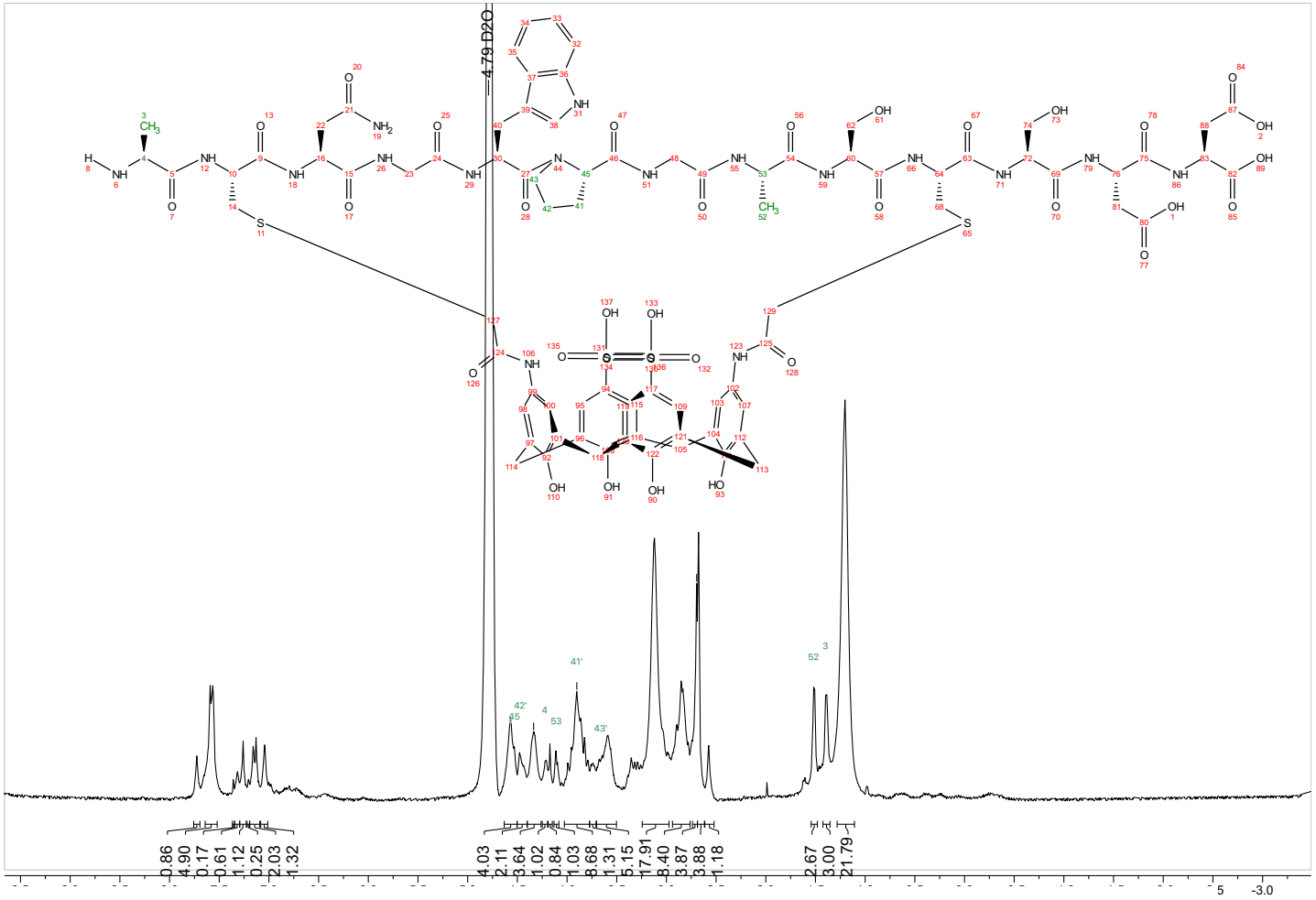


Figure S4.25.[1]H-NMR **sCx4-P1**

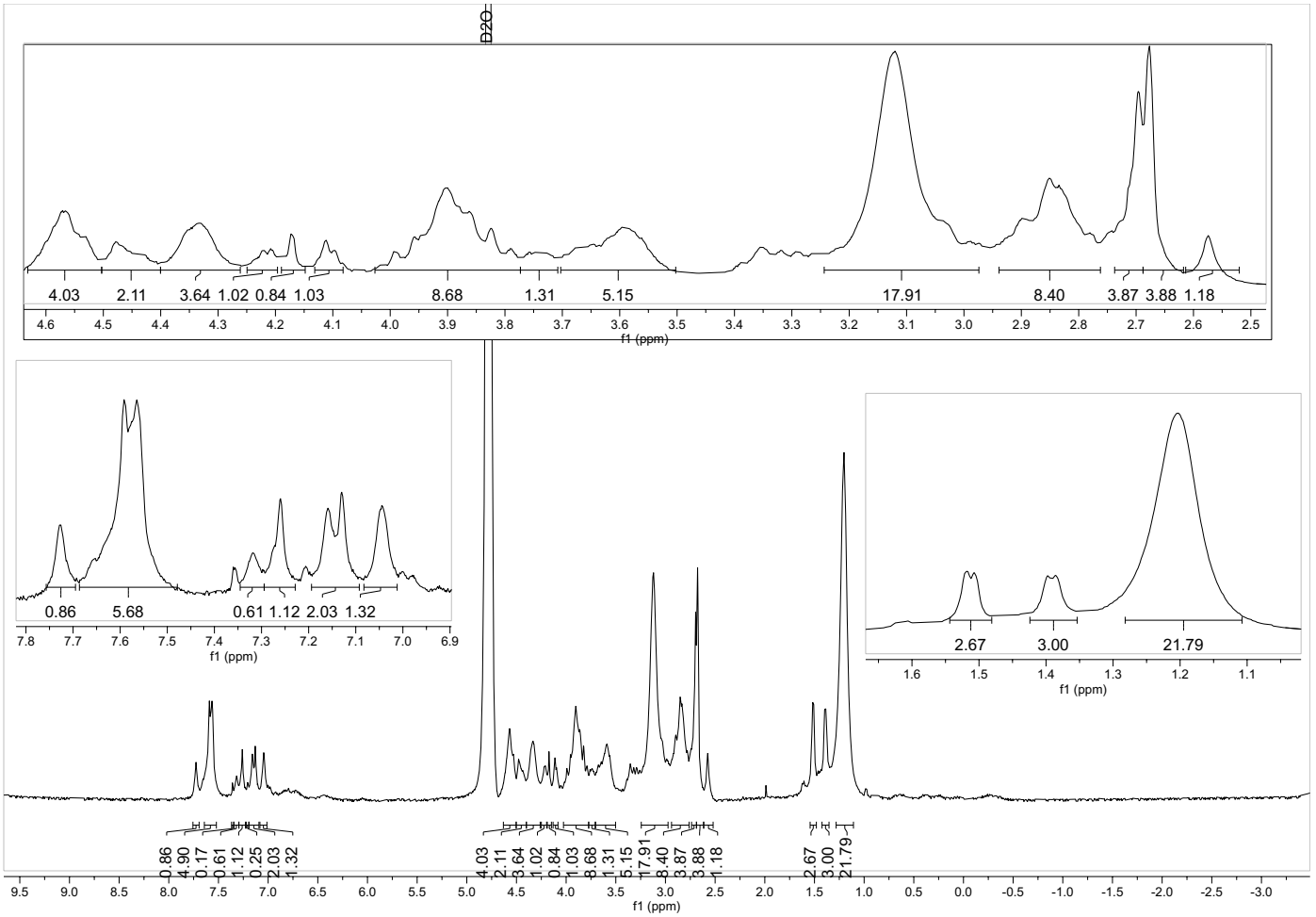


Figure S4.26. ^{1}H-NMR **sCx4-P1** with expansions

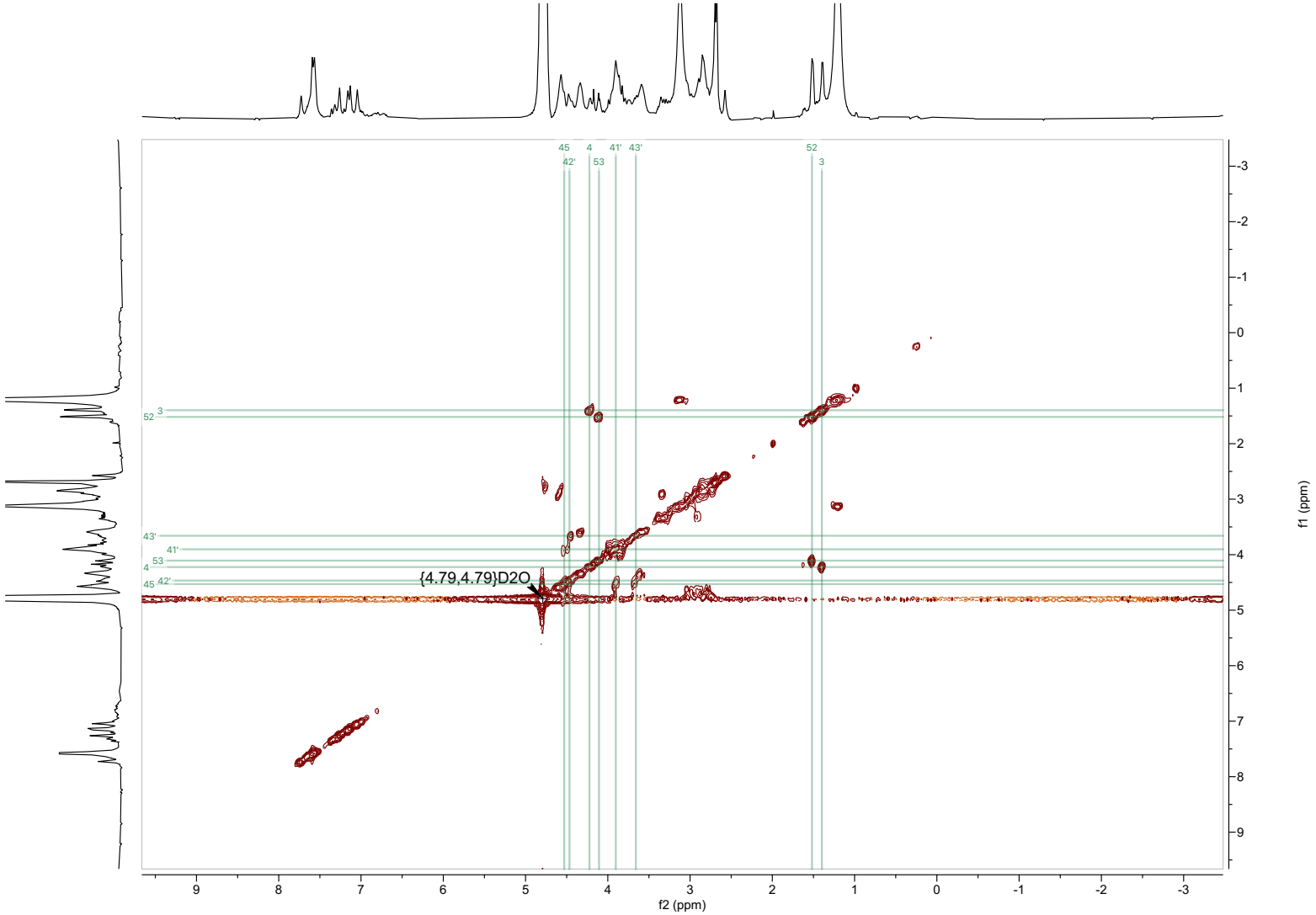


Figure S4.27. COSY **sCx4-P1**, full spectra

