Synthesis and Exploration of Large-Ring Cyclodextrins using Enzyme-Mediated Dynamic Combinatorial Chemistry

Erichsen, Andreas Kjær

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Synthesis and Exploration of Large-Ring Cyclodextrins using Enzyme-Mediated Dynamic Combinatorial Chemistry

PhD Thesis

Andreas Kjær Erichsen

December 2022

Department of Chemistry

Technical University of Denmark

Supervisor: Associate Professor Sophie R. Beeren

Co-supervisor: Associate Professor Sebastian Meier

Preface

This PhD thesis presents research carried out from November 2019 to December 2022 at the Department of Chemistry, Technical University of Denmark, under the supervision of Associate Professor Sophie R. Beeren and co-supervision of Associate Professor Sebastian Meier. An external research stay of four months was carried out at the Department of Chemistry, University of Cambridge, United Kingdom, in the research group of Professor Jonathan R. Nitschke.

Coursework of 30 ECTS and mandatory teaching activities were completed as required. Funding for this project was provided by the Technical University of Denmark.

Acknowledgements

First of all, I would like to thank my supervisor Associate Professor Sophie R. Beeren. Thank you for allowing me to work with and learn from you. Thank you for all of your guidance and for truly caring about your students. Your dedication to and passion about your research is inspiring.

Thank you to Professor Günther H. J. Peters, who agreed to supervise my work with molecular dynamics simulations. Thank you for helping me learn and explore a different and fascinating research area and for all the helpful discussions and feedback.

Thank you to my co-supervisor Sebastian Meier for your help with NMR spectroscopy throughout the years. Thank you to all the people at the NMR Center DTU, with a special thank you to Kasper Enemark-Rasmussen for all the help with acquiring high field NMR spectra.

Thank you to all the present and former group members of the Beeren group who I have had to pleasure to work alongside. You have all contributed to making the group a friendly, helpful and dynamic place. A special thanks to Dennis Larsen, for your help in many different regards and for functioning as another co-supervisor. Also a special thanks to Thomas and Mie for your work with and synthesis of compounds relevant to this thesis. Thank you to all the technical staff at DTU chemistry for various help throughout the years.

Thank you to Professor Jonathan R. Nitschke for welcoming me in your research group at the University of Cambridge. Thank you to all of the Nitschke group members for your help and for making me feel welcome, with a special thanks to Jaehyoung Koo for your help and the collaboration on research.

Andreas Kjær Erichsen
Copenhagen, December 2022

Abstract

In this thesis, the synthesis and exploration of cyclodextrins (CDs) using enzyme-mediated dynamic combinatorial chemistry (DCC) is presented (Figure 1). CDs are macrocycles consisting of α-(1→4)-linked glucopyranose units (α-1,4-glucans). The 'native' CDs, with 6, 7 or 8 glucopyranose units, are ubiquitous hosts both within the field of supramolecular chemistry and in the industry. Large-ring cyclodextrins (LRCDs), with more than 8 glucopyranose units, are largely unexplored due to synthetic inaccessibility. When an α-1,4-glucan starting material is treated with the enzyme cyclodextrin glucanotransferase (CGTase), a mixture of interconverting CDs is obtained. In this mixture, also called a dynamic combinatorial library (DCL), the distribution of products is determined by their relative stabilities, which means that the addition of a guest (template) that binds to specific CDs will lead to the selective production (amplification) of the bound CDs (Figure 1).

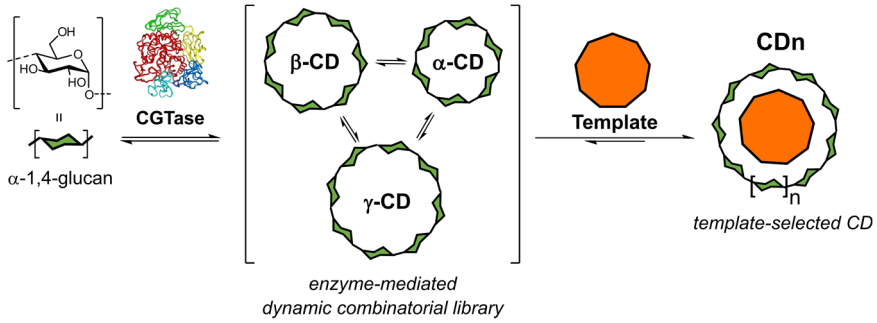


Figure 1. Concept of enzyme-mediated dynamic combinatorial chemistry with cyclodextrins (CDs).

This thesis explores the use of a variety of different templates for use in enzyme-mediated DCLs of CDs. In the first part of the thesis, it is shown how the addition of a range of simple chaotropic and kosmotropic anions lead to either direct template effects or to increased template effects of hydrophobic guests, respectively.

The main part of this thesis deals with LRCDs, including the exploration of their host-guest chemistry and as synthetic targets in enzyme-mediated DCLs of CDs. Using bolaamphiphile templates, the LRCD with 9 glucopyranose units, δ-CD, was amplified and synthesized on a preparatory scale in unprecedented yields, with efficient recycling of the template. NMR titrations revealed how δ-CD forms [2]-, [3]-, and [4]-*pseudo*rotaxanes with the bolaamphiphiles. The host-guest chemistry of LRCDs with 12–16 glucopyranose units was explored using adamantanecarboxylic acid as a hydrophobic probe in molecular dynamics (MD) simulations and NMR titrations, where cooperative 1:2 binding with figure-of-eight-like geometries was found in some cases. The synthesis of very large LRCDs was explored, where it was found that the addition of (bola)amphiphiles to DCLs with 'cycloamylose' as starting material led to the kinetic trapping of LRCDs with 26, 27 and 38 glucopyranose units. Finally, a high-yielding, chromatography-free, scalable templated enzymatic synthesis of δ-CD using a superchaotropic dodecaborate cluster as template is presented.

Table of contents

List of Abbreviations

ACA	Adamantane carboxylate
AF	Amplification factor
aq.	Aqueous
Asp	Aspartic acid
Boc	*tert*-butyloxycarbonyl protecting group
br	Broad signal
CDn	Cyclodextrin with n = number of glucopyranose units
COSY	Correlation spectroscopy
d	Doublet
DCC	Dynamic combinatorial chemistry
DCL	Dynamic combinatorial library
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DOSY	Diffusion-ordered spectroscopy
EC Number	Enzyme Commission Number
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
eq.	Equivalent(s)
ESI	Electrospray ionisation
Et	Ethyl
Fmoc	Fluorenylmethyloxycarbonyl protecting group
G	Guest
GAFF	General Amber Force Field
Glc-1-P	α-D-glucose-1-phosphate
Glu	Glutamic acid
Gn	Linear α-1,4-glucan with n = number of glucopyranose units
H	Host
HILIC	Hydrophilic interaction liquid chromatography
HMBC	Heteronuclear multiple-bond correlation
HPLC	High-performance liquid chromatography
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single quantum correlation
J	Coupling constant
K	Equilibrium constant
K_a	Association/binding constant
K_{dim}	Dimerisation constant
LCMS	Liquid chromatography–mass spectrometry
LRCD	Large-ring cyclodextrin
m	Multiplet
M.P.	Melting point
m/z	mass-to-charge ratio

MALDI	Matrix-Assisted Laser Desorption/Ionisation
MD	Molecular dynamics
Me	Methyl
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
Oxyma	Ethyl cyanohydroxyiminoacetate
p	Pentet/Quintet
ppm	Parts per million
q	Quartet
rt.	Room temperature
rpm	Revolutions per minute
s	Singlet
SDS	Sodium dodecyl sulfate
SHS	Sodium hexadecyl sulfate
t	Triplet
TBAI	Tetra-*n*-butylammonium iodide
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TOF	Time-of-flight
U	Enzyme activity unit
δ	Chemical shift
Δδ	Change in chemical shift
ΔG	Change in Gibbs free energy
ΔH	Change in Enthalpy
ΔS	Change in Entropy

Chapter 1. Introduction

Abstract

The following chapter, which is divided into two parts, serves as a general introduction to this thesis. The first part introduces cyclodextrins (CDs), both native CDs and large-ring cyclodextrins (LRCDs), including their structure, properties, host-guest chemistry, enzymatic synthesis and applications. The second part of the chapter introduces systems chemistry and dynamic combinatorial chemistry, primarily focusing on enzyme-mediated dynamic combinatorial chemistry.

1.1. Cyclodextrins

1.1.1. α-Glucans

Sugars, carbohydrates, saccharides — the most abundant and structurally diverse class of biomolecules also boast a diverse and complex soup of names and classifications. The term 'glycan', usually synonymous with the terms 'polysaccharide' (many sugars) and 'oligosaccharide' (few sugars), covers the class of compounds formed from monosaccharide units linked together through glycosidic bonds.[1,2] The focus of this thesis is on the class of glycans called 'glucans', the most abundant glycan type consisting entirely of glucose monosaccharide units.[3] Glucose (Figure 1.1) in its pyranose form has two different configurations, the α- and β-configuration, where for D-glucose, these correspond to an axial and an equatorial hydroxyl group at the anomeric center, respectively.

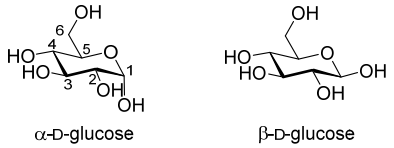


Figure 1.1. Configurations of D-glucopyranose.

From just this one building block with two configurations, a diverse array of linear and branched glucans are formed in Nature with a wide variety of biological functions.[3,4] In most cases the glucose units are linked together through (1→4) or (1→6) glycosidic linkages, but (1→2) and (1→3) linkages are also found.[3] The most abundant glucans in Nature are the two components of starch — amylose and amylopectin[3,5] — and the primary component of cell walls, cellulose (Figure 1.2).[5,6]

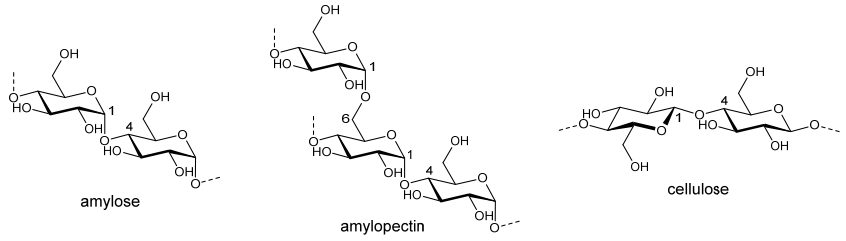


Figure 1.2. Structure of the most abundant glucans found in Nature, amylose, amylopectin and cellulose.

Amylose is a linear polymer, where glucose units are joined together only by α-(1→4) glycosidic linkages (an α-1,4-glucan), while amylopectin is a branched polymer with mostly α-(1→4) and a few α-(1→6) glycosidic linkages. Cellulose is a linear β-(1→4)-linked polymer (a β-(1,4)-glucan) that self-assembles into strong, sheet-like microfibrils held together by intra- and intermolecular hydrogen bonds.[7] Amylopectin contains branch points (α-(1→6) linkages) every 24 to 30 glucose units and can consists of up to more than a million glucose units, while amylose usually consist of hundreds to thousands of glucose units.[8] In the crystalline form, amylose adopts a double helical supramolecular structure, with two different native allomorphs (A and B form).[9–12] Amylose is able to bind hydrophobic guests by adapting a third form — the 'V-form'[12–15] (Figure 1.3), where the polysaccharide adopts a left-handed helical

conformation with guests bound to the hydrophobic cavity formed by the helix, exemplified with the well-known starch-iodine complex. The V-amylose helix is stabilized by intramolecular hydrogen bonds between glucose units from sequential turns of the helix. V-amylose is also quite flexible and can adapt to the size of the hydrophobic guests by varying the diameter of the cavity.[12,16]

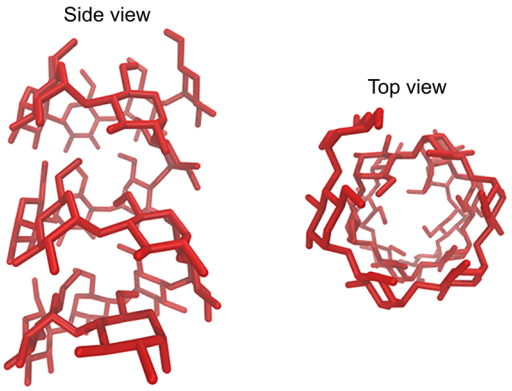


Figure 1.3. Illustration of a V-amylose helix with 6 glucose units per helical turn. Adapted with permission from ref.[12]

Relative orientations of neighboring glucose units in α-1,4-glucans can be defined as *syn, anti* or *kink* (Figure 1.4). All glucose units in crystalline V-amylose are in *syn* relative orientations,[12,13] and solution NMR studies have also shown that the *syn* relative orientation is the major conformation in solution.[17] However, the experimentally observed transition between extended and coiled structures of single-stranded amylose[17] has been attributed to transitions from the *syn* to the *anti* conformation based on molecular dynamics (MD) simulations,[18,19] and three out of four of the major carbohydrate force fields allow this transition.[19] The *anti* conformation also plays a major role in the structure of large-ring cyclodextrins, as will be discussed below.

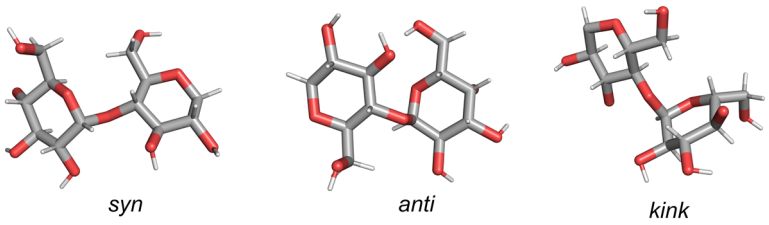


Figure 1.4. Relative conformations of neighboring glucose units.

1.1.2. Structure of cyclodextrins

Cyclodextrins (CDs) are a family of macrocyclic oligo- and polysaccharides formed from α-1,4-linked glucopyranose units. The 'native' or 'classical' CDs, α-CD, β-CD and γ-CD, with 6, 7 and 8 glucose units (CD6, CD7 and CD8) were discovered more than a century ago.[20] Since then, the nomenclature of CDs have been continuously up to debate. French, who was one of the pioneers in the early studies of CDs and discovered the molecular weights of α-CD and β-CD, argued that the prefixes α and β served no structural significance, and proposed the

nomenclature ‘cycloamylose’ for CDs (cyclohexaamylose for α-CD, cycloheptaamylose for β-CD and so on) due to their structural similarity with amylose.[21] Despite this, the Greek prefixes have remained the preferred nomenclature for the native CDs. Around the same time that the structures of the native CDs were elucidated (1940s–1970s), large-ring cyclodextrins (LRCDs),[22,23] with more than 8 glucopyranose units, were discovered,[24–26] but they remained unexplored until the successful isolation and characterization of a range of LRCDs (CD9–CD31) on milligram scales in the 1990s.[27–31]

The structure and properties of the native CDs have been rigorously studied throughout the years, especially since the 1970s when the industrial scale production of these CDs became possible.[20,32] In contrast, the knowledge of LRCDs has remained relatively sparse due to the limited quantities available.[22,23] So far, crystal structures of CD9 (δ-CD),[27,28] CD10 (ε-CD)[33], CD14[33] and CD26[34–36] have been reported. Figure 1.5a shows a generalized structure of CDs, and Figure 1.5b and 1.5c shows crystal structures of CD6–CD9. The native CDs, CD6–CD8 exhibit rigid, cone-shaped structures with a hydrophobic cavity. All glucopyranose units are in *syn* relative orientations and adhere to the schematic representation shown in Figure 1.5a, where a narrow ‘primary’ rim contains the O6 hydroxyl groups and a wider ‘secondary’ rim contains the O2 and O3 hydroxyl groups.

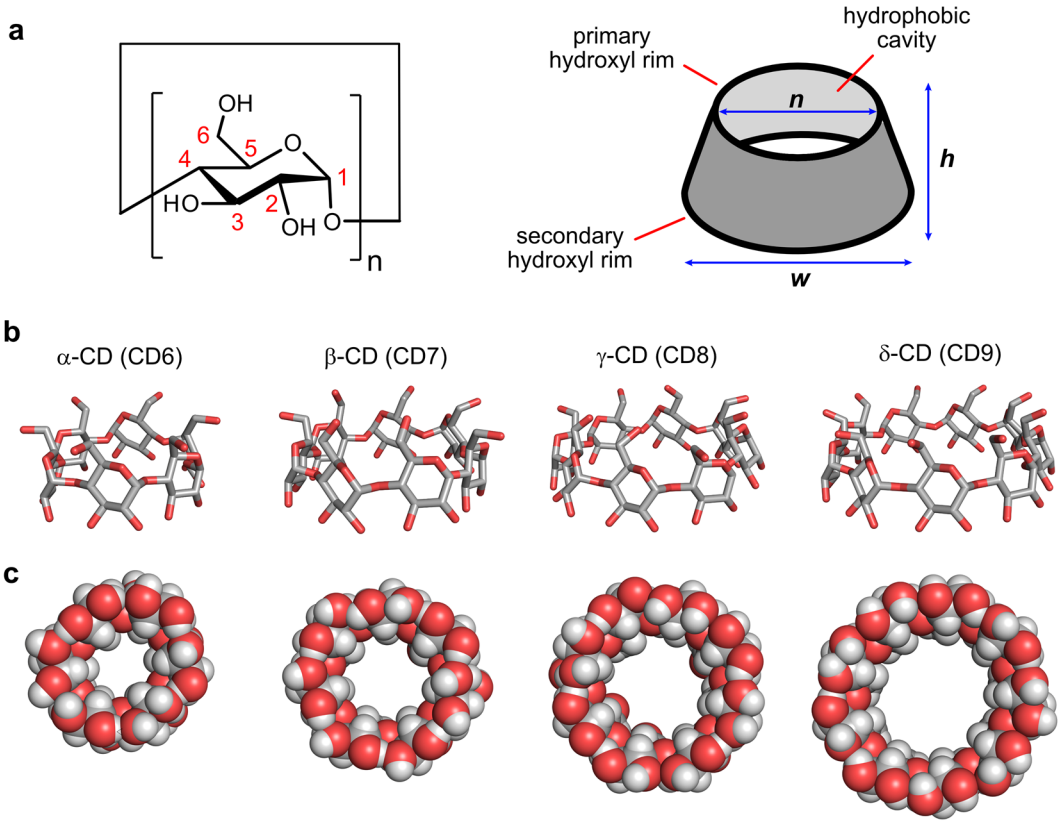


Figure 1.5. Structure of CD6–CD9. (a) General structure of CDs. (b) Crystal structures of CD6,[37] CD7,[38] CD8,[39] and CD9[40] (co-crystallized guests omitted, hydrogens removed for clarity). (c) Space-filling models of the same crystal structures.

In the first crystal structure of CD9 that was reported, the structure did not adhere to the strict cone-shape depicted in Figure 1.5, but rather a slightly elliptic boat-shape (Figure 1.6).[27] Later, a crystal structure of CD9 in complex with cycloundecanone was reported,[40] where the structure of CD9 adhered to the 'classical' cone-shape. This structure of CD9, with cycloundecanone omitted, is depicted in Figure 1.5. The structural parameters for CD6–CD9,[41,42] where the parameters for CD9 are linearly extrapolated from CD6–CD8 assuming an annular cone-shape, are shown in Table 1.1, where the increase in size (diameter parameters *w* and *n* and volume of cavity V_{cavity}) of the CDs with each extra glucose is apparent.

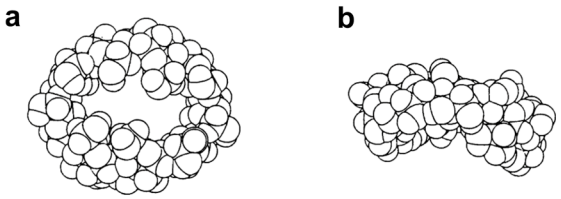


Figure 1.6. Crystal structure of CD9. (a) View from O6 side. (b) Side view. Adapted with permission from ref.[27]

Table 1.1. Structural parameters for CD6–CD9 assuming an annular CD shape.[41,42]

CD	*w* / Å	*n* / Å	*h* / Å	V_{cavity} / Å^3
CD6	8.0	5.0	9.0	174
CD7	9.7	5.6	9.0	262
CD8	10.7	7.0	9.0	427
CD9	12.6	7.8	9.0	541

From the two reported crystal structures of CD9, it is clear that this CD, the smallest of the LRCDs, has some flexibility to both deviate from the cone-shape to alleviate strain with no guest present and also to adjust its geometry to the binding of a guest. This feature unsurprisingly becomes more prominent with even larger LRCDs. The crystal structures of CD10,[33] CD14[33] and CD26[36] are shown in Figure 1.7. CD10 and CD14 were crystallized without any guests, while the structure of CD26 shown here was crystallized with undecanoic acid. All of these structures exhibit 'band flips', a term introduced by the Saenger and co-workers to describe the *anti* oriented glucose units observed in these CDs.[33,43] It has commonly been argued that the band flips are *induced* by conformational strain in the macrocycles,[33,34,43] while others have argued that the band flips instead constitute an *allowed* structure due to the higher flexibilities of these larger CDs.[22] Regardless, the structures obtained for these LRCDs are very different from the rigid cone-shapes of the small CDs. CD10 and CD14 exhibit elliptically distorted, open bent-boat like structures. In CD26, the two band flips connect two left-handed V-amylose-like helices with 6 glucoses per turn of the helices, with one molecule of undecanoic acid bound to each helical segment, analogous to the host-guest chemistry observed for V-amylose.[12]

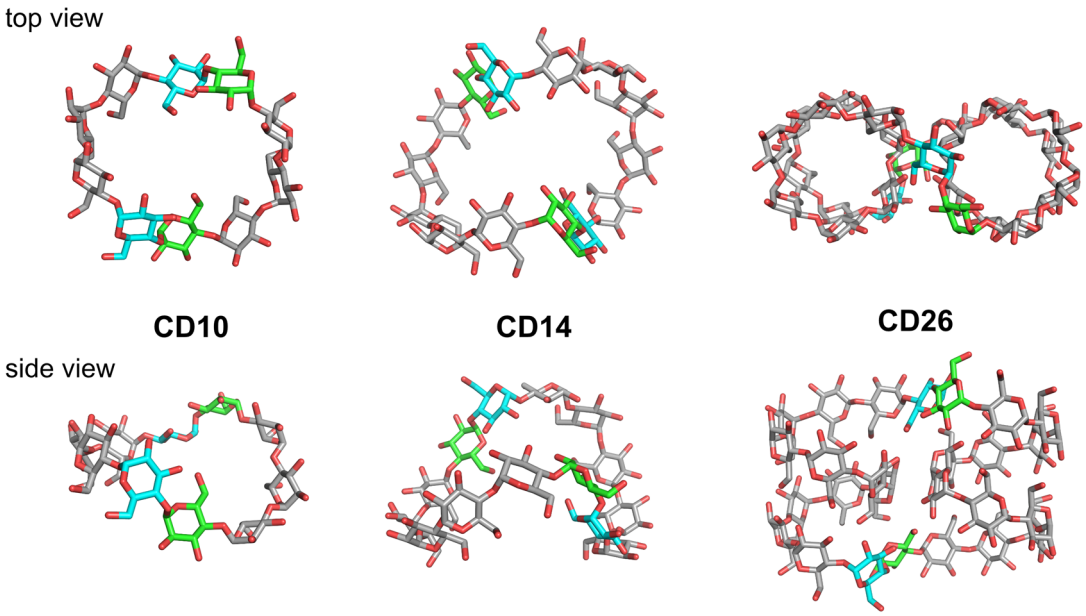


Figure 1.7. Crystal structures of CD10,[33] CD14[33] and CD26.[36] Guest (undecanoic acid) omitted for CD26. Glucose units involved in 'band flips' colored cyan and green.

In the decade following the discovery of the crystal structures of LRCDs, Ivanov, Jaime and co-workers performed molecular dynamics (MD) simulations with a wide range of LRCDs.[44–51] A summary of some of their results, with average structures obtained from the MD simulations with CD10–CD17 and CD24, CD26 and CD30, is shown in Figure 1.8. From here, one general trend is quite clear: larger CDs exhibit higher flexibilities. Going from CD10–CD13, the CDs adhere to a bent-boat-like macrocyclic structure,[46,49] similar to the crystal structures of CD10 and CD14.[33] From CD14 to CD17, the higher flexibilities of these CDs leads to the appearance of figure-of-eight-like geometries.[46,49] Simulations with CD24, CD26 and CD30, whose starting geometries were all based on the crystal structure of CD26, show that these CDs are very flexible,[46] and the structures obtained with CD26 diverge from the symmetrical nature of the crystal structure of CD26.

The flexibility of the LRCDs is also apparent in their physiochemical properties. For all CDs, ^{13}C NMR analysis showed only six sharp peaks (one for each carbon in the glucose units), which indicates rapid transitions and structural changes in the CDs well below the ~millisecond timescale of these NMR experiments.[23,31,43] Except for CD9, CD10, CD14 and CD26, the aqueous solubility of the LRCDs is generally high (>100 mg/ml), which has also been related to the flexibility and lack of well-defined rigid structures of these CDs.[23,52,53]

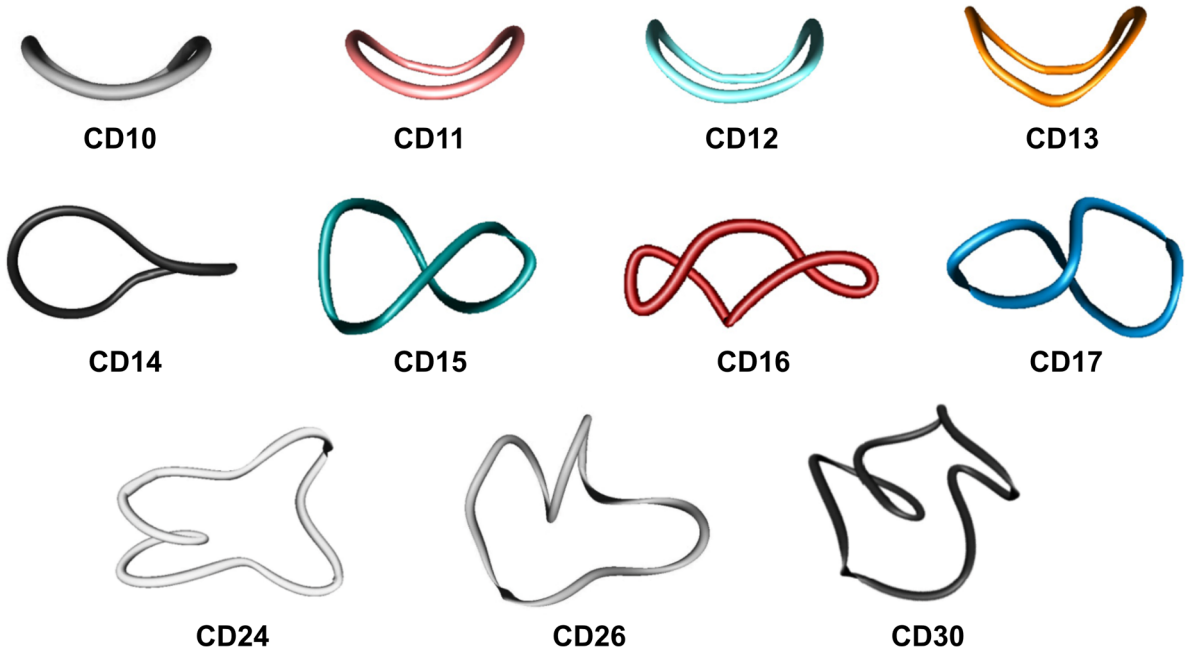


Figure 1.8. Average structures of CD10–CD17 and CD24, CD26 and CD30 derived from the MD simulations with these CDs carried out by Ivanov, Jaime and co-workers. Adapted with permission from ref.[46,49]

1.1.3. Host-guest chemistry of cyclodextrins

The host-guest chemistry of the native cyclodextrins, CD6–CD8, has been the subject of rigorous studies throughout the years.[20,41] Starting in the 1970s, Wolfram Saenger pioneered the study of cyclodextrin inclusion complexes by studying the crystal structures of the complexes formed.[54,55] The complexation thermodynamics of cyclodextrins in aqueous solution have since been studied with a wide variety of techniques, such as UV-Vis, NMR, fluorescence, circular dichroism, solubility determinations, pH potentiometry and calorimetry methods.[41]

The hydrophobic cavities of CD6–CD8[55] lead to the favorable inclusion of hydrophobic guest molecules in water driven in general by the hydrophobic effect. The nature and origin of the hydrophobic effect is controversial and has been the subject of continuous debate.[56–61] A well-known model for the hydrophobic effect, which began with the work of Frank and Evans,[62] invokes highly structured water, or 'icebergs' around nonpolar solutes. In this model, the aggregation of these nonpolar solutes leads to the release of structured water, giving a process characterized by a favorable entropic term, $\Delta S > 0$, a hallmark of what is generally known as 'the classical hydrophobic effect'.[63] As to the validity of this model — well, in the words of one the current significant voices in the debate, Bruce C. Gibb, "*the jury is still out*".[60] The 'nonclassical' hydrophobic effect is a designation used to describe processes primarily driven by a favorable enthalpic term, $\Delta H < 0$, a phenomenon commonly observed with cavitands such as curcubiturils and cyclodextrins.[61] To explain this phenomenon, a model involving the release of 'high-energy' water is sometimes invoked (Figure 1.9),[61,64] but the usefulness of this nomenclature has been questioned by others.[60]

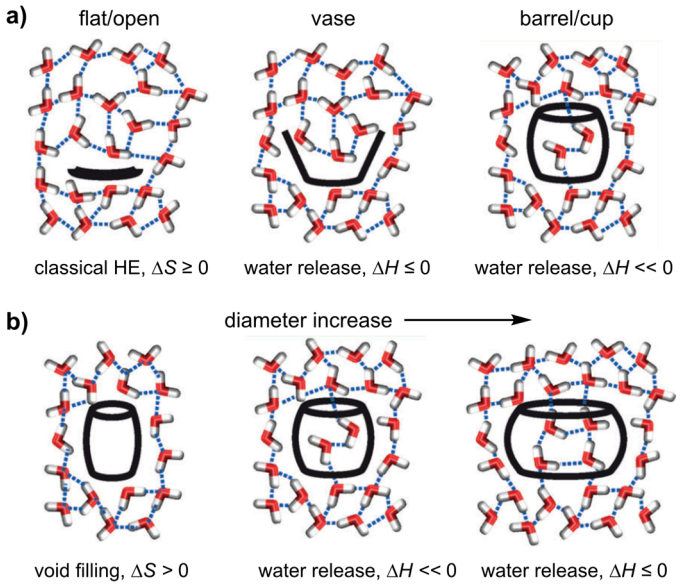


Figure 1.9. The classical ($\Delta S > 0$) or nonclassical ($\Delta H < 0$) hydrophobic effect (HE) in hydrophobic hosts with different geometries. Adopted with permission from ref.[61]

In this model,[61] concave hosts ('barrel/cup') (Figure 1.9a) such as cyclodextrins and especially curcubiturils contain 'high-energy' water molecules that do not form the standard number of hydrogen bonds as seen in bulk water (~3.6). Expulsion of this 'high-energy' water by the binding of hydrophobic guests then leads to the (highly) favorable enthalpic terms associated with the non-classical hydrophobic effect. With more open non-polar surfaces, such as the 'flat/open' and 'vase'-type geometries, the water molecules close to the non-polar surfaces are not as 'frustrated' due to a higher number of hydrogen bonds formed, and thus for the flat/open solutes, the classical hydrophobic effect ($\Delta S > 0$) drives binding of other non-polar solutes. Highly favorable enthalpic terms ($\Delta H << 0$) is seen for cavities with an 'optimal' size (Figure 1.9b), that effectively encapsulate and 'frustrate' several water molecules. Larger cavities generally lead to less frustrated water molecules with more hydrogen bonds.

