Experimental Characterization of the Association of Nine Novel Cyclodextrin Derivatives with Two Guest Compounds

K. Kellett, D. R. Slochower, M. Schauperl, B. M. Duggan, and M. K. Gilson*

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0751, USA

* mgilson@health.ucsd.edu

Keywords: SAMPL, Host-guest, cyclodextrin, ITC, NMR

1 ABSTRACT

We investigate the binding of native β-cyclodextrin (β-CD) and eight novel β-CD derivatives with two different guest compounds, using isothermal calorimetry (ITC) and 2D NOESY NMR. In all cases, the stoichiometry is 1:1 and binding is exothermic. Overall, modifications at the 3' position of β-CD, which is at the secondary face, weaken binding by several kJ/mol relative to native β-CD, while modifications at the 6' position (primary face) maintain or somewhat reduce the binding affinity. The variations in binding enthalpy are larger than the variations in binding free energy, so entropy-enthalpy compensation is observed. Characterization of the bound conformations with NOESY NMR shows that the polar groups of the guests may be situated at either face, depending on the host molecule, and, in some cases, both orientations are populated. The present results were used in the SAMPL7 blinded prediction challenge whose results are detailed in the same special issue of JCAMD.

2 INTRODUCTION

An early requirement of many drug discovery projects is the design or discovery of a small organic molecule that binds a disease-related protein with high affinity; i.e., with strongly negative free energy. Computational methods of estimating protein-small molecule binding free energies may be used to help at this project stage. However, current computational methods are not yet accurate enough to obviate the need for synthesis and testing of hundreds or thousands of small molecules to discover a potent drug candidate. Researchers in academia and industry are therefore working to improve accuracy and thus reduce the time and costs of drug discovery.

As new computational methods and parameters are developed, they must be tested against experimental data. An obvious and valuable approach is to test their ability to replicate measured protein-ligand binding affinities. However, achieving adequately converged results can be time-consuming and in some cases intractable, because the conformational fluctuations of proteins can occur on microsecond and longer time scales that are difficult to access with current computational technologies.

As a consequence, tests of computational methods often make use of low molecular weight receptor molecules, called hosts, in place of proteins[1]–[5]. A typical host, beta-cyclodextrin (β-CD; Figure 1), is more than an order of magnitude less massive than even a small globular protein comprising ~100 residues, and many proteins of pharmaceutical interest are 5-10 times larger. Furthermore, hosts tend to be comparatively rigid, because they are usually both macrocyclic and polycyclic and therefore have few freely rotatable bonds per unit mass. Nonetheless,

the same basic forces and physical chemistry are at work in host-guest binding as in protein-small molecule binding. For example, both proteins and hosts can form hydrogen bonds and hydrophobic interactions with the molecules they bind, and the same statistical thermodynamics framework applies to them[6].

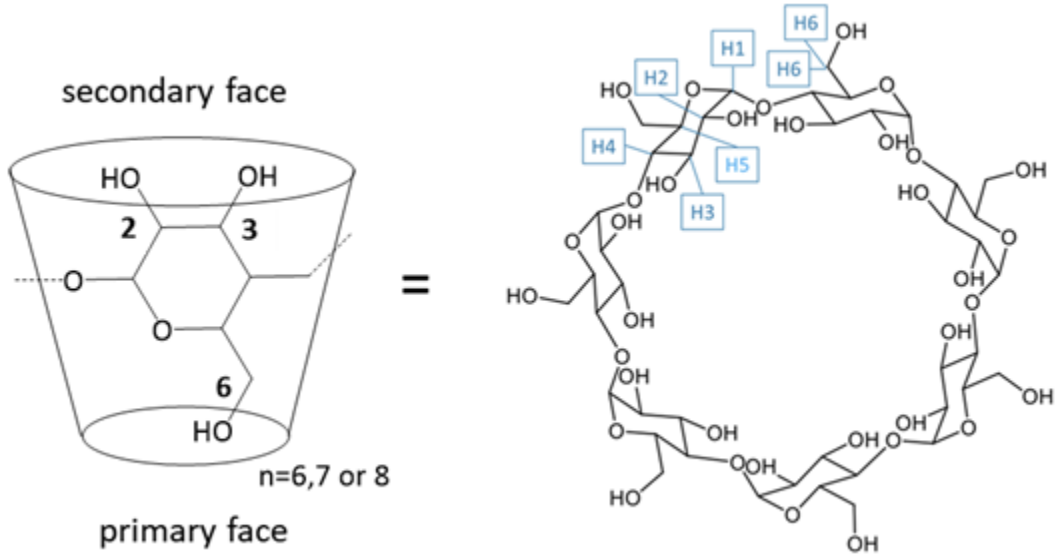

Figure 1. Representations of β-cyclodextrin (β-CD), highlighting the secondary and primary faces (left) and atom numbering relevant for the present NMR studies.

The cyclodextrins represent the least expensive and most abundant class of host molecules, as they can be synthesized from starch using readily available enzymes[7]. The most common cyclodextrins, α-cyclodextrin, β-cyclodextrin (β-CD), and γ-cyclodextrin, are macrocycles of 5, 6 or 7 glucose units respectively (Figure 1). They possess a hydrophobic cavity which is open at two ends. The rim of the larger opening is called the secondary face because of its 14 secondary hydroxyl groups, and the rim of the smaller opening is called the primary face because of its 7 primary hydroxyl groups. The polar hydroxyl groups contribute to the water solubility of these nonionic molecules, and aqueous binding affinities have been measured for native cyclodextrins with a variety of guest molecules[8]–[12]. These data are useful for testing computational models of binding but are sharply limited in the types of noncovalent interactions they probe, because although one may vary the guest compounds used, the host is uniformly a simple glucose polymer. Greater diversity in binding interactions may be accessed by using derivatives of the cyclodextrins. Binding data have also been generated for some cyclodextrin derivatives[12]–[14], but the range of noncovalent interactions these probe falls short of that routinely explored by medicinal chemists. Other families of hosts, such as those based on the glycouril unit[15], the deep cavity cavitands[16], and molecular clips and tweezers[17], [18], offer valuable additional diversity but still fail to explore a full variety of drug-related interactions.