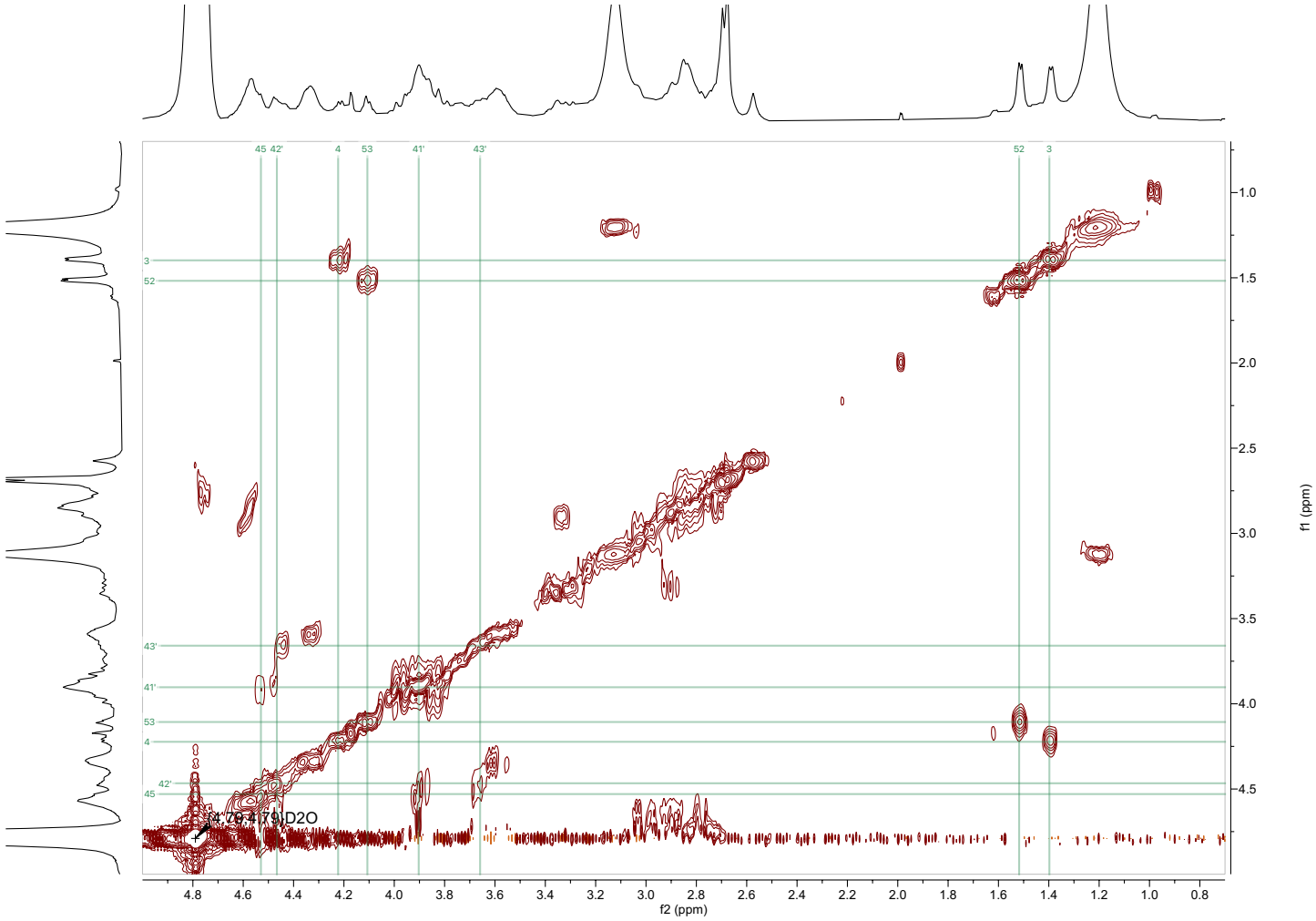


Figure S4.28. COSY **sCx4-P1**, zoom on assigned protons

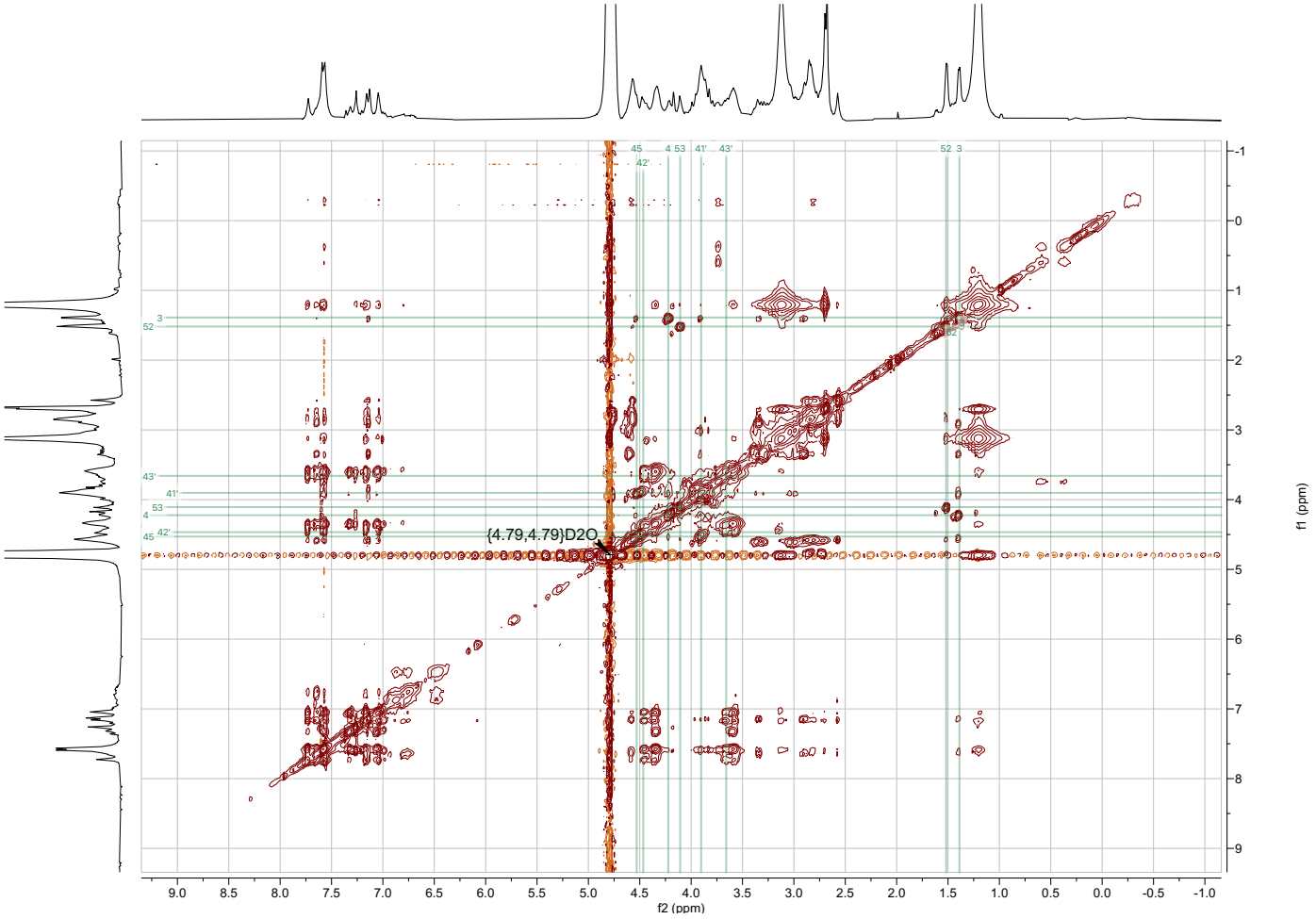


Figure S4.29. NOESY **sCx4-P1**, full spectrum

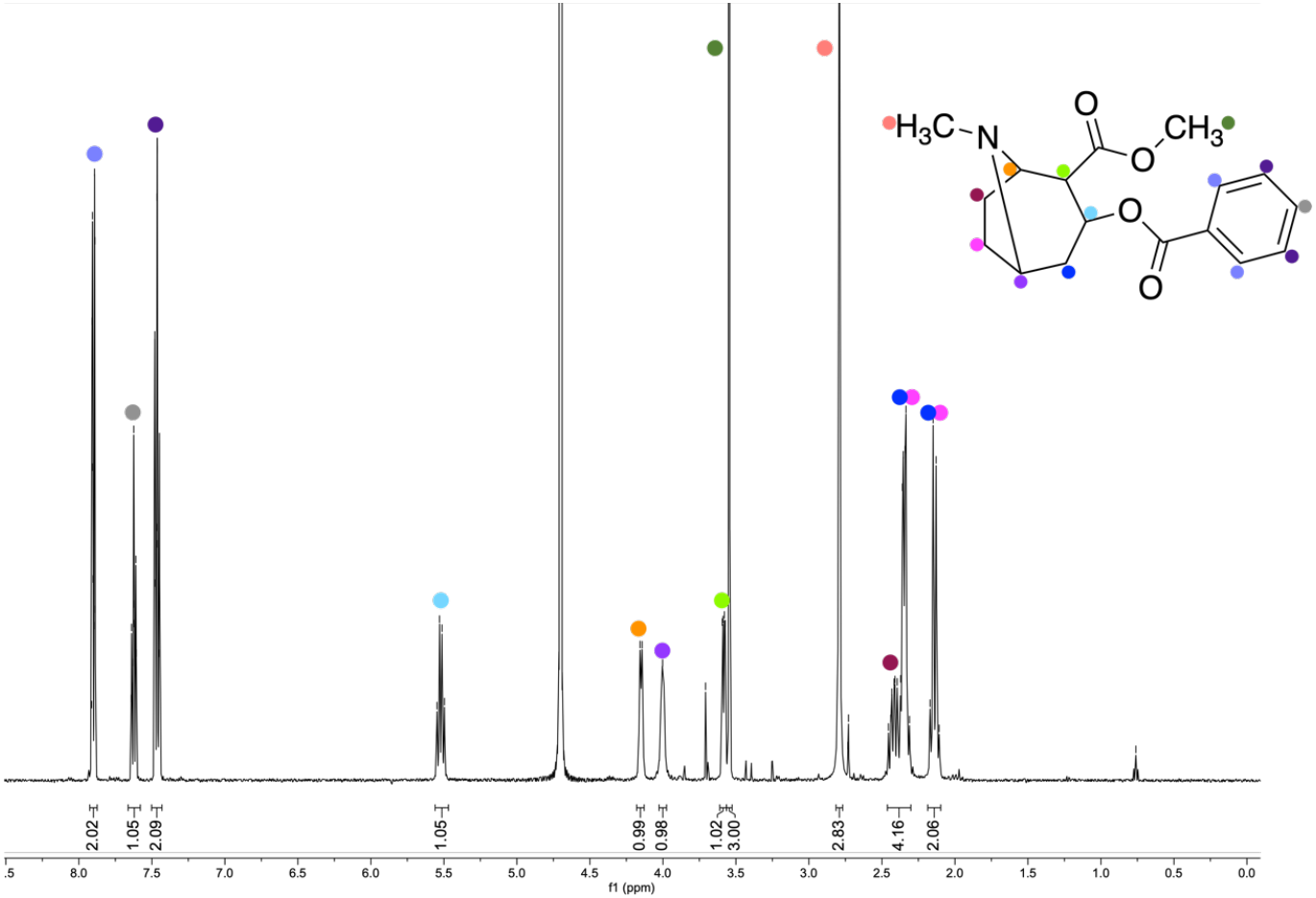


Figure S4.30. ^{1}H-NMR cocaine