For CDs, the binding of predominantly hydrophobic guests is quite complex, and specific interactions, such as hydrogen bonds and polar interactions (with polar guests) and dispersive interactions, as well as binding modes with guests binding to the outside of the CDs, make the picture more intricate that the simplified model of Figure 1.9.[41,61] For the binding of hydrophobic guests to the native cyclodextrins (CD6–CD8), favorable enthalpies (i.e. non-classical hydrophobic effect) are generally observed.[41] The entropic contributions are generally unfavorable for CD6 with the smallest cavity, while for CD7 and CD8, the entropic contributions can be either favorable or unfavorable.[41] Sometimes, significant favorable entropic contributions can be seen for CD7 and CD8, where less confined and precise binding is involved.[41,61,65] Generally, high binding affinities to CD6–CD8 are observed in the cases where there is good size complementarity between the guest and the cavity of the CD. Alkyl-based guests generally bind strongest to CD6 ($\log(K_a) \sim 3$–4), while adamantyl-based guests have high affinity towards CD7 ($\log(K_a) \sim 3$–5) and larger guests such as pyrene prefer the CD8 cavity ($\log(K_a) \sim 3$).[41] Guests with long hydrophobic axles can form 'channel' inclusion

complexes where several CDs are threaded along the axle ([n]-rotaxanes and [n]-*pseudo*rotaxanes), and using polymers such as polyethylene glycol-derivatives as guests leads to the formation of poly-rotaxanes and poly-*pseudo*rotaxanes.[66–68] When CD8 is used in combination with guests with long hydrophobic axles, two axles can be incorporated into the 'channel' of CDs.[67]

The binding of non-hydrophobic solutes, such as chaotropic anions, to the native CDs, is a well-known phenomenon.[69–74] Recently, Nau and co-workers characterized the binding of chaotropes to hydrophobic surfaces as a generic driving force under the term 'the chaotropic effect'.[75] The thermodynamic fingerprint of chaotropic binding is a favorable enthalpy ($\Delta H < 0$) and an unfavorable entropy ($\Delta S < 0$),[74,75] in contrast to the entropy-driven classical hydrophobic effect. While 'standard' chaotropic anions such as ClO_4^- and SCN^- generally exhibit low binding affinities to cyclodextrins (< 30 M^{-1}), high affinity binding to CD8 (in the range of 10^4–10^6 M^{-1}) was observed for the 'superchaotropic' dodecaborate cluster anions ($B_{12}X_{12}^{2-}$, where X is a halide).[76] This study was later extended to large-ring cyclodextrins, CD9–CD11, where binding affinities of similar sizes were observed.[42]

1.1.4. Host-guest chemistry of large-ring cyclodextrins

While the host-guest chemistry of the native CDs is exceedingly well-studied, the knowledge of the host-guest chemistry of large-ring cyclodextrins (LRCDs) is relatively sparse due to synthetic inaccessibility, with only a relatively small number of research groups having worked with isolated LRCDs.[23,42] The most well-studied of the LRCDs, CD9, has been shown to solubilize a number of hydrophobic drugs, albeit significantly less well than CD7 and CD8.[28,77] From these solubility studies, binding constants between CD9 and the steroid-derived drugs digitoxin and spironolactone were determined to be 1700 M^{-1} and 820 M^{-1}, respectively.[77] CD9 has also been shown to bind cycloundecanone in the solid state[40,78] and increase the water solubility of fullerenes C_{60}[79] and C_{70}.[80,81] Binding studies between CD9–CD17 and a number of anionic substituted benzenes were carried by K. Larsen *et al*. using capillary electrophoresis.[82,83] The results obtained with 4-*tert*-butyl benzoate and ibuprofen are shown in Figure 1.10. The binding constants obtained for the LRCDs were no greater than 50 M^{-1}, and generally worse or similar to the affinities of these guests towards the native CDs.

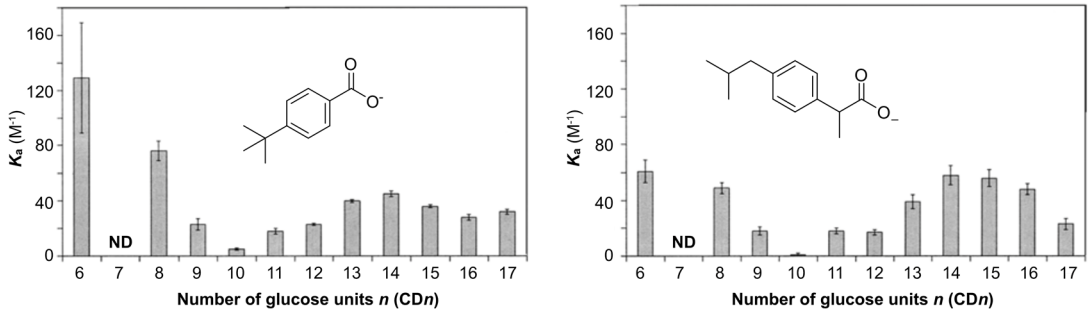


Figure 1.10. Binding constants (K_a) assuming 1:1 binding between CDs and 4-*tert*-butyl benzoate (left) and ibuprofen (right) obtained using capillary electrophoresis. ND: 'not determined'. Adopted with permission from ref.[22]

Some general trends can be extracted from the results. Going from CD6 to CD8 to CD9 and then CD10, the affinities become sequentially worse. This is presumably due to an increasing size mismatch between the CD cavities and the guest, where each additional glucose unit makes the CD cavities increasingly too large for the guests. Then, going from CD10 to CD14, the affinities increase again, now presumably due to the increased flexibility of these CDs, where they are able to adopt conformations that more closely match the size of the guests.

Okada and co-workers used isothermal titration calorimetry (ITC) to determine binding constants and thermodynamic parameters between I_3^- and CD21–CD32. The binding isotherms showed clear evidence of cooperative higher order binding, and the data were fitted to a 1:2 binding model with two identical interacting sites giving first (K_1) and second (K_2) binding constants and thermodynamic parameters (Figure 1.11). In general, the second binding was found to be enthalpically driven with an entropic cost. One trend that stands out in the series, is that enthalpy-entropy compensation[84] is at work for the second binding. Relatively large variations in the binding enthalpies and entropies lead to relatively small changes in the Gibbs free energy across the series. The largest outlier in the thermodynamic parameters is CD26, for which crystal structures with undecanoic acid, dodecanol,[36] and I_3^-[35] have been reported. Here, a highly favorable enthalpic term is balanced out by a highly unfavorable entropic term. In agreement with the reported crystal structures, it seems that the binding of I_3^- to CD26 leads to a highly ordered structure in comparison with the other CDs.

In a different ITC study with CD25–CD28 and sodium dodecyl sulfate (SDS), clear differences between CD26 and the other LRCDs were also observed.[85] For CD26, cooperative 1:2 binding of SDS was found, with fitted binding constants of $K_{a1} = 1.3 \times 10^4$ M^{-1} and $K_{a1} = 1.3 \times 10^5$ M^{-1}. For CD25, CD27, and CD28, 1:1 binding was instead observed, with binding constants of 7.2×10^3 M^{-1}, 8.3×10^3 M^{-1} and 2.6×10^3 M^{-1}, respectively, again highlighting how CD26 is a clear structural outlier is this range of LRCDs.

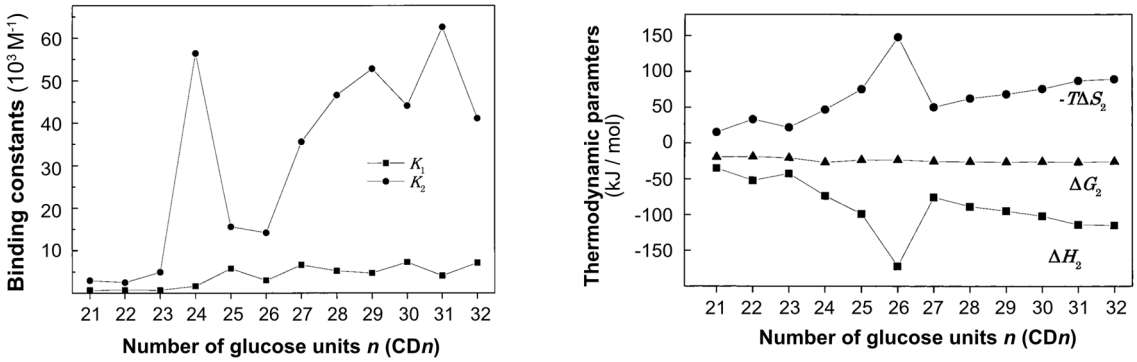


Figure 1.11. Binding constants (left) and thermodynamic parameters for the second binding event (right) of the binding of I_3^- to CD21–CD32 determined using ITC. Adopted with permission from ref.[86]

1.1.5. Cooperativity

As seen in the examples above, cooperativity can be an important factor in the binding of guests to large-ring cyclodextrins. Ternary complexes, with two guests included in the cavity, are also seen for the native CDs, with most observed cases being with CD8.[67,87–90] For example,

Abrahamian and co-workers reported that the *E*-isomer of an aromatic hydrazone-based photoswitch exhibited 1:2 binding with CD8 (Figure 1.12) with binding constants of $K_1 = 320\ M^{-1}$ and $K_2 = 8970\ M^{-1}$.[90]

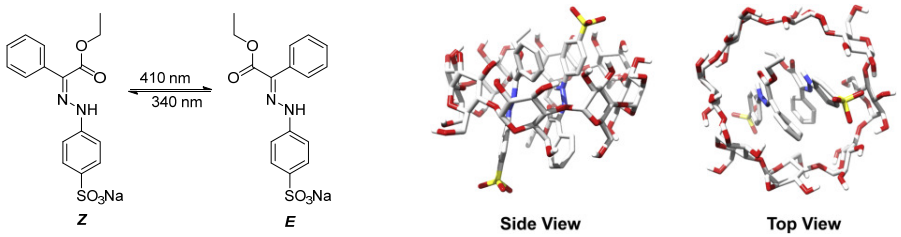


Figure 1.12. Cooperative 1:2 binding between hydrazone photoswitch and CD8. *Left: Z* and *E* isomers of photoswitch. *Right:* Structure of 1:2 complex between *E*-photoswitch and CD8 from MD simulation. Figures adapted with permission from ref.[90]

A few notes on cooperativity are warranted here. In a system where a host contains two binding sites, cooperativity exists when the binding of a guest to one site influences the binding of a guest to the second binding site. This type of cooperativity is called allosteric cooperativity.[91] For the 1:2 binding of a guest (G) to a host (H) with two identical binding sites, ($H + 2G \rightleftharpoons HG_2$) and where the binding sites are independent (no cooperativity), the second binding constant (K_2) will be four times smaller than the first binding constant (K_1), so $K_1 = 4K_2$.[92] Conceptually, this can be explained with the following: In the first binding event ($H + 2G \rightleftharpoons HG + G$), there are two guests and two binding sites available, while for the second binding ($HG + G \rightleftharpoons HG_2$), there is only one binding site and one guest available. As such, the probability of binding is four times smaller for the second binding event.[92] Thus, positive allosteric cooperativity exists in cases where $K_2 > 4K_1$ and negative allosteric cooperativity exists for cases where $K_2 < 4K_1$.[92]

1.1.6. CGTase and the enzymatic synthesis of cyclodextrins

The native CDs, CD6–CD8, are produced industrially by the enzymatic breakdown of starch[93] using the enzyme cyclodextrin glucanotransferase (CGTase, EC 2.4.1.19).[94,95] CGTase is an enzyme in the α-amylase family of enzymes,[96] which comprises a range of enzymes that in many cases catalyze the hydrolysis of α-(1→4) and α-(1→6) glycosidic bonds and in some cases catalyze the production of oligosaccharides with specific compositions, like cyclodextrins. CGTase acts on α-1,4-glucans to catalyze two different reaction types, reversible transglycosylations and slow hydrolysis (Figure 1.13).[94,97] The transglycosylation reaction can be either intermolecular, between two linear α-1,4-glucans (disproportionation), or intramolecular, where a linear α-1,4-glucan cyclizes to form a cyclodextrin and a shorter, linear α-1,4-glucan.

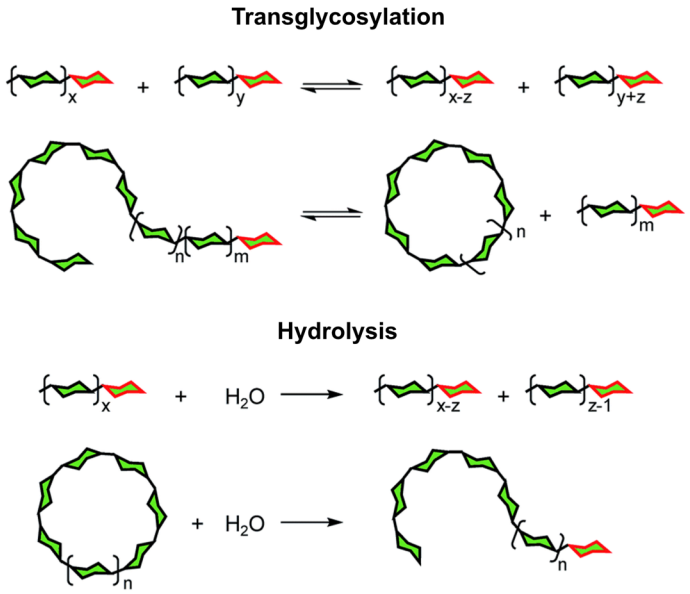


Figure 1.13. Reactions catalyzed by the enzyme CGTase. Green hexagons correspond to α-(1→4) linked glucopyranose units. Hydrolysis leads to the formation of a new reducing-end glucose unit (in red). n, m = [0,1,2,…]; x, y, x = [1,2,3,…]. Adapted with permission from ref.[98]

The mechanism for the transglycosylation reaction is shown in Figure 1.14.[95] Here, it is shown how two catalytic residues, a glutamate (Glu257) and an aspartate (Asp229) are involved in the mechanism. The Asp229 residue functions as a nucleophile while the Glu257 residue functions as an acid/base catalyst. A linear α-1,4-glucan (in blue) serves as leaving group and the O4 of the non-reducing end of another α-1,4-glucan (in red) serves as nucleophile in the final step to form a new α-(1→4) glycosidic bond. If the red and the black glucose residues belong to the same α-1,4-glucan, a cyclodextrin is formed (intramolecular transglycosylation).

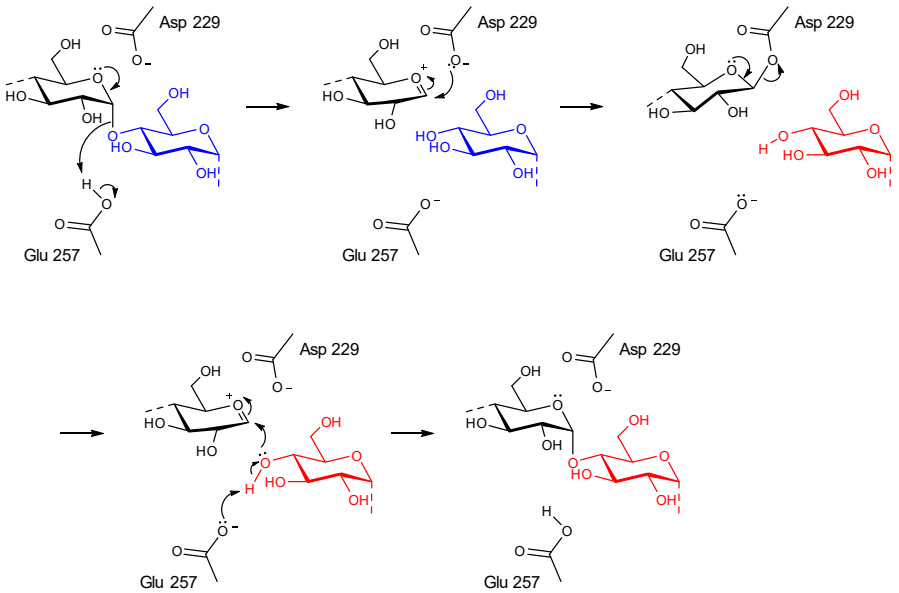


Figure 1.14. Mechanism for transglycosylation reactions with the enzyme CGTase. Adapted with modifications with permission from ref.[95]

A wide range of different processes for the industrial scale production of CD6, CD7 and CD8 exist.[93] Most of them involve the initial liquefaction of starch (usually by jet-cooking) followed by cooling of the solution to a temperature suitable for CGTase (usually around 60 °C). Then CGTase and 'complexing agents' in form of organic solvents such as toluene, ethanol or acetone are added, where they serve to increase the yield by precipitating the CDs. Depending on the complexing agent added, the reaction conditions, and the type of CGTase used, different ratios of CD6, CD7 and CD8 are obtained.[93] Purification typically involves removal of organic complexing agents by steam distillation, treatment with activated carbon and filtration, which yields a mixture of CDs according to the ratio obtained in the initial reaction. Due to its low aqueous solubility compared to CD6 and CD8, CD7 is easily obtained in high purity by recrystallization. The isolation of CD6 and CD8 often involves different types of chromatography, adsorption procedures or further precipitations with other complexing agents.[93]

Large-ring cyclodextrins are not produced industrially in this process described above. They are however present in small amounts in the reaction mixtures obtained when starch is treated with CGTase.[25] A range of LRCDs (CD9–CD13) were successfully isolated from 'cyclodextrin powder' in the 1990s in milligram quantities.[27–29,99] Cyclodextrin powder was a commercially available mixture of mainly CD6–CD8 that also contained small amounts of CD9–CD13, presumably made by the treatment of starch with CGTase. The isolation of CD9–CD13 involved the breakdown of non-cyclic glucans using enzymes and yeast, followed by precipitation of CD6–CD8 using complexing agents and ultimately column chromatography to isolate the LRCDs. Using this method, CD9 was isolated first in 0.04% yield,[28] and later in 0.24% yield using a different 'cyclodextrin powder'.[99] The discovery in 1997 that the initial reaction of CGTase with synthetic amylose produces LRCDs[97] (CD9–CD60) resulted in the subsequent successful isolation and characterization of CD9–CD31 using repeated semi-preparative reverse-phase chromatography.[31,100,101] The organic synthesis of CD9 was achieved in 2002, using a 26-step synthesis involving intramolecular transglycosylations.[102] In recent years, Zimmermann and co-workers have developed engineered CGTases with increased selectivity for LRCDs and isolated CD10–CD12 in a single chromatographic step.[103–105] Despite these advances, LRCDs are still generally unavailable, with CD10 being the only LRCD advertised as commercially available.[106]

1.1.7. Applications of cyclodextrins

The native CDs, CD6–CD8, and modified versions of these CDs have found widespread use in numerous industries, such as the pharmaceutical, food and cosmetics industries.[107] In 1998, the annual global consumption of CDs was 6000 tons, with an annual growth rate of 15–20%.[93] The general role of CDs in the industries is to solubilize, stabilize, mask or passivate certain ingredients in products by forming host-guest complexes with these ingredients/compounds.[108] In the food industry, this could be to protect certain unstable or volatile aromas, flavors or plant oils from oxidation or evaporation during processing or storage.[108,109] CDs can also be used to mask certain unwanted flavors or aromas, such as the

branched chain fatty acids (like 4-methyloctanoic acid) present in goat milk and responsible for the 'goaty' flavor.[110] In pharmaceuticals, CDs are likewise used to mask the bitterness of drugs for oral consumption.[108] Native CDs or modified CDs with hydrophilic substituents such as hydroxypropyl-β-CD are used to increase the bioavailability of lipophilic drugs by increasing the solubility and dissolution rate of these drugs.[111]

1.2. Systems Chemistry

In Nature, life emerges from the immensely complex interplay between myriads of molecules that interact, fold, assemble and react. For a long time, chemists have avoided the study of complex mixtures, and relied on the isolation of pure compounds to study their properties and structures.[112–115] Chemists of the past were limited by the available analytical methods, but the vast strides made in the development of these methods, such as high-performance liquid chromatography (HPLC) systems coupled to various detectors, have since facilitated the study of complex systems.[115] The field of supramolecular chemistry is now more than 50 years old.[116] The deeper understanding of the nature of interactions between molecules gained in this field has further enabled the study of complex systems of interacting molecules. The developing field of Systems Chemistry studies the 'emergent properties' in systems of interconverting and interacting molecules. Such 'emergent properties' could be the biological functions found in Nature, or in synthetic supramolecular systems, amplification of specific molecules due to self-assembly or stimuli within complex mixtures.[115] Many such systems operate under thermodynamic control (Figure 1.15a), which is the case for dynamic combinatorial libraries (DCLs).[117–121] Kinetically controlled systems (Figure 1.15b) also exist,[122–124] where structures and properties exist transiently as systems are trapped in local energy minima.[114] Far-from-equilibrium systems (Figure 1.15c) have also been developed,[125,126] where a continuous energy supply pushes the system towards otherwise unseen assemblies and properties.[114]

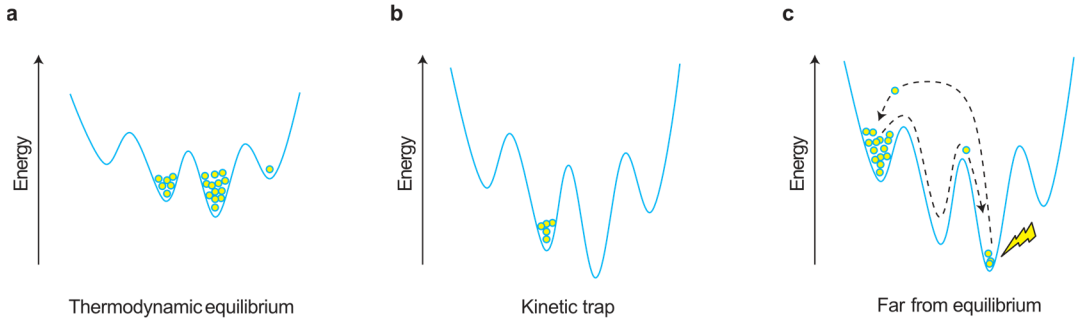


Figure 1.15. Schematic representations of chemical systems. (a) Thermodynamic equilibrium. (b) Kinetic control. (c) Far-from-equilibrium. Adopted with permission from ref.[114]

1.2.1. Dynamic combinatorial chemistry

In dynamic combinatorial chemistry (DCC),[118,121,127,128] building blocks are linked together using reversible chemical bonds (Figure 1.16) generating a dynamic combinatorial library (DCL) of interconverting oligomers. Reversible reactions commonly employed in DCLs include imine exchange, hydrazone exchange, acetal exchange, disulfide exchange, alkene cross metathesis, borate ester exchange and metal-ligand coordination.[118,121,128] In DCLs under thermodynamic control, the equilibrium distribution of oligomers in the DCL will depend on the intrinsic stabilities of each library member. In these systems, a change in the experimental conditions such as a change in temperature, concentrations or solvent will lead to the shift in distributions of oligomers to reflect the changed equilibrium distribution under the new conditions.[127] Another way of influencing a DCL involves the addition of a 'template', a

compound that can bind to one or more library members leading to the stabilization of the bound library members. The shift in distribution towards these bound library members is called an amplification, which can be quantified using the *amplification factor* (AF), where, for a library member A, the amplification factor upon the addition of a template is defined as $AF = [A]_{templated}/[A]_{untemplated}$.[129]

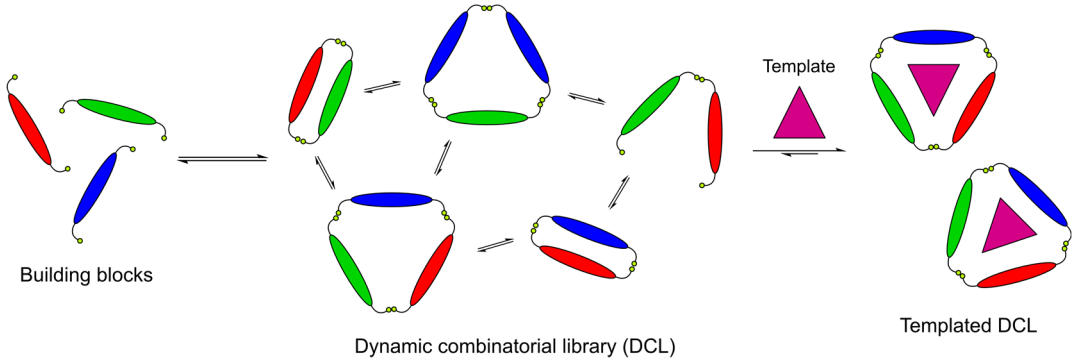
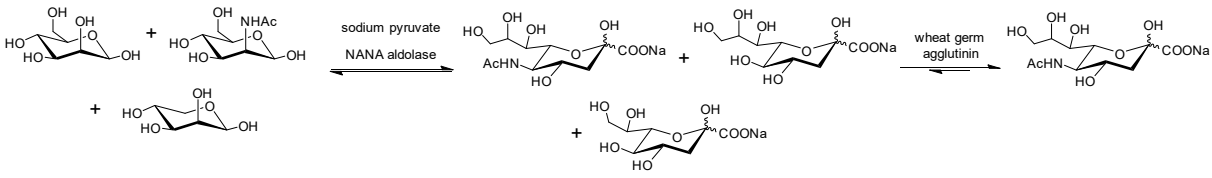


Figure 1.16. Schematic representation of a dynamic combinatorial library (DCL) and a templated DCL, where the addition of a template leads to amplification of the best binder.

By employing templates in DCC, a diverse range of complex and even surprising structures have been discovered throughout the years, such as macrocycles,[117] anion receptors,[130,131] cages,[132] knots and catenanes,[120,133,134], foldamers[135] and ligands and receptors for biomolecules.[136]

1.2.2. Enzyme-mediated DCC

While most studies using dynamic combinatorial chemistry have employed the conventional reversible covalent reactions mentioned previously, enzymatic reactions are also great candidates for use in DCLs. One of the advantages of enzymatic reactions is that they work well under mild conditions at physiological pH in aqueous solution, which means that targeting biologically relevant, water-soluble compounds is feasible. Enzymatic reactions are also often substrate-, chemo- and stereoselective, which, in principle, enables the selective exchange with biological building blocks with several functional groups with similar reactivity and complex stereochemistry such as saccharides. However, only a few enzyme-mediated DCC systems have been reported thus far.[98,126,137–141] Brown and co-workers used the enzyme *N*-acetylneuraminic acid aldolase (NANA aldolase, EC 4.1.3.3) to generate a small DCL of sialic acid analogues and showed that the addition of wheat germ agglutinin as a template led to the amplification of one of the library members (Scheme 1.1).[137,138]



Scheme 1.1. Enzyme-mediated DCL of sialic acid analogues catalysed by the enzyme NANA aldolase using wheat germ agglutinin as a template to amplify a library member.