Synthetic modification of commercially available cyclodextrins offers a straightforward way of increasing the diversity of host-guest interactions. However, modifications of cyclodextrins have focused largely on persubstitution with a small set of substituents; random substitution, which yields products not suited for rigorous comparison with calculations; and mono-substitution of only the more synthetically accessible 6’ hydroxyls of the smaller primary face (Figure 1). Mono-substitution of the less reactive 2’ and 3’ hydroxyls of the larger secondary face has been regarded as relatively challenging. We recently addressed this with synthetic approaches to modify not only the 6’ but also the 3’ position with a variety of substituents, and to simplify purification of the desired products. We hope ultimately to generate a dense matrix of thermodynamically characterized host-guest interactions (Figure 2) that will be of ongoing value to researchers seeking to test or parameterize computational models of binding and thus to advance the accuracy of computer-aided drug design.

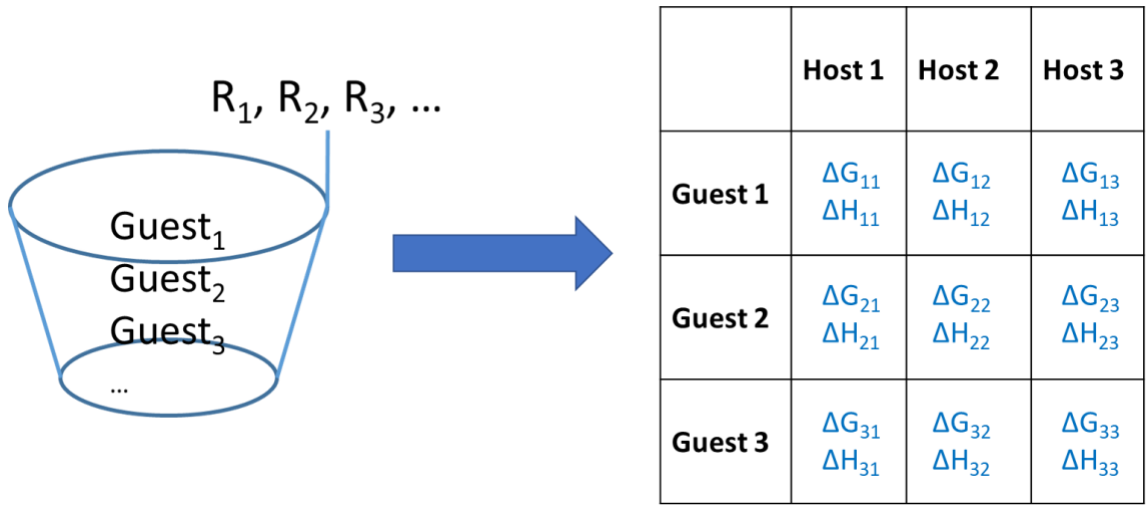

Figure 2. Concept of a matrix of host-guest binding thermodynamics designed for use in parameterization and testing simulation methodologies.

As a next step in this direction, we report here the binding thermodynamics, measured with isothermal titration calorimetry, of two guest compounds (Figure 3) with β-CD and eight diverse monosubstituted β-CD derivatives (Figure). The β-CD substituents include a steroid nucleus; cationic, neutral and anionic groups; an aromatic group; and a cysteine-containing peptidic chain. The conformational preferences of selected complexes were further characterized by 2D NOESY NMR experiments. These cases were adopted for use in the SAMPL7 blinded prediction challenge detailed in other papers of this special issue.

3 EXPERIMENTAL

3.1 HOSTS, GUESTS, AND OTHER MATERIALS

We studied the association of two guest molecules (Figure 3), trans-4-methylcyclohexanol (henceforth methylcyclohexanol) and R-rimantadine hydrochloride (henceforth rimantidine), with nine different hosts. Rimantadine was purchased from Enamine (Monmouth, NJ) and methylcyclohexanol was purchased from TCI Chemicals (Portland, OR). Both guests were used as received, without further purification.

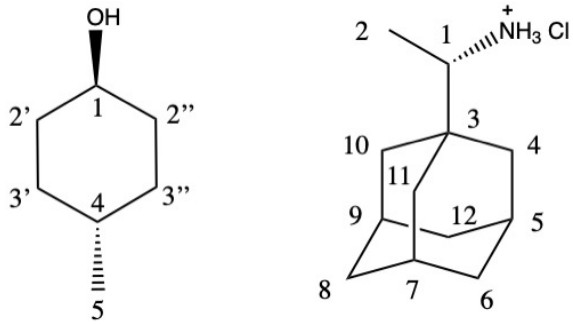

Figure 3. The two guest molecules used in this study, with ionization states expected at the experimental pH of 6.8, and atom numbering relevant for the NMR studies. Left: trans-methylcyclohexanol. Right: R-rimantidine.

The nine host molecules are listed in Table 1 and illustrated in Figure 4; machine-readable representations are provided in the SI. Unmodified β-CD was purchased from Sigma-Aldrich Company (St. Louis, MO) and all host molecules other than β-CD were synthesized as previously detailed [19]. Their purities, assessed with LC-MS and

NMR, were found to be ≥ 95%. The buffer components deuterium oxide, mono-sodium phosphate and di-sodium phosphate were purchased from Sigma-Aldrich Company (St. Louis, MO) and also were used as purchased.