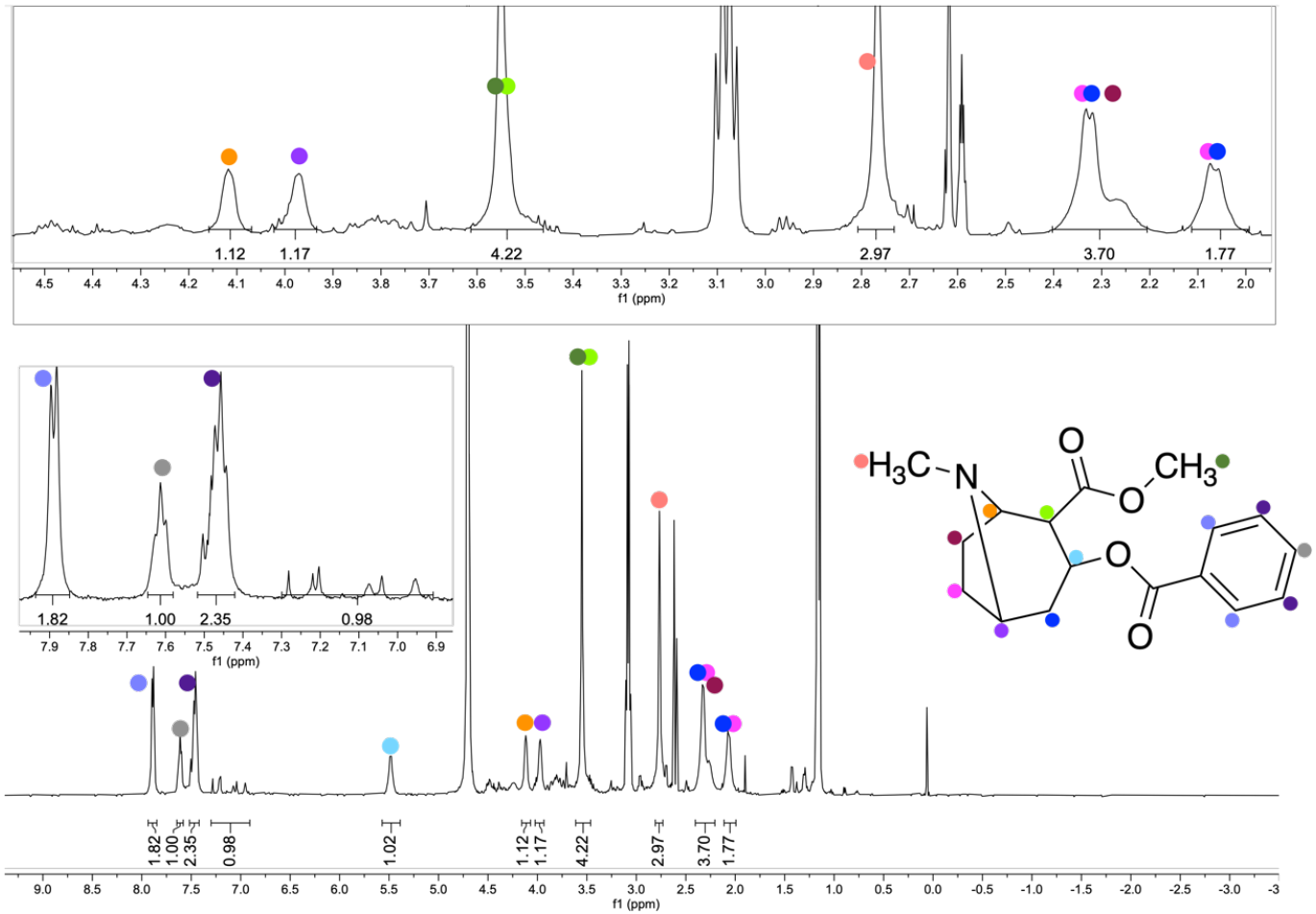


Figure S4.31. ^{1}H-NMR **sCx4-P1** with cocaine. Since the cocaine is in excess, its protons are the only ones integrated

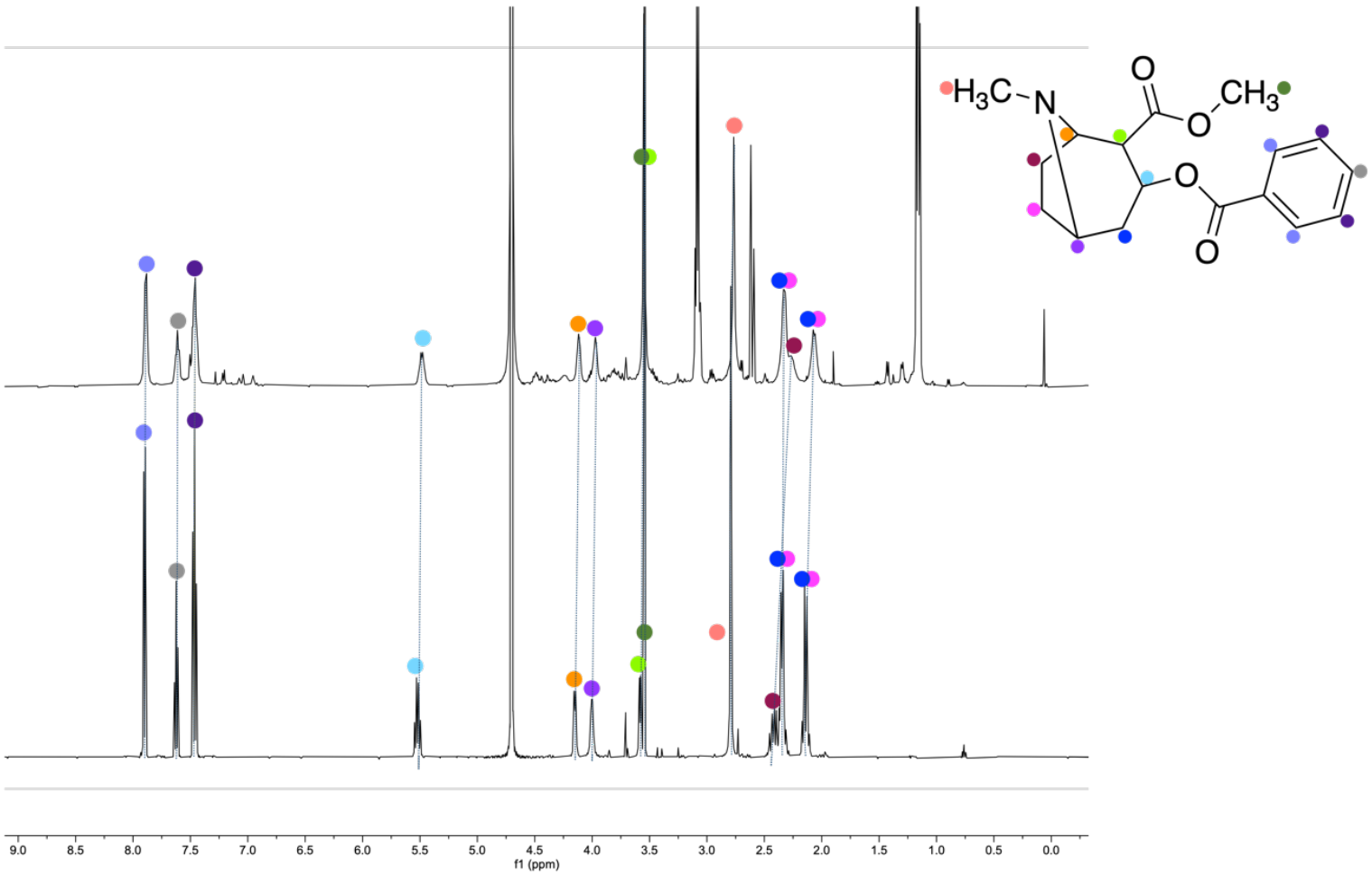


Figure S4.32. Stack plot of cocaine (bottom) and **sCx4-P1** with cocaine (top)