Ulijn and co-workers have explored the use of the enzyme thermolysin (EC 3.4.24.27) for the generation of peptide-based DCLs.[126,139,140] Thermolysin is an endopeptidase that catalyzes the reversible hydrolysis and formation of amide bonds, with an equilibrium that lies far towards the hydrolysed products under normal conditions (Figure 1.17). By employing a peptide design using Fmoc-protected peptides that could self-assemble into beta-sheet structures, the DCLs were pushed towards the most stable self-assembled structures while operating under thermodynamic control.[139] Later, Ulijn and co-workers further developed the system for use with unprotected peptides to form peptide nanostructures.[140]

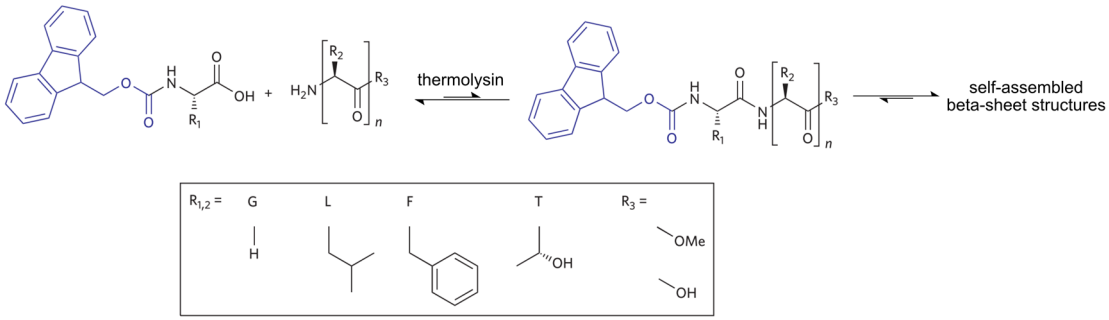


Figure 1.17. Enzyme-mediated DCC using the enzyme thermolysin and various peptides to generate self-assembled beta-sheet structures under thermodynamic control. Adapted with permission from ref.[139]

1.2.3. Enzyme-mediated DCC with oligosaccharides

The Beeren group has recently explored enzyme-mediated DCLs with α-1,4-glucans. The enzyme phosphorylase (EC 2.4.1.1) has been employed in enzyme-mediated DCLs of linear α-1,4-glucans. In Nature, phosphorylase catalyzes the breakdown of glycogen to glucose-1-phosphate (Glc-1-P) using inorganic phosphate.[142,143] Using the reversible nature of this enzymatic reaction, former PhD student Charlotte Nybro Dansholm showed how a reaction mixture with an excess of glucose-1-phosphate, fluorescently labelled maltohexaose, and phosphorylase leads to DCLs of linear α-1,4-glucans (Figure 1.18).[144] It was shown that the system operated under thermodynamic control, and that the addition of a bolaamphiphile template led to a significant amplification of certain α-1,4-glucans with specific lengths due to the binding of the α-1,4-glucans to the template in a V-amylose-like binding mode.[144]

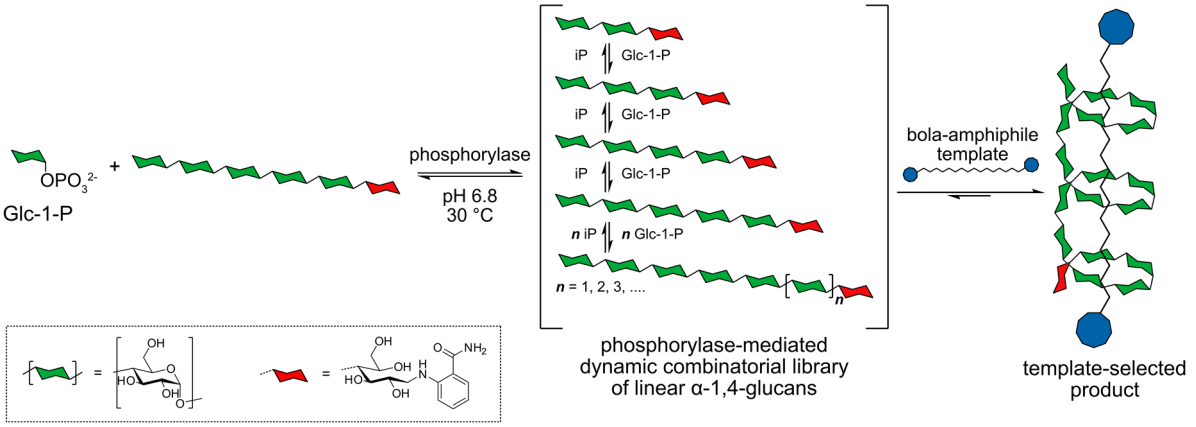


Figure 1.18. Phosphorylase-mediated DCC using fluorescently labelled maltohexaose, glucose-1-phosphate (Glc-1-P) and phosphorylase. Addition of a suitable template leads to changes in the product distribution.

In the Beeren group, Larsen *et al*. showed in 2019 how the enzyme CGTase could be used to generate DCLs of CDs under *pseudo*-thermodynamic control.[98] By treating an α-1,4-glucan starting material, such as α-CD, with CGTase at neutral pH and at room temperature, a mixture of interconverting native CDs (α-, β-, and γ-CD) and linear α-1,4-glucans (**Gn**, where n is the number of glucopyranose units) was obtained. The authors showed that the CDs form as the main products in a transient DCL, a kinetically trapped dynamic subsystem (Figure 1.19). As CGTase catalyzes both fast, reversible transglycosylations and slow hydrolysis,[94,95] the DCL is slowly converted into the true thermodynamic product, glucose (**G1**).

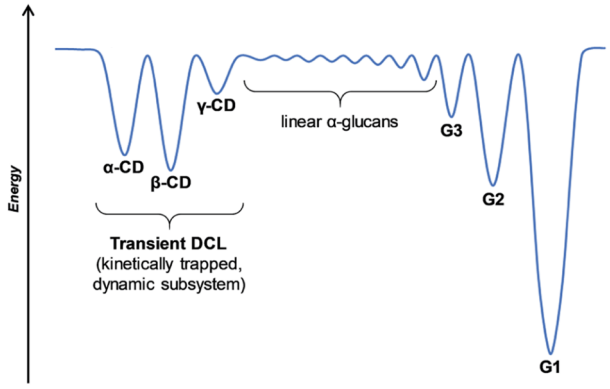


Figure 1.19. Energy diagram showing the members of CGTase-mediated DCLs of α-1,4-glucans, with CDs and their linear counterparts.

By adding different templates with selective binding to each of the CDs, the authors obtained β-CD and γ-CD with 99% selectivity, using adamantane carboxylate and tetraphenylborate, respectively (Figure 1.20). α-CD was obtained with 83% selectivity using sodium dodecyl sulfate (SDS). The lesser selectivity using SDS was due to strong binding of SDS to both α-CD and β-CD. The addition of superchaotropic dodecaborate cluster $B_{12}I_{12}^{2-}$, which binds to γ-CD, CD9 and CD10 with high affinities (6.7×10^4 M^{-1}, 6.8×10^5 M^{-1}, and 2.1×10^6 M^{-1}, respectively),[42] led to the amplification of CD9 and CD10. Here, γ-CD forms as the primary product despite the higher binding constants for CD9 and CD10 due to the intrinsically higher stability of γ-CD.

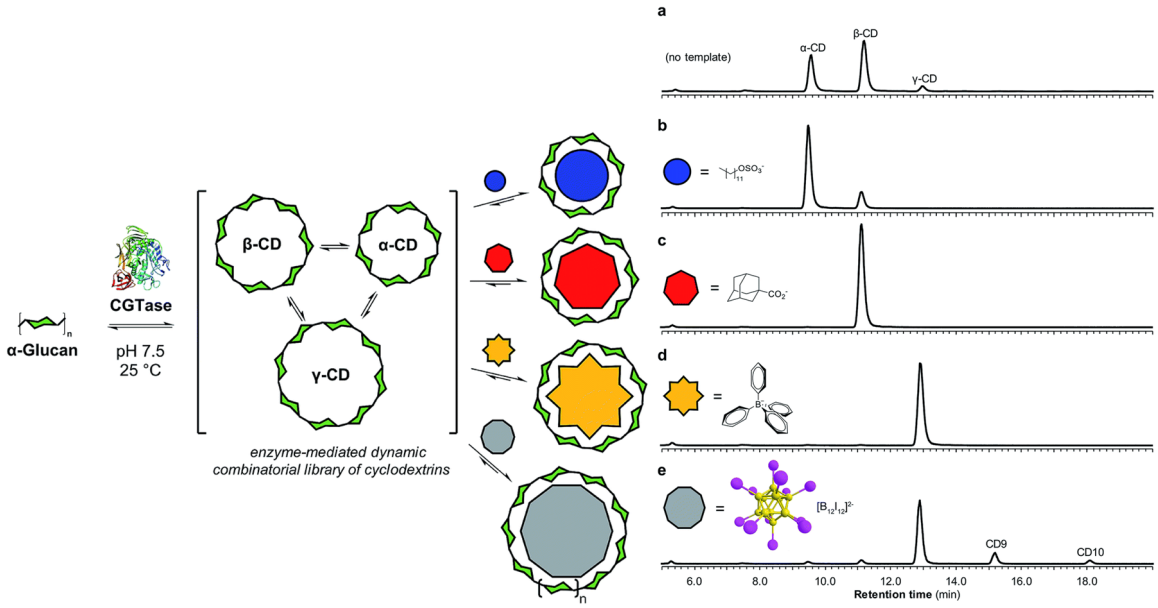


Figure 1.20. CGTase-mediated DCL of CDs. *Right:* Distributions obtained with (a) no template, (b) sodium dodecyl sulfate, (c) adamantanecarboxylic acid, (d) $NaBPh_4$, and (e) $Na_2B_{12}I_{12}$. Green hexagons correspond to α-(1→4) linked glucopyranose units. Reproduced with permission from ref.[98]

In a later study, Larsen *et al.* showed how the tuning of the reaction conditions in CGTase-mediated DCLs could change the product distribution. In one of their findings, the authors obtained 99% selectivity for α-CD using SDS as a template by lowering the temperature. α-CD binds SDS in a 2:1 stoichiometry with two high binding affinities of $K_{a1} = 1.9 \times 10^4$ M^{-1} and $K_{a2} = 2.3 \times 10^4$ M^{-1}, while β-CD binds SDS in a 1:1 stoichiometry with one high binding affinity of $K_{a1} = 1.6 \times 10^4$ M^{-1} (Figure 1.21).[98,145] Thus, by lowering the temperature, the formation of the 2:1 complex, which is entropically disfavored, becomes *less* entropically disfavored leading to a higher selectivity for α-CD in the DCL. It was also shown how the addition of up to 30% ethanol to the reaction mixture led to a higher production of α-CD (from 34% α-CD in water to 79% α-CD in a 30/70 ethanol/water mixture). This change was ascribed to the low affinity binding of ethanol to α-CD ($K_a = 10$ M^{-1})[41] leading to a regular template effect.[145]

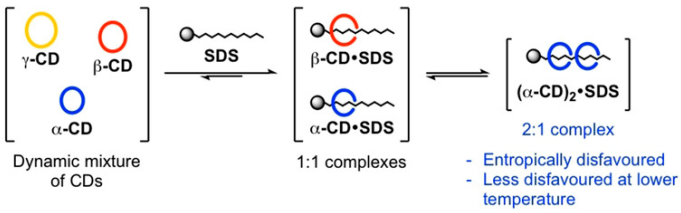


Figure 1.21. Schematic representation of the complexes (1:1 and 2:1) formed in a DCL of CDs in the presence of sodium dodecyl sulfate (SDS). Reproduced with permission from ref.[145]

The use of photoswitches as templates in DCLs of CDs have also been explored in the Beeren group. In a study from 2019, Larsen *et al.* used various azobenzenes to generate light-responsive DCLs.[146] Here, the *trans* isomers led to the amplification of α-CD and the *cis* isomers led to the amplification of β-CD, and it was shown that the composition of the DCLs

could be continuously switched back and forth in consecutive cycles. In a different study carried out in collaboration with the Abrahamian group, a hydrazone-based photoswitch that binds γ-CD cooperatively in a 1:2 stoichiometry (shown previously in Figure 1.12) was used to amplify and template the dynamic enzymatic synthesis of γ-CD.[90]

In a recent study, Larsen *et al.* showed how the CGTase-mediated DCC of CDs could be used to synthesize β-CD from maltose in high yields (70%) using adamantanecarboxylic acid as a template.[147] A different investigation looked into the synthesis of modified CDs using CGTase-mediated DCLs.[148] In this study, it was shown that when employing mono-6-deoxy-α-CD as a starting material, CDs with up to three modifications could be obtained. It was also shown that it was possible to template the amplification of specific sizes of modified CDs (α-, β-, γ-CD), and that the system (CGTase) also tolerated other modified CDs as substrates, specifically mono-6-azido-6-deoxy-β-CD and mono-6-deoxy-6-iodo-β-CD.

1.3. Motivation and thesis outline

Large-ring cyclodextrins (LRCDs) are largely unexplored, and have remained an academic curiosity since their discovery, due to synthetic inaccessibility. With the multitude of academic and industrial applications that have been found for the native CDs, it is hard to imagine that LRCDs would not follow the same trend, were they to become available in significant quantities. Enzyme-mediated DCC with oligosaccharides and CDs, while still in its infancy, has shown great promise, and the fundamentals have been laid out by the previous groundwork carried out in the Beeren group. As such, the main goals of this thesis were twofold: the design and synthesis of templates that bind to LRCDs, and the use of these templates, both to probe the host-guest chemistries of LRCDs and to direct the templated enzymatic synthesis of LRCDs. Finding efficient syntheses of LRCDs is, of course, highly desirable. In addition, exploring the host-guest chemistry of LRCDs could lead to the discovery of better templates/guests for LRCDs and inspire possible applications of LRCDs.

In a different project, the effects of chaotropic and kosmotropic anions on CGTase-mediated DCLs of native CDs were explored to further expand the knowledge of and toolbox available for use in these enzymatic DCLs.

The contents of this thesis is divided into the following chapters:

Chapter 2 deals with the influence of chaotropic and kosmotropic anions on CGTase-mediated dynamic combinatorial libraries of native CDs.

Chapter 3 presents the exploration of novel synthetic bolaamphiphiles as templates for the templated enzymatic synthesis of the first large-ring cyclodextrin CD9 (δ-CD). This chapter also discusses the formation of [n]-*pseudo*rotaxanes with these bolaamphiphiles and CD6–CD9.

Chapter 4 considers large-ring cyclodextrins with 10–20 glucose units and describes the exploration of the host-guest chemistry of these LRCDs using molecular dynamics (MD) simulations and NMR spectroscopy.

Chapter 5 presents the use of more well-established binding motifs to target the templated enzymatic synthesis of LRCDs. Part one of this chapter presents the efforts towards the synthesis of CD26 using (bola)amphiphiles as templates, while part two of this chapter presents the scalable synthesis of CD9 (δ-CD) using a superchaotropic dodecaborate cluster as template.

Finally, a general conclusion and outlook for the work presented in this thesis is provided.

1.4. Bibliography

[1] A. D. McNaught, A. Wilkinson, *IUPAC. Compendium of Chemical Terminology, 2nd Ed. (the 'Gold Book')*, Blackwell Scientific Publications, Oxford, **1997**.

[2] R. A. Dwek, *Biochem. Soc. Trans.* **1995**, *23*, 1–25.

[3] A. Synytsya, M. Novak, *Ann. Transl. Med.* **2014**, *2*, 1–14.

[4] A. Varki, *Glycobiology* **2017**, *27*, 3–49.

[5] T. K. Lindhorst, *Essentials of Carbohydrate Chemistry and Biochemistry*, Wiley-VCH, Weinheim, Germany, **2007**.

[6] K. Kamide, *Cellulose Products and Cellulose Derivatives: Molecular Characterization and Its Applications*, Elsevier, Amsterdam, **2005**.

[7] M. C. Jarvis, *Food Hydrocoll.* **2011**, *25*, 257–262.

[8] D. L. Nelson, M. M. Cox, *Principles of Biochemistry*, W. H. Freeman And Company, New York, USA, **2013**.

[9] W. Hinrichs, G. Büttner, M. Steifa, C. Betzel, V. Zabel, B. Pfannemüller, W. Saenger, *Science* **1987**, *238*, 205–208.

[10] A. Imberty, H. Chanzy, S. Perez, A. Buleon, V. Tran, *J. Mol. Biol.* **1988**, *201*, 365–378.

[11] A. Imberty, S. Perez, *Biopolymers* **1988**, *27*, 1205–1221.

[12] G. Fittolani, P. H. Seeberger, M. Delbianco, *Pept. Sci.* **2019**, e24124.

[13] R. E. Rundle, F. C. Edwards, *J. Am. Chem. Soc.* **1943**, *65*, 2200–2203.

[14] Y. Yamashita, *J. Polym. Sci. A.* **1965**, *3*, 3251–3260.

[15] T. Hirai, S. Hayashi, M. Hirai, T. Ueki, *Macromolecules* **1992**, *25*, 6699–6702.

[16] J. A. Putseys, L. Lamberts, J. A. Delcour, *J. Cereal Sci.* **2010**, *51*, 238–247.

[17] N. W. H. Cheetham, L. Tao, *Carbohydr. Polym.* **1998**, *35*, 287–295.

[18] B. M. Sattelle, A. Almond, *Phys. Chem. Chem. Phys.* **2014**, *16*, 8119–8126.

[19] J. K. Koneru, X. Zhu, J. Mondal, *J. Chem. Theory Comput.* **2019**, *15*, 6203–6212.

[20] G. Crini, *Chem. Rev.* **2014**, *114*, 10940–10975.

[21] D. French, R. E. Rundle, *J. Am. Chem. Soc.* **1942**, *64*, 1651–1653.

[22] K. L. Larsen, *J. Incl. Phenom.* **2002**, *43*, 1–13.

[23] T. Endo, *Trends Glycosci. Glycotechnol.* **2011**, *23*, 79–92.

[24] K. Freudenberg, F. Cramer, *Z. Naturforsch.* **1948**, *3b*, 464.

[25] A. O. Pulley, D. French, *Biochem. Biophys. Res. Commun.* **1961**, *5*, 11–15.

[26] D. French, A. Pulley, J. A. Effenberger, M. A. Rougvie, M. Abdullah, *Arch. Biochem. Biophys.* **1965**, *111*, 153–160.

[27] T. Fujiwara, N. Tanaka, S. Kobayashi, *Chem. Lett.* **1990**, *19*, 739–742.

[28] I. Miyazawa, H. Ueda, H. Nagase, T. Endo, S. Kobayashi, T. Nagai, *Eur. J. Pharm. Sci.* **1995**, *3*, 153–162.

[29] T. Endo, H. Ueda, S. Kobayashi, T. Nagai, *Carbohydr. Res.* **1995**, *269*, 369–373.

[30] T. Endo, H. Nagase, H. Ueda, S. Kobayashi, T. Nagai, *Chem. Pharm. Bull.* **1997**, *45*, 532–536.

[31] K. Koizumi, H. Sanbe, Y. Kubota, Y. Terada, T. Takaha, *J. Chromatogr. A* **1999**, *852*, 407–416.

[32] E. M. M. Del Valle, *Process Biochem.* **2004**, *39*, 1033–1046.

[33] J. Jacob, K. Gessler, D. Hoffmann, H. Sanbe, K. Koizumi, S. M. Smith, T. Takaha, W. Saenger, *Angew. Chem. Int. Ed.* **1998**, *37*, 605–609.

[34] K. Gessler, I. Usón, T. Takaha, N. Krauss, S. M. Smith, S. Okada, G. M. Sheldrick, W. Saenger, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 4246–4251.

[35] O. Nimz, K. Geßler, I. Usón, S. Laettig, H. Welfle, G. M. Sheldrick, W. Saenger, *Carbohydr. Res.* **2003**, *338*, 977–986.

[36] O. Nimz, K. Gessler, I. Usón, G. M. Sheldrick, W. Saenger, *Carbohydr. Res.* **2004**, *339*, 1427–1437.

[37] K. Harata, *Bull. Chem. Soc. Jpn.* **1977**, *50*, 1416–1424.

[38] K. Harata, *Bull. Chem. Soc. Jpn.* **1984**, *57*, 2596–2599.

[39] K. Harata, *Bull. Chem. Soc. Jpn.* **1987**, *60*, 2763–2767.

[40] K. Harata, H. Akasaka, T. Endo, H. Nagase, H. Ueda, *Chem. Commun.* **2002**, 1968–1969.

[41] M. V. Rekharsky, Y. Inoue, *Chem. Rev.* **1998**, *98*, 1875–1917.

[42] K. I. Assaf, D. Gabel, W. Zimmermann, W. M. Nau, *Org. Biomol. Chem.* **2016**, *14*, 7702–7706.

[43] W. Saenger, J. Jacob, K. Gessler, T. Steiner, D. Hoffmann, H. Sanbe, K. Koizumi, S. M. Smith, T. Takaha, *Chem. Rev.* **1998**, *98*, 1787–1802.

[44] P. M. Ivanov, C. Jaime, *J. Phys. Chem. B* **2004**, *108*, 6261–6274.

[45] I. Maestre, I. Beà, P. M. Ivanov, C. Jaime, *Theor. Chem. Acc.* **2007**, *117*, 85–97.

[46] M. G. Gotsev, P. M. Ivanov, *J. Phys. Chem. B* **2009**, *113*, 5752–5759.

[47] P. M. Ivanov, *J. Phys. Chem. B* **2010**, *114*, 2650–2659.

[48] M. G. Gotsev, P. M. Ivanov, *Int. J. Quantum Chem.* **2007**, *107*, 1657–1672.

[49] P. Ivanov, *J. Mol. Struct.* **2012**, *1009*, 3–10.

[50] P. Ivanov, E. Atanassov, C. Jaime, *J. Mol. Struct.* **2014**, *1056–1057*, 238–245.

[51] P. M. Ivanov, E. J. Atanassov, C. Jaime, *Org. Biomol. Chem.* **2015**, *13*, 1680–1689.

[52] H. Ueda, M. Wakisaka, H. Nagase, T. Takaha, S. Okada, *J. Incl. Phenom.* **2002**, *44*, 403–405.

[53] H. Taira, H. Nagase, T. Endo, H. Ueda, *J. Incl. Phenom. Macrocycl. Chem.* **2006**, *56*, 23–28.

[54] R. K. McMullan, W. Saenger, J. Fayos, D. Mootz, *Carbohydr. Res.* **1973**, *31*, 37–46.

[55] W. Saenger, *Angew. Chem. Int. Ed.* **1980**, *19*, 344–362.

[56] W. Blokzijl, J. B. F. N. Engberts, *Angew. Chem. Int. Ed.* **1993**, *32*, 1545–1579.

[57] N. T. Southall, K. A. Dill, A. D. J. Haymet, *J. Phys. Chem. B* **2002**, *106*, 521–533.

[58] D. Chandler, *Nature* **2005**, *437*, 640–647.

[59] P. Ball, *Chem. Rev.* **2008**, *108*, 74–108.

[60] N. E. Ernst, B. C. Gibb, *Water runs deep* in *Supramolecular Chemistry in Water* (Edited by S. Kubik) Wiley-VCH Verlag GmbH, Weinheim, Germany, **2019**.

[61] F. Biedermann, W. M. Nau, H. J. Schneider, *Angew. Chem. Int. Ed.* **2014**, *53*, 11158–11171.

[62] H. S. Frank, M. W. Evans, *J. Chem. Phys.* **1945**, *13*, 507–532.

[63] E. V. Anslyn, D. A. Dougherty, *Modern Physical Organic Chemistry*, University Science Books, **2006**.

[64] F. Biedermann, V. D. Uzunova, O. A. Scherman, W. M. Nau, A. De Simone, *J. Am. Chem. Soc.* **2012**, *134*, 15318–15323.

[65] A. V. Odinokov, S. V. Titov, V. A. Tikhomirov, M. V. Basilevsky, M. V. Alfimov, *Mol. Simul.* **2013**, *39*, 442–452.

[66] S. A. Nepogodiev, J. F. Stoddart, *Chem. Rev.* **1998**, *98*, 1959–1976.

[67] G. Wenz, B. H. Han, A. Müller, *Chem. Rev.* **2006**, *106*, 782–817.

[68] A. Hashidzume, H. Yamaguchi, A. Harada, *European J. Org. Chem.* **2019**, *2019*, 3344–3357.

[69] Á. Buvári, L. Barcza, *Inorganica Chim. Acta* **1979**, *33*, 179–180.

[70] I. Sanemasa, M. Fujiki, T. Deguchi, *Bull. Chem. Soc. Jpn.* **1988**, *61*, 2663–2665.

[71] Y. Matsui, M. Ono, S. Tokunaga, *Bull. Chem. Soc. Jpn.* **1997**, *70*, 535–541.

[72] R. I. Gelb, L. M. Schwartz, M. Radeos, D. A. Laufer, *J. Phys. Chem.* **1983**, *87*, 3349–3354.

[73] J. F. Wojcik, R. P. Rohrbach, *J. Phys. Chem.* **1975**, *79*, 2251–2253.

[74] M. R. Sullivan, W. Yao, D. Tang, H. S. Ashbaugh, B. C. Gibb, *J. Phys. Chem. B* **2018**, *122*, 1702–1713.

[75] K. I. Assaf, W. M. Nau, *Angew. Chem. Int. Ed.* **2018**, *57*, 13968–13981.

[76] K. I. Assaf, M. S. Ural, F. Pan, T. Georgiev, S. Simova, K. Rissanen, D. Gabel, W. M. Nau, *Angew. Chem. Int. Ed.* **2015**, *54*, 6852–6856.

[77] H. Ueda, A. Wakamiya, T. Endo, H. Nagase, K. Tomono, T. Nagai, *Drug Dev. Ind. Pharm.* **1999**, *25*, 951–954.

[78] H. Akasaka, T. Endo, H. Nagase, H. Ueda, S. Kobayashi, *Chem. Pharm. Bull.* **2000**, *48*, 1986–1989.

[79] T. Furuishi, Y. Ohmachi, T. Fukami, H. Nagase, T. Suzuki, T. Endo, H. Ueda, K. Tomono, *J. Incl. Phenom. Macrocycl. Chem.* **2010**, *67*, 233–239.

[80] T. Furuishi, T. Endo, H. Nagase, H. Ueda, T. Nagai, *Chem. Pharm. Bull.* **1998**, *46*, 1658–1659.

[81] T. Furuishi, T. Fukami, H. Nagase, T. Suzuki, T. Endo, H. Ueda, K. Tomono, *Pharmazie* **2008**, *63*, 54–57.

[82] K. L. Larsen, T. Endo, H. Ueda, *Carbohydr. Res.* **1998**, *309*, 153–159.

[83] K. L. Larsen, W. Zimmermann, *J. Chromatogr. A* **1999**, *836*, 3–14.

[84] L. Liu, Q. Guo, *Chem. Rev.* **2001**, *101*, 673–695.

[85] S. Kitamura, *Japanese Soc. Appl. Glycosci.* **2003**, *50*, 321–325.

[86] S. Kitamura, K. Nakatani, T. Takaha, S. Okada, *Macromol. Rapid Commun.* **1999**, *20*, 612–615.

[87] J. Andreaus, J. Draxler, R. Marr, A. Hermetter, *J. Colloid Interface Sci.* **1997**, *16*, 8–16.

[88] K. Kasatani, M. Ohashi, H. Sato, *Carbohydr. Res.* **1989**, *192*, 197–214.

[89] M. Pessêgo, N. Basilio, M. Parajó, *Org. Biomol. Chem.* **2015**, *13*, 1213–1224.

[90] S. Yang, D. Larsen, M. Pellegrini, S. Meier, D. F. Mierke, S. R. Beeren, I. Aprahamian, *Chem* **2021**, *7*, 2190–2200.

[91] C. A. Hunter, H. L. Anderson, *Angew. Chem. Int. Ed.* **2009**, *48*, 7488–7499.

[92] P. Thordarson, *Chem. Soc. Rev.* **2011**, *40*, 1305–1323.

[93] A. Biwer, G. Antranikian, E. Heinzle, *Appl. Microbiol. Biotechnol.* **2002**, *59*, 609–617.

[94] Y. B. Tewari, R. N. Goldberg, M. Sato, *Carbohydr. Res.* **1997**, *301*, 11–22.