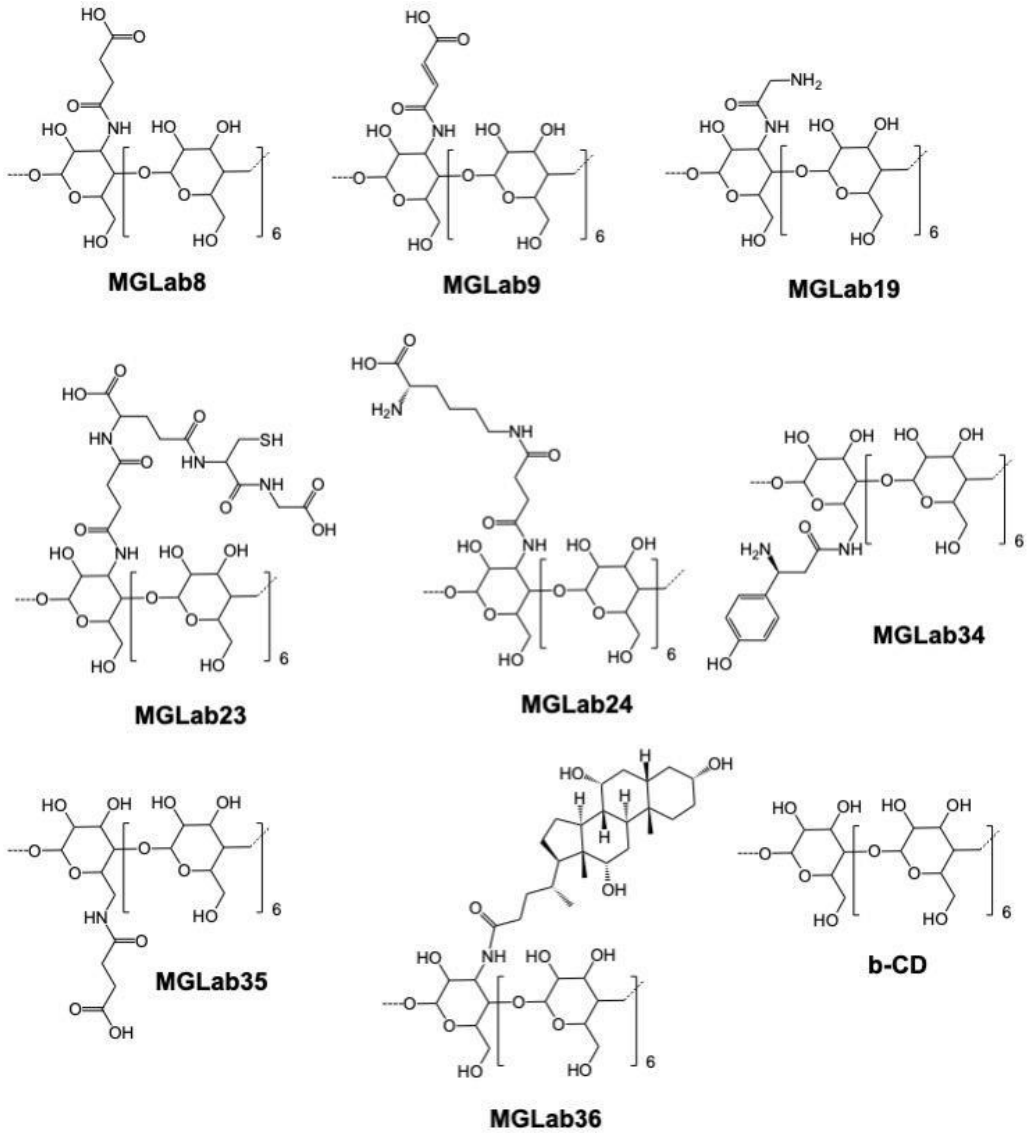

Figure 4. Structures of the nine β-CD derivatives studied here. Two substituents are at the primary face (position 6) and the rest are at the secondary face (position 3)(see Figure 1).

Table 1. The nine host molecules used in this study, with the charge states expected for the experimental pH of 6.8.

ID	Compound	Charges
β-CD	β-cyclodextrin	None
MGLab8	Mono-3-(3-carboxypropionamido)-β-cyclodextrin	-1
MGLab9	Mono-3-(4-oxobut-2-enoic acid)-α-cyclodextrin	-1
MGLab19	Mono-3-glycine-β-cyclodextrin	+1
MGLab23	Mono-3-(3-carboxypropionamido-glutathione)-β-cyclodextrin	-1, -1
MGLab24	Mono-3-(3-carboxypropionamido-lysine)-β-cyclodextrin	+1, -1
MGLab34	Mono-6-β-tyrosine-β-cyclodextrin	+1
MGLab35	Mono-6-(carboxypropionamido)-β-cyclodextrin	-1
MGLab36	Mono-3-cholic acid-β-cyclodextrin	None

3.2 ITC METHODS AND ANALYSIS

All ITC results were collected in duplicate, with fresh solutions for each run, on a Microcal ITC200. Solutions were made up in 25 mM phosphate buffer (sodium counterion, no potassium), at pH 6.8. Experiments were run at 27°C and used 16 injections of 2.42 µL each. In all cases, the host was in the cell and the guest in the syringe. The concentrations were adjusted for each complex to provide values of C in the preferred range of 5 to 12 [20]. Here $C = K_d/C_{cell}$, where K_d is the host-guest dissociation constant and C_{cell} is the concentration of the cell reactant, here the host. Compound purities were taken into account when calculating the concentrations. The commercially sourced rimantadine and methylcyclohexanol have reported purities of 95% and 98%, respectively, and the host molecules all have purities of ≥ 95%. β-CD has a reported 8 hydration water molecules, and we assumed the same for the derivatives. The stoichiometry, N, was treated as a floating parameter during curve fitting. Experimental uncertainties were computed by Monte Carlo resampling, accounting for estimated concentration and heat uncertainties[21]. The ITC data were analyzed with code written within the Gilson group, which is publicly available at https://github.com/GilsonLabUCSD/itc_fit (commit ID e9eabb0). The reported values of binding free energy, binding enthalpy, binding entropy, and their respective uncertainties, are averages of the values obtained for the duplicates. The free energy, enthalpy and entropy results from the duplicates consistently fall within the reported uncertainty ranges. The only exception is MGLab 34 with rimantidine, for which the duplicates yielded less consistent results. In this case, the reported uncertainty is half the deviation between the two values.

3.3 NMR

NMR spectra of selected host-guest complexes were collected at 298 K on a 600 MHz Bruker Avance III spectrometer fitted with a 5 mm triple resonance cryoprobe with z-axis gradients. All NMR studies were run in phosphate buffer (0.1 M, pH 6.8), made up with 90% H_2O and 10% D_2O. ^{1}H-NMR was collected with presaturation of the water peak, with 16 scans. 2D NOESY spectra were run with water suppression using excitation sculpting with gradients, States-TPPI acquisition mode, and a 500 ms mixing time.

4 RESULTS AND DISCUSSION

4.1 BINDING THERMODYNAMICS

Isothermal calorimetry (ITC) was used to determine the binding free energy, enthalpy, and entropy of binding for the 18 binding reactions defined by our nine hosts and two guests. The same experimental conditions were used throughout, except that the concentrations of the reactants were adjusted individually to reach C values in a suitable range[20]. A low ionic strength buffer (25 mM) was used to reduce the complexity of the overall systems for computational chemists, particularly as high ionic concentrations can significantly alter binding affinities, especially for charged systems[22]. The compiled affinity data for all host-guest complexes are shown in Table 2. All of the fitted stoichiometries (N) are in the range 0.70 – 1.18. Further experimental details (e.g., sample concentrations and enthalpograms) are available in the Supporting Information.