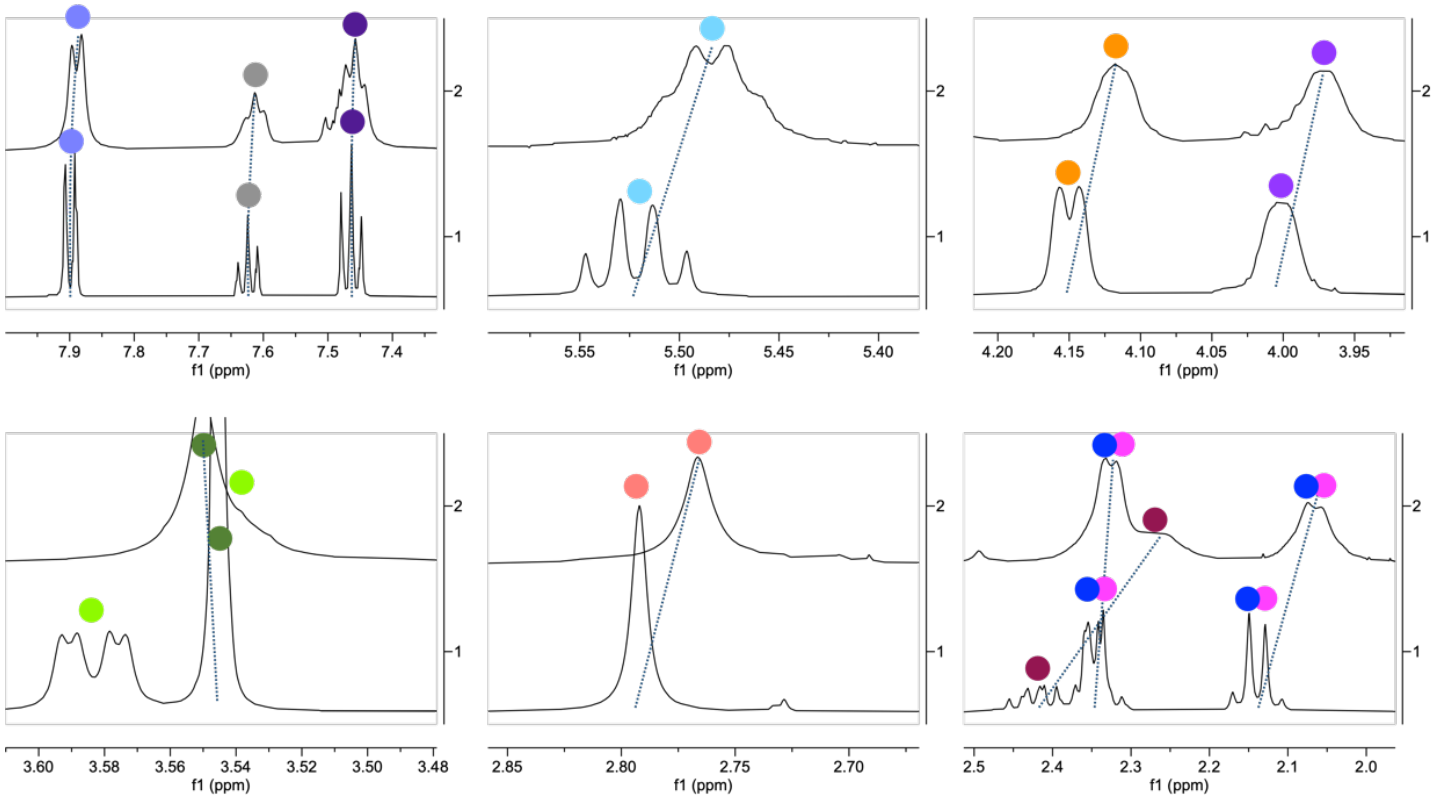


Figure S4.33. Expansion stack plot of cocaine (bottom) and **sCx4-P1** with cocaine (top).

Chapter 5: Conclusion and future directions

5.1 Contributions to the field of supramolecular chemistry

Host-guest chemistry in water and biological solutions is challenging. The prediction of chemical behaviours in water is one of the major challenges in supramolecular chemistry today. The rational design of hosts that are strong and specific in water is hard, and computational predictions are generally poor. In the literature, the structure-function lessons are plentiful, but are often performed by comparing a few related structures within a single class of hosts, or (more commonly), comparing many different guest molecules binding to a single host. Even though this approach is useful, making comparisons across host classes and across even more diverse sets of hosts within a single class could lead to new insights. At the start of this thesis work there were few examples of diversity-oriented synthesis of hosts. Chapter 1 reviews the literature to date but shows that there is still lot of progress that can be made.

We have synthesized a set of novel calix[4]arene hosts and studied their binding properties, adding to the fundamental understanding of host-guest binding. In Chapter 2 nicotine and its metabolites nornicotine and cotinine were studied. This contributed to the fundamental knowledge of binding motifs and strengths. From this study we learned that different substitution patterns on the upper-rim of the calixarene lead to a difference in binding strengths and selectivity. In Chapter 3 we continued this study, but this time expanding the hosts to include acyclic cucurbiturils and cleft-like hosts, with a set of closely related amino acids as guests. This contributed to a deeper understanding of how

host geometry influence host-guest binding and how it can mostly outweigh the more obvious differences in functional group identity and arrangement that typically dominate our thinking about host-guest binding.

We also advanced the field of high-throughput host synthesis by using phage display to create a large library calix[4]arene hosts. In Chapter 4 phage display was used to synthesize a large library (10^9) of macrobicyclic calix[4]arene hosts. With the knowledge of Chapter 3, that host geometry is a major influence in host guest binding, we created a host library consisting of macrobicyclic hosts that can encapsulate the target. Using phage display we stepped away from designing the peptide sequences, where previously a level of design was still needed when modifying calixarenes with peptides. However, improvements are needed for the selection of hosts from this library, as we were unable to select a tight binding host with the affinity pull-down.

5.2 Future directions

Although great progress has been made by the development of a large calix[4]arene library by utilising phage displayed peptides, improvements are still needed. The first improvement to be made is in the affinity pull-down step itself. As covered in Chapter 4, the affinity pull-down did not lead to the desired outcome of a strong binding host selected with a target of interest. The fact that calixarene-peptide hybrids were selected that did not have strong or specific affinity suggests that some amount of weak or non-specific binding is involved in the selection process. To continue this work, multiple rounds of bio-panning should be included in the affinity pull-down, to increase

the stringency for selections of strong binders.[214] After achieving a successful affinity pull-down, this work could be extended on by different avenues.

The first such avenue could be the calixarene hosts that get grafted onto the phage displayed peptides. Modifying these calixarenes could alter the reactivity and binding pocket in a favourable way. One could synthesize different cysteine reactive handles such as acrylamide or iodoacetamide handles (Figure 5.1a). Changing the reactivity of the handles with the cysteines could change the alkylation efficiency, possibly increasing the number of hosts in the library and/or increasing the level of phage viability after alkylation. The binding pocket size of the macrobicyclic hosts also could be altered by changing the length of the peptide or the number of phenols in the calixarenes (Figure 5.1b). A smaller binding cavity could be more selective for the smaller targets where a bigger cavity could accommodate larger targets or prove to be more selective by the increased number of interactions. Shortening the peptide or changing the calixarene structure could avoid the hydrophobic collapse that we observed for the structures created in Chapter 4. The overall topology of the host-peptide hybrid could be altered by tethering only one end of the peptide to the calixarene host (Figure 5.1c). This way the peptide will be less restricted in its movements, less likely to experience hydrophobic collapse, and possible able to make more favourable interactions by the peptide folding over the target.

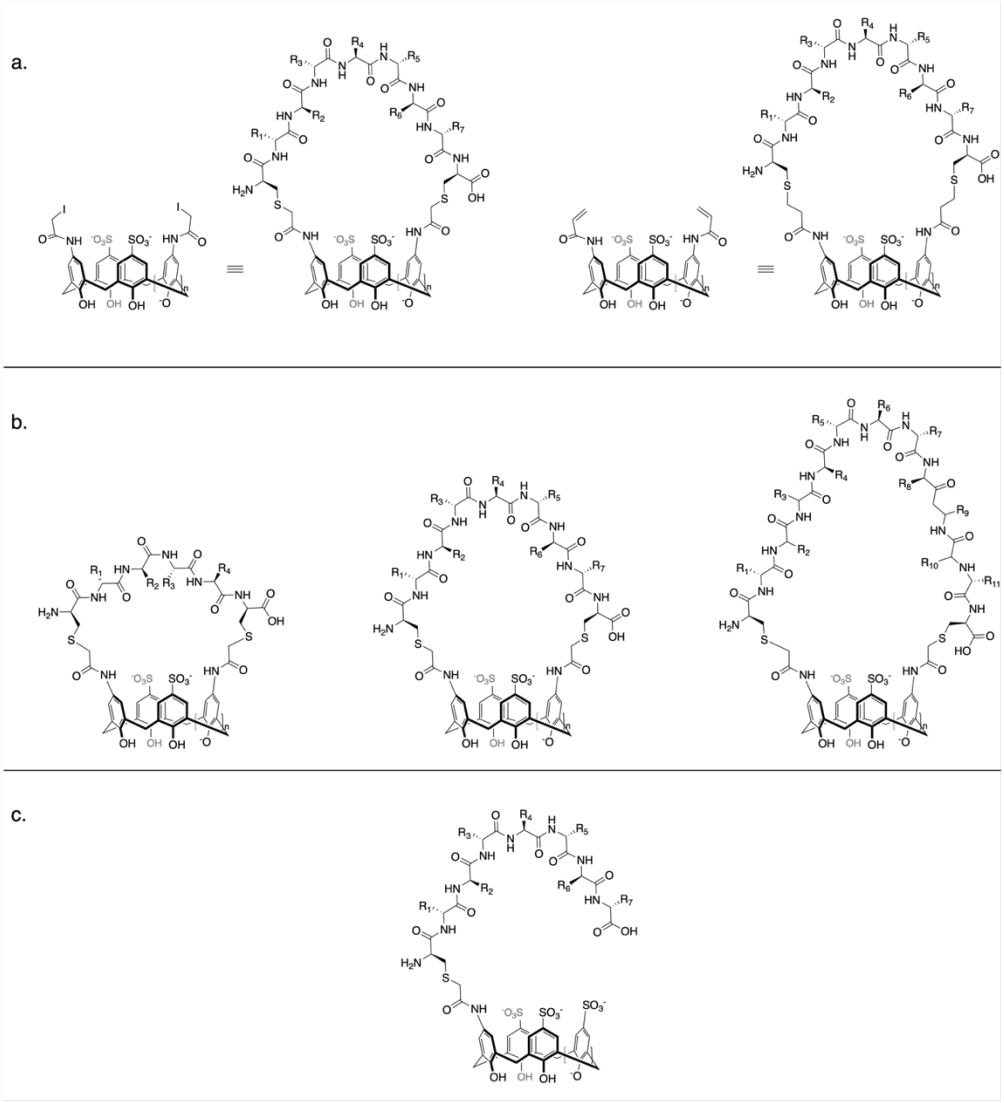


Figure 5.1. Possible calixarene modifications. a) iodoacetamide (left) and acrylamide (right) handles on calix[n]arene host and resulting macrobicyclic hosts. b) Calix[n]arene hosts with different length peptides CX_nC where n = 4, 7, 11 left to right. c) Calix[n]arene host with peptide tethered onto one position only.

The second avenue would be extending the list of targets, as the real power of the phage display approach is the number of hosts present in the library all with slightly different properties. Within this massive library, hosts should be present that are a good

fit for many different guests. In Chapter 4 ADMA and cocaine were targeted, however the range of small molecules could be the targets as long as they can be attached to a solid support to make a suitable bait for pull-down from the phage-displayed library. For example other amino acids that are involved in diseases such as phenylalanine (phenylketonuria[148,217]) and leucine, isoleucine and valine (maple syrup urine disease[149]) or different drugs and their metabolites.

It is clear that doing chemistry on phage particles is challenging, so one very different option would be to make a more modestly sized library of calixarene-peptide hybrids by parallel synthesis methods instead of using phage display. This could easily be achieved on (for example) a 96-well plate format in an automated peptide synthesizer, by making 96 different peptides and using the same alkylation chemistry to make the macrobicyclic hosts. This would dramatically decrease the structural diversity present with the host library, but in exchange would give a much-improved ability to track synthetic reaction progress and understand the results of follow-up binding assays.

The big picture goal: combining all these approaches could result in selective macrobicyclic calix[4]arene hosts for a wide array of targets which eventually could be developed into detection methods for metabolic diseases, drug detection, therapeutics and/or reversal agents.

Bibliography

(1) Oshovsky, G. V.; Reinhoudt, D. N.; Verboom, W. Supramolecular chemistry in water. *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 2366-2393.

(2) Katsuhiko, A.; Kanittake, T.: *Supramolecular Chemistry - Fundamentals and Applications*; Springer, 2006.