[95] J. C. M. Uitdehaag, B. A. Van Der Veen, L. Dijkhuizen, B. W. Dijkstra, *Enzyme Microb. Technol.* **2002**, *30*, 295–304.

[96] B. Svensson, *Plant Mol. Biol.* **1994**, *25*, 141–157.

[97] Y. Terada, M. Yanase, H. Takata, T. Takaha, S. Okada, *J. Biol. Chem.* **1997**, *272*, 15729–15733.

[98] D. Larsen, S. R. Beeren, *Chem. Sci.* **2019**, *10*, 9981–9987.

[99] A. Wakamiya, T. Endo, H. Nagase, H. Ueda, S. Kobayashi, T. Nagai, *Yakuzaigaku* **1997**, *57*, 220–223.

[100] Q. Qi, X. She, T. Endo, W. Zimmermann, *Tetrahedron* **2004**, *60*, 799–806.

[101] Q. Qi, M. N. Mokhtar, W. Zimmermann, *J. Incl. Phenom. Macrocycl. Chem.* **2007**, *57*, 95–99.

[102] M. Wakao, K. Fukase, S. Kusumoto, *J. Org. Chem.* **2002**, *67*, 8182–8190.

[103] C. Sonnendecker, S. Melzer, W. Zimmermann, *Microbiologyopen* **2019**, *8*, 1–8.

[104] C. Sonnendecker, W. Zimmermann, *Catalysts* **2019**, *9*, 1–12.

[105] C. Sonnendecker, S. Thürmann, C. Przybylski, F. D. Zitzmann, N. Heinke, Y. Krauke, K. Monks, A. A. Robitzki, D. Belder, W. Zimmermann, *Angew. Chem. Int. Ed.* **2019**, *58*, 6411–6414.

[106] "epsilon-Cyclodextrin, Catalog NO.: 156510-98-4, Boc Sciences," can be found under https://www.bocsci.com/product/epsilon-cyclodextrin-cas-156510-98-4-2986.html?nid=856, **2022**.

[107] A. R. Hedges, *Chem. Rev.* **1998**, *98*, 2035–2044.

[108] Q. Wang, *Handbook of Macrocyclic Supramolecular Assembly*, Springer Nature Singapore Pte Ltd, Singapore, **2020**.

[109] G. Astray, J. C. Mejuto, J. Morales, R. Rial-Otero, J. Simal-Gándara, *Food Res. Int.* **2010**, *43*, 1212–1218.

[110] O. A. Young, R. B. Gupta, S. Sadooghy-Saraby, *J. Food Sci.* **2012**, *77*, 122–127.

[111] J. Pitha, J. Milecki, H. Fales, L. Pannell, K. Uekama, *Int. J. Pharm.* **1986**, *29*, 73–82.

[112] R. F. Ludlow, S. Otto, *Chem. Soc. Rev.* **2008**, *37*, 101–108.

[113] J. R. Nitschke, *Nature* **2009**, *462*, 736–738.

[114] E. Mattia, S. Otto, *Nat. Nanotechnol.* **2015**, *10*, 111–119.

[115] G. Ashkenasy, T. M. Hermans, S. Otto, A. F. Taylor, *Chem. Soc. Rev.* **2017**, *46*, 2543–2554.

[116] C. J. Pedersen, *J. Am. Chem. Soc.* **1967**, *89*, 2495–2496.

[117] S. Otto, R. L. E. Furlan, J. K. M. Sanders, *Science* **2002**, *297*, 590–593.

[118] P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J. L. Wietor, J. K. M. Sanders, S. Otto, *Chem. Rev.* **2006**, *106*, 3652–3711.

[119] J. M. Lehn, *Chem. Soc. Rev.* **2007**, *36*, 151–160.

[120] N. Ponnuswamy, F. B. L. Cougnon, J. M. Clough, G. D. Pantoş, J. K. M. Sanders, *Science* **2012**, *338*, 783–785.

[121] J. Li, P. Nowak, S. Otto, *J. Am. Chem. Soc.* **2013**, *135*, 9222–9239.

[122] J. M. A. Carnall, C. A. Waudby, A. M. Belenguer, M. C. A. Stuart, J. J. P. Peyralans, S. Otto, *Science* **2010**, *327*, 1502–1507.

[123] P. A. Korevaar, S. J. George, A. J. Markvoort, M. M. J. Smulders, P. A. J. Hilbers, A. P. H. J. Schenning, T. F. A. De Greef, E. W. Meijer, *Nature* **2012**, *481*, 492–496.

[124] S. Yang, G. Schaeffer, E. Mattia, O. Markovitch, K. Liu, A. S. Hussain, J. Ottelé, A. Sood, S. Otto, *Angew. Chem. Int. Ed.* **2021**, *60*, 11344–11349.

[125] J. Boekhoven, A. M. Brizard, K. N. K. Kowlgi, G. J. M. Koper, R. Eelkema, J. H. Van Esch, *Angew. Chem. Int. Ed.* **2010**, *49*, 4825–4828.

[126] S. Debnath, S. Roy, R. V. Ulijn, *J. Am. Chem. Soc.* **2013**, *135*, 16789–16792.

[127] F. B. L. Cougnon, J. K. M. Sanders, *Acc. Chem. Res.* **2012**, *45*, 2211–2221.

[128] P. Frei, R. Hevey, B. Ernst, *Chem. Eur. J.* **2019**, *25*, 60–73.

[129] R. F. Ludlow, S. Otto, *J. Am. Chem. Soc.* **2010**, *132*, 5984–5986.

[130] M. Lisbjerg, B. M. Jessen, B. Rasmussen, B. E. Nielsen, A. Madsen, M. Pittelkow, *Chem. Sci.* **2014**, *5*, 2647–2650.

[131] S. R. Beeren, J. K. M. Sanders, *J. Am. Chem. Soc.* **2011**, *133*, 3804–3807.

[132] S. Zarra, D. M. Wood, D. A. Roberts, J. R. Nitschke, *Chem. Soc. Rev.* **2015**, *44*, 419–432.

[133] J. F. Ayme, J. E. Beves, D. A. Leigh, R. T. McBurney, K. Rissanen, D. Schultz, *Nat. Chem.* **2012**, *4*, 15–20.

[134] D. A. Leigh, R. G. Pritchard, A. J. Stephens, *Nat. Chem.* **2014**, *6*, 978–982.

[135] C. G. Pappas, P. K. Mandal, B. Liu, B. Kauffmann, X. Miao, D. Komáromy, W. Hoffmann, C. Manz, R. Chang, K. Liu, et al., *Nat. Chem.* **2020**, *12*, 1180–1186.

[136] A. Herrmann, *Chem. Soc. Rev.* **2014**, *43*, 1899–1933.

[137] R. J. Lins, S. L. Flitsch, N. J. Turner, E. Irving, S. A. Brown, *Angew. Chem. Int. Ed.* **2002**, *41*, 3405–3407.

[138] R. J. Lins, S. L. Flitsch, N. J. Turner, E. Irving, S. A. Brown, *Tetrahedron* **2004**, *60*, 771–780.

[139] R. J. Williams, A. M. Smith, R. Collins, N. Hodson, A. K. Das, R. V. Ulijn, *Nat. Nanotechnol.* **2009**, *4*, 19–24.

[140] C. G. Pappas, R. Shafi, I. R. Sasselli, H. Siccardi, T. Wang, V. Narang, R. Abzalimov, N. Wijerathne, R. V. Ulijn, *Nat. Nanotechnol.* **2016**, *11*, 960–967.

[141] P. A. Swann, R. A. Casanova, A. Desai, M. M. Frauenhoff, M. Urbancic, U. Slomczynska, A. J. Hopfinger, G. C. Le Breton, D. L. Venton, *Biopolymers* **1996**, *40*, 617–625.

[142] E. C. O'Neill, R. A. Field, *Carbohydr. Res.* **2015**, *403*, 23–37.

[143] M. Yanase, T. Takaha, T. Kuriki, *J. Sci. Food Agric.* **2006**, *86*, 1631–1635.

[144] C. N. Dansholm, Phosphorylase-Mediated Dynamic Combinatorial Chemistry with Linear α-1,4-Glucans, Technical University of Denmark, **2022**.

[145] D. Larsen, S. R. Beeren, *Chem. Eur. J.* **2020**, *26*, 11032–11038.

[146] D. Larsen, P. M. Bjerre, S. R. Beeren, *Chem. Commun.* **2019**, *55*, 15037–15040.

[147] D. Larsen, S. R. Beeren, *Chem. Commun.* **2021**, *57*, 2503–2506.

[148] D. Larsen, M. Ferreira, S. Tilloy, E. Monflier, S. R. Beeren, *Chem. Commun.* **2022**, *58*, 2287–2290.

Chapter 2. Chaotropic and Kosmotropic Anions Regulate the Outcome of Enzyme-Mediated Dynamic Combinatorial Libraries of Cyclodextrins in Two Different Ways

The following chapter is based on the article "Chaotropic and Kosmotropic Anions Regulate the Outcome of Enzyme-Mediated Dynamic Combinatorial Libraries of Cyclodextrins in Two Different Ways" which was published in Frontiers in Chemistry in August 2021.[1] The contents of the article were copied as is, and then a few changes were made: The "Materials, Instrumentation and Methods" section has been moved to the experimental section of this chapter, and text, references and tables have been formatted to fit the style of this thesis. The 'Supplementary Material' of the publication can be found in the supporting information for chapter 2 of this thesis. Unlike the rest of this thesis, the focus of this chapter is not on large-ring cyclodextrins. Instead, this chapter focuses on expanding the toolbox available in enzyme-mediated dynamic combinatorial chemistry by exploring the use of chaotropic and kosmotropic anions.

Andreas Erichsen, Dennis Larsen and Sophie R. Beeren*

Department of Chemistry, Technical University of Denmark, Kongens Lyngby, Denmark

Abstract

We demonstrate how different anions from across the Hofmeister series can influence the behavior of enzyme-mediated dynamic combinatorial libraries of cyclodextrins (CDs). Using cyclodextrin glucanotransferase to catalyze reversible transglycosylation, dynamic mixtures of interconverting cyclodextrins can be formed wherein the relative concentrations of α-CD, β-CD and γ-CD is determined by their intrinsic stabilities and any stabilizing influences of added template (guest) molecules. Here, we find that addition of high concentrations of kosmotropic anions can be used to enhance the effects of added hydrophobic templates, while chaotropic anions can themselves act as templates, causing predictable and significant changes in the cyclodextrin composition due to weak, but specific, binding interactions with α-CD.

2.1. Introduction

Since the groundbreaking work of Franz Hofmeister more than a century ago into how salts affect the solubility of proteins,[2] countless studies have repeatedly revealed the Hofmeister series of anions: F^-, SO_4^{2-}, AcO^-, Cl^-, Br^-, NO_3^-, ClO_3^-, I^-, ClO_4^-, and SCN^-. Kosmotropes, such as F^- and SO_4^{2-}, generally decrease the solubility of proteins and other solutes (salting out), and while the rules governing these phenomena are not fully understood, these ions are said to enhance the hydrophobic effect.[3] Chaotropes, such as ClO_4^- and SCN^-, generally increase the solubility of proteins and other solutes (salting in), cause the denaturation of proteins at high concentrations and diminish the hydrophobic effect. The affinity of chaotropic anions towards hydrophobic surfaces allow them to compete with the interactions between hydrophobic solutes.[4] Chaotropic anions form complexes with hosts that have hydrophobic cavities, such as cavitands and cyclodextrins (CDs).[5] The binding of chaotropic anions to hydrophobic solutes is associated with a certain thermodynamic fingerprint — a favorable enthalpy and an entropic penalty — which has recently been characterized as a generic driving force under the term "the chaotropic effect" by Nau and coworkers.[6,7] These authors stressed that the chaotropic effect should be distinguished from the classical hydrophobic effect, where the thermodynamic signature is a favorable entropic term. Understanding the different influences of kosmotropes, chaotropes and hydrophobes on self-assembly processes in aqueous solution, both as modulators of solvent effects[3,8–10] and as recognition motifs,[7,11,12] is key to the successful design of supramolecular systems. In this work, we examine how the interplay between kosmotropic, chaotropic and hydrophobic effects modulates the behavior of an enzyme-mediated dynamic system of cyclodextrins.

Dynamic combinatorial chemistry (DCC) is a powerful method to explore molecular self-assembly and the templated synthesis of complex molecular architectures using reversible bond formation under thermodynamic control.[13–17] We have previously described how dynamic combinatorial libraries (DCLs) of interconverting cyclodextrins can be generated by employing an enzyme that enables reversible transglycosylation.[18–20] Cyclodextrins are macrocycles formed from α-1,4-linked glucopyranose units. The native cyclodextrins, α-CD, β-CD and γ-CD, with six, seven, and eight glucopyranose units, respectively, exhibit truncated cone-like structures and are widely utilized hosts for the encapsulation of hydrophobic molecules in the foods, pharmaceutical and cosmetics industries.[21–23] Cyclodextrin glucanotransferase (CGTase) catalyzes both fast, reversible inter- and intramolecular transglycosylation and slow hydrolysis of α (1–4)-glycosidic bonds.[24–26] Exposure of an α-1,4-glucan source to CGTase, therefore, generates a dynamic mixture of linear α-1,4-glucans (maltooligosaccharides) and cyclic α-1,4-glucans (cyclodextrins).[18] As α-CD, β-CD and γ-CD are intrinsically more stable than their linear counterparts, a complex dynamic system is formed in which α-CD, β-CD and γ-CD are kinetically trapped and transiently form as the primary products before being eventually converted to glucose.

We previously showed that even though α-CD, β-CD and γ-CD form out-of-equilibrium in this enzyme-mediated dynamic system, they exist in a subsystem that operates under *pseudo*-thermodynamic control.[18] The distribution of products formed can be controlled by addition

of a template that binds selectively to specific cyclodextrins. We were able to produce α-CD, β-CD or γ-CD with 99% selectivity using sodium dodecyl sulfate, 1-adamantanecarboxylic acid and sodium tetraphenylborate as templates.[18,19] While investigating different reaction conditions, it was also found that CGTase could not only tolerate very high concentrations of $NaNO_3$ (up to 7.5 M) but that the presence of $NaNO_3$ altered the distribution of cyclodextrins in the DCL that was formed. These results encouraged us to explore how the addition of a range of sodium salts in high concentrations would influence the cyclodextrin distribution in our dynamic system. Here, we present how the addition of salts to CGTase-mediated DCLs of cyclodextrins can either enhance the template effects of added guests, in the case of kosmotropes, or lead to direct template effects, in the case of chaotropes (Figure 2.1).

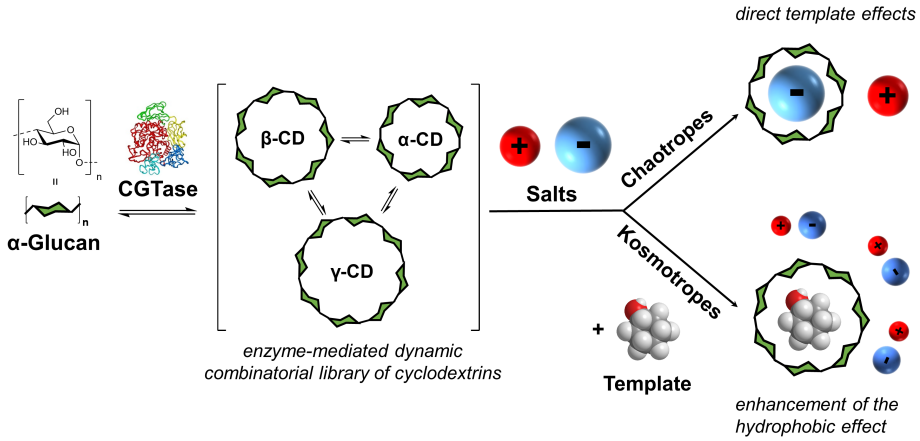


Figure 2.1. Concept of enzyme-mediated dynamic combinatorial chemistry with α-1,4-glucans and the effect of added salts. Cyclodextrin glucanotransferase (CGTase) acts on α-1,4-glucans to generate a dynamic mixture of cyclodextrins. Addition of salts leads to either direct template effects (with chaotropic salts) or to the enhancement of the template effect of an added guest (with kosmotropic salts).

2.2. Results and discussion

2.2.1. The Influence of a Series of Anions on Cyclodextrin DCLs

To explore the influence of anions on CGTase-mediated DCLs of cyclodextrins, we examined a series of DCLs prepared in the presence of different sodium salts at concentrations up to 4 M. The following series of anions was investigated, ranked according to the Hofmeister series: $HPO_4^{2-}/H_2PO_4^- > Cl^- > NO_3^- > Br^- > ClO_4^- > SCN^-$. The libraries were prepared by addition of CGTase (50 µl stock solution per ml reaction mixture) to solutions of α-CD (10 mg/ml) with the desired sodium salts at various concentrations up to 4 M concentration in phosphate buffer (50 mM, pH 7.5). Despite the lower solubility of sodium phosphates, we chose also to include phosphate buffer in this study albeit at a maximum of 0.45 M. The library compositions were monitored as the dynamic system evolved over time using hydrophilic interaction liquid chromatography (HILIC) with an evaporative light scattering detector (ELSD), which enabled the separation and quantification of the chromophore-lacking glucan mixtures. The influence of varying concentrations of different salts on the equilibrium cyclodextrin distribution and the time taken to reach this steady distribution is summarized for all anions tested in Table 2.1.

Table 2.1. Summary of results (relative CD yield at *pseudo*-equilibrium and time to *pseudo*-equilibrium) for CGTase-mediated Dynamic Combinatorial Libraries (DCLs) of cyclodextrins in the presence of different sodium salts.

Entry #	Salt [a]	Salt Concentration (M)	Time to *pseudo*-equilibrium (h) [b]	Relative CD yield at *pseudo*-equilibrium (% by weight) [c]		
				α-CD	β-CD	γ-CD
1	no salt		1	32	57	11
2	HPO_4^{2-} / $H_2PO_4^-$	0.45	1	36	54	10
3	NaCl	1	1	37	54	10
4	NaCl	2	2	37	54	9
5	NaCl	3	2	37	53	9
6	NaCl	4	4	40	52	8
7	$NaNO_3$	1	1	48	45	7
8	$NaNO_3$	2	1.5	49	45	6
9	$NaNO_3$	3	2	51	43	6
10	$NaNO_3$	4	2	55	40	5
11	NaBr	1	1	41	50	8
12	NaBr	2	1.5	44	49	7
13	NaBr	3	2.5	48	46	6
14	NaBr	4	4	54	40	5
15	$NaClO_4$	1	3	62	36	2
16	$NaClO_4$	2	6	65	34	1
17	$NaClO_4$	3	[d]	—	—	—
18	NaSCN	1	2	72	26	2
19	NaSCN	2	8	75	23	2
20	NaSCN	3	[d]	—	—	—

[a] Conditions: α-CD (10 mg/mL) in sodium phosphate buffer (50 mM, pH 7.5) treated with CGTase at room temperature. [b] Estimated time (to the nearest half hour) until a steady distribution of CDs was obtained. [c] Values taken from a single data point after a steady distribution of CDs was obtained. (Instrumental uncertainty of about ±2 % points in the relative CD yields). [d] Zero or close to zero enzyme activity, presumably due to enzyme denaturation.

Figure 2.2 shows representative data obtained for DCLs of cyclodextrins prepared in the presence of increasing amounts of NaBr. Figure 2.2A depicts chromatograms showing the distributions of α-1,4-glucan products formed after *pseudo*-equilibrium is obtained (2–6 h) in the absence and presence of different concentrations of NaBr. It is immediately evident that the relative concentration of α-CD increased at the expense of β-CD and γ-CD in the presence of increasing concentrations of NaBr. Amplifications of α-CD were seen for all the anions tested, but the magnitude of the effect was anion-dependent (Table 2.1). The evolution of each DCL was monitored over time (Supporting Figures S2.1–S2.6). For all DCLs, the CD yield decreased gradually overtime, due to background hydrolysis and the consequent build-up of short linear α-1,4-glucan and glucose. For the DCL without salt, a steady distribution of α-CD, β-CD and γ-CD was obtained after approximately one hour, when a *pseudo*-thermodynamic

equilibrium of cyclodextrins was reached. At this point >90% of the glucan material was still present as cyclodextrins. With increasing concentrations of NaBr, the evolution of the DCL became slower, and it took up to approximately four hours (with 4 M NaBr) to reach a steady distribution of cyclodextrins (Figure 2.2C). In all cases, addition of anions led to a slower evolution of the dynamic enzymatic system (Supporting Figures S2.1–S2.6). With the chaotropic anions ClO_4^- and SCN^- (Table 2.1, entries 15–20) this retardation was quite significant at 1–2 M concentrations, and at concentrations of 3 M and higher, the activity of the enzyme was zero or close to zero within just 30 min of being exposed to the salt solutions (Supporting Figure S2.8), presumably due to denaturation of the enzyme under these conditions.

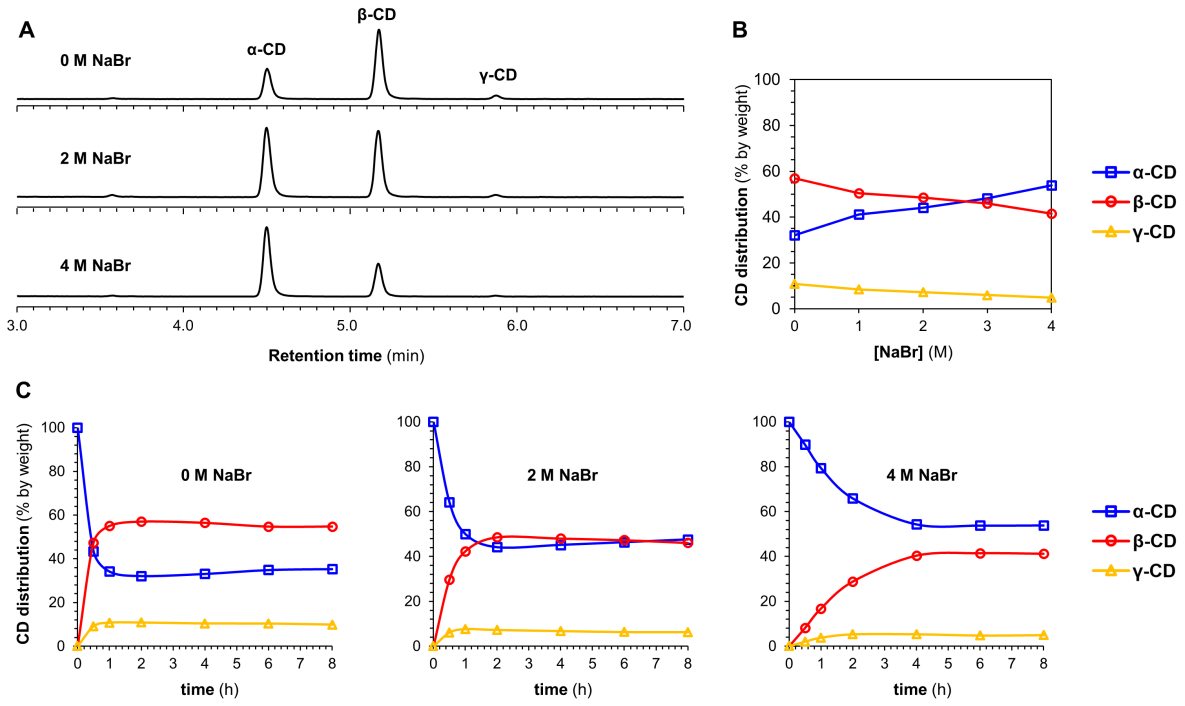


Figure 2.2. The effect of NaBr on CGTase-mediated DCLs. Conditions: α-CD (10 mg/ml) in sodium phosphate buffer (50 mM, pH 7.5) treated with CGTase at room temperature. (A) Chromatograms (HPLC-ELS) showing the *pseudo*-equilibrium distribution of CDs produced in the absence or presence of NaBr (2 and 4 M). The chromatograms show data obtained after 2 h (0 and 2 M NaBr) or 6 h (4 M NaBr). (B) *Pseudo*-equilibrium distribution of CDs as a function of NaBr concentration. (C) CD distribution monitored over time in the absence or presence of NaBr (2 and 4 M) (lines are included only to guide the eye).

The equilibrium cyclodextrin distributions obtained in the presence of different anions at 2 M concentration are summarized in Figure 2.3. Following a largely systematic trend, the relative yield of α-CD increases upon moving from the most kosmotropic anions (HPO_4^{2-}/$H_2PO_4^-$,Cl^-, minor changes) to the most chaotropic anions (ClO_4^-, SCN^-, large changes). These results suggest that all the tested anions have a specific, albeit weak affinity for α-CD and function as templates in the enzyme-mediated DCL of cyclodextrins. Selected binding constants for Cl^-, NO_3^-, Br^-, SCN^-, and ClO_4^- interacting with α-CD, β-CD and γ-CD have previously been reported (Table 2.2). These binding constants were determined using a variety of techniques: ^{1}H-NMR spectroscopy,[27] conductance,[28] potentiometry,[29]

spectrophotometry,[30] volatilization[31] and isothermal titration calorimetry.[5] While the numeric values of the binding constants vary somewhat depending on the method, the trends are certainly clear. In the kosmotropic end of the Hofmeister series, the binding of Cl^- to α-CD was found to be negligible in most cases, which corresponds well with our data, where only minor changes in the cyclodextrin distribution occur upon addition of Cl^- (Table 2.1, entries 3–6). The minor increase in the relative yield of α-CD, from 32 to 37%, does, however, indicate that there could be a very weak binding between α-CD and Cl^-, as found in some cases in the literature.[28] Next in the series, both NO_3^- and Br^- bind to α-CD with small and similar binding constants (1–4 M^{-1}), which matches the small but significant amplification of α-CD observed with these anions (Table 2.1, entries 7–14). The chaotropic anions SCN^- and ClO_4^- have significantly higher affinities for α-CD (16–46 M^{-1}) and led to much larger changes in the cyclodextrin distribution (Table 2.1, entries 15–20). The fact that the amplification of α-CD is larger with SCN^- than ClO_4^- can be explained by the relatively higher competing affinity of ClO_4^- for β-CD, as the distribution obtained in a DCL is influenced by the binding interaction of the template with all members of a library. It is noteworthy that each of the small anions tested amplified and bound most strongly to α-CD, whereas we have previously observed the amplification of larger CDs with 8, 9, and 10 glucopyranose units in the presence of the large superchaotropic anion $B_{12}I_{12}^{2-}$ (diameter 11.7 Å).[18] There is clearly, thus, a relationship between the size of the anion (Table 2.2, column 3) and the preferential formation of the CD(s) with a suitable size cavity. Overall, we found that the addition of anions in high concentrations to the CGTase-mediated DCLs of cyclodextrins leads to changes characteristic of template effects, and remarkably, the system remains dynamic even at 2 M concentrations of the denaturing salts NaSCN and $NaClO_4$.

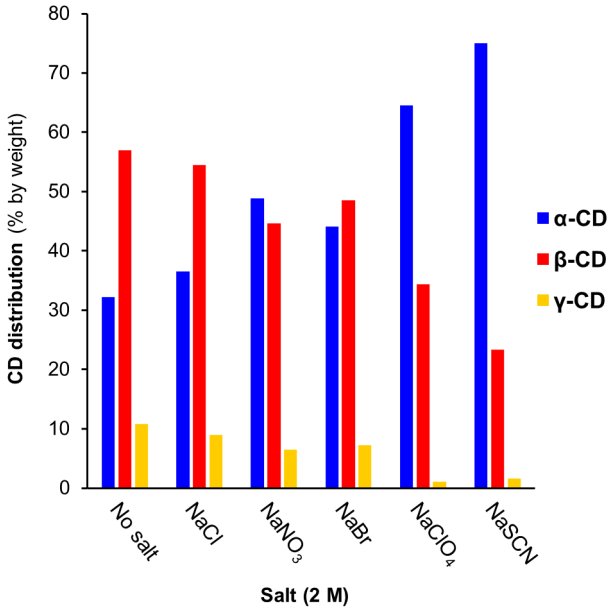


Figure 2.3. The effect of different anions (at 2 M concentration) on *pseudo*-equilibrium CD distributions obtained in CGTase-mediated DCLs. Conditions: α-CD (10 mg/ml) in sodium phosphate buffer (50 mM, pH 7.5) treated with CGTase at room temperature.