Table 2. Experimentally measured binding thermodynamics for host-guest complexes, in kJ/mol (free energy and enthalpy) and J/mol/K (entropy). Face: site of the host's monosubstitution; i.e., primary (1o) or secondary (2o). Expected host charges are as listed, and the fitted values of the stoichiometry (N) range from 0.70-1.18.

			Methyl cyclohexanol			**Rimantadine**		
Host	**Face**	**Charges**	**ΔG**	**ΔH**	**ΔS**	**ΔG**	**ΔH**	**ΔS**
β-CD	n/a	None	**-19.0 ± 0.1**	-10.9 ± 0.4	27.0 ± 1.5	**-26.2 ± 0.1**	-43.3 ± 1.8	-56.8 ± 5.8
MGLab8	2o	-1	**-13.7 ± 0.1**	-9.2 ± 1.4	15.1 ± 4.7	**-16.9 ± 0.1**	-30.5 ± 1.5	-45.2 ± 4.9
MGLab9	2o	-1	**-13.5 ± 0.1**	-10.5 ± 1.8	9.9 ± 6.0	**-16.3 ± 0.1**	-38.1 ± 1.7	-72.7 ± 5.8
MGLab19	2o	+1	**-13.2 ± 0.1**	-9.1 ± 0.7	13.5 ± 2.5	**-14.4 ± 0.1**	-47.6 ± 3.4	-110 ± 11
MGLab23	2o	-1, -1	**-13.6 ± 0.1**	-9.9 ± 1.7	12.6 ± 5.6	**-18.1 ± 0.1**	-32.1 ± 1.4	-46.5 ± 1.6
MGLab24	2o	+1, -1	**-14.0 ± 0.1**	-7.8 ± 0.8	20.9 ± 2.8	**-17.6 ± 0.1**	-35.5 ± 1.5	-59.5 ± 5.1
MGLab34	1o	+1	**-19.1 ± 0.1**	-17.9 ± 0.7	4.1 ± 2.4	**-23.1 ± 1.5**	-37.1 ± 6.6	-46.6 ± 17
MGLab35	1o	-1	**-15.8 ± 0.1**	-15.1 ± 0.7	2.3 ± 2.5	**-25.5 ± 0.1**	-31.6 ± 1.3	-20.2 ± 4.3
MGLab36	2o	None	**-14.0 ± 0.1**	-12.2 ± 0.9	5.6 ± 3.4	**-14.8 ± 0.1**	-45.5 ± 3.7	-102 ± 12

4.1.1 Trans-4-methylcyclohexanol

Methylcyclohexanol, is a neutral molecule whose hydrophobic aliphatic ring can form hydrophobic interactions with the nonpolar interior of the cyclodextrins, and whose hydroxyl group could form hydrogen bonds with the hydroxyls of the primary or secondary face of the hosts or with the substituents added to the cyclodextrin derivatives. However, its affinity is likely limited by the fact that its hydrophobic group is too small to fit snugly into the binding cavity and interact favorably with the entire inner wall of the host at one time. Rekharsky et al. reported binding of

this compound to β-CD with ΔG = -19.00 ± 0.07[10] kJ/mol, and our result is in excellent agreement with this prior measurement.

None of the eight cyclodextrin derivatives bind methyl cyclohexanol more tightly than native β-CD, despite the added interaction opportunities afforded by the side chains. However, MGLab 34, with a tyrosine side chain at the primary face, matches the affinity of native β-CD. Interestingly, the other host that is substituted at the primary face, MGLab 35, has the second strongest affinity, ΔG = -15.80 ± 0.1. The other hosts, whose substituents are at the secondary face, all have lower affinities in a rather narrow range, -13.2 to -14.0 kJ/mol. These results suggest that secondary face substitution hinders binding, rather than increasing it likely, by sterically hindering binding at the secondary face.

All of these complexes exhibit exothermic binding, with hosts β-CD and MGLab 8 – MGLab 24 showing ΔH values in the rather narrow range (-9.1 to -10.9 kJ/mol). All of the binding entropies also are positive and hence contribute further to favorable binding, though hosts MGLab 34-MGLab 36 exhibit relatively small entropy increases. A scatter plot of binding entropy vs enthalpy shows entropy-enthalpy compensation (Figure 5, blue points), with the higher-affinity hosts lying to the left (lower binding free energy) of the other compounds. The structural basis of the observed thermodynamics cannot be determined from these purely thermodynamic data, and we are not aware of prior studies reporting whether the guest's hydroxyl resides at the primary or at the secondary face.