(3) Cerny, J.; Hobza, P. Non-covalent interactions in biomacromolecules. *Phys Chem Chem Phys* **2007**, *9*, 5291-5303.

(4) Galloway, W. R.; Spring, D. R. Is synthesis the main hurdle for the generation of diversity in compound libraries for screening? *Expert Opin Drug Discov* **2009**, *4*, 467-472.

(5) Tiwari, G.; Tiwari, R.; Rai, A. K. Cyclodextrins in delivery systems: Applications. *J Pharm Bioallied Sci* **2010**, *2*, 72-79.

(6) Gibaud, S.; Zirar, S. B.; Mutzenhardt, P.; Fries, I.; Astier, A. Melarsoprol-cyclodextrins inclusion complexes. *Int J Pharm* **2005**, *306*, 107-121.

(7) Nag, K.; Singh, D. R.; Shetti, A. N.; Kumar, H.; Sivashanmugam, T.; Parthasarathy, S. Sugammadex: A revolutionary drug in neuromuscular pharmacology. *Anesth Essays Res* **2013**, *7*, 302-306.

(8) Puglisi, A.; Yagci, Y. Cyclodextrin-Based Macromolecular Systems as Cholesterol-Mopping Therapeutic Agents in Niemann-Pick Disease Type C. *Macromol. Rapid Commun.* **2019**, *40*, e1800557.

(9) Deng, C. L.; Murkli, S. L.; Isaacs, L. D. Supramolecular hosts as in vivo sequestration agents for pharmaceuticals and toxins. *Chem Soc Rev* **2020**, *49*, 7516-7532.

(10) Yin, H.; Zhang, X.; Wei, J.; Lu, S.; Bardelang, D.; Wang, R. Recent advances in supramolecular antidotes. *Theranostics* **2021**, *11*, 1513-1526.

(11) Ma, X.; Zhao, Y. Biomedical Applications of Supramolecular Systems Based on Host-Guest Interactions. *Chem Rev* **2015**, *115*, 7794-7839.

(12) Hof, F. Host-guest chemistry that directly targets lysine methylation: synthetic host molecules as alternatives to bio-reagents. *Chem Commun (Camb)* **2016**, *52*, 10093-10108.

(13) 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**, *50*, 4812-4832.

(14) Bosshard, H. R.; Marti, D. N.; Jelesarov, I. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J Mol Recognit* **2004**, *17*, 1-16.

(15) Sippel, K. H.; Quiocho, F. A. Ion-dipole interactions and their functions in proteins. *Protein Sci* **2015**, *24*, 1040-1046.

(16) Ganesan, S. J.; Matysiak, S. Role of Backbone Dipole Interactions in the Formation of Secondary and Supersecondary Structures of Proteins. *J Chem Theory Comput* **2014**, *10*, 2569-2576.

(17) Riera, T. V.; Zheng, L.; Josephine, H. R.; Min, D.; Yang, W.; Hedstrom, L. Allosteric activation via kinetic control: potassium accelerates a conformational change in IMP dehydrogenase. *Biochemistry* **2011**, *50*, 8508-8518.

(18) Kumar, R.; Schmidt, J. R.; Skinner, J. L. Hydrogen bonding definitions and dynamics in liquid water. *J Chem Phys* **2007**, *126*, 204107.

(19) Steiner, T. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 49-76.

(20) Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. Aromatic interactions. *J. Chem. Soc. Perk. T2.* **2001**, 651-669.

(21) Dougherty, D. A. The cation-pi interaction. *Acc. Chem. Res.* **2013**, *46*, 885-893.

(22) Escobar, L.; Ballester, P. Molecular Recognition in Water Using Macrocyclic Synthetic Receptors. *Chem Rev* **2021**, *121*, 2445-2514.

(23) Camilloni, C.; Bonetti, D.; Morrone, A.; Giri, R.; Dobson, C. M.; Brunori, M.; Gianni, S.; Vendruscolo, M. Towards a structural biology of the hydrophobic effect in protein folding. *Sci Rep* **2016**, *6*, 28285.

(24) Doyle, R. J. Contribution of the hydrophobic effect to microbial infection. *Microbes Infect* **2000**, *2*, 391-400.

(25) Gibb, B. C. Hydrophobia! *Nat Chem* **2010**, *2*, 512-513.

(26) Biedermann, F.; Nau, W. M.; Schneider, H. J. The hydrophobic effect revisited--studies with supramolecular complexes imply high-energy water as a noncovalent driving force. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 11158-11171.

(27) Hillyer, M. B.; Gibb, B. C. Molecular Shape and the Hydrophobic Effect. *Annu Rev Phys Chem* **2016**, *67*, 307-329.

(28) Jacobson, L. C.; Molinero, V. A methane-water model for coarse-grained simulations of solutions and clathrate hydrates. *J Phys Chem B* **2010**, *114*, 7302-7311.

(29) Bogunia, M.; Makowski, M. Influence of Ionic Strength on Hydrophobic Interactions in Water: Dependence on Solute Size and Shape. *J Phys Chem B* **2020**, *124*, 10326-10336.

(30) Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640-647.

(31) Barnett, J. W.; Sullivan, M. R.; Long, J. A.; Tang, D.; Nguyen, T.; Ben-Amotz, D.; Gibb, B. C.; Ashbaugh, H. S. Spontaneous drying of non-polar deep-cavity cavitand pockets in aqueous solution. *Nat Chem* **2020**, *12*, 589-594.

(32) Kunz, W.; Henle, J.; Ninham, B. W. 'Zur Lehre von der Wirkung der Salze' (about the science of the effect of salts): Franz Hofmeister's historical papers. *Current Opinion in Colloid & Interface Science* **2004**, *9*, 19-37.

(33) Jungwirth, P.; Cremer, P. S. Beyond Hofmeister. *Nat Chem* **2014**, *6*, 261-263.

(34) *Supramolecular Chemistry in Water*; Wiley-VCH, 2019.

(35) Galloway, W. R.; Isidro-Llobet, A.; Spring, D. R. Diversity-oriented synthesis as a tool for the discovery of novel biologically active small molecules. *Nat Commun* **2010**, *1*, 80.

(36) Spandl, R. J.; Diaz-Gavilan, M.; O'Connell, K. M.; Thomas, G. L.; Spring, D. R. Diversity-oriented synthesis. *Chem Rec* **2008**, *8*, 129-142.

(37) Mondal, M.; Radeva, N.; Fanlo-Virgos, H.; Otto, S.; Klebe, G.; Hirsch, A. K. Fragment Linking and Optimization of Inhibitors of the Aspartic Protease Endothiapepsin: Fragment-Based Drug Design Facilitated by Dynamic Combinatorial Chemistry. *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 9422-9426.

(38) Mondal, M.; Radeva, N.; Koster, H.; Park, A.; Potamitis, C.; Zervou, M.; Klebe, G.; Hirsch, A. K. Structure-based design of inhibitors of the aspartic protease endothiapepsin by exploiting dynamic combinatorial chemistry. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 3259-3263.

(39) Crini, G. Review: a history of cyclodextrins. *Chem Rev* **2014**, *114*, 10940-10975.

(40) Szente, L.; Szeman, J.; Sohajda, T. Analytical characterization of cyclodextrins: History, official methods and recommended new techniques. *J. Pharm. Biomed. Anal.* **2016**, *130*, 347-365.

(41) Larsen, D.; Beeren, S. R. Enzyme-mediated dynamic combinatorial chemistry allows out-of-equilibrium template-directed synthesis of macrocyclic oligosaccharides. *Chem Sci* **2019**, *10*, 9981-9987.

(42) Larsen, D.; Beeren, S. R. Tuning the Outcome of Enzyme-Mediated Dynamic Cyclodextrin Libraries to Enhance Template Effects. *Chemistry* **2020**, *26*, 11032-11038.

(43) Larsen, D.; Bjerre, P. M.; Beeren, S. R. Light-controlled out-of-equilibrium assembly of cyclodextrins in an enzyme-mediated dynamic system. *Chem Commun (Camb)* **2019**, *55*, 15037-15040.

(44) Larsen, D.; Beeren, S. R. Building up cyclodextrins from scratch - templated enzymatic synthesis of cyclodextrins directly from maltose. *Chem Commun (Camb)* **2021**, *57*, 2503-2506.

(45) Erichsen, A.; Larsen, D.; Beeren, S. R. Chaotropic and Kosmotropic Anions Regulate the Outcome of Enzyme-Mediated Dynamic Combinatorial Libraries of Cyclodextrins in Two Different Ways. *Front Chem* **2021**, *9*, 721942.

(46) Larsen, D.; Ferreira, M.; Tilloy, S.; Monflier, E.; Beeren, S. R. Unnatural cyclodextrins can be accessed from enzyme-mediated dynamic combinatorial libraries. *Chem Commun (Camb)* **2022**, *58*, 2287-2290.

(47) Pedersen, C. J. The discovery of crown ethers. *Science* **1988**, *241*, 536-540.

(48) van Karnebeek, C. D.; Stockler, S. Treatable inborn errors of metabolism causing intellectual disability: a systematic literature review. *Mol Genet Metab* **2012**, *105*, 368-381.

(49) Pedersen, C. J. Cyclic Polyethers and Their Complexes with Metal Salts. *J. Am. Chem. Soc.* **1967**, *89*, 7017-&.

(50) Monvisade, P.; Hodge, P.; Ruddick, C. L. Synthesis of soluble combinatorial libraries of crown ether-ester analogues via the cyclodepolymerisation of linear polyesters. *Chem Commun* **1999**, 1987-1988.

(51) Fukunaga, K.; Hatanaka, T.; Ito, Y.; Minami, M.; Taki, M. Construction of a crown ether-like supramolecular library by conjugation of genetically-encoded peptide linkers displayed on bacteriophage T7. *Chem Commun (Camb)* **2014**, *50*, 3921-3923.

(52) Bazan, J.; Calkosinski, I.; Gamian, A. Phage display--a powerful technique for immunotherapy: 2. Vaccine delivery. *Hum Vaccin Immunother* **2012**, *8*, 1829-1835.

(53) Mochizuki, K.; Matsukura, L.; Ito, Y.; Miyashita, N.; Taki, M. A medium-firm drug-candidate library of cryptand-like structures on T7 phage: design and selection of a strong binder for Hsp90. *Org Biomol Chem* **2021**, *19*, 146-150.

(54) Ngola, S. M.; Dougherty, D. A. Evidence for the Importance of Polarizability in Biomimetic Catalysis Involving Cyclophane Receptors. *J Org Chem* **1996**, *61*, 4355-4360.

(55) Otto, S.; Furlan, R. L.; Sanders, J. K. Selection and amplification of hosts from dynamic combinatorial libraries of macrocyclic disulfides. *Science* **2002**, *297*, 590-593.

(56) Brisig, B.; Sanders, J. K.; Otto, S. Selection and amplification of a catalyst from a dynamic combinatorial library. *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 1270-1273.

(57) Corbett, P. T.; Tong, L. H.; Sanders, J. K.; Otto, S. Diastereoselective amplification of an induced-fit receptor from a dynamic combinatorial library. *J Am Chem Soc* **2005**, *127*, 8902-8903.