Table 2.2. Binding constants (K_a) for anions with α-CD, β-CD and γ-CD reported by various authors and the size (diameter, *d*) of the anions.

host	guest	*d* (Å) [c]	K_a (M^{-1}) [a]					
			I	**II**	**III**	**IV**	**V**	**VI**
α-CD	Cl^-	3.6	—	<1	—	~0	3	no binding
	NO_3^-	4.0	1.4	1.4	2.31	—	4	[b]
	Br^-	3.9	1.6	3.5	0.96	—	—	—
	SCN^-	4.3	28.4	18.7	33.5	—	—	16
	ClO_4^-	4.8	33.0	28.9	45.8	—	35	23
β-CD	NO_3^-	—	—	—	—	0.2	—	—
	Br^-	—	—	—	0.45	1.1	—	—
	SCN^-	—	9.2	9.9	9.2	5.7	—	—
	ClO_4^-	—	13.6	—	—	9.0	—	—
γ-CD	SCN^-	—	4.1	—	—	—	—	—

[a] Binding constants for anions (as Na^+ or K^+ salts) to α, β and γ-CD measured in H_2O or D_2O at 20 °C or 25 °C with various techniques: **I)** ^{1}H-NMR spectroscopy[27]; **II)** Conductance[28]; **III)** Potentiometry[29]; **IV)** Spectrophotometry[30]; **V)** Volatilization[31]; **VI)** Isothermal titration calorimetry[5]. [b] Binding too weak to determine a binding constant. [c] From reference.[32]

2.2.2. DCL Simulation to Support Templating Effects of Chaotropes

To gain further support for our conclusion that chaotropic anions influence the production of specific cyclodextrins in CGTase-mediated dynamic systems through direct template effects, we sought to simulate the DCL generated in the presence of increasing concentrations of NaSCN. *DCLSim* is a software developed in the Otto group[33] that enables the prediction of product distributions in templated DCLs operating under thermodynamic control. To simulate the DCLs, binding constants for the interaction of the template with each library member (α-CD, β-CD and γ-CD) is required, and this was available for SCN^- (Table 2.2, column I).[27] The relative formation constants K_f for α-CD, β-CD and γ-CD are also needed for the simulation and these could be calculated from the relative concentrations of α-CD, β-CD and γ-CD generated at equilibrium in the DCL without salt. We simulated DCLs of cyclodextrins with SCN^- at 0, 1 and 2 M concentrations. The results, summarized in Figure 2.4, show that the simulations correlate well with the experimental results, supporting the conclusion that chaotropic anions function as templates in this system, and at high concentrations can strongly influence the product selectivity in CGTase-mediated cyclodextrin synthesis.

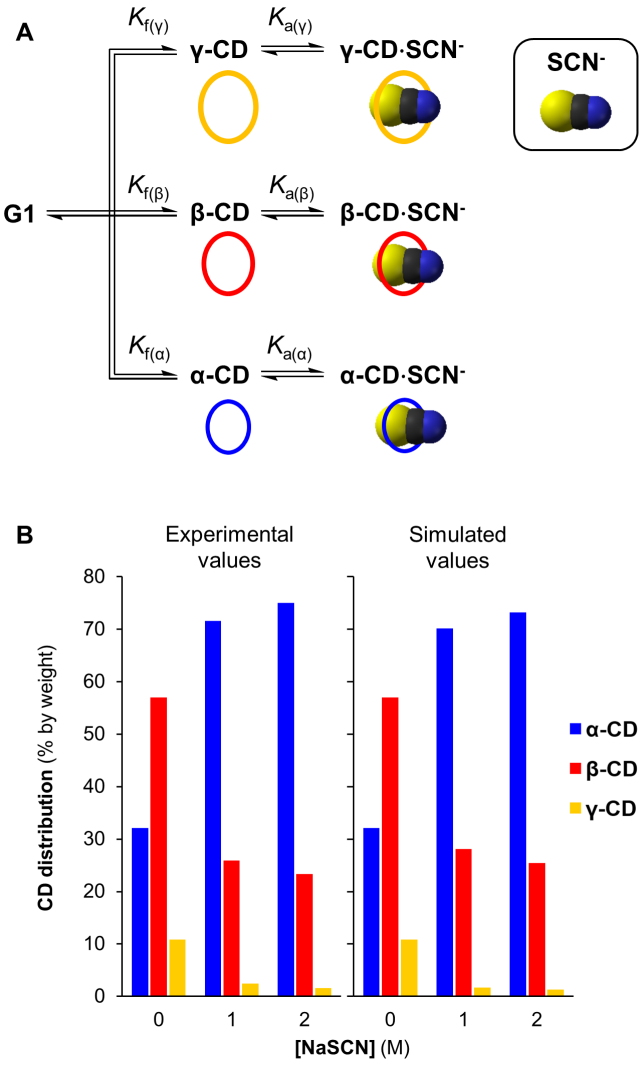


Figure 2.4. *DCLSim* simulations of DCLs of cyclodextrins with and without the NaSCN. (A) Model employed to simulate DCLs of α-, β-, and γ-CD in the presence of NaSCN. (B) Comparison between the simulated and experimentally determined CD distributions with NaSCN (0–2 M).

2.2.3. Increasing the Template Effect of Hydrophobic Guests by Addition of a Kosmotropic Salt

As kosmotropic salts can lead to stronger binding between cyclodextrins and hydrophobic guests,[30,34] we sought to investigate whether we could enhance the template effect of hydrophobic templates in this dynamic cyclodextrin system by using high concentrations of NaCl. For these experiments, we chose to employ cyclohexanol (template **1**) and cyclohexanecarboxylate (template **2**) as templates, as both guests have a relatively low affinity for β-CD in the absence of salts (ca. 700 and ca. 300 M^{-1}, respectively),[35,36] thus giving room for a possible improvement in templating effect upon addition of NaCl. A series of CGTase-mediated DCLs were set-up starting from α-CD (10 mg/ml) with template (10 mM) in phosphate buffer (50 mM, pH 7.5) with 0–3 M NaCl (Supporting Figure S2.7). It was found that the addition of NaCl in increasing concentrations up to 3 M did in fact lead to a moderate increase in the selectivity for β-CD obtained in the presence of each template (from 80 to 87%

for template 1 and from 75 to 82% for template 2) (Figure 2.5). It is worth noting that increasing the concentration of NaCl from 0 to 3 M in the absence of template also changes the cyclodextrin distribution, but in the opposite direction, with a decrease in the relative yield of β-CD from 57 to 53%, thus counteracting the observed increase in the template effect with 1 and 2. Accordingly, the actual kosmotrope-induced increase in the template effect is potentially larger than the effect observed here.

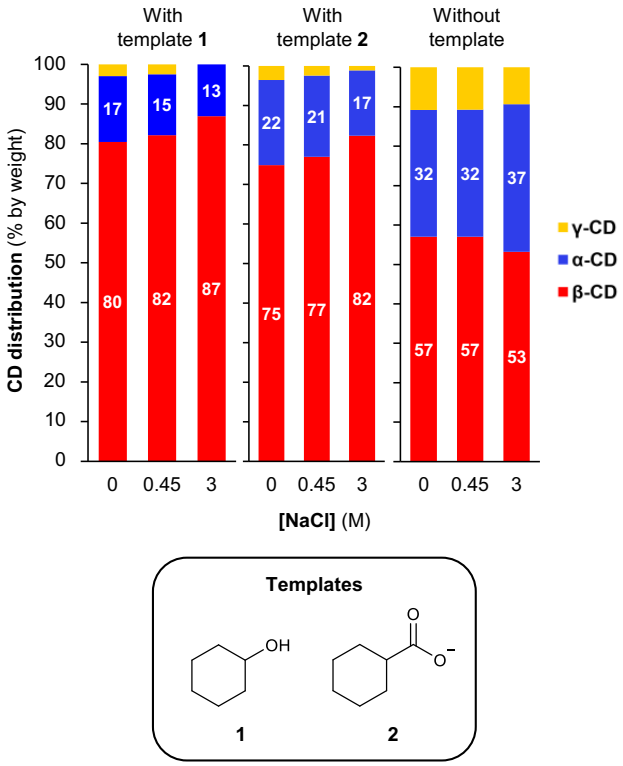


Figure 2.5. Enhancing the template effect of cyclohexanol and cyclohexane carboxylate by the addition of NaCl. CD distributions from DCLs generated in the absence or presence of template (10 mM) and in the absence or presence of NaCl (up to 3 M). Conditions: α-CD (10 mg/ml) in sodium phosphate buffer (50 mM, pH 7.5) treated with CGTase at room temperature.

2.3. Conclusion

In this work, we have systematically explored the influence of different anions on the behavior of CGTase-mediated dynamic combinatorial libraries of cyclodextrins and distinguished two key effects. One the one hand, we observed that NaCl (which lies towards the kosmotropic end of the Hofmeister series) could subtly enhance the templating effects of hydrophobic guest added to the DCLs. On the other hand, we observed direct templating effects due to specific interactions between chaotropes and cyclodextrins. These interactions are very weak ($K_a < 50$ M^{-1}), but, when present in high concentrations (2–4 M), chaotropes can, nevertheless, cause significant changes in the distribution of α-CD, β-CD and γ-CD generated. For example, addition of 2 M NaSCN led to a shift in selectivity from 32% α-CD to 75% α-CD. The observed amplifications of α-CD correlated well with reported binding constants, and simulation of DCLs templated with NaSCN matched well to the experimental results, which is further evidence that the observed effects are due to specific anion-cyclodextrin binding interactions rather than the global influence of the high salt concentration on the bulk solvent. In fact, this dynamic system potentially provides a new method to identify very weak binding of guests to CDs, which would be very difficult to detect otherwise. Finally, we note that the CGTase used in these experiments was remarkably stable at high salt concentrations. Our study showcases how the interplay between kosmotropes, chaotropes and hydrophobes in dynamic supramolecular systems can be utilized to alter the outcome of these systems in predictable and systematic ways.

2.4. Author Contributions

AE performed the experiments. All authors contributed to the experiment design. DL and SB conceived the concept and supervised the research. AE prepared figures and the original draft of the article. All authors contributed to the preparation and editing of the article.

2.5. Experimental

2.5.1. Materials

All chemicals and solvents of HPLC (high performance liquid chromatography) grade were obtained from commercial suppliers and used as received. High purity water used both in reactions and chromatographic analysis was obtained using a Merck Millipore Synergy UV water purification system. Colorless Corning CoStar 0.65 ml centrifuge tubes were used for enzymatic reactions and sample preparation (dilution and centrifugation), while colorless 2 ml glass vials with PTFE-lined screw-cap septa and 0.2 ml glass inserts were used for short-term sample storage and injection on HPLC equipment. A stock solution of the enzyme CGTase derived from *Bacillus macerans* was received as a kind gift from Amano Enzyme, Inc., Nagoya, Japan. The stock solution was stored at 5°C and used as received.

2.5.2. Instrumentation

Chromatographic analysis was performed on a Thermo Scientific Dionex Ultimate 3000 HPLC (ultra-high pressure) system equipped with a Waters Acquity UPLC BEH Amide 1.7 µm 2.1 × 150 mm column maintained at 30° C, and an autosampler module maintained at 20° C. Detection was carried out using an Agilent Technologies 1,260 Infinity ELSD (evaporative light scattering detector), operating with the evaporator at 90° C, nebulizer at 70° C, and a N_2 gas flow of 1.0 L/min. The ELSD enables the detection of the chromophore-lacking oligosaccharides. Calibration curves from 0.01 mg/ml to 10 mg/ml for α-, β- and γ-CD and linear α-1,4-glucans up to maltooctaose were used to correct for differences in the ELS detector response for different oligosaccharides. The calibrations were based on masses injected (0.018–3.66 µg) and the resulting response curves were fitted to a simple power equation $M = kA^p$ (where M is the injected mass of compound, A is the area under the peak in the chromatogram and k and p are fitted parameters) using non-linear curve fitting (in OriginPro 2018b from OriginLab Corp.) See recent paper for details.[20] The gradient profile for HPLC runs was a linear gradient from 75% acetonitrile in water to 55% acetonitrile in water over 8 min with a flow rate of 0.6 ml/min. Both eluents contained 0.1% formic acid by volume.

2.5.3. Enzymatic Reactions and Analysis

Reaction mixtures with the desired concentrations of salts were prepared by mixing appropriate amounts of two types of stock solutions in buffered water (50 mM sodium phosphate at pH 7.5): 1) a solution containing α-CD (10 mg/ml) and various salts (4.0 M); and 2) a solution containing α-CD (10 mg/ml). In templated experiments, cyclohexanol or cyclohexanecarboxylic acid was dissolved in these resulting mixtures at a concentration of 10 mM. For the experiments with a higher concentration of sodium phosphate buffer, a stock solution of α-CD (10 mg/ml) in 0.45 M sodium phosphate buffer at pH 7.5 was prepared. The starting mixtures containing salt, α-CD and template (if any) in buffer were then aliquoted (165–365 µl) into reaction vessels and kept at ambient temperature. All reactions were then initiated by adding CGTase stock solution (50 µl per ml of starting mixture) to the starting mixtures followed by thorough mixing. The reactions were then monitored at various time

points: Aliquots for analysis (4–5 µl) were taken out and rapidly diluted (31 fold) in a 1% trifluoroacetic acid (TFA) solution in 3:1 acetonitrile/water with 10 mM ammonium chloride to stop the enzymatic reaction. For experiments with NaCl, aliquots for analysis (20 µl) were taken out and rapidly diluted (six fold) in a 1% TFA solution in water. Samples were then centrifuged (10,000 RPM for 4 min) to prevent column blockage by insoluble species such as enzyme and salts, and the top fractions (leaving behind 20 µl) were then transferred to 2 ml glass vials with 0.2 ml glass inserts, kept at 20° C and then injected on the HPLC instrument within 48 h. Injection volumes were 10 µl or 2 µl (for 31-fold and 6-fold diluted samples, respectively). Peaks in the chromatograms corresponding to α-CD, β-CD and γ-CD and linear α-1,4-glucans up to maltooctaose were identified by comparison with authentic samples obtained from commercial suppliers.

2.5.4. Simulations of Dynamic Combinatorial Libraries

Simulations of dynamic combinatorial libraries (DCLs) with and without the anion SCN^- were carried out using the *DCLSim* software developed in the Otto group[33] and kindly made available to us. The program requires input of the concentration of the building block (glucopyranose units in this case), the composition of the oligomers (library members) formed in the DCL (α-CD, β-CD and γ-CD, with six, seven or eight glucose units in this case), the relative formation constants K_f of the library members (determined from the *pseudo*-equilibrium composition of α-, β-, and γ-CD in an untemplated library), the binding constants (K_a) of each library member to the template (SCN^-) and the concentration of the template. Details about how the relative formation constants K_f of α-, β-, and γ-CD were calculated can be found in the Supporting Info. The binding constants used were obtained from a study carried out by Tokunaga and coworkers, where the authors used ^{1}H-NMR spectroscopy to determine binding constants between inorganic anions and cyclodextrins.[27]

2.6. Bibliography

[1] A. Erichsen, D. Larsen, S. R. Beeren, *Front. Chem.* **2021**, *9*, 1–8.

[2] F. Hofmeister, *Arch. für Exp. Pathol. und Pharmakologie* **1888**, *25*, 1–30.

[3] B. C. Gibb, *Isr. J. Chem.* **2011**, *51*, 798–806.

[4] C. L. D. Gibb, B. C. Gibb, *J. Am. Chem. Soc.* **2011**, *133*, 7344–7347.

[5] M. R. Sullivan, W. Yao, D. Tang, H. S. Ashbaugh, B. C. Gibb, *J. Phys. Chem. B* **2018**, *122*, 1702–1713.

[6] K. I. Assaf, M. S. Ural, F. Pan, T. Georgiev, S. Simova, K. Rissanen, D. Gabel, W. M. Nau, *Angew. Chem. Int. Ed.* **2015**, *54*, 6852–6856.

[7] K. I. Assaf, W. M. Nau, *Angew. Chem. Int. Ed.* **2018**, *57*, 13968–13981.

[8] W. Kunz, P. Lo Nostro, B. W. Ninham, *Curr. Opin. Colloid Interface Sci.* **2004**, *9*, 1–18.

[9] B. C. Gibb, *Nat. Chem.* **2019**, *11*, 963–965.

[10] N. F. A. Van Der Vegt, D. Nayar, *J. Phys. Chem. B* **2017**, *121*, 9986–9998.

[11] N. Busschaert, C. Caltagirone, W. Van Rossom, P. A. Gale, *Chem. Rev.* **2015**, *115*, 8038–8155.

[12] M. B. Hillyer, B. C. Gibb, *Annu. Rev. Phys. Chem.* **2016**, *67*, 307–329.

[13] S. Otto, R. L. E. Furlan, J. K. M. Sanders, *Science* **2002**, *297*, 590–593.

[14] P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J. L. Wietor, J. K. M. Sanders, S. Otto, *Chem. Rev.* **2006**, *106*, 3652–3711.

[15] J. M. Lehn, *Chem. Soc. Rev.* **2007**, *36*, 151–160.

[16] N. Ponnuswamy, F. B. L. Cougnon, J. M. Clough, G. D. Pantoş, J. K. M. Sanders, *Science* **2012**, *338*, 783–785.

[17] J. Li, P. Nowak, S. Otto, *J. Am. Chem. Soc.* **2013**, *135*, 9222–9239.

[18] D. Larsen, S. R. Beeren, *Chem. Sci.* **2019**, *10*, 9981–9987.

[19] D. Larsen, S. R. Beeren, *Chem. Eur. J.* **2020**, *26*, 11032–11038.

[20] D. Larsen, S. R. Beeren, *Chem. Commun.* **2021**, *57*, 2503–2506.

[21] A. R. Hedges, *Chem. Rev.* **1998**, *98*, 2035–2044.

[22] E. M. M. Del Valle, *Process Biochem.* **2004**, *39*, 1033–1046.

[23] N. Sharma, A. Baldi, *Drug Deliv.* **2016**, *23*, 739–757.

[24] Y. Terada, M. Yanase, H. Takata, T. Takaha, S. Okada, *J. Biol. Chem.* **1997**, *272*, 15729–15733.

[25] Y. B. Tewari, R. N. Goldberg, M. Sato, *Carbohydr. Res.* **1997**, *301*, 11–22.

[26] J. C. M. Uitdehaag, B. A. Van Der Veen, L. Dijkhuizen, B. W. Dijkstra, *Enzyme Microb. Technol.* **2002**, *30*, 295–304.

[27] Y. Matsui, M. Ono, S. Tokunaga, *Bull. Chem. Soc. Jpn.* **1997**, *70*, 535–541.

[28] J. F. Wojcik, R. P. Rohrbach, *J. Phys. Chem.* **1975**, *79*, 2251–2253.

[29] R. I. Gelb, L. M. Schwartz, M. Radeos, D. A. Laufer, *J. Phys. Chem.* **1983**, *87*, 3349–3354.

[30] Á. Buvári, L. Barcza, *Inorganica Chim. Acta* **1979**, *33*, 179–180.

[31] I. Sanemasa, M. Fujiki, T. Deguchi, *Bull. Chem. Soc. Jpn.* **1988**, *61*, 2663–2665.

[32] Y. Marcus, *Ion Properties*, Marcel Dekker, Inc., New York, **1997**.

[33] P. T. Corbett, S. Otto, J. K. M. Sanders, *Chem. Eur. J.* **2004**, *10*, 3139–3143.

[34] R. Holm, C. Schönbeck, P. Somprasirt, P. Westh, H. Mu, *J. Incl. Phenom. Macrocycl. Chem.* **2014**, *80*, 243–251.

[35] R. I. Gelb, L. M. Schwartz, *J. Incl. Phenom. Mol. Recognit. Chem.* **1989**, *7*, 465–476.

[36] M. V. Rekharsky, F. P. Schwarz, Y. B. Tewari, R. N. Goldberg, M. Tanaka, Y. Yamashoji, *J. Phys. Chem.* **1994**, *98*, 4098–4103.

Chapter 3. Templated dynamic enzymatic synthesis of δ-CD by [n]-*pseudo*rotaxane formation

Abstract

In this chapter, a range of bolaamphiphile templates with different sizes of hydrophilic head groups and different lengths of alkyl chains were synthesized and explored for the templated enzymatic synthesis of large-ring cyclodextrins (LRCDs). One of the templates, **T1**, led to the amplification of δ-CD (CD9) in CGTase-mediated dynamic combinatorial libraries (DCLs) of CDs. Using this template **T1**, δ-CD was synthesized and isolated on a preparatory scale in unprecedented yields. The template **T1** was readily isolated from the reaction mixture and reused in subsequent syntheses. The different sizes of the head groups and the lengths of the alkyl chains of the various bolaamphiphiles led to the formation of a diverse array of different [n]-*pseudo*rotaxanes of the templates and β-, γ-, and δ-CD, as revealed by NMR spectroscopy titrations. The formation of [2]-, [3]-, and [4]-*pseudo*rotaxanes with δ-CD were observed, with complexation processes in both the fast and slow exchange regimes of the NMR chemical shift timescale. Equations to fit data extracted from NMR titrations in mixed fast/slow exchange with 1:2 and 1:3 binding were derived. The binding was thus determined to be cooperative for the second threading of **T1** through δ-CD, leading to formation of a [3]-*pseudo*rotaxane.

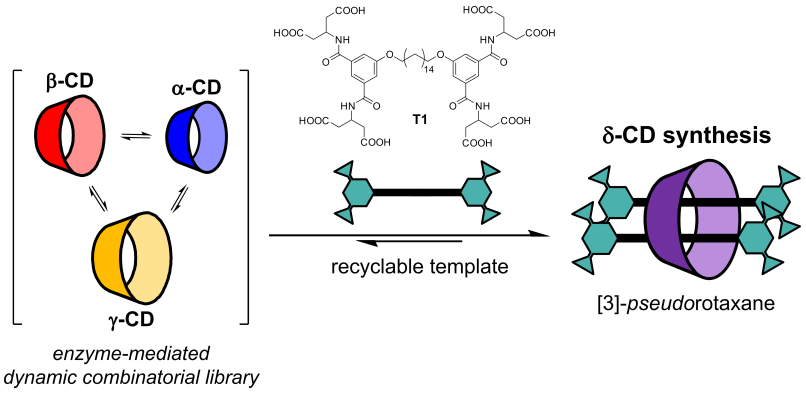


Figure 3.1. Templated synthesis of δ-CD in CGTase-mediated DCLs of CDs using template **T1**.

The equations for fitting of 1:2 and 1:3 binding in mixed fast/slow exchange were derived by Associate Professor Sophie R. Beeren

3.1. Introduction

Large-ring cyclodextrins (LRCDs),[1,2] with more than 8 glucopyranose units, are fascinating macrocycles that have remained relatively unexplored compared to the native CDs due to synthetic inaccessibility. A general feature of CDs is their water solubility and negligible toxicity,[1,3] and the small CDs have found industrial applications in many fields, such as the foods, pharmaceutical and cosmetics industries.[3–5] Similar applications have been proposed for LRCDs.[1,2] As such, efficient synthetic procedures for the production of LRCDs are of great interest.

To target the synthesis of large-ring cyclodextrins (LRCDs) using enzyme-mediated dynamic combinatorial chemistry, a template designed to selectively bind LRCDs over the native CDs was required. As highlighted in Chapter 1 (section 1.1), the use of the superchaotropic anion $B_{12}I_{12}^{2-}$ as a template did not lead to high selectivity for CD9 (δ-CD) and CD10[6] despite the significantly higher affinity of binding to these LRCDs compared to γ-CD.[7] Instead, γ-CD was obtained as the primary product due to the higher intrinsic stability of γ-CD. To achieve selective binding to LRCDs, a bolaamphiphile template design was introduced.

3.1.1. Template design

The bolaamphiphile template design employed in this chapter is shown in Figure 3.2, both as schematic illustrations and with the chemical structure of the six target templates **T1**–**T6**. In this design, alkyl chains serve as hydrophobic recognition motifs for the LRCDs. The hypothesis was, that while a single staggered alkyl chain is too small to fill out the cavity of a 'cone-shaped LRCD', the flexibilities of both the LRCDs and the alkyl chains would lead to both binding partners adopting complementary conformations, perhaps even with higher binding stoichiometries. Hydrophilic head groups of three different sizes provide water solubility and serve as bulky stoppers to prevent the binding of the native CDs to the alkyl chains, thus potentially providing the desired selectivity for the LRCDs over the native CDs. The head groups are based on 5-hydroxyisophthalic acid, and the bulkier versions of the head groups are made by introducing either one or two hydrophilic arms with carboxylic acids, giving a medium-size or large-size head group, respectively.

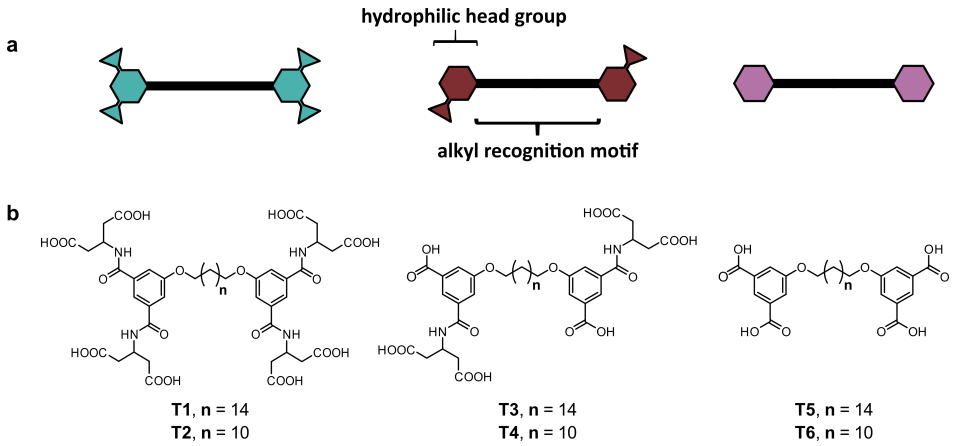
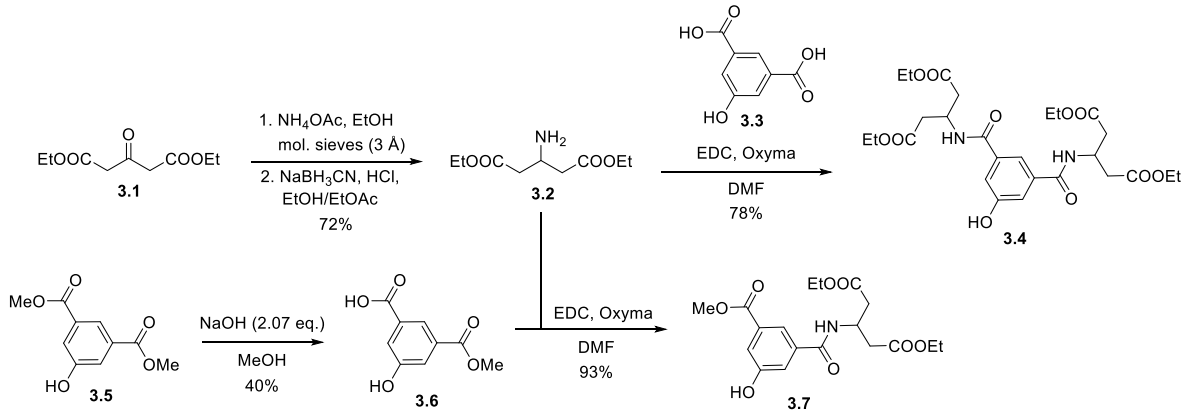


Figure 3.2. Design of bolaamphiphile templates **T1**–**T6** with (a) schematic representations and (b) chemical structures.

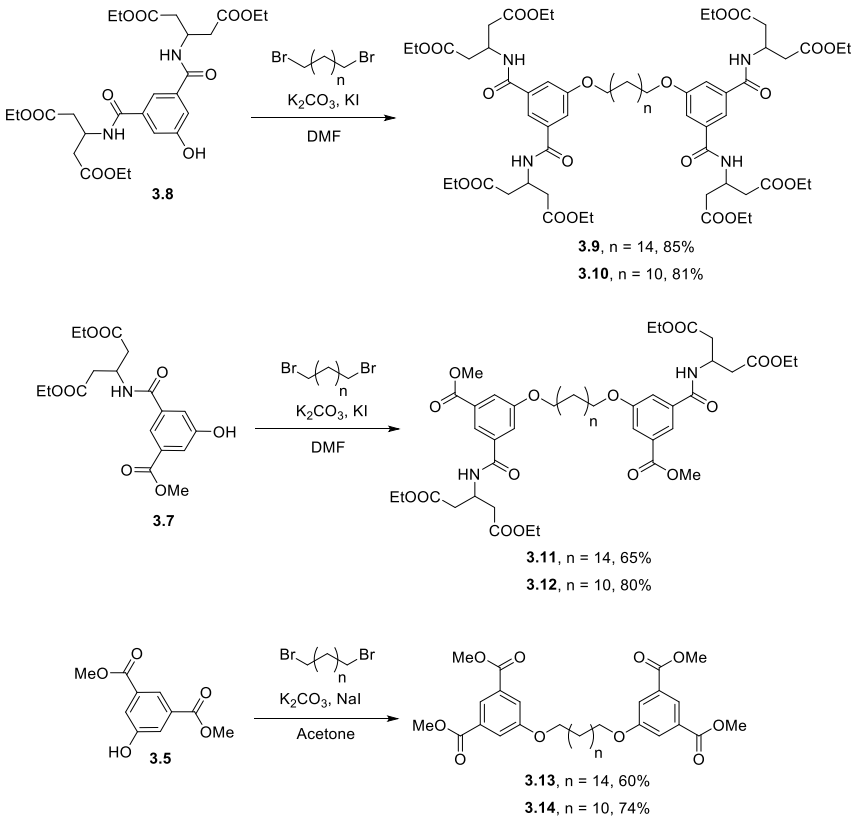
3.2. Synthesis of bolaamphiphile templates T1–T6

The approach for the synthesis of templates **T1**–**T6** involved the initial synthesis of the hydrophilic head groups with esters as protection groups for the carboxylic acids. The synthesis is shown Scheme 3.1. Adapting a published procedure,[8] commercially available diethyl 3-oxoglutarate **3.1** was reacted with ammonium acetate, in a reductive amination with sodium cyanoborohydride as reducing agent, yielding diethyl 3-aminoglutarate **3.2**. A double peptide coupling of **3.2** with 5-hydroxyisophthalic acid **3.3,** using the coupling reagents EDC and Oxyma, yielded ester-protected large-size head group **3.4**. To provide the medium-size head group **3.7**, a singly ester-protected 5-hydroxyisophthalic acid was required. By reacting commercially available dimethyl 5-hydroxyisophthalate **3.5** with two equivalents of NaOH in methanol (one equivalent to deprotonate the phenol, one to hydrolyze one of the esters), a roughly statistical mixture of the unreacted starting material **3.5**, singly hydrolyzed product **3.6** and twice hydrolyzed product **3.3** were obtained. The desired product **3.6** was isolated using dry column vacuum chromatography on silica gel. A peptide coupling with **3.6** and amine **3.2** using EDC and Oxyma then yielded ester-protected medium size head group **3.7**.