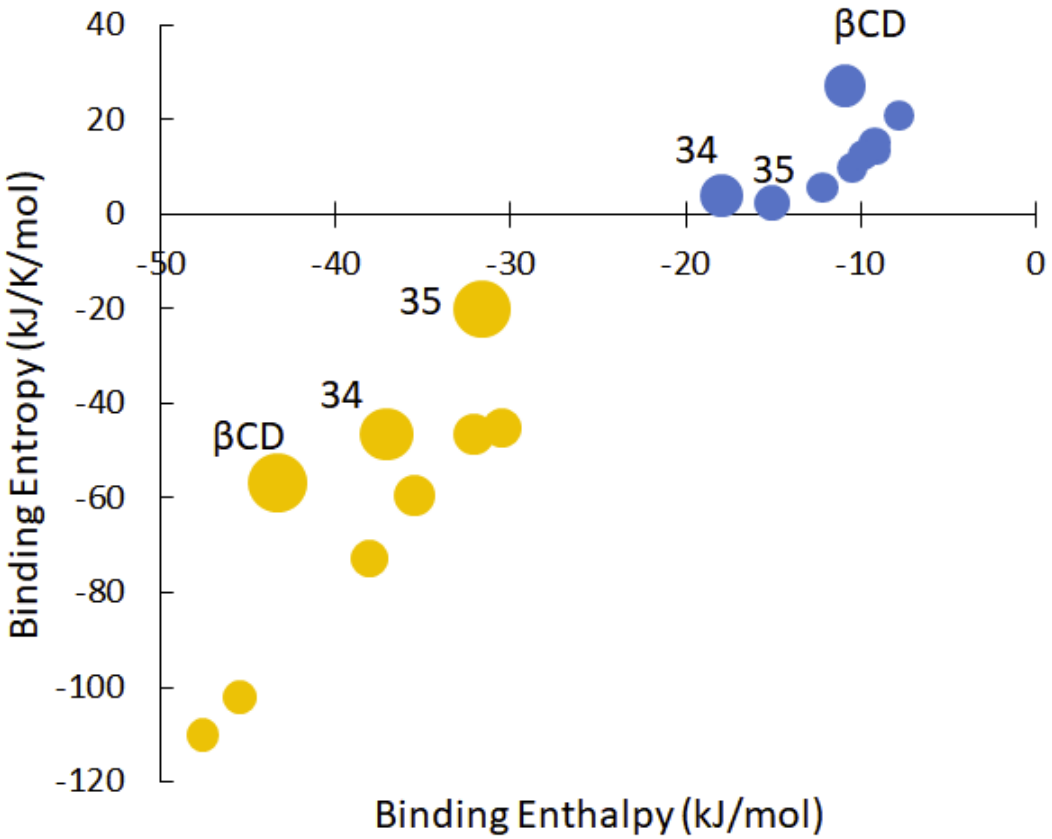

Figure 5. Entropy-enthalpy scatter plot of data from Table 2. Blue: methylcyclohexanol. Orange: rimantidine. Radius of each data point is proportional to binding free energy.

4.1.2 R-Rimantadine

Rimantadine, a chiral, cationic molecule, has a large hydrophobic adamantyl moiety, which fills the hydrophobic β-CD cavity more completely than does methylcyclohexanol. Accordingly, the binding free energies of rimantidine across β-CD and the β-CD derivatives are 1 to 10 kJ/mol more favorable than those of methylcyclohexanol for the same host molecules. Note that our results for this guest with native β-CD agree well with prior measurements reporting a binding free energy of -26.4 ± 0.08 kJ/mol[23]. Perhaps unexpectedly, this cationic guest does not bind more strongly to CD derivatives with anionic substituents, such as MGLab 8, 9, and 23, than it does to native β-CD. However, the cationic MGLab 19 host does show the weakest binding to rimantidine. Possibly the clearest trend in this series is that binding of rimantidine to hosts with substituents on the secondary face is about 10 kJ/mol weaker than binding to hosts with derivatives on the primary face. This is similar to, but stronger than, the corresponding trend for guest methylcyclohexanol (Section 4.1.1).

Rimantidine binds exothermically to all of the hosts tested here, and its binding enthalpies are 2-5 times greater (more negative) than those of methylcyclohexanol for the corresponding hosts. However, rimantidine binding is disfavored by negative entropy changes in all cases, in striking contrast to methylcyclohexanol, for which the entropy change uniformly favors binding. The combination of rimantidine's more favorable binding enthalpies and more unfavorable binding entropies, relative to methylcyclohexane, suggests further entropy-enthalpy compensation, and this is confirmed by the scatter plot in Figure 5 (orange points).

4.2 DETERMINATION OF BINDING POSES USING 2D NOESY NMR

Experimental NOESY studies of the bound conformations of these host-guest complexes can identify atoms of the host and guest that reside near each other in the bound complex and thus provide information about the bound conformations. This information offers a further check of computational methods and could help guide the design of additional host-guest systems with desired properties. Here, the overall orientation of the guest is considered primary if its polar functional group (hydroxyl for methylcyclohexanol, ammonium for rimantidine) resides at the primary face, and secondary if it resides at the secondary face. This determination is based chiefly on NOEs between hydrogens at each end of the guest (Figure 3) with hydrogen 3 of the host (Figure 1), which projects into the cavity from the secondary rim, and with hydrogen 6, which is at the primary rim. Interactions with hydrogen 5 of the host also help to confirm binding, since it is within the cavity, but are less informative about orientation because of this hydrogen's central location. In all cases studied here, shifting of internal protons confirmed binding into the host's cavity. Some complexes showed NOEs interactions associated with both orientations, which implies that both are populated. Also, for a few of the β-CD derivatives, NOEs were observed between the added substituent and atoms of the guest. Finally, it is worth noting that absence of an NOE between two atoms does not necessarily mean that the atoms do not interact. The NMR spectra for the present host-guest complexes are provided in the Supporting Information.

4.2.1 Trans-4-methylcyclohexanol

Table 3 summarizes the key NOEs observed between all nine hosts and methylcyclohexanol, noting as well the implications of these data for the bound conformations. A mixture of primary and secondary binding (Figure 6) is found with four of the nine hosts studied (Table 3). We conjecture that the interconversion between these two states occurs on unbinding and rebinding, because rotation within the binding cavity would lead to steric obstructions and would require at least partial desolvation of the guest's polar hydroxyl group. It is not possible to determine the lifetime of the bound states from the NOESY data as the conformations could exchange rapidly while recording the spectra, or longer lived states may be present as two separate populations. For the remaining five hosts, three generate primary binding and two generate secondary binding. We did not notice any correlations between the binding orientation and the location or nature of the novel functionalities introduced to either the primary or secondary face. In combination with the reduced binding affinity seen from ITC compared to β-CD, this would suggest that there are no strong interactions between the guests and the added β-CD functionalities, as they are not contributing to increased binding affinities. Methylcyclohexanol's small size means that, when bound, it is not likely

to protrude much from the cavity (Figure 6). This may reduce its ability to form stabilizing interactions with the CD substitutions. A larger guest molecule might offer more opportunities to probe such interactions.

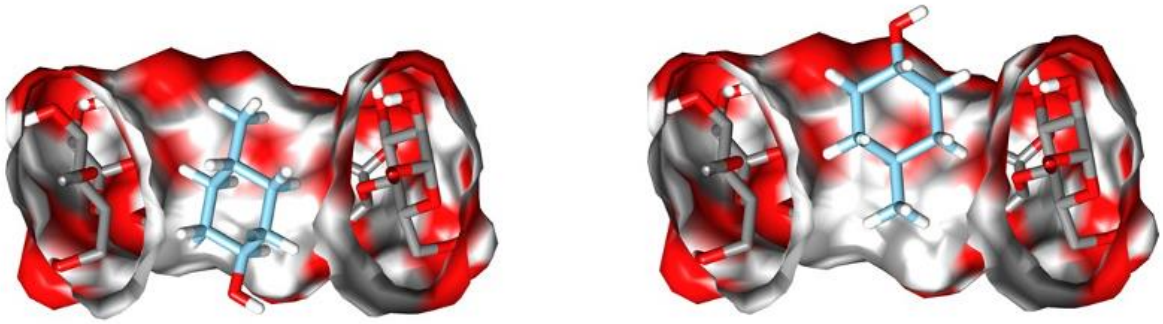

Figure 6. Two possible binding orientations of methylcyclohexanol in β-cyclodextrin. Secondary face is at the top. Left: secondary orientation. Right: primary orientation

Table 3. Binding orientations of methylcyclohexanol complexes deduced from 2D NOESY data. Substituent face and charges from prior tables. NOE assignments are given as Hn-Gm here n and m indicate the proton numbers, as shown in prior figures.