(58) Corbett, P. T.; Sanders, J. K.; Otto, S. Exploring the relation between amplification and binding in dynamic combinatorial libraries of macrocyclic synthetic receptors in water. *Chemistry* **2008**, *14*, 2153-2166.

(59) Ingerman, L. A.; 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 (Camb)* **2010**, *46*, 1839-1841.

(60) 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.

(61) Pinkin, N. K.; Waters, M. L. Development and mechanistic studies of an optimized receptor for trimethyllysine using iterative redesign by dynamic combinatorial chemistry. *Org Biomol Chem* **2014**, *12*, 7059-7067.

(62) Gober, I. N.; Waters, M. L. Optimization of a synthetic receptor for dimethyllysine using a biphenyl-2,6-dicarboxylic acid scaffold: insights into selective recognition of hydrophilic guests in water. *Org Biomol Chem* **2017**, *15*, 7789-7795.

(63) 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.

(64) Mullins, A. G.; St Louis, L. E.; Waters, M. L. Using changes in speciation in a dynamic combinatorial library as a fingerprint to differentiate the methylation states of arginine. *Chem Commun (Camb)* **2020**, *56*, 3947-3950.

(65) Harrison, E. E.; Carpenter, B. A.; St Louis, L. E.; Mullins, A. G.; Waters, M. L. Development of "Imprint-and-Report" Dynamic Combinatorial Libraries for Differential Sensing Applications. *J Am Chem Soc* **2021**, *143*, 14845-14854.

(66) Hanna, T. A.; Liu, L. H.; Angeles, A.; Gutsche, D. C. Partially metallated calixarenes. *Abstr Pap Am Chem S* **2002**, *223*, A63-A63.

(67) Neri, P.; Sessler, J. L.; Wang, M.-X.: *Calixarenes and Beyond*, 2016.

(68) Baekeland, L. H. Method of making insoluble products of phenol and formaldehyde. *Patent* **1909**, 942, 699.

(69) Gutsche, C. D.; Bauer, L. J. Calixarenes .13. The Conformational Properties of Calix[4]Arenes, Calix[6]Arenes, Calix[8]Arenes, and Oxacalixarenes. *J Am Chem Soc* **1985**, *107*, 6052-6059.

(70) Harvey, P. D. Wide-rim and outer-face functionalizations of calix[4]arene. *Coordin Chem Rev* **2002**, *233*, 289-309.

(71) 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.

(72) Tabet, S.; Douglas, S. F.; Daze, K. D.; Garnett, G. A.; Allen, K. J.; Abrioux, E. M.; Quon, T. T.; Wulff, J. E.; Hof, F. Synthetic trimethyllysine receptors that bind histone 3, trimethyllysine 27 (H3K27me3) and disrupt its interaction with the epigenetic reader protein CBX7. *Bioorg Med Chem* **2013**, *21*, 7004-7010.

(73) Ali, M.; Daze, K. D.; Strongin, D. E.; Rothbart, S. B.; Rincon-Arano, H.; Allen, H. F.; Li, J.; Strahl, B. D.; Hof, F.; Kutateladze, T. G. Molecular Insights into Inhibition of the Methylated Histone-Plant Homeodomain Complexes by Calixarenes. *J. Biol. Chem.* **2015**, *290*, 22919-22930.

(74) Allen, H. F.; Daze, K. D.; Lai, A.; Musselman, C. A.; Sims, J. K.; Wade, P. A.; Hof, F.; Kutateladze, T. G. Inhibition of histone readers by supramolecular hosts. *Biochem. J.* **2014**, *459*, 505–512.

(75) Wang, Z.: *Encyclopedia of Physical Organic Chemistry*; John Wiley & Sons, Inc, 2016.

(76) Arduini, A.; Pochini, A.; Reverberi, S.; Ungar, R. p-t-Butyl-calix[4]areneTetracarboxylic Acid. A Water Soluble Calixarene in a Cone Structure. *J. Chem. Soc. Chem. Commun.* **1984**, 981 - 982.

(77) Shinkai, S.; Mori, S.; Tsubaki, T.; Sone, T.; Manabe, O. New Water-Soluble Host Molecules Derived from Calix[6]Arene. *Tetrahedron Letters* **1984**, *25*, 5315-5318.

(78) Perret, F.; Lazar, A. N.; Coleman, A. W. Biochemistry of the para-sulfonato-calix[n]arenes. *Chem Commun (Camb)* **2006**, 2425-2438.

(79) Selinger, A. J.; Cavallin, N. A.; Yanai, A.; Birol, I.; Hof, F. Template-Directed Synthesis of Bivalent, Broad-Spectrum Hosts for Neuromuscular Blocking Agents. *Angew. Chem. Int. Ed. Engl.* **2022**, *61*, e202113235.

(80) Nimse, S. B.; Kim, T. Biological applications of functionalized calixarenes. *Chem Soc Rev* **2013**, *42*, 366-386.

(81) Coleman, A. W.; Jebors, S.; Cecillon, S.; Perret, P.; Garin, D.; Marti-Battle, D.; Moulin, M. Toxicity and biodistribution of para-sulfonato-calix[4] arene in mice. *New J Chem* **2008**, *32*, 780-782.

(82) Shinkai, S.; Araki, K.; Manabe, O. Does the Calixarene Cavity recognise the Size of Guest Molecules? On the 'Hole-size Selectivity' in Water-soluble Calixarenes. *J. Am. Chem. Soc.* **1998**, 3.

(83) Hioki, H.; Yamada, T.; Fujioka, C.; Kodama, M. Peptide library based on calix[4]arene. *Tetrahedron Letters* **1999**, *40*, 6821-6825.

(84) Hioki, H.; Kubo, M.; Yoshida, H.; Bando, M.; Ohnishi, Y.; Kodama, M. Synthesis of fluorescence-labeled peptidocalix[4]arene library and its peptide sensing ability. *Tetrahedron Letters* **2002**, *43*, 7949-7952.

(85) Hioki, H.; Ohnishi, Y.; Kubo, M.; Nashimoto, E.; Kinoshita, Y.; Samejima, M.; Kodama, M. Synthesis of calix[4]arene library substituted with peptides at the upper rim. *Tetrahedron Letters* **2004**, *45*, 561-564.

(86) Kubo, M.; Nashimoto, E.; Tokiyo, T.; Morisaki, Y.; Kodama, M.; Hioki, H. Development of calixarene-based host molecules for peptides in aqueous media. *Tetrahedron Letters* **2006**, *47*, 1927-1931.

(87) Hioki, H.; Nishimoto, R.; Kawaguchi, K.; Kubo, M.; Harada, K.; Fukuyama, Y. Discovery of hydrolytic catalysts in a peptidocalixarene library by binding assay with a transition state analogue for the hydrolysis. *Chem Commun (Camb)* **2009**, 7194-7196.

(88) Brewster, R. E.; Dalton, B. G.; Shuker, S. B. Synthesis of a peptidocalix[4]arene library and identification of compounds with hydrolytic activity. *Bioorg Chem* **2005**, *33*, 16-21.

(89) Kuhnert, N.; Le-Gresley, A. The use of deep cavity tetraformyl calix[4]arenes in the synthesis of static and dynamic macrocyclic libraries. *Tetrahedron Letters* **2005**, *46*, 2059-2062.

(90) Le-Gresley, A.; Kuhnert, N. The design and synthesis of acrylato and imino derivatives of calix[4]arene for applications in static and dynamic combinatorial libraries. *Journal of Chemical Research* **2010**, 61-67.

(91) Beatty, M. A.; Selinger, A. J.; Li, Y.; Hof, F. Parallel Synthesis and Screening of Supramolecular Chemosensors That Achieve Fluorescent Turn-on Detection of Drugs in Saliva. *J Am Chem Soc* **2019**, *141*, 16763-16771.

(92) Benowitz, N. L. Clinical pharmacology of nicotine: implications for understanding, preventing, and treating tobacco addiction. *Clin Pharmacol Ther* **2008**, *83*, 531-541.

(93) Zhou, Y. Y.; Yu, H. P.; Zhang, L.; Xu, H. W.; Wu, L. A.; Sun, J. Y.; Wang, L. A new spectrofluorometric method for the determination of nicotine base on the inclusion interaction of methylene blue and cucurbit[7]uril. *Microchim Acta* **2009**, *164*, 63-68.

(94) Minami, T.; Esipenko, N. A.; Zhang, B.; Kozelkova, M. E.; Isaacs, L.; Nishiyabu, R.; Kubo, Y.; Anzenbacher, P., Jr. Supramolecular sensor for cancer-associated nitrosamines. *J Am Chem Soc* **2012**, *134*, 20021-20024.

(95) Berglund, J.; Cedergren, L.; Andersson, S. B. Determination of the stability constant for the inclusion complex between beta-cyclodextrin and nicotine using capillary electrophoresis. *Int J Pharm* **1997**, *156*, 195-200.

(96) Tlusty, M.; Slavik, P.; Kohout, M.; Eigner, V.; Lhotak, P. Inherently Chiral Upper-Rim-Bridged Calix[4]arenes Possessing a Seven Membered Ring. *Org Lett* **2017**, *19*, 2933-2936.

(97) Vanloon, J. D.; Arduini, A.; Coppi, L.; Verboom, W.; Pochini, A.; Ungaro, R.; Harkema, S.; Reinhoudt, D. N. Selective Functionalization of Calix[4]Arenes at the Upper Rim. *J Org Chem* **1990**, *55*, 5639-5646.

(98) Gutsche, C. D.; Lin, L. G. Calixarenes .12. The Synthesis of Functionalized Calixarenes. *Tetrahedron* **1986**, *42*, 1633-1640.

(99) Berthalon, S.; deVains, J. B. R.; Lamartine, R. Selective mono-dealkylation of tetra-p-tert-butyl-calix[4]arene at the upper rim. *Synthetic Communications* **1996**, *26*, 3103-3108.

(100) Daze, K. D.; Ma, M. C. F.; Pineux, F.; Hof, F. Synthesis of New Trisulfonated Calix[4]arenes Functionalized at the Upper Rim, and Their Complexation with the Trimethyllysine Epigenetic Mark. *Organic Letters* **2012**, *14*, 1512-1515.

(101) Tabet, S.; Douglas, S. F.; Daze, K. D.; Garnett, G. A. E.; Allen, K. J. H.; Abrioux, E. M. M.; Quon, T. T. H.; Wulff, J. E.; Hof, F. Synthetic trimethyllysine receptors that bind histone 3, trimethyllysine 27 (H3K27me3) and disrupt its interaction with the epigenetic reader protein CBX7. *Bioorgan Med Chem* **2013**, *21*, 7004-7010.

(102) 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.

(103) Lynch, J. A.; Mestayer, J. J.; Blanda, M. T. Efficient syntheses of calix[4]arenes in the 1,2-alternate conformation via intramolecular benzoate ester migrations. *Journal of Supramolecular Chemistry* **2001**, *1*, 139-145.