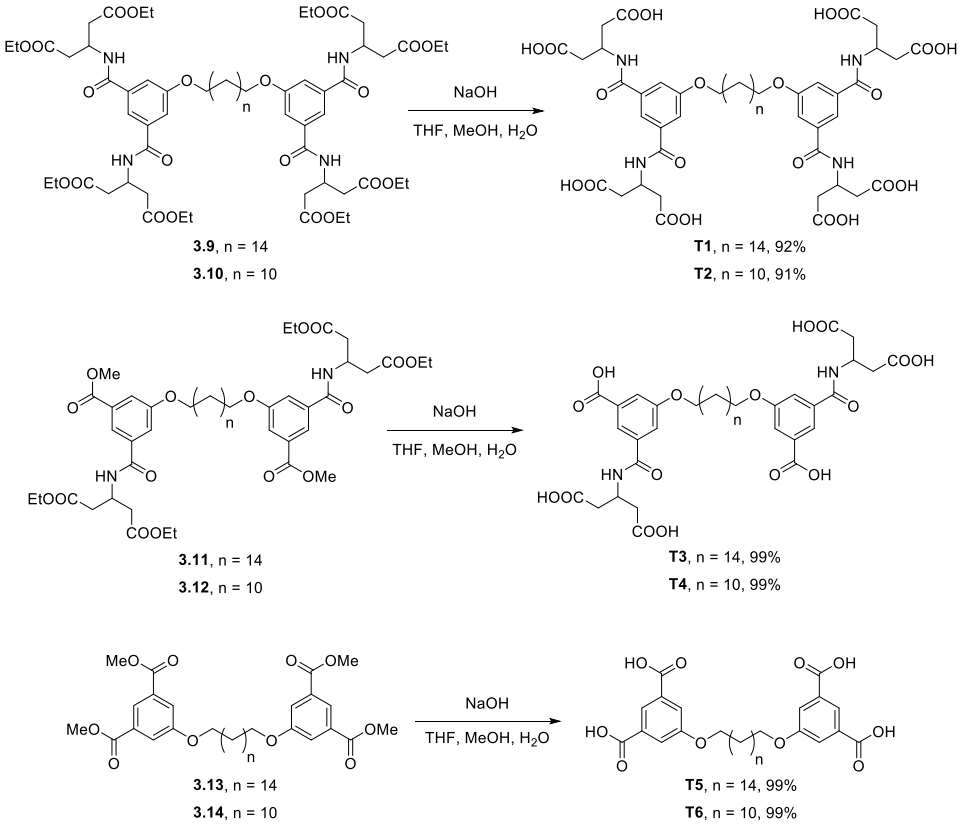


Scheme 3.1. Synthesis of ester-protected head groups **3.4** and **3.7**.

Ester protected small- (**3.5**), medium- (**3.7**) and large-size (**3.4**) head groups were then reacted twice with either 1,12-dibromododecane (commercially available) or 1,16-dibromohexadecane (synthesized from 1,16-hexadecanediol according to a published procedure)[9] as shown in Scheme 3.2, which yielded ester-protected templates **3.9**–**3.14**. Hydrolysis of the esters using NaOH in THF/methanol/water mixtures (to solubilize the starting materials) then yielded the desired templates **T1**–**T6** (Scheme 3.3). For **T3**–**T6**, evaporation of the organic solvents followed by acidification of the resulting basic aqueous solutions led to the immediate precipitation of **T3**–**T6**. For **T1** and **T2**, acidification of the resulting aqueous solutions did not lead to immediate precipitation of the products. In these two cases, the acidic aqueous solution had to be left overnight at room temperature, which then led to the slow precipitation of the products.



Scheme 3.2. Synthesis of ester-protected templates **3.9**–**3.14**.



Scheme 3.3. Hydrolysis reactions to yield templates **T1**–**T6**.

3.3. Template effects in CGTase-mediated DCLs

To investigate whether the template design of **T1**–**T6** leads to the selective binding to and production of LRCDs in CGTase-mediated dynamic combinatorial libraries (DCLs) of CDs, a series of DCLs were set up. It was found that **T5** and **T6** did not have high enough solubility (≤ 2 mM) in aqueous buffer at neutral pH for standard DCLs with CDs, which are usually carried out with α-1,4-glucan concentrations of 10 mg/ml and template concentrations of 10 mM.[6] As such, only DCLs in the presence of templates **T1**–**T4** were set up. The library compositions of the dynamic system were monitored over time using hydrophilic interaction liquid chromatography (HILIC) with evaporative light scattering detection (ELSD), which enabled the separation and detection of the chromophore-lacking glucans. The results from an untemplated DCL (10 mg/ml α-CD as starting material) are shown in Figure 3.3, with (a) selected chromatograms obtained from the reaction and (b) the distribution of CDs as a function of time. Here, a stable *pseudo*equilibrium distribution of α-, β- and γ-CD was obtained within the first hour of reaction which remained stable for several hours. At the same time, the total yield of CDs was decreasing, due to background hydrolysis, where the CDs were converted into linear α-1,4-glucans (**G2**, **G3**, etc.) and ultimately glucose (**G1**). In an untemplated library, α-CD and β-CD were the primary products after 4 hours, with only a small amount of γ-CD produced. Libraries with **T1**–**T4** were then set up with α-CD as the starting material (10 mg/ml), templates **T1**–**T4** (10 mM), and CGTase in sodium phosphate buffer (0.2 M, pH 7.5). Selected chromatograms from the DCLs are shown in Figure 3.4, while the distribution of products, as a function of time, are shown in Figure 3.5.

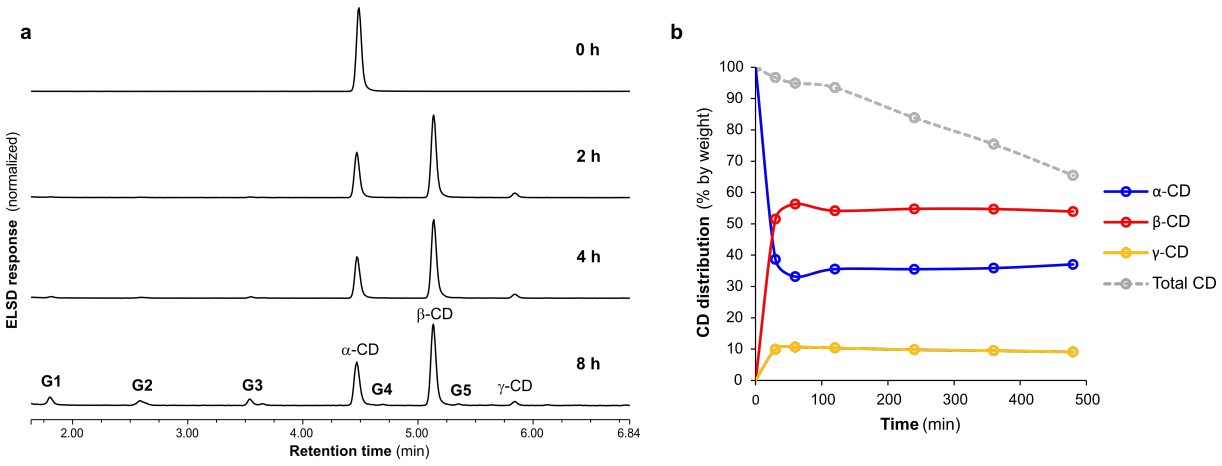


Figure 3.3. Untemplated DCL formed when α-CD (10 mg/ml) is treated with commercial stock CGTase (50 µl per ml reaction mixture) in sodium phosphate buffer (50 mM, pH 7.5). Reaction mixture diluted in 3:1 acetonitrile/water denaturing mixture before injection on HPLC. (a) HPLC-ELSD chromatograms from the reaction with time points as indicated on figure. (b) CD distributions (% by weight) and total fraction of CDs (% by weight out of all glucans in the mixture) as a function of time.

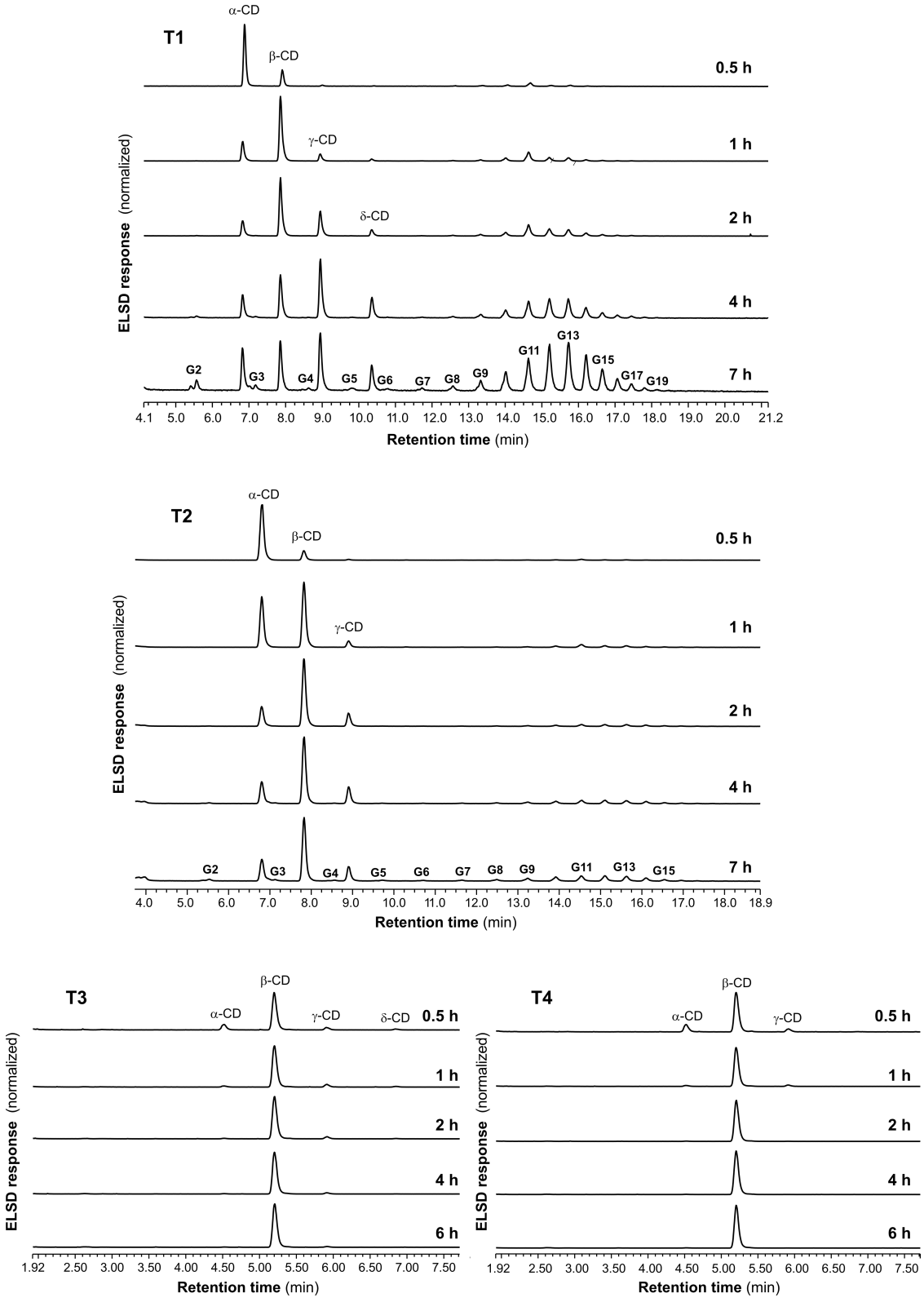


Figure 3.4. HPLC-ELSD chromatograms from templated enzymatic reactions with α-CD (10 mg/ml) and templates **T1**–**T4** (10 mM) in sodium phosphate buffer (0.2 M, pH 7.5). For reactions with **T1** and **T2**, glycerol-free CGTase (65 µl per ml reaction mixture) was used and aliquots of the reactions were diluted in aqueous denaturing mixture before injection on HPLC. For reactions with **T3** and **T4**, commercial CGTase (50 µl per ml reaction mixture) was used and aliquots from the reaction were diluted in in 3:1 acetonitrile/water denaturing mixture before injection on HPLC.

In the DCL in the presence of **T1** (C_{16} alkyl chain, large head group), remarkably, a significant production of the LRCD δ-CD was seen. In contrast, in the DCL in the presence of **T2** (C_{12} alkyl chain, large head group) only a small amplification of γ-CD and a very small amplification of δ-CD was seen in comparison with the untemplated library. In addition, in the DCLs in the presence of either **T1** or **T2**, it was surprising to see the amplification of linear α-1,4-glucans with 8 to 20 glucose units (**G8**–**G20**), which had not previously been observed in CGTase-mediated DCLs of CDs in the Beeren group.[6,10–14] In the DCL with **T1**, the concentration of δ-CD built up slowly (Figure 3.5a and b), and it took roughly 2–4 hours before a somewhat stable distribution of CDs was obtained. Concurrently, the total yield of CDs was decreasing significantly as the fraction of both shorter (**G2**–**G7**) and longer (**G8**–**G20**) linear α-1,4-glucans was increasing in the library.

The standard procedure for the sampling and analysis of DCLs of CDs in the Beeren group involves the dilution of aliquots from the reaction in a 3:1 mixture of acetonitrile and water before the injection on the HPLC equipment. Using this procedure with the DCLs in the presence of **T1** and **T2** led to the precipitation of **G8**–**G20**. To avoid this, aliquots from the DCLs were instead diluted in water before injection of the HPLC equipment. However, this led to overlap of peaks corresponding to salts and glucose (**G1**) in the chromatograms, so distributions of α-1,4-glucans in DCLs in the presence of **T1** and **T2** are thus reported without **G1**. The commercially available stock solution of CGTase contains glycerol, so to avoid the formation of glyceryl glycosides of linear α-1,4-glucans, which had been observed in a previous study in the Beeren group,[14] analytical scale reactions with **T1** and **T2** were carried out using a filtered, solvent-exchanged CGTase (see section 3.7.1 for details).

In the DCLs in the presence of the templates with the smaller head groups, **T3** and **T4**, β-CD was produced with high selectivity. In the DCL in the presence of **T3** (C_{16} alkyl chain) some γ-CD was also produced, and δ-CD was produced in the early stages of the reaction. In contrast, in the DCL in the presence of **T4** (C_{12} alkyl chain), β-CD was produced almost exclusively and no δ-CD was produced at any time.

From these results with DCLs in the presence of **T1**–**T4,** it is evident that both the size of the hydrophilic head groups and the length of the hydrophobic alkyl chain are important factors in determining the binding between the templates and the CDs, and, consequently, the outcome of the DCLs with **T1**–**T4**.

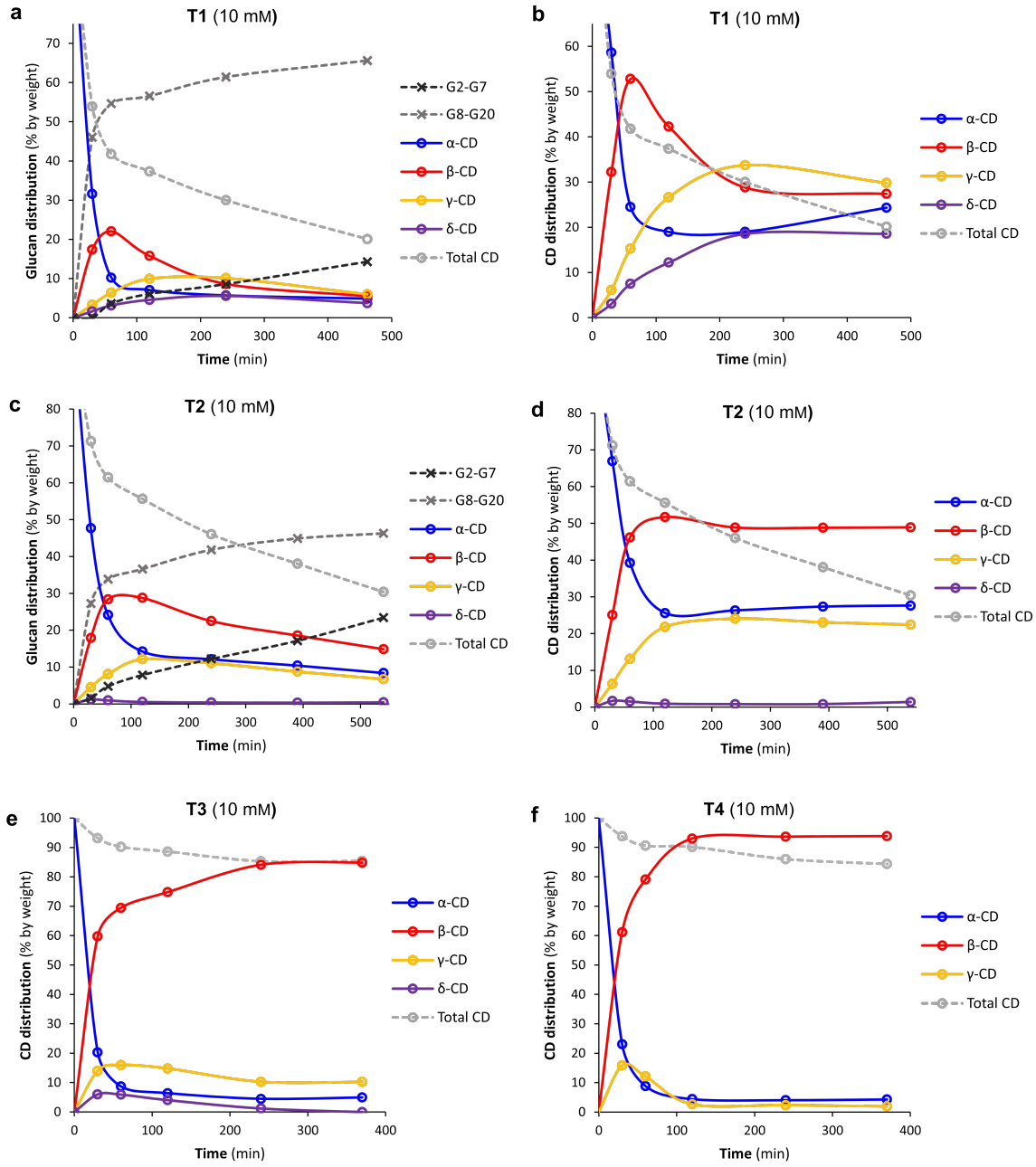


Figure 3.5. Glucan distributions in templated enzymatic reactions with α-CD (10 mg/ml) and templates **T1**–**T4** (10 mM) in sodium phosphate buffer (0.2 M, pH 7.5). For reactions with **T1** and **T2**, glycerol-free CGTase (65 µl per ml reaction mixture) was used and aliquots of the reactions were diluted in aqueous denaturing mixture before injection on HPLC. For reactions with **T3** and **T4**, commercial CGTase (50 µl per ml reaction mixture) was used and aliquots from the reaction were diluted in in 3:1 acetonitrile/water denaturing mixture before injection on HPLC. (a) and (c): Distribution of all glucans, CDs (α, β, γ, δ-CD) and linear α-1,4-glucans (**G2**–**G7** and **G8**–**G20**) and % of CDs in total library. (b), (d), (e) and (f): Distribution within CD subsystem (α, β, γ, δ-CD) and % of CDs in total library.

3.4. NMR spectroscopy studies of [n]-*pseudo*rotaxane formation

To investigate the interplay between head group size and alkyl chain length and its effect on the binding of templates **T1**–**T4** to α-, β-, γ-, and δ-CD, a series of ^{1}H NMR spectroscopy titrations in phosphate buffered D_2O at pH 7.5 or 7.0 were performed. It is well established that the formation of axial inclusion complexes (*pseudo*rotaxanes) of CDs is dependent on the size of the guest and the breadth of the CD, where the narrowest part of the CD a guest must penetrate to bind, designated d_{min}, is 4.4, 5.8 and 7.4 Å for α-, β-, and γ-CD, respectively.[15]

3.4.1. Titrations with α-CD and β-CD

Titrations with templates **T3**, **T4** and α-CD showed no binding to α-CD, which can be attributed to the head groups blocking the threading of α-CD (Supporting Figures S3.1–S3.2). From these results, it was then assumed that **T1** and **T2** with the larger head groups, also would not bind to α-CD, which is corroborated by the lack of any amplification of α-CD in the DCLs with **T1**–**T4**.

The titration with **T4** and β-CD shows binding in the slow exchange regime of the NMR chemical shift timescale (Figure 3.6 and Supporting Figure S3.4). It seems that the head group size of **T4** (and by extension **T3**) only just allows for the binding of β-CD to the alkyl chain, leading to a relatively slow binding event to form a [2]-*pseudo*rotaxane. Evidence for *pseudo*rotaxane formation is seen in the splitting of the ^{1}H NMR signals of the template upon complexation with β-CD. This splitting is caused by the non-symmetrical nature of the *pseudo*rotaxane formed, with the primary and secondary rim of the CD providing different environments.[16,17] With more than two equivalents of β-CD added, additional signals appear which can be attributed to [3]-*pseudo*rotaxanes (β-CD_2•**T4**), with two CDs threaded by **T4**.

For a 1:1 binding event between a CD and a template **T**, the binding constant K_a can be determined using the concentrations of unbound CD, [CD], unbound template, [**T**], and the complex, [CD•**T**] (Equation 3.1). For an NMR titration with complexation in the slow exchange regime, the concentration of each species can be determined from the integrals. Plotting [CD•**T**] vs [CD][**T**] for each data point in the titration should then give a linear relationship with K_a as the slope of the linear fit according to Equation (3.2).

$$K_a = \frac{[\mathrm{CD}\bullet\mathbf{T}]}{[\mathrm{CD}][\mathbf{T}]} \tag{3.1}$$

$$[\mathrm{CD}][\mathbf{T}]K_a = [\mathrm{CD}\bullet\mathbf{T}] \tag{3.2}$$

Using the integrals from early parts of the titration (0.2–0.9 equivalents β-CD), the 1:1 binding constant between **T4** and β-CD was estimated ($K_a = (7.8 \pm 0.9) \times 10^3\ M^{-1}$), disregarding any higher order binding (Figure 3.6). This relatively strong binding corresponds well to other known binding constants between β-CD and alkyl chains of similar length[6,18] and to the high selectivity for β-CD observed for the DCLs in the presence of **T3** and **T4**.

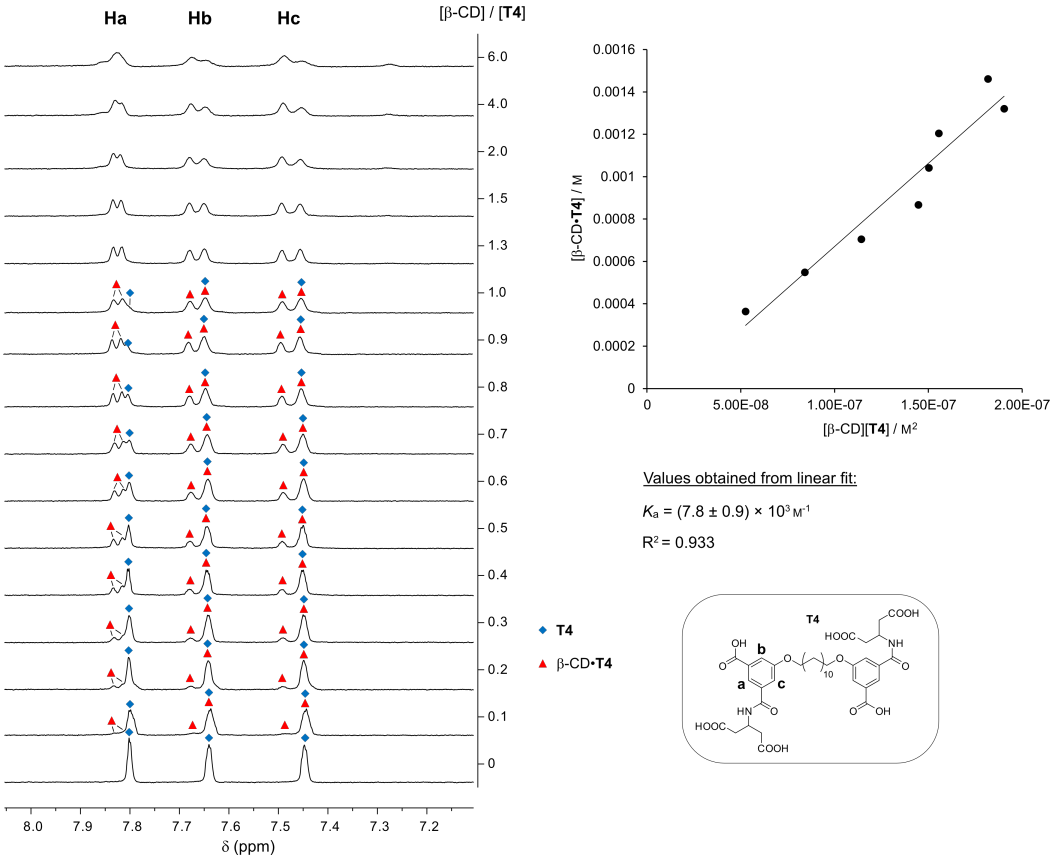


Figure 3.6. ^{1}H NMR spectroscopy titration of **T4** with β-CD in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T4** at 2.0 mM. *Left*: Partial ^{1}H NMR spectra (400 MHz) of **T4** with increasing concentrations of β-CD. *Right:* [β-CD•**T4**] as a function of [β-CD][**T4**], calculated using integrals of data points from 0.2–0.9 equivalents β-CD and disregarding any higher order binding. The reported 1:1 binding constant K_a corresponds to the slope of the linear fit and the uncertainty reported is the standard deviation of the slope of linear fit.

The effect of the size of the head group on the binding of CDs was again clearly seen in the titration of **T2** (large head group, C_{12} alkyl chain) with β-CD (Supporting Figure S3.3), which shows no binding. The large head groups of **T1** and **T2** are thus too big to allow *pseudo*rotaxane formation with β-CD.

3.4.2. Titrations with γ-CD

The binding of γ-CD to templates **T1**–**T4** revealed a rather complex picture, with binding in both the fast and slow exchange regimes and various [n]-*pseudo*rotaxanes with 1:1, 1:2, 2:1 and 2:2 stoichiometries formed. The results from the titrations are summarized in Figure 3.7.

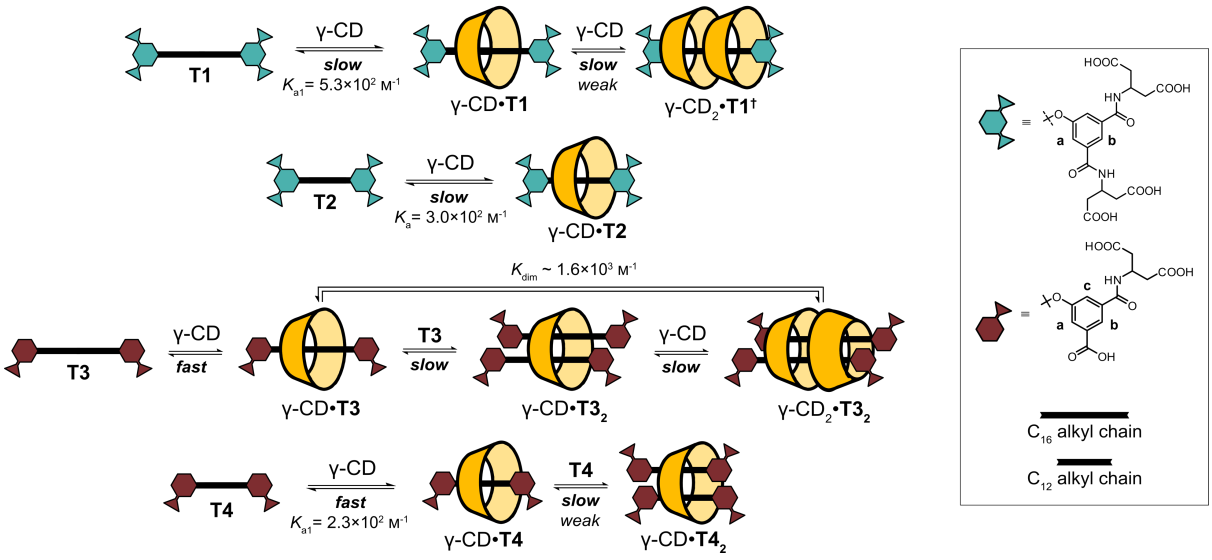


Figure 3.7. Schematic representation of the equilibria where [2]-, [3]-, and [4]-*pseudo*rotaxanes are formed when templates **T1**–**T4** bind to γ-CD, including binding constants (K_a) and whether each equilibrium corresponds to a process in slow or fast exchange on the NMR chemical shift timescale. † Orientation of the CDs is unknown.