Host	Substituent Face	Host Charges	NOEs observed (δ ppm)	Binding Orientation
β-CD	n/a	None	H6-G1 (3.423-1.73) H6-G2'2"/3'3" (3.438-1.09/0.86)	Primary and Secondary
MGLab8	2º	-1	H6-G4 (0.561-3.36) H6-G3'3" (0.83-3.39) H3/5-G1 (1.726-3.85/3.46)	Secondary
MGLab9	2º	-1	H6-G3’3’’ (3.51-0.875) H5-G4 (3.78-0.90) H6-G2’2’’ (3.45-1.155) H3-G1 (3.98-1.804)	Primary and Secondary
MGLab19	2º	1	H6-G3'3" (3.514-0.894) H6-G2'2" (3.522-1.146) H3-G1 (1.834-3.98)	Primary
MGLab23	2º	-1, -1	H3-G1 (1.732-3.85) H6-G4 (1.546-3.47)	Primary and Secondary
MGLab24	2º	+1, -1	HCH2-G4 (2.862-1.56) H6-G1 (3.422-1.76) H5-G4 (1.562-3.49)	Primary
MGLab34	1º	1	H5 – G2’2’’ (3.42-1.08) H6 – G3’3’’ (3.426-0.82)	Secondary
MGLab35	1º	-1	H6-G2'2" (3.396-1.13) H3-G4 (3.762-1.52) H6-G3'3" (3.45-0.815)	Primary
MGLab36	2º	None	H3-G5 (3.74-0.75) H6-G1 (3.424-1.78)	Primary and Secondary

4.2.2 R-Rimantadine

Like methylcylohexanol, rimantadine can bind the cyclodextrins in either a primary or secondary orientation (Figure 8), where secondary binding means that its amine functionality is at the larger secondary opening of cyclodextrin, and primary binding means the amine is at the smaller primary opening. The near-symmetry of rimantadine can make it difficult to distinguish among its proton peaks and thus to determine its binding orientation. However, on binding to β-CD, the CH (H3, H5, H7, H9) and CH_2 (C4, C6, C8, C10) peaks shift to give two separate quartets, with the CH groups consistently more deshielded[23]. This small difference makes it possible to determine binding through NOESY interactions with the host’s protons.

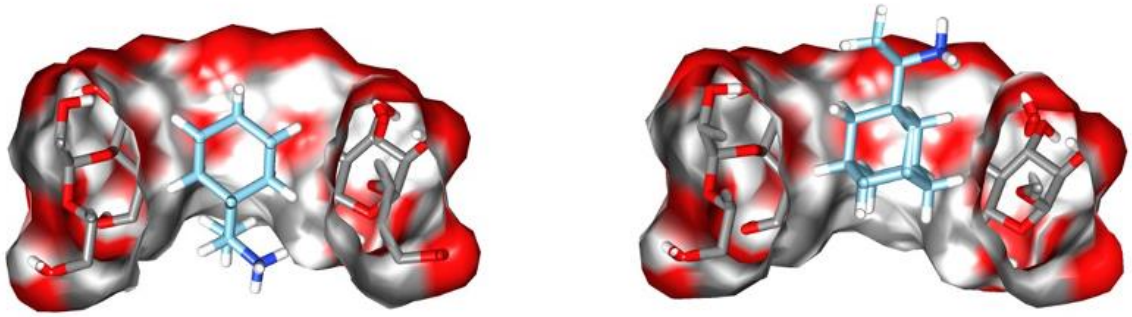

Figure 8. Two possible binding orientations of rimantadine in β-cyclodextrin. Secondary face is at the top. Left: secondary orientation. Right: primary orientation.

Table 4 summarizes the key NOEs observed between the present hosts and rimantidine, noting as well the implications of these data for the bound conformations. Cases of primary and secondary binding are observed across the different derivatives, and, unlike methylcyclohexanol, there are no cases where NOEs corresponding to both orientations were observed. For MGLab 9, no intermolecular NOEs were observed, and for MGLab 35, the large substituent led to overlapping proton peaks with rimantadine, making it hard to determine the binding orientation. There is no clear correlation between the orientational preferences of rimantidine and those of methylcylohexanol across these hosts.

It might be expected that the orientation of rimantidine would be influenced by electrostatic interactions between its cationic ammonium group and the charges of the various cyclodextrin substituents. Thus, a substituent at the secondary face with a negative or positive charge might lead to secondary or primary orientation binding, respectively, with analogous logic holding for charges at the primary face. Although MGLab8 fits this pattern, any attractive electrostatic interactions that may have led rimantindine to adopt the secondary orientation in this host are not obviously manifest in the binding affinity (Table 2), which is not especially strong. Unfortunately, the other negatively charged hosts, MGLab9 and 35, gave inconclusive NOESY results. MGLab19, with a positive charge at the secondary face and primary binding orientation, also seems to fit the expected pattern, but MGLab34, does not, as it has a positive charge at the primary face, but shows a primary binding orientation.