(104) Olah, G. A.; Narang, S. C.; Olah, J. A.; Lammertsma, K. Recent Aspects of Nitration - New Preparative Methods and Mechanistic Studies (a Review). *P Natl Acad Sci USA* **1982**, *79*, 4487-4494.

(105) Guo, D. S.; Uzunova, V. D.; Su, X.; Liu, Y.; Nau, W. M. Operational calixarene-based fluorescent sensing systems for choline and acetylcholine and their application to enzymatic reactions. *Chem Sci* **2011**, *2*, 1722-1734.

(106) Dsouza, R. N.; Pischel, U.; Nau, W. M. Fluorescent Dyes and Their Supramolecular Host/Guest Complexes with Macrocycles in Aqueous Solution. *Chem Rev* **2011**, *111*, 7941-7980.

(107) Thordarson, P. Determining association constants from titration experiments in supramolecular chemistry (vol 40, pg 1305, 2011). *Chem. Soc. Rev.* **2011**, *40*, 5922-5923.

(108) Hibbert, D. B.; Thordarson, P. The death of the Job plot, transparency, open science and online tools, uncertainty estimation methods and other developments in supramolecular chemistry data analysis. *Chem Commun* **2016**, *52*, 12792-12805.

(109) Laughrey, Z. R.; Gibb, C. L. D.; Senechal, T.; Gibb, B. C. Guest binding and orientation within open nanoscale hosts. *Chem-Eur J* **2003**, *9*, 130-139.

(110) Twum, K.; Rautiainen, J. M.; Yu, S. L.; Truong, K. N.; Feder, J.; Rissanen, K.; Puttreddy, R.; Beyeh, N. K. Host-Guest Interactions of Sodiumsulfonatomethyleneresorcinarene and Quaternary Ammonium Halides: An Experimental-Computational Analysis of the Guest Inclusion Properties. *Crystal Growth & Design* **2020**, *20*, 2367-2376.

(111) 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.

(112) Guo, D. S.; Wang, K.; Liu, Y. Selective binding behaviors of p-sulfonatocalixarenes in aqueous solution. *J Incl Phenom Macro* **2008**, *62*, 1-21.

(113) Selkti, M.; Coleman, A. W.; Nicolis, I.; Douteau-Guevel, N.; Villain, F.; Tomas, A.; de Rango, C. The first example of a substrate spanning the calix[4]arene bilayer: the solid state complex of p-sulfonatocalix[4]arene with L-lysine. *Chem Commun* **2000**, 161-162.

(114) 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.

(115) DeBerardinis, R. J.; Thompson, C. B. Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* **2012**, *148*, 1132-1144.

(116) 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.

(117) 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.

(118) 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.

(119) Kungulovski, G.; Jeltsch, A. Quality of histone modification antibodies undermines chromatin biology research. *F1000Res* **2015**, *4*, 1160.

(120) 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.

(121) 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.

(122) Martins, J. N.; Lima, J. C.; Basilio, N. Selective Recognition of Amino Acids and Peptides by Small Supramolecular Receptors. *Molecules* **2020**, *26*.

(123) Douteau-Guevel, N.; Coleman, A. W.; Morel, J. P.; Morel-Desrosiers, N. Complexation of the basic amino acids lysine and arginine by three sulfonatocalix[n]arenes (n = 4, 6 and 8) in water: microcalorimetric determination of the Gibbs energies, enthalpies and entropies of complexation. *J Chem Soc Perk T 2* **1999**, 629-633.

(124) Douteau-Guevel, N.; Coleman, A. W.; Morel, J. P.; Morel-Desrosiers, N. Complexation of basic amino acids by water-soluble calixarene sulphonates as a study of the possible mechanisms of recognition of calixarene sulphonates by proteins. *J Phys Org Chem* **1998**, *11*, 693-696.

(125) Hessz, D.; Badogos, S.; Bojtar, M.; Bitter, I.; Drahos, L.; Kubinyi, M. Complexes of carboxylato pillar[6]arene with Brooker-type merocyanines: Spectral properties, pK(a) shifts and the design of a displacement assay for trimethyl lysine. *Spectrochim Acta A Mol Biomol Spectrosc* **2021**, *252*, 119455.

(126) Li, C.; Ma, J.; Zhao, L.; Zhang, Y.; Yu, Y.; Shu, X.; Li, J.; Jia, X. Molecular selective binding of basic amino acids by a water-soluble pillar[5]arene. *Chem Commun (Camb)* **2013**, *49*, 1924-1926.

(127) Gamal-Eldin, M. A.; Macartney, D. H. Selective molecular recognition of methylated lysines and arginines by cucurbit[6]uril and cucurbit[7]uril in aqueous solution. *Org Biomol Chem* **2013**, *11*, 488-495.

(128) Biedermann, F.; Nau, W. M. Noncovalent chirality sensing ensembles for the detection and reaction monitoring of amino acids, peptides, proteins, and aromatic drugs. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 5694-5699.

(129) Bailey, D. M.; Hennig, A.; Uzunova, V. D.; Nau, W. M. Supramolecular tandem enzyme assays for multiparameter sensor arrays and enantiomeric excess determination of amino acids. *Chemistry* **2008**, *14*, 6069-6077.

(130) Roy, M. N.; Ekka, D.; Saha, S.; Roy, M. C. Host-guest inclusion complexes of alpha and beta-cyclodextrins with alpha-amino acids. *Rsc Adv* **2014**, *4*, 42383-42390.

(131) 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.

(132) 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.

(133) Hadrovic, I.; Rebmann, P.; Klarner, F. G.; Bitan, G.; Schrader, T. Molecular Lysine Tweezers Counteract Aberrant Protein Aggregation. *Front Chem* **2019**, *7*, 657.

(134) 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.

(135) McQuinn, K.; McIndoe, J. S.; Hof, F. Insights into the post-translational methylation of arginine from studies of guanidinium-water nanodroplets. *Chemistry* **2008**, *14*, 6483-6489.

(136) Cooke, J. P. ADMA: its role in vascular disease. *Vasc Med* **2005**, *10 Suppl 1*, S11-17.

(137) Sibal, L.; Agarwal, S. C.; Home, P. D.; Boger, R. H. The Role of Asymmetric Dimethylarginine (ADMA) in Endothelial Dysfunction and Cardiovascular Disease. *Curr Cardiol Rev* **2010**, *6*, 82-90.

(138) Kielstein, J. T.; Salpeter, S. R.; Bode-Boeger, S. M.; Cooke, J. P.; Fliser, D. Symmetric dimethylarginine (SDMA) as endogenous marker of renal function--a meta-analysis. *Nephrol. Dial. Transplant.* **2006**, *21*, 2446-2451.

(139) Tutarel, O.; Denecke, A.; Bode-Boger, S. M.; Martens-Lobenhoffer, J.; Schieffer, B.; Westhoff-Bleck, M.; Kielstein, J. T. Symmetrical dimethylarginine outperforms CKD-EPI and MDRD-derived eGFR for the assessment of renal function in patients with adult congenital heart disease. *Kidney Blood Press Res* **2011**, *34*, 41-45.

(140) Schepers, E.; Barreto, D. V.; Liabeuf, S.; Glorieux, G.; Eloot, S.; Barreto, F. C.; Massy, Z.; Vanholder, R.; European Uremic Toxin Work, G. Symmetric dimethylarginine as a proinflammatory agent in chronic kidney disease. *Clin J Am Soc Nephrol* **2011**, *6*, 2374-2383.

(141) Tain, Y. L.; Hsu, C. N. Toxic Dimethylarginines: Asymmetric Dimethylarginine (ADMA) and Symmetric Dimethylarginine (SDMA). *Toxins (Basel)* **2017**, *9*.

(142) Nemeth, B.; Ajtay, Z.; Hejjel, L.; Ferenci, T.; Abram, Z.; Muranyi, E.; Kiss, I. The issue of plasma asymmetric dimethylarginine reference range - A systematic review and meta-analysis. *PLoS One* **2017**, *12*, e0177493.

(143) Erre, G. L.; Mangoni, A. A.; Castagna, F.; Paliogiannis, P.; Carru, C.; Passiu, G.; Zinellu, A. Meta-Analysis of Asymmetric Dimethylarginine Concentrations in Rheumatic Diseases. *Sci Rep* **2019**, *9*, 5426.

(144) Fleck, C.; Janz, A.; Schweitzer, F.; Karge, E.; Schwertfeger, M.; Stein, G. Serum concentrations of asymmetric (ADMA) and symmetric (SDMA) dimethylarginine in renal failure patients. *Kidney Int Suppl* **2001**, *78*, S14-18.

(145) 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.

(146) Gruber, T. Synthetic Receptors for the Recognition and Discrimination of Post-Translationally Methylated Lysines. *Chembiochem* **2018**, *19*, 2324-2340.

(147) 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.

(148) Adler-Abramovich, L.; Vaks, L.; Carny, O.; Trudler, D.; Magno, A.; Caflisch, A.; Frenkel, D.; Gazit, E. Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria. *Nat Chem Biol* **2012**, *8*, 701-706.

(149) 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.

(150) 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.

(151) Dsouza, R. N.; Pischel, U.; Nau, W. M. Fluorescent dyes and their supramolecular host/guest complexes with macrocycles in aqueous solution. *Chem Rev* **2011**, *111*, 7941-7980.

(152) 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*, 88-96.

(153) 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.

(154) Ma, D.; Zavalij, P. Y.; Isaacs, L. Acyclic cucurbit[n]uril congeners are high affinity hosts. *J Org Chem* **2010**, *75*, 4786-4795.

(155) Cao, L.; Sekutor, M.; Zavalij, P. Y.; Mlinaric-Majerski, K.; Glaser, R.; Isaacs, L. Cucurbit[7]uril·guest pair with an attomolar dissociation constant. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 988-993.

(156) Fokkens, M.; Schrader, T.; Klarner, F. G. A molecular tweezer for lysine and arginine. *J Am Chem Soc* **2005**, *127*, 14415-14421.

(157) Jordan, J. H.; Ashbaugh, H. S.; Mague, J. T.; Gibb, B. C. Buffer and Salt Effects in Aqueous Host-Guest Systems: Screening, Competitive Binding, or Both? *J Am Chem Soc* **2021**, *143*, 18605-18616.

(158) Jordan, J. H.; Gibb, C. L. D.; Wishard, A.; Pham, T.; Gibb, B. C. Ion-Hydrocarbon and/or Ion-Ion Interactions: Direct and Reverse Hofmeister Effects in a Synthetic Host. *J Am Chem Soc* **2018**, *140*, 4092-4099.

(159) Carnegie, R. S.; Gibb, C. L.; Gibb, B. C. Anion complexation and the Hofmeister effect. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 11498-11500.

(160) Dutt, S.; Wilch, C.; Gersthagen, T.; Talbiersky, P.; Bravo-Rodriguez, K.; Hanni, M.; Sanchez-Garcia, E.; Ochsenfeld, C.; Klarner, F. G.; Schrader, T. Molecular tweezers with varying anions: a comparative study. *J Org Chem* **2013**, *78*, 6721-6734.