The titration of **T1** with γ-CD gave binding in the slow exchange regime, with signal splitting showing [2]-*pseudo*rotaxane formation (Figure 3.8 and Supporting Figure S3.5). With more than five equivalents of γ-CD added, more signals appear, which can be attributed to [3]-*pseudo*rotaxanes (γ-CD$_2$•**T1**) formed. The 1:1 binding constant was estimated (K_a = (5.3 ± 0.8) × 10^2 M^{-1}) using integrals from early parts of the titration (0.5–4 equivalents γ-CD, disregarding any higher order binding).

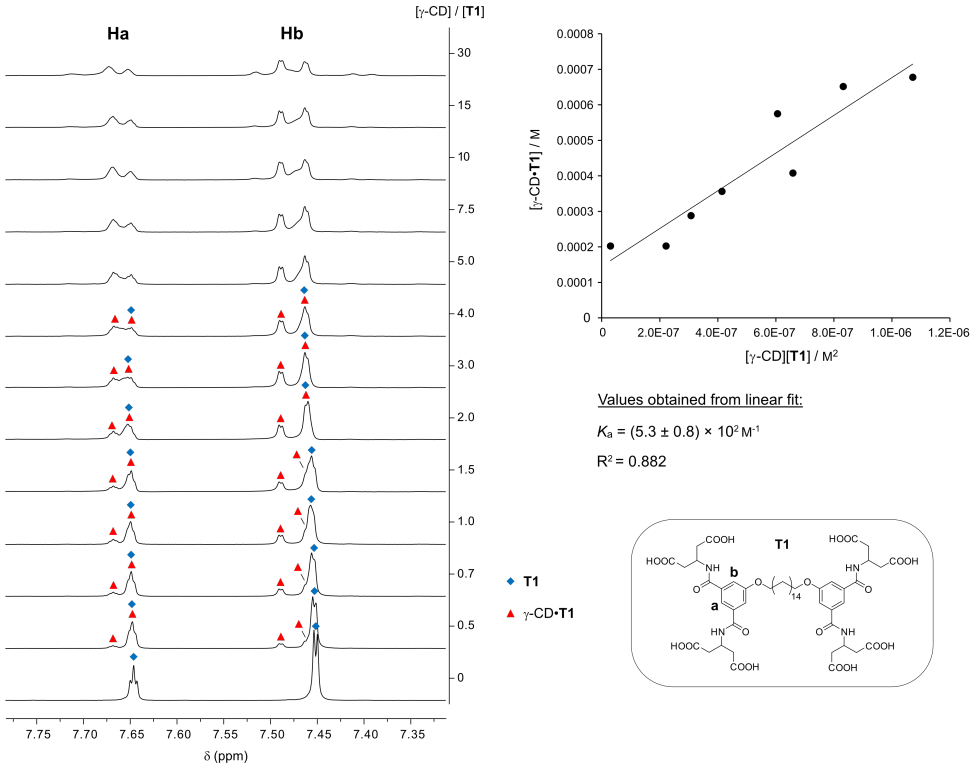


Figure 3.8. ^{1}H NMR spectroscopy titration of **T1** with γ-CD in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T1** at 1.0 mM. *Left*: Partial ^{1}H NMR spectra (400 MHz) of **T1** with increasing concentrations of γ-CD. *Right:* Plot of [γ-CD•**T1**] as a function of [γ-CD][**T1**], calculated using integrals of data points with 0.5–4.0 mM γ-CD and disregarding any higher order binding. The 1:1 binding constant K_a is estimated from the slope of the linear fit and the uncertainty reported is the standard deviation of the slope of linear fit.

Moving to **T2**, an effect of the shorter alkyl chain length in comparison with **T1** was clearly seen. With **T2**, only 1:1 binding (no 1:2 binding) in the slow exchange regime was observed ([2]-*pseudo*rotaxane formation), with a binding constant of $K_a = (3.0 \pm 0.09) \times 10^2\ M^{-1}$, (Figure 3.9 and Supporting Figure S3.6). Apparently, the C_{12} alkyl chain of **T2** does not accommodate the binding of two γ-CDs. While the full length of a completely staggered C_{12} alkyl chain should be able to accommodate two γ-CDs (just as the C_{12} alkyl chain of **T4** accommodates two β-CDs), the requirement for a more folded conformation of the alkyl chain, with gauche kinks to effectively fill out the large cavity of γ-CD, means the long, staggered conformation of the alkyl chain wil be disfavoured. This effect is seen even for the smaller cavity of β-CD upon complexation with bolaamphiphiles with alkyl chains as hydrophobic binding motifs.[17,19] This effect has also been observed previously in the Beeren group for the binding of sodium dodecyl sulfate (SDS) to α-, β-, and γ-CD, where α-CD binds twice to SDS with two strong binding interactions, as the alkyl chain can be staggered and still effectively bind to the small α-CD cavity. Conversely, β-CD also binds twice, but exhibits a strong binding followed by weak binding, while γ-CD exhibits only one weaker binding to SDS,[6] showing the requirement for more folded, 'shorter' conformations of the alkyl chain with larger CD cavities and the importance of size complementarity between CDs and hydrophobic guests for strong binding.

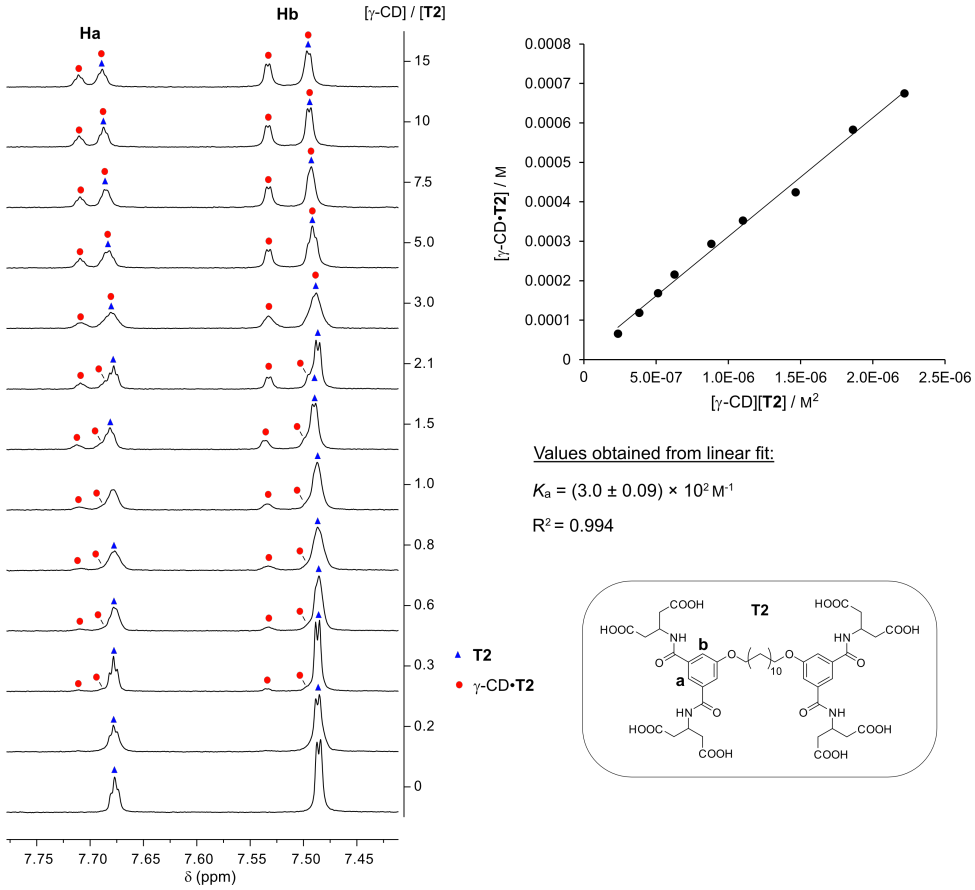


Figure 3.9. ^{1}H NMR spectroscopy titration of **T2** with γ-CD in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T2** at 1.0 mM. *Left*: Partial ^{1}H NMR spectra (400 MHz) of **T2** with increasing concentrations of γ-CD. *Right:* Plot [γ-CD•**T2**] as a function [γ-CD][**T2**], calculated using integrals of data points from 0.3–7.5 mM γ-CD. The 1:1 binding constant K_a was calculated as the slope of the linear fit and the uncertainty reported is the standard deviation of the slope of linear fit.

In the titration of **T4** with γ-CD (Figure 3.10 and Supporting Figure S3.12), the effect of the smaller head group was apparent, with binding occurring in the fast exchange regime in contrast to the binding in the slow exchange regime of **T1** / **T2** and γ-CD. The binding in the fast exchange regime can be attributed to 1:1 complex formation (γ-CD•**T4**), presumably [2]-*pseudo*rotaxane formation. For binding in the fast exchange regime, there is also the possibility of contributions from 'U-shape' binding, where the alkyl chain folds and binds to the γ-CD cavity without threading through the cavity, a binding mode reported recently for γ-CD.[20]

Evidence of γ-CD•**T4**$_2$ binding in the slow exchange regime is also seen, with the appearance of small signals in the spectra with about 0.8–2 equivalents of added γ-CD – signals that disappear again with higher concentrations of γ-CD, where the 1:1 stoichiometry is favored over the 2:1 (**T4**/γ-CD) stoichiometry. This second binding could correspond to the threading of a second **T4** through the γ-CD cavity, forming a [3]-*pseudo*rotaxane, where the fact the binding occurs in the slow exchange regime can be explained by the increased steric demands for the threading of a second **T4**, when the first **T4** is already occupying the γ-CD cavity. It is well established that γ-CD can thread two guest molecules through its cavity.[15,21]

For NMR titrations where binding occurs in the fast exchange regime, the binding constant can be determined by fitting the change in chemical shift to a suitable binding model. For a 1:1 complexation between a host (H) and a guest (G):

$$H + G \rightleftharpoons HG \tag{3.3}$$

The binding constant (K_a) is defined as:

$$K_a = \frac{[HG]}{[H][G]} \tag{3.4}$$

and for an NMR titration under fast exchange, K_a can be determined by fitting the binding isotherm to the following equation (3.5):[22]

$$\Delta\delta_{obs} = \frac{\Delta\delta_{max}}{2[H]_0}\left(\left([H]_0+[G]_0 + \frac{1}{K_a}\right) - \sqrt{\left([H]_0+[G]_0 + \frac{1}{K_a}\right)^2 - 4[H]_0[G]_0}\right) \tag{3.5}$$

where $\Delta\delta_{obs}$ is the observed change in chemical shift of the host upon addition of guest, $\Delta\delta_{max}$ is the maximum possible change in chemical shift for the host upon addition of guest, i.e. $\delta_{HG} - \delta_H$, $[H]_0$ is the total concentration of host and $[G]_0$ is the total concentration of added guest.

The 1:1 binding constant between γ-CD and **T4** was then estimated by fitting the change in chemical shift to the 1:1 binding model (Equation 3.5), which gave $K_a = (2.3 \pm 0.4) \times 10^2\ M^{-1}$, disregarding any higher order binding.

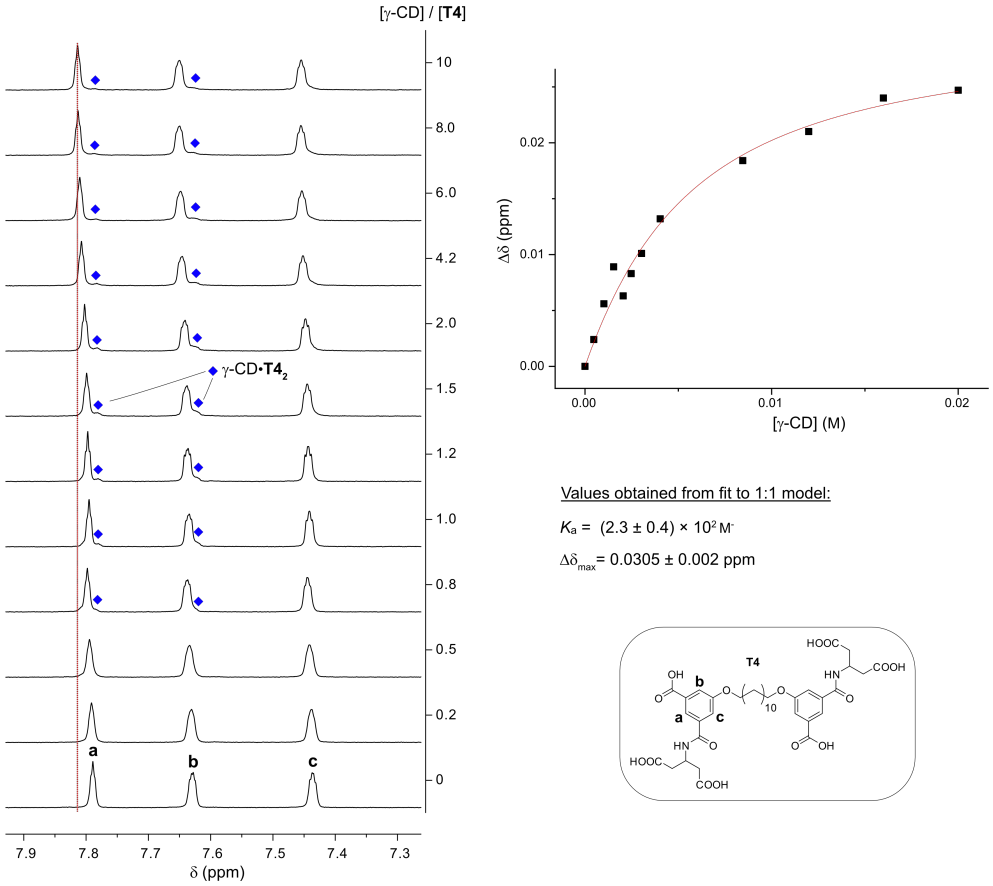


Figure 3.10. ^{1}H NMR spectroscopy titration of **T4** with γ-CD in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T4** at 2.0 mM. *Left*: Partial ^{1}H NMR spectra (400 MHz) of **T4** with increasing concentrations of γ-CD. *Right:* The change in chemical shift of the proton at 7.8 ppm and the resulting fit to a 1:1 binding model.

The titration of **T3** with γ-CD also revealed binding both in the fast and slow exchange regimes (Figure 3.11 and Supporting Figures S3.7–S3.11). Similar to the titration of **T4** with γ-CD, the fast exchange process can be attributed to a 1:1 complexation (γ-CD•**T3**), and signals that appear as small signals and then disappear again with higher concentrations of γ-CD can be attributed to a slow [3]-*pseudo*rotaxane formation (γ-CD•**T3**$_2$), with two templates threading γ-CD. A third set of signals also appear and grow with higher concentrations of γ-CD, while the set of moving signals corresponding to the 1:1 complexation process shrink. At high concentrations of γ-CD (>8 equivalents), the intensities of these two sets of signals stop changing (Figure 3.11b), indicating that these two sets of signals correspond to complexes with the same stoichiometries. The growing set of signals can thus be attributed to the formation of a 2:2 (γ-CD$_2$•**T3**$_2$) complex, or a [4]-*pseudo*rotaxane. The integrals of the set of signals corresponding to the 2:2 complex confirm the stoichiometry (1.0 : 0.97, **T3**/γ-CD, based on an average of integrals of data with 1–6.5 equivalents of γ-CD).

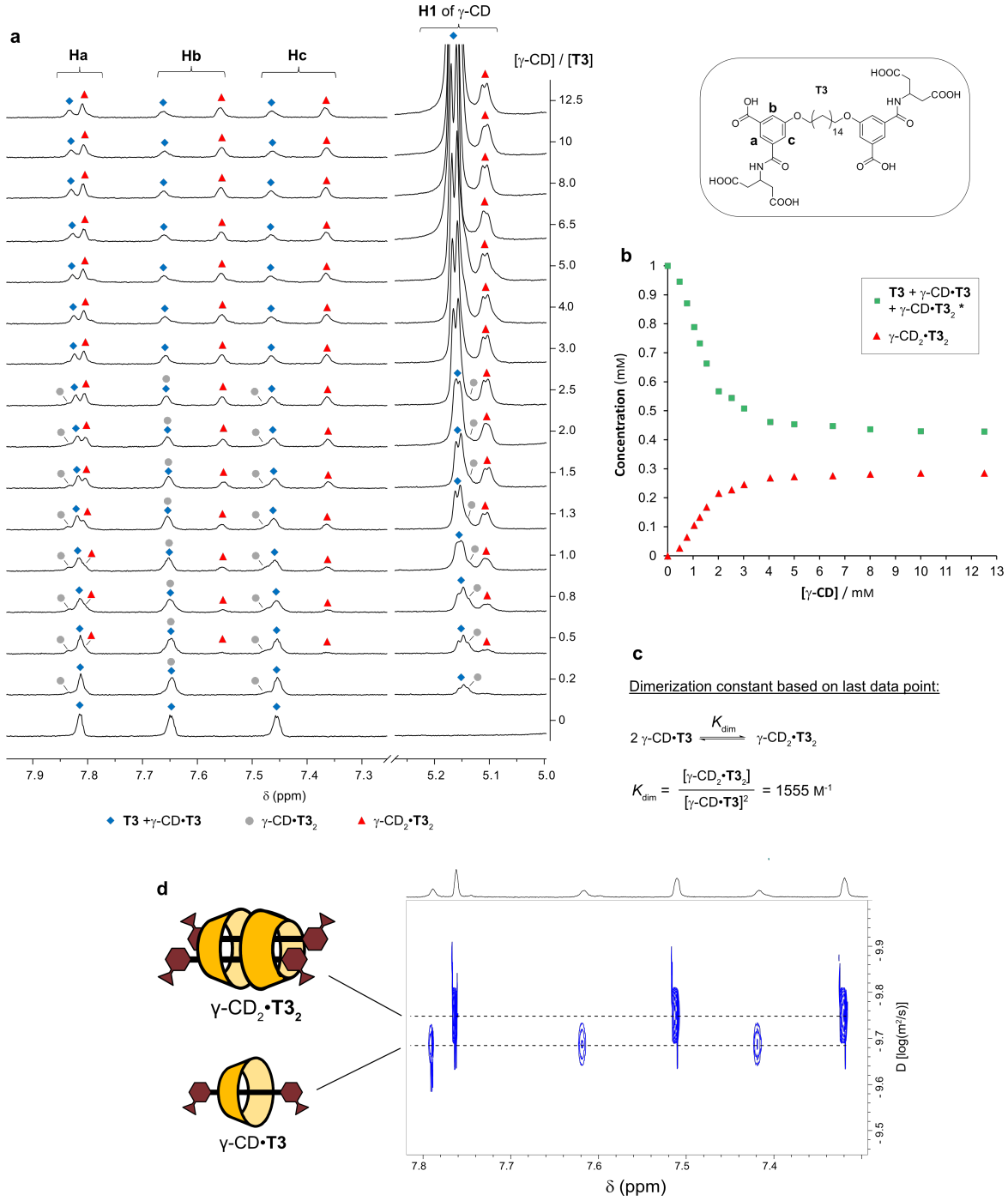


Figure 3.11. ^{1}H NMR spectroscopy titration of **T3** with γ-CD in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T3** at 1.0 mM. (a) Partial ^{1}H NMR spectra (400 MHz) of **T3** with increasing concentrations of γ-CD. (b) Speciation plot showing molar fractions of the species indicated, calculated using the integrals from the titration. *Concentrations calculated from integrals assuming one **T3** per species, as the ratios between **T3**, γ-CD•T3 and γ-CD•**T3**$_2$ are unknown. (c) Estimated 'dimerization constant' calculated using the last data point (12.5 equivalents γ-CD) where the concentrations of unbound **T3** and γ-CD•**T3**$_2$ are assumed to be negligible. (d) DOSY (800 MHz) spectrum of the last data point (12.5 equivalents γ-CD).

The ^{1}H diffusion-ordered spectroscopy (DOSY) spectrum of the titration endpoint (Figure 3.11d and Supporting Figure S3.7) confirmed the presence of two different species of **T3**, with the signals corresponding to the bigger γ-CD$_2$•**T3**$_2$ complex displaying the smaller diffusion

constant. From the integrals of the titration endpoint a 'dimerization constant' K_{dim} = [γ-CD_2•**T3**$_2$]/[γ-CD•**T3**]^2 = 1.6 × 10^3 M^{-1} was estimated, under the assumption that these are the only two species containing **T3** at this point, indicating that the formation of the [4]-*pseudo*rotaxane is favored over the formation of the [2]-*pseudo*rotaxane. It seems that **T3**, with the smaller head group and the longer C_{16} alkyl chain, has the right dimensions to enable the effective binding of two templates to two γ-CDs, thus more effectively filling out the cavity of γ-CD while burying the entire length of the alkyl chains inside the hydrophobic cavities.

No splitting of the signals of the protons of **T3** in the γ-CD_2•**T3**$_2$ complex was observed, which indicates that the structure is one of the two possible symmetrical γ-CD_2•**T3**$_2$ [4]-*pseudo*rotaxanes, with the two γ-CDs binding in either a *head-to-head* and *tail-to-tail* mode. In the ^{1}H-^{1}H nuclear Overhauser effect spectroscopy (NOESY) spectrum of a concentrated sample (5 mM) of **T3** and γ-CD (Figure 3.12 and S3.8), the γ-CD_2•**T3**$_2$ species exhibits relatively strong crosspeaks between two of the aromatic protons of **T3** (Hb and Hc) and the H5, H6 and H4 protons of γ-CD. In contrast, comparatively weak crosspeaks between these aromatic protons and the H3 and H2 protons of γ-CD are observed. This indicates that the symmetrical [4]-*pseudo*rotaxane is the *head-to-head* structure, with the narrow, primary rim of the two γ-CDs (which contains the H5 and H6 protons) pointing towards the aromatic head groups (Figure 3.12, b and c).

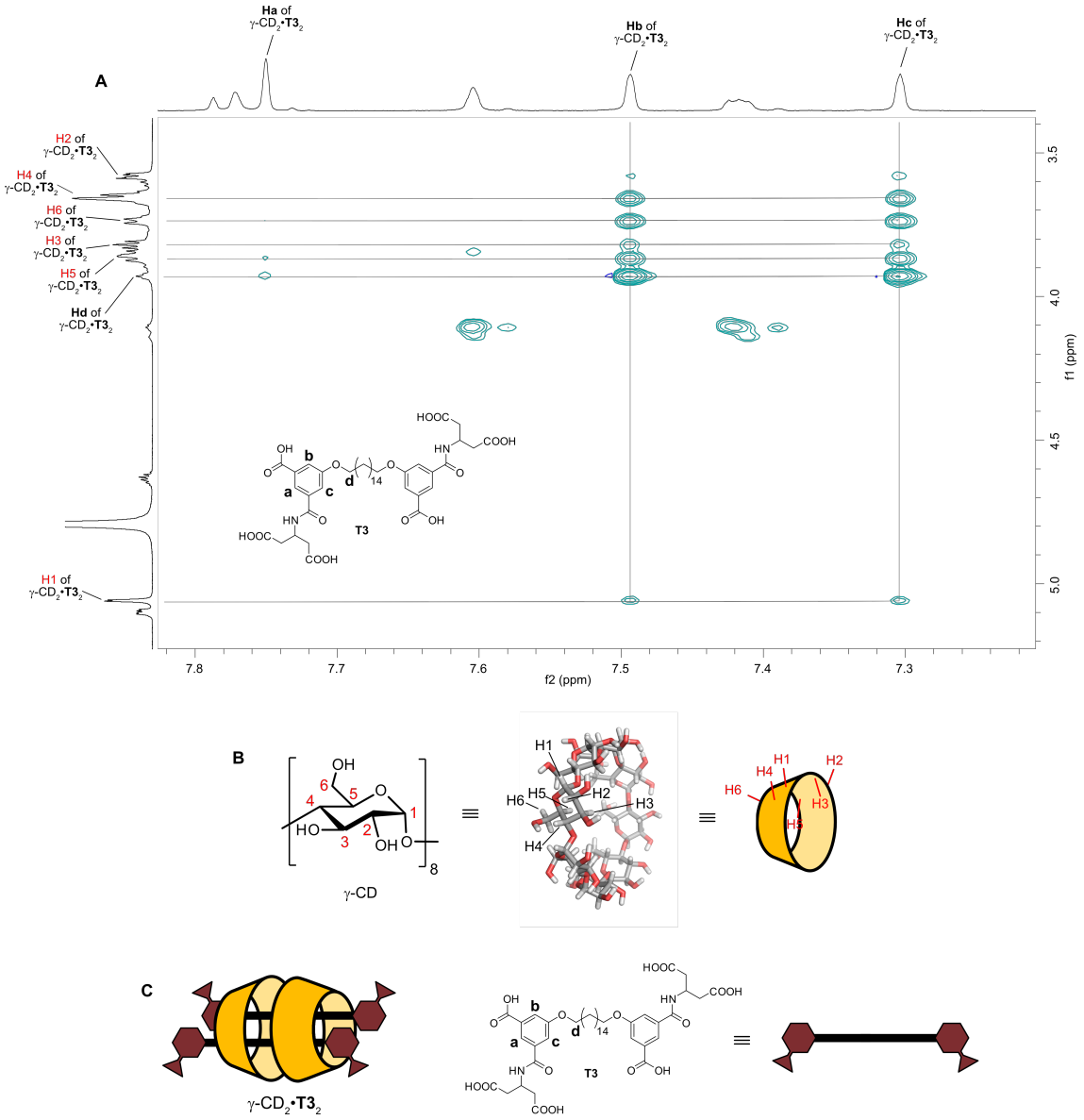


Figure 3.12. Structural assignment of the [4]-*pseudo*rotaxane γ-CD_2•**T3**$_2$. (A) Partial ^{1}H-^{1}H NOESY spectrum (800 MHz) of **T3** (5 mM) and γ-CD (5 mM) in D_2O phosphate buffer (pH 7.5, 100 mM) showing stronger crosspeaks between Hb and Hc of **T3** and H5, H6 and H4 of γ-CD and weaker crosspeaks to H3 and H2 of γ-CD. Signals assigned using 2D NMR techniques (HSQC, COSY) (Supporting Figures S3.9–S3.11). (B) Structure of γ-CD showing the positions of each proton. (C) Schematic representation of the proposed binding mode based on the NOEs, showing the symmetrical *head-to-head* [4]-*pseudo*rotaxane.

3.4.3. Titrations with δ-CD

δ-CD for use in the NMR titrations described here was synthesized enzymatically using **T1** as a template and then isolated using preparative HPLC as described in section 3.6. The NMR titrations with **T1–T4** and δ-CD revealed another rather complex picture. In some cases, only fast exchange was observed, while in other cases, combinations of fast and slow exchange were seen. The results are summarized in Figure 3.13.