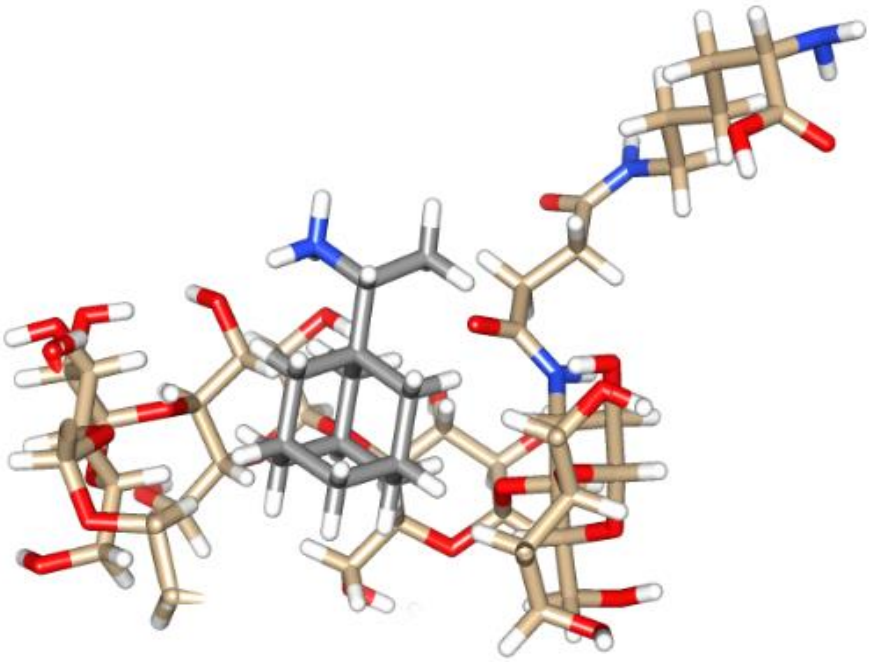

Figure 8. Schematic bound conformation of rimantadine in MGLab24, with the guest CH3 directed toward the succinic acid CH2 group, as indicated by the NOE data. A glucose unit has been hidden from view for clarity.

For the host molecules that have succinic acid linkers, MGLab8, MGLab 23 and MGLab 24, NOEs are observed between the CH_3 (H2) on the guest and the succinic acid side chain CH_2 groups for those complexes which show secondary face binding. This means that the rimantidine's amino group is directed away from the amide linker and thus that no hydrogen bond is formed between this amino group and the carbonyl group of the amine linker (Figure 8).

Table 4. Binding orientations of rimantidine complexes deduced from 2D NOESY data. One complex was inconclusive due to lack of NOEs or overlapping peaks. Substituent face and charges from prior tables. NOE assignments are given as Hn-Gm here n and m indicate the proton numbers, as shown in prior figures

Host	Substituent Face	Host Charges	NOEs observed (δ ppm)	Binding orientation
β-CD	n/a	None	H3-GNH3 (2.04-3.79) H3-G2 (1.17-3.76) H5-G3/7 (3.63-1.71)	Secondary
MGLab8	2°	-1	HCH_2-G2 (2.35-8.00) H3-G1 (2.04-1.633)	Secondary
MGLab9	2°	-1	No cross peaks observed	Inconclusive
MGLab19	2°	1	H6-G8/6 (3.70-1.46) H6-G7 (3.69-1.68) H5-G8/6 (3.78-1.51)	Primary
MGLab23	2°	-1, -1	H6-G7 (3.69 -1.65)	Primary
MGLab24	2°	+1, -1	H3-G2 (2.88-1.14)	Secondary

			HCH_2-G2 (2.35-8.00)	
MGLab34	1°	1	H6-G9/7/5 (3.728-1.635) H5-G8/6 (3.7716-1.525)	Primary
MGLab35	1°	-1	No usable NOEs	Inconclusive
MGLab36	2°	None	H3-G9/7/5 (3.93-1.56) H3-G2 (3.79-1.88)	Secondary

5 CONCLUSIONS

This article provides a new thermodynamic dataset spanning 18 host-guest binding systems and including eight novel β-CD derivatives. These data have already been used in the SAMPL7 blinded prediction challenge (see other articles in this issue) and should continue to be useful in future work employing host-guest binding data to test or adjust computational methods of estimating binding thermodynamics. Because we used ITC to characterize binding, the dataset includes not only free energies but also binding enthalpies (and hence binding entropies), and these, too, can be used as references for computational work [2], [24], [25]. The present study furthermore provides information on the bound orientations of most of the 18 cases, through the use of 2D NMR. This allows users to flag cases where computational methods may be giving the right thermodynamics for the wrong reason; i.e., with an incorrect bound pose.

Although adding substituents to either face of β-CD, as done here, creates the possibility of new, attractive interactions that might in principle lead to stronger binding, we found that adding these substituents either reduced or at best maintained the affinities, relative to β-CD. We speculate that the reductions in affinity stem, at least in part, from partial blockage of binding by the substituents, without compensating stabilizing interactions. Guests with substituents that project further from the binding cavity may be better suited to forming stabilizing interactions with these β-CD substituents.

6 DECLARATIONS

6.1 FUNDING

MKG acknowledges funding from National Institute of General Medical Sciences (GM61300). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the funders.

6.2 CONFLICTS OF INTEREST/COMPETING INTERESTS

MKG has an equity interest in and is a cofounder and scientific advisor of VeraChem LLC.

6.3 AVAILABILITY OF DATA AND MATERIAL

All data and raw data files for ITC provided in supplementary information.

6.4 CODE AVAILABILITY

https://github.com/GilsonLabUCSD/itc_fit (commit ID e9eabb0).

7 SUPPORTING INFORMATION

- SAMPL7_Kellett_Gilson_SMILES.xlsx: SMILES strings for host compounds.
- SAMPL7_Kellett_Gilson_SAMPL_ITC_complexes.xlsx: Results of analysis of ITC data for duplicate runs for each host-guest system.
- SAMPL7_Kellett_Gilson_ITC_raw_data.zip: compressed file with numerical data from the ITC runs and an explanatory README file.
- SAMPL7_Kellett_Gilson_enthalpograms_SAMPL.pdf: images of representative ITC enthalpograms.
- See also https://github.com/GilsonLabUCSD/itc_fit for the software used to fit the ITC data
- SAMPL7_Kellett_Gilson_NMR-R-Rimantidine.pdf: 2D NMR spectra of host-rimantidine complexes
- SAMPL7_Kellett_Gilson_NMR-trans-methyl-cyclohexanol.pdf: 2D NMR spectra of host-methylcyclohexanol complexes

8 REFERENCES

[1] C. Oostenbrink, "Efficient free energy calculations on small molecule host-guest systems—A combined linear interaction energy/one-step perturbation approach," *J. Comput. Chem.*, vol. 30, no. 2, pp. 212–221, 2009.

[2] A. T. Fenley, N. M. Henriksen, H. S. Muddana, and M. K. Gilson, "Bridging Calorimetry and Simulation through Precise Calculations of Cucurbituril–Guest Binding Enthalpies," *J. Chem. Theory Comput.*, vol. 10, no. 9, pp. 4069–4078, 2014.