(161) Ndendjio, S. A. Z.; Isaacs, L. Molecular recognition properties of acyclic cucurbiturils toward amino acids, peptides, and a protein. *Supramol Chem* **2019**, *31*, 432-441.

(162) Gibb, C. L.; Oertling, E. E.; Velaga, S.; Gibb, B. C. Thermodynamic profiles of salt effects on a host-guest system: new insight into the Hofmeister effect. *J Phys Chem B* **2015**, *119*, 5624-5638.

(163) Evich, M.; Stroeva, E.; Zheng, Y. G.; Germann, M. W. Effect of methylation on the side-chain pKa value of arginine. *Protein Sci* **2016**, *25*, 479-486.

(164) 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.

(165) Mallon, M.; Dutt, S.; Schrader, T.; Crowley, P. B. Protein Camouflage: Supramolecular Anion Recognition by Ubiquitin. *Chembiochem* **2016**, *17*, 774-783.

(166) 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.

(167) Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 host-guest blind prediction challenge: an overview. *J. Comput. Aided Mol. Des.* **2014**, *28*, 305-317.

(168) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D. Overview of the SAMPL6 host-guest binding affinity prediction challenge. *J. Comput. Aided Mol. Des.* **2018**, *32*, 937-963.

(169) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *J. Comput. Aided Mol. Des.* **2020**, *34*, 601-633.

(170) Amezcua, M.; El Khoury, L.; Mobley, D. L. SAMPL7 Host-Guest Challenge Overview: assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations. *J. Comput. Aided Mol. Des.* **2021**, *35*, 1-35.

(171) Amezcua, M.; Setiadi, J.; Ge, Y.; Mobley, D. L. An overview of the SAMPL8 host-guest binding challenge. *J. Comput. Aided Mol. Des.* **2022**, *36*, 707-734.

(172) Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M. Synthetic therapeutic peptides: science and market. *Drug Discov. Today* **2010**, *15*, 40-56.

(173) Houghten, R. A. General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids. *Proc Natl Acad Sci U S A* **1985**, *82*, 5131-5135.

(174) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* **1991**, *37*, 487-493.

(175) Monnee, M. C.; Brouwer, A. J.; Verbeek, L. M.; van Wageningen, A. M.; Liskamp, R. M. Bio-inspired synthetic receptor molecules towards mimicry of vancomycin. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1521-1525.

(176) Flint, A. J.; Davis, A. P. Vancomycin mimicry: towards new supramolecular antibiotics. *Org Biomol Chem* **2022**, *20*, 7694-7712.

(177) Hioki, H. [Synthesis of peptidocalix[4]arene libraries and their application to the development of chemical sensors for oligopeptides]. *Yakugaku Zasshi* **2005**, *125*, 263-270.

(178) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, *354*, 82-84.

(179) Ohashi, H.; Kanamori, T.; Osada, E.; Akbar, B. K.; Ueda, T. Peptide screening using PURE ribosome display. *Methods Mol Biol* **2012**, *805*, 251-259.

(180) Valencia, C. A.; Zou, J.; Liu, R. In vitro selection of proteins with desired characteristics using mRNA-display. *Methods* **2013**, *60*, 55-69.

(181) Eldridge, B.; Cooley, R. N.; Odegrip, R.; McGregor, D. P.; Fitzgerald, K. J.; Ullman, C. G. An in vitro selection strategy for conferring protease resistance to ligand binding peptides. *Protein Eng. Des. Sel.* **2009**, *22*, 691-698.

(182) Reiersen, H.; Lobersli, I.; Loset, G. A.; Hvattum, E.; Simonsen, B.; Stacy, J. E.; McGregor, D.; Fitzgerald, K.; Welschof, M.; Brekke, O. H.; Marvik, O. J. Covalent antibody display--an in vitro antibody-DNA library selection system. *Nucleic Acids Res.* **2005**, *33*, e10.

(183) Hudson, E. P.; Uhlen, M.; Rockberg, J. Multiplex epitope mapping using bacterial surface display reveals both linear and conformational epitopes. *Sci Rep* **2012**, *2*, 706.

(184) Boder, E. T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **1997**, *15*, 553-557.

(185) Sohrabi, C.; Foster, A.; Tavassoli, A. Methods for generating and screening libraries of genetically encoded cyclic peptides in drug discovery. *Nat Rev Chem* **2020**, *4*, 90-101.

(186) Sulakvelidze, A.; Alavidze, Z.; Morris, J. G., Jr. Bacteriophage therapy. *Antimicrob. Agents Chemother.* **2001**, *45*, 649-659.

(187) Wittebole, X.; De Roock, S.; Opal, S. M. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. *Virulence* **2014**, *5*, 226-235.

(188) Nobrega, F. L.; Costa, A. R.; Kluskens, L. D.; Azeredo, J. Revisiting phage therapy: new applications for old resources. *Trends Microbiol* **2015**, *23*, 185-191.

(189) Smith, G. P. Phage Display: Simple Evolution in a Petri Dish (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **2019**, *58*, 14428-14437.

(190) Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **1985**, *228*, 1315-1317.

(191) Hess, G. T.; Cragnolini, J. J.; Popp, M. W.; Allen, M. A.; Dougan, S. K.; Spooner, E.; Ploegh, H. L.; Belcher, A. M.; Guimaraes, C. P. M13 bacteriophage display framework that allows sortase-mediated modification of surface-accessible phage proteins. *Bioconjug Chem* **2012**, *23*, 1478-1487.

(192) Hoogenboom, H. R.; de Bruine, A. P.; Hufton, S. E.; Hoet, R. M.; Arends, J. W.; Roovers, R. C. Antibody phage display technology and its applications. *Immunotechnology* **1998**, *4*, 1-20.

(193) Heinis, C.; Winter, G. Encoded libraries of chemically modified peptides. *Curr Opin Chem Biol* **2015**, *26*, 89-98.

(194) Mohan, K.; Weiss, G. A. Chemically Modifying Viruses for Diverse Applications. *ACS Chem Biol* **2016**, *11*, 1167-1179.

(195) Ekanayake, A. I.; Sobze, L.; Kelich, P.; Youk, J.; Bennett, N. J.; Mukherjee, R.; Bhardwaj, A.; Wuest, F.; Vukovic, L.; Derda, R. Genetically Encoded Fragment-Based Discovery from Phage-Displayed Macrocyclic Libraries with Genetically Encoded Unnatural Pharmacophores. *J Am Chem Soc* **2021**, *143*, 5497-5507.

(196) Derda, R.; Ng, S. Genetically encoded fragment-based discovery. *Curr Opin Chem Biol* **2019**, *50*, 128-137.

(197) Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R. Discovery of light-responsive ligands through screening of a light-responsive genetically encoded library. *ACS Chem Biol* **2014**, *9*, 443-450.

(198) Ng, S.; Derda, R. Phage-displayed macrocyclic glycopeptide libraries. *Org Biomol Chem* **2016**, *14*, 5539-5545.

(199) Tian, F.; Tsao, M. L.; Schultz, P. G. A phage display system with unnatural amino acids. *J Am Chem Soc* **2004**, *126*, 15962-15963.

(200) Milosevich, N.; Vizely, K. H.; Nikolic, R. P.; McCallum, J. F.; Treanor, L. M.; Khan, S. S.; Derda, R.; Hof, F. Installation of cysteine-derived methyllysine mimics on phage-dis-played peptide libraries: optimization of reaction conditions for conversion and phage viability. **2022**.

(201) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat Chem Biol* **2009**, *5*, 502-507.

(202) Kalhor-Monfared, S.; Jafari, M. R.; Patterson, J. T.; Kitov, P. I.; Dwyer, J. J.; Nuss, J. M.; Derda, R. Correction: Rapid biocompatible macrocyclization of peptides with decafluoro-diphenylsulfone. *Chem Sci* **2017**, *8*, 807.

(203) Chen, S.; Bertoldo, D.; Angelini, A.; Pojer, F.; Heinis, C. Peptide ligands stabilized by small molecules. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 1602-1606.

(204) Deyle, K.; Kong, X. D.; Heinis, C. Phage Selection of Cyclic Peptides for Application in Research and Drug Development. *Acc. Chem. Res.* **2017**, *50*, 1866-1874.

(205) Boger, R. H.; Bode-Boger, S. M.; Szuba, A.; Tsao, P. S.; Chan, J. R.; Tangphao, O.; Blaschke, T. F.; Cooke, J. P. Asymmetric dimethylarginine (ADMA): a novel risk factor for endothelial dysfunction: its role in hypercholesterolemia. *Circulation* **1998**, *98*, 1842-1847.

(206) Cardounel, A. J.; Cui, H.; Samouilov, A.; Johnson, W.; Kearns, P.; Tsai, A. L.; Berka, V.; Zweier, J. L. Evidence for the pathophysiological role of endogenous methylarginines in regulation of endothelial NO production and vascular function. *J Biol Chem* **2007**, *282*, 879-887.

(207) Kampman, K. M. The treatment of cocaine use disorder. *Sci Adv* **2019**, *5*, eaax1532.

(208) Florea, M.; Kudithipudi, S.; Rei, A.; Gonzalez-Alvarez, M. J.; Jeltsch, A.; Nau, W. M. A fluorescence-based supramolecular tandem assay for monitoring lysine methyltransferase activity in homogeneous solution. *Chemistry* **2012**, *18*, 3521-3528.

(209) Bakirci, H.; Nau, W. M. Fluorescence regeneration as a signaling principle for choline and carnitine binding: A refined supramolecular sensor system based on a fluorescent azoalkane. *Adv. Funct. Mater.* **2006**, *16*, 237-242.

(210) Warmerdam, Z.; Kamba, B. E.; Le, M. H.; Schrader, T.; Isaacs, L.; Bayer, P.; Hof, F. Binding Methylarginines and Methyllysines as Free Amino Acids: A Comparative Study of Multiple Host Classes**. *Chembiochem* **2022**, *23*.

(211) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 6974-6998.

(212) Kalhor-Monfared, S.; Jafari, M. R.; Patterson, J. T.; Kitov, P. I.; Dwyer, J. J.; Nuss, J. M.; Derda, R. Rapid biocompatible macrocyclization of peptides with decafluoro-diphenylsulfone. *Chem Sci* **2016**, *7*, 3785-3790.

(213) Matochko, W. L.; Derda, R. Next-generation sequencing of phage-displayed peptide libraries. *Methods Mol Biol* **2015**, *1248*, 249-266.

(214) Pande, J.; Szewczyk, M. M.; Grover, A. K. Phage display: concept, innovations, applications and future. *Biotechnol. Adv.* **2010**, *28*, 849-858.

(215) Laughrey, Z. R.; Gibb, C. L.; Senechal, T.; Gibb, B. C. Guest binding and orientation within open nanoscale hosts. *Chemistry* **2003**, *9*, 130-139.

(216) Bryant, R. G. The Nmr Time Scale. *J. Chem. Educ.* **1983**, *60*, 933-935.

(217) Gazit, E. Metabolite amyloids: a new paradigm for inborn error of metabolism disorders. *J Inherit Metab Dis* **2016**, *39*, 483-488.