[3] H. S. Muddana, A. T. Fenley, D. L. Mobley, and M. K. Gilson, "The SAMPL4 host-guest blind prediction challenge: An overview," *J. Comput. Aided. Mol. Des.*, vol. 28, no. 4, pp. 305–317, 2014.

[4] D. R. Bell *et al.*, "Calculating binding free energies of host-guest systems using AMOEBA polarizable force field," *Phys. Chem. Chem. Phys.*, vol. 18, no. 44, pp. 30261–30269, Nov. 2016.

[5] D. L. Mobley and M. K. Gilson, "Predicting Binding Free Energies: Frontiers and Benchmarks," *Annu. Rev. Biophys.*, vol. 46, no. 1, pp. 531–558, 2017.

[6] M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, "The Statistical-Thermodynamic Basis for Computation of Binding Affinities:A CriticalReview," *Biophys. J.*, vol. 72, pp. 1047–1069, 1997.

[7] K. N. Houk, A. G. Leach, S. P. Kim, and X. Zhang, "Thermodynamic Organic Complexes Binding Affinities of Host – Guest , Protein – Ligand , and Protein – Transition-State Complexes Angewandte," *Angew. Chem. Int. Ed.*, vol. 42, pp. 4872–4897, 2003.

[8] E. Alvira, "Capacity of small molecules to form b -cyclodextrin inclusion complexes," *Supramol. Chem.*, vol. 22, no. 3, pp. 156–162, 2010.

[9] M. V Rekharsky, M. P. Mayhew, R. N. Goldberg, P. D. Ross, Y. Yamashoji, and Y. Inoue, "Thermodynamic and nuclear magnetic resonance study of the reactions of alpha- and beta-cyclodextrin with acids, aliphatic amines, and cyclic alcohols," *J. Phys. Chem. B*, vol. 101, no. 1, pp. 87–100, 1997.

[10] M. V. Rekharsky, M. P. Mayhew, R. N. Goldberg, P. D. Ross, Y. Yamashoji, and Y. Inoue, "Thermodynamic

and Nuclear Magnetic Resonance Study of the Reactions of α- and β-Cyclodextrin with Acids, Aliphatic Amines, and Cyclic Alcohols," *J. Phys. Chem. B*, vol. 101, no. 1, pp. 87–100, 1997.

[11] M. Rekharsky and Y. Inoue, "Chiral Recognition Thermodynamics of beta-Cyclodextrin: The Thermodynamic Origin of Enantioselectivity and the enthalpa-Entropy Compensation Effect," *Am. Chem. Soc.*, vol. 122, no. 18, pp. 4418–4435, 2000.

[12] M. V Rekharsky and Y. Inoue, "Complexation Thermodynamics of Cyclodextrins," *Chem. Rev.*, vol. 98, no. 5, pp. 1875–1917, 1998.

[13] K. Kellett, S. A. Kantonen, B. M. Duggan, and M. K. Gilson, "Toward Expanded Diversity of Host-Guest Interactions via Synthesis and Characterization of Cyclodextrin Derivatives," *J. Solution Chem.*, vol. 47, pp. 597–1608, 2018.

[14] M. V. Rekharsky and Y. Inoue, "Solvent and Guest Isotope Effects on Complexation Thermodynamics of Alpha-, Beta-, and 6-amino-6-deoxy-beta-cyclodextrins," *J. Am. Chem. Soc.*, vol. 124, no. 41, pp. 12361–12371, 2002.

[15] J. Lagona, P. Mukhopadhyay, S. Chakrabarti, and L. Isaacs, "The Cucurbit[n]uril Family Angewandte," *Angew. Chem. Int. Ed.*, vol. 44, pp. 4844–4870, 2005.

[16] M. B. Hillyer, C. L. D. Gibb, P. Sokkalingam, J. H. Jordan, S. E. Ioup, and B. C. Gibb, "Synthesis of Water-Soluble Deep-Cavity Cavitands," *Org. Lett.*, vol. 18, no. 16, pp. 4048–4051, 2016.

[17] M. Hardouin-Lerouge, P. Hudhomme, and M. Salle, "Molecular clips and tweezers hosting neutral guests," *Chem. Soc. Rev.*, vol. 40, pp. 30–43, 2011.

[18] F. G. Klarner and B. Kahlert, "Molecular Tweezers and Clips as Synthetic Receptors . Molecular Recognition and Dynamics in Receptor - Substrate Complexes," *Acc. Chem. Res.*, vol. 36, no. 12, pp. 919–932, 2003.

[19] K. Kellett, B. M. Duggan, and M. Gilson, "Facile Synthesis of a Diverse Library of Mono-3-substituted β - Cyclodextrin Analogues," *Supramol. Chem.*, vol. 31, no. 4, pp. 251–259, 2018.

[20] T. Wiseman, S. Williston, J. F. Brandts, and L. N. Lin, "Rapid measurement of binding constants and heats of binding using a new titration calorimeter," *Anal. Biochem.*, vol. 179, no. 1, pp. 131–137, 1989.

[21] S. A. Kantonen, N. M. Henriksen, and M. K. Gilson, "Evaluation and Minimization of Uncertainty in ITC Binding Measurements: Heat Error, Concentration Error, Saturation, and Stoichiometry," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1861, no. 2, pp. 485–498, 2017.

[22] A. C. Rodrigo, E. Laurini, M. P. Vieira, S. Pricl, and D. K. Smith, "Effect of buffer at nanoscale molecular recognition interfaces – electrostatic binding of biological polyanions," *Chem. Commun.*, vol. 53, no. 84, pp. 11580–11583, 2017.

[23] J. Carrazana, A. Jover, F. Meijide, V. H. Soto, and J. V. Tato, "Complexation of Adamantyl Compounds by beta-cyclodextrin and Monoaminoderivatives," *J. Phys. Chem. B*, vol. 109, no. 19, pp. 9719–9726, 2005.

[24] B. Lai and C. Oostenbrink, "Binding free energy , energy and entropy calculations using simple model systems," *Theor. Chem. Acc.*, vol. 131, p. 1272, 2012.

[25] N. M. Henriksen, A. T. Fenley, and M. K. Gilson, "Computational Calorimetry : High-Precision Calculation of Host − Guest Binding Thermodynamics," *J. Chem. Theory Comput.*, vol. 11, pp. 4377–4394, 2015.