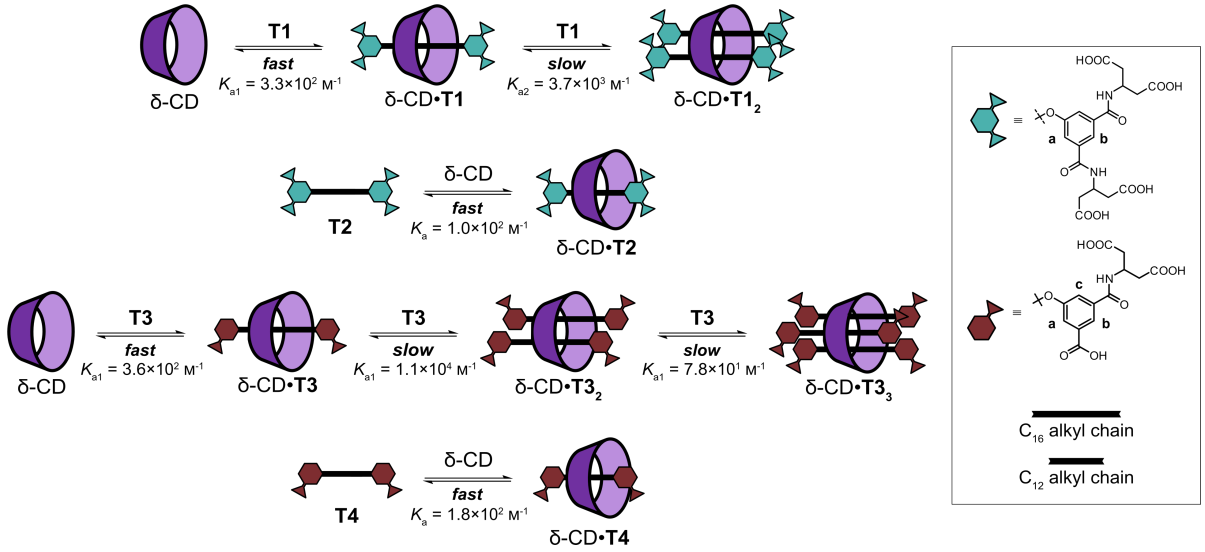


Figure 3.13. Schematic representation of the equilibria where [2]-, [3]-, and [4]-*pseudo*rotaxanes are formed when templates **T1**–**T4** bind to δ-CD, including binding constants (K_a) and whether each equilibrium corresponds to a process in slow or fast exchange on the NMR chemical shift timescale.

The titrations with δ-CD and the templates with the shorter C_{12} alkyl chains (**T2** and **T4**) showed only weak binding in the fast exchange regime (Supporting Figures S3.15, S3.16, S3.18, S3.19). The changes in chemical shift were fitted to a 1:1 binding model (Equation 3.5), yielding binding constants of $K_a = (1.0 \pm 0.05) \times 10^2\ M^{-1}$ for δ-CD•**T2** and $K_a = (1.8 \pm 0.1) \times 10^2\ M^{-1}$ for δ-CD•**T4**. Again, the U-shape binding mode cannot be ruled out as a possibility here. In contrast to these results with **T2** and **T4**, the titrations with δ-CD and the templates with the longer C_{16} alkyl chains (**T1** and **T3**) gave dramatically different results.

<u>δ-CD and **T1**</u>

Initially, a titration with a constant concentration of **T1** and increasing concentrations of δ-CD was carried out (Supporting Figure S3.13). This titration showed binding in both the fast and the slow exchange regimes, with evidence of the formation of a ternary complex with two **T1** molecules bound to δ-CD (δ-CD•$\mathbf{T1}_2$). To better observe the sequential formation of the δ-CD•**T1** and δ-CD•$\mathbf{T1}_2$ complexes, the reverse titration was carried out, with a constant concentration of δ-CD and increasing concentrations of **T1** (Figure 3.14 and Supporting Figure S3.14). In this titration, the binding in the fast exchange regime, clearly observed with the movement of the H1 proton of δ-CD, was attributed to 1:1 complexation or [2]-*pseudo*rotaxane formation. The set of signals appearing and growing at higher concentrations of **T1** were then

attributed to the formation a [3]-*pseudo*rotaxane (δ-CD•**T1**$_2$). Like in the previously described results with γ-CD, the threading of a second template leads to increased steric demands, and thus a complexation process in the slow exchange regime. The splitting of the signals of **T1** in the δ-CD•**T1**$_2$ complex confirms the formation of a *pseudo*rotaxane. The DOSY spectrum of the data point with 2.7 equivalents of **T1** confirms the presence of two different species of δ-CD, with the signals corresponding to the bigger 1:2 complex δ-CD•**T1**$_2$ displaying the smaller diffusion constant.

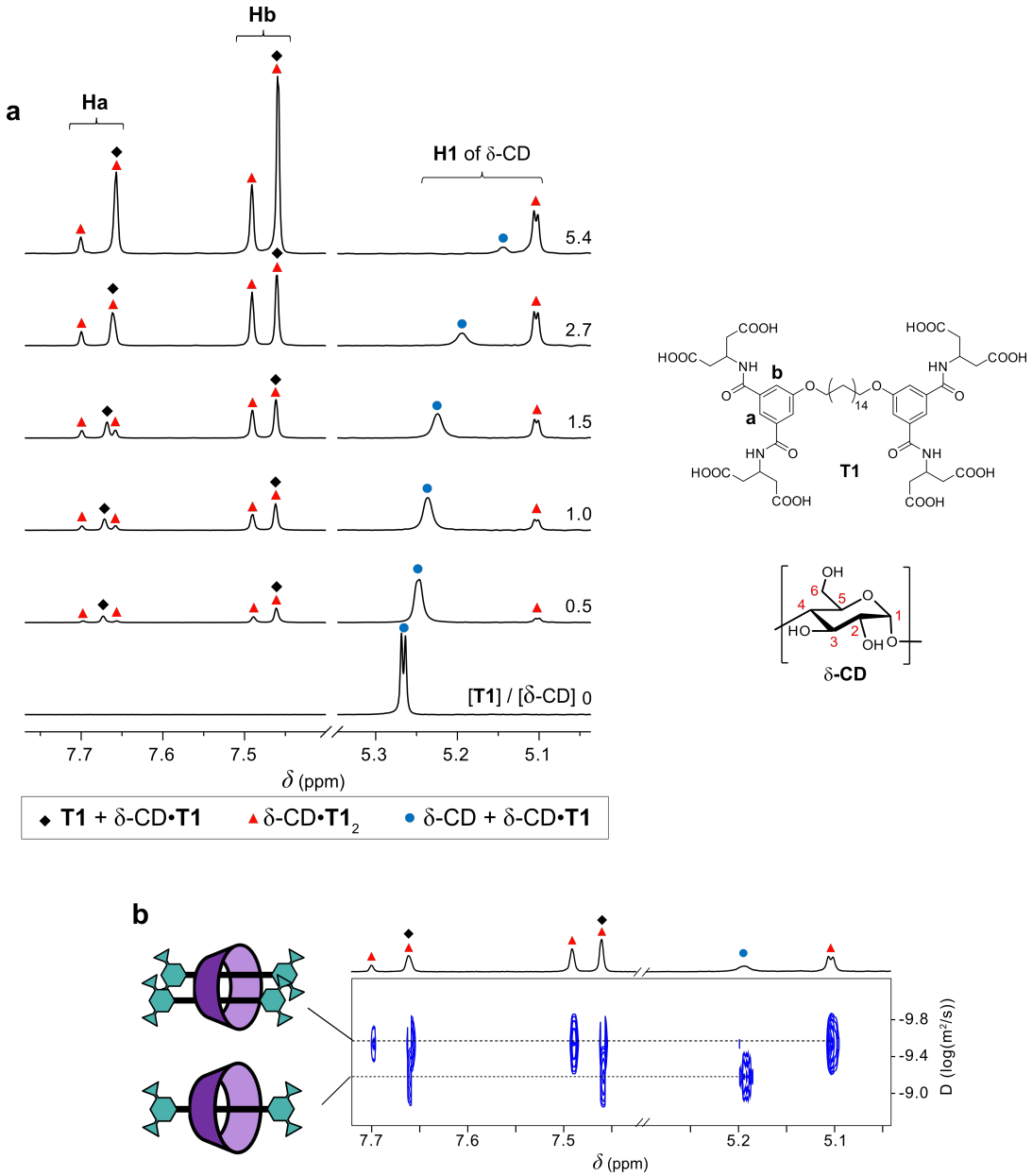


Figure 3.14. Titration with δ-CD and **T1** in sodium phosphate buffer (100 mM) at pH 7.0 in D_2O with a constant concentration of δ-CD at 1.0 mM. (a) Partial ^{1}H NMR spectra (800 MHz) of δ-CD with increasing concentrations of **T1**. (d) DOSY (800 MHz) spectrum of a solution of δ-CD (1.0 mM) and **T1** (2.7 mM).

In this NMR titration with **T1** and δ-CD, the simultaneous occurrence of complexation processes in both the fast and slow exchange regimes meant that standard equations or fitting

procedures could not be applied to determine binding constants. From the titration, an observed change in chemical shift ($\Delta\delta_{obs}$) for the H1 proton of δ-CD (due to the fast 1:1 complexation process between unbound δ-CD and δ-CD•**T1)** could be monitored. Also, the concentrations of [δ-CD] + [δ-CD•**T1]** and [δ-CD•$\mathbf{T1}_2$**]** could be determined by integration of the peaks corresponding to these species. Equations 3.6 and 3.7 were then derived to relate these observables to the concentration of added host ($[H]_t$), guest ($[G]_t$), K_{a1}, K_{a2}, and the chemical shifts of δ-CD (δ_H) and δ-CD• **T1** (δ_{HG}) (see supporting information section S3.1.1 for derivation of equations).

$$\Delta\delta_{obs}\left(\frac{[H]+[HG]}{[H]_t}\right) = \frac{\Delta\delta_{HG}K_{a1}[G]}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2} \tag{3.6}$$

$$\frac{[HG_2]}{[H]_t} = \frac{K_{a1}K_{a2}[G]^2}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2} \tag{3.7}$$

where [G] is related to $[G]_t$ according to equation (3.8)

$$A[G]^3 + B[G]^2 + C[G] + D = 0 \tag{3.8}$$

wherein

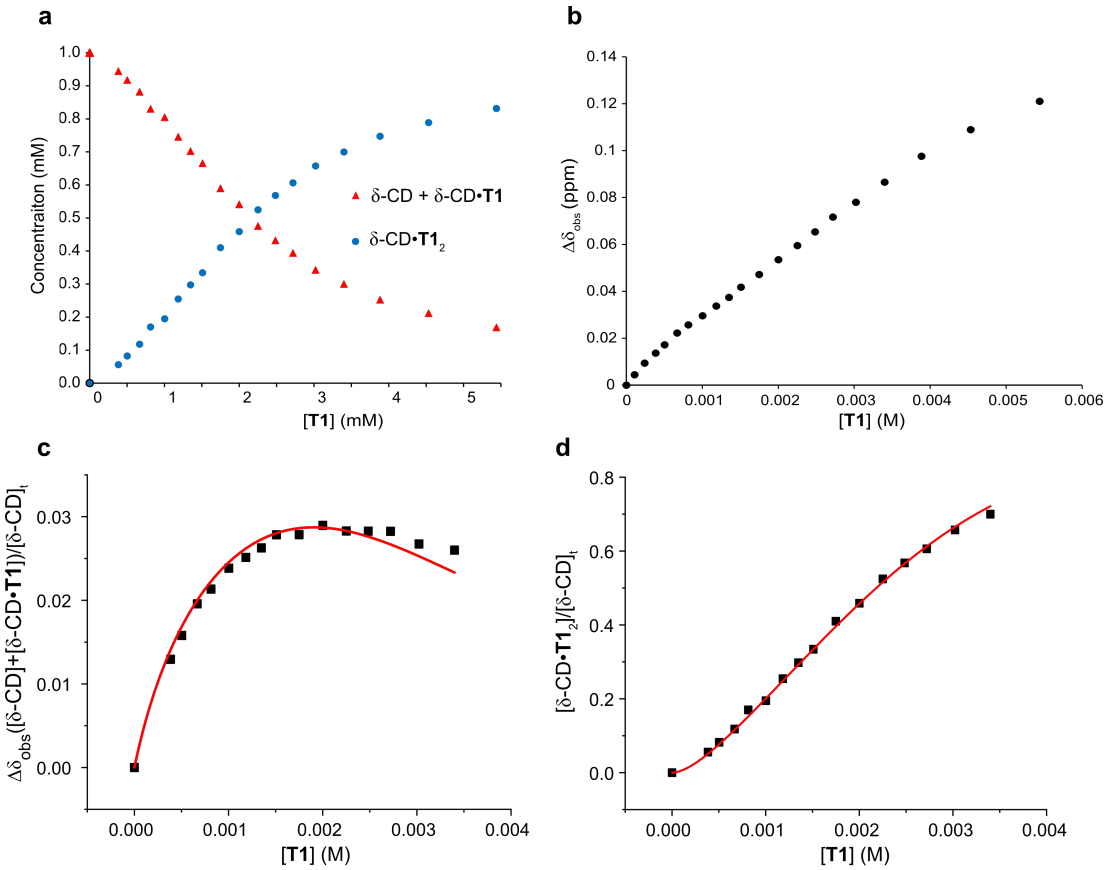
$$A = K_{a1}K_{a2}$$

$$B = K_{a1} + 2K_{a1}K_{a2}[H]_t - K_{a1}K_{a2}[G]_t$$

$$C = 1 + K_{a1}[H]_t - K_1[G]_t$$

$$D = -[G]_t$$

The NMR data was then simultaneously fitted to (3.6) and (3.7) in the non-linear curve fitter of Origin to determine K_{a1}, K_{a2} and $\Delta\delta_{HG}$. The data and the resulting fits are shown in Figure 3.15. Binding constants on the order of K_{a1} = $(3.3 \pm 2.7) \times 10^2\ M^{-1}$ and $K_{a2} = (3.7 \pm 1.8) \times 10^3\ M^{-1}$ were thus determined, showing that the binding is cooperative and that the [3]-*pseudo*rotaxane is the more stable species. The fits to the data were generally good, albeit with a slight systematic error in the fit to equation (3.6) (Figure 3.15c) most apparent with higher concentrations of **T1**. This error with higher concentrations of **T1** could be due to beginning overlap between the two signals of the H1 proton of δ-CD (Figure 3.14a) leading to errors in integration of the signals. This overlap is also the reason why only data with up to 3.4 mM of **T1** added were used in the fitting procedure.



Values obtained from fitting to equations (3.6) and (3.7):

K_{a2} = (3.3 ± 2.7) × 10^2 M^{-1}

K_{a2} = (3.7 ± 1.8) × 10^3 M^{-1}

$\Delta\delta_{HG}$ = 0.22 ± 0.13 ppm

Figure 3.15. Analysis of NMR titration with δ-CD and **T1**. (a) Concentrations obtained from integration of peaks in the NMR spectra. (b) The observed change in chemical shift ($\Delta\delta_{obs}$) of the H1 proton of δ-CD. (c) Fit to equation (3.6). (d) Fit to equation (3.7).

δ-CD and **T3**

The titration with δ-CD and increasing concentrations of **T3** showed a very similar picture to the titration with **T1**, with binding in both the fast and slow exchange regimes (Figure 3.16 and Supporting Figure S3.17). However, with high concentrations of **T3**, a third set of signals appeared, which is presumably the threading of a third **T3**, forming the δ-CD•**T3**$_3$ complex, a [4]-*pseudo*rotaxane. The DOSY spectrum of the data point with 3.4 equivalents of **T3** confirms the presence of three different species of **T3**, with diffusion constants in the expected order with the δ-CD•**T3**$_3$ species displaying the smallest diffusion constant (Figure 3.16b).

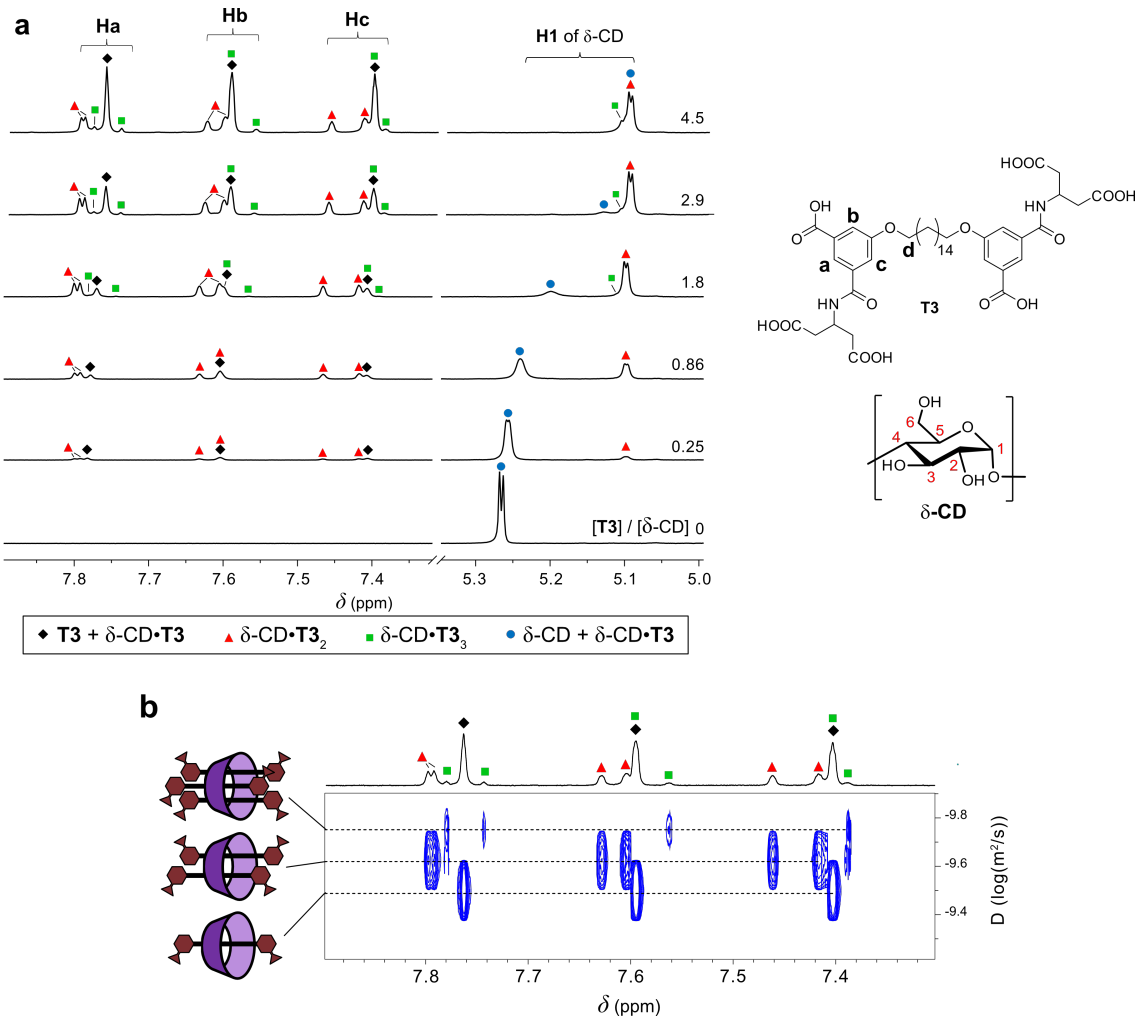


Figure 3.16. Titration with δ-CD and **T3** in sodium phosphate buffer (100 mM) at pH 7.5 in D_2O with a constant concentration of δ-CD at 1.0 mM. (a) Partial ^{1}H NMR spectra (800 MHz) of δ-CD with increasing concentrations of **T1**. (d) DOSY (800 MHz) spectrum of a solution of δ-CD (1.0 mM) and **T3** (3.4 mM).

From this titration with δ-CD and **T3** it was possible to monitor both the observed change in chemical shift ($\Delta\delta_{obs}$) for the H1 proton of δ-CD (due to the fast 1:1 complexation process between unbound δ-CD and δ-CD•**T3)** and the concentrations of [δ-CD] + [δ-CD•**T3**], [δ-CD•**T3**$_2$] and [δ-CD•**T3**$_3$] from the integration of relevant peaks. Equations (3.9) and (3.10) were then derived to relate these observables to the concentration of added host ($[H]_t$), guest

($[G]_t$), K_{a1}, K_{a2}, K_{a3} and the chemical shifts of δ-CD (δ_H) and δ-CD•**T3** (δ_{HG}) in a 1:3 binding process (see supporting information section S3.1.2 for derivation).

$$\Delta\delta_{obs}\left(\frac{[H]+[HG]}{[H]_t}\right) = \frac{\Delta\delta_{HG}K_{a1}[G]}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2+K_{a1}K_{a2}K_{a3}[G]^3} \quad (3.9)$$

$$\frac{[HG_3]}{[H]_t} = \frac{K_{a1}K_{a2}K_{a3}[G]^3}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2+K_{a1}K_{a2}K_{a3}[G]^3} \quad (3.10)$$

where [G] is related to $[G]_t$ according to equation (3.11)

$$A[G]^4 + B[G]^3 + C[G]^2 + D[G] + E = 0 \quad (3.11)$$

wherein

$$A = K_{a1}K_{a2}K_{a3}$$

$$B = K_{a1}K_{a2} + 3K_{a1}K_{a2}K_{a3}[H]_t - K_{a1}K_{a2}K_{a3}[G]_t$$

$$C = K_{a1} + 2K_{a1}K_{a2}[H]_t - K_{a1}K_{a2}[G]_t$$

$$D = 1 + K_{a1}K_{a2} - K_{a1}[G]_t$$

$$E = -[G]_t$$

The NMR data was then fitted simultaneously to equation (3.9) and (3.10) to determine K_{a1}, K_{a2}, K_{a3}, and $\Delta\delta_{HG}$ (Figure 3.17). Good fits were obtained, and binding constants on the order of $K_{a1} = (3.6 \pm 1.6) \times 10^2\ M^{-1}$, $K_{a2} = (1.1 \pm 0.5) \times 10^4\ M^{-1}$ and $K_{a3} = (7.8 \pm 0.2) \times 10^1\ M^{-1}$ were thus determined. These results show that the first two templates are bound cooperatively to form the [3]-*pseudo*rotaxane δ-CD•**T3**$_2$, with binding constants of comparable magnitudes to the cooperative 1:2 binding obtained with **T1** and δ-CD. In contrast, the third binding of **T3** to form the [4]-*pseudo*rotaxane δ-CD•**T3**$_3$ was found to be comparatively weak.

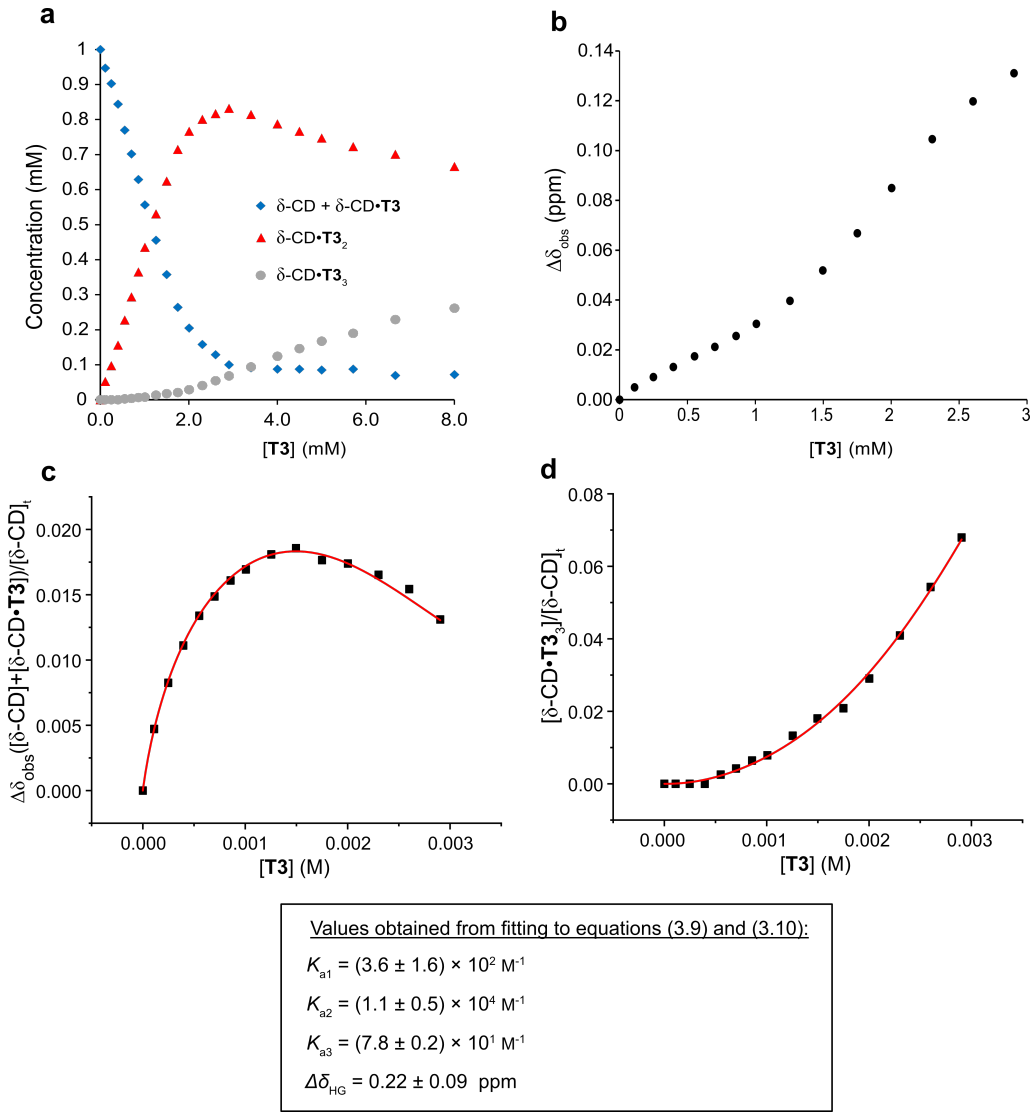


Figure 3.17. Analysis of NMR titration with δ-CD and **T3**. (a) Concentrations obtained from integration of peaks in the NMR spectra. (b) The observed change in chemical shift ($\Delta\delta_{obs}$) of the H1 proton of δ-CD. (c) Fit to equation (3.9). (d) Fit to equation (3.10).

From the results of the titrations with δ-CD and **T1**–**T4**, it can be concluded that δ-CD forms higher order complexes with bolaamphiphiles containing alkyl chains when possible, threading two and even up to three bolaamphiphiles to form [n]-*pseudo*rotaxanes. The bolaamphiphiles need to have a certain minimum length for this type of complexation to occur, as shown in the lack of any detectable higher order binding with **T2** and **T4** (C_{12} alkyl chain), where the threading of two of these bolaamphiphiles through δ-CD presumably leads to steric and/or electronic clashes between the bulky, charged head groups. In contrast, the longer alkyl chains of **T1** and **T3** allow the head groups to be further apart in the 1:2 δ-CD/template complexes, alleviating any repulsive clashes. The ability of **T3** to form a 1:3 δ-CD/template [4]-*pseudo*rotaxane is presumably due to the smaller head group (in comparison with **T1**) making the third complexation possible, whereas the larger head group of **T1** would disfavor formation of the 1:3 δ-CD/template [4]-*pseudo*rotaxane, either due to kinetics (not possible to thread a third template with a large head group) or due to thermodynamics (steric/electronic clashes

between the large head groups). The binding between δ-CD and templates **T1** and **T3** are the strongest affinities of hydrophobic guests to δ-CD reported to date.

3.4.4. Molecular dynamics simulation of δ-CD•T1$_2$

To obtain further information about the nature of the δ-CD•**T1**$_2$ [3]-*pseudo*rotaxane, a molecular dynamics (MD) simulation of the complex was performed. The simulation was carried out using the AMBER20 suite[23] and employing the GLYCAM-06 force field[24] for δ-CD and the general amber force field (GAFF)[25] for **T1** (for computational details see supporting section 3.7.3). The starting geometry was manually generated with two **T1**s (as the fully deprotonated octa-carboxylate anion) pre-threaded through the cavity of δ-CD. The resulting structure was then solvated with TIP3P water,[26] and the system was neutralized with sodium ions, energy minimized, heated to 300 K and equilibrated as described in experimental section 3.7.3. An MD production run (500 ns) was then carried out at 300 K in the *NPT* ensemble, and the resulting trajectory was subjected to clustering analysis using the 'kmeans' algorithm,[27] and a representative complex structure extracted from the most dominant cluster is displayed in Figure 3.18. The structure clearly shows how the two alkyl chains fold and wrap around each other, forming a bisecting U-motif in order to effectively fill out the cavity of δ-CD.

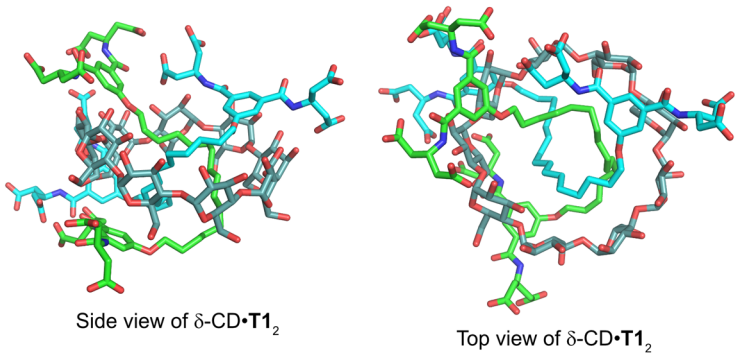


Figure 3.18. Cluster representative from clustering analysis of the MD simulation (500 ns) of complex δ-CD•**T1**$_2$ showing the proposed [3]-*pseudo*rotaxane binding mode.

3.4.5. Comparison of NMR results with DCLs of T1–T4

With these results in hand, it was now possible to understand the DCLs formed in the presence of **T1**–**T4** (Figure 3.4 and 3.5). **T3** and **T4** lead to the amplification of β-CD, as the smaller head group of these templates allows for the strong complexation of β-CD to the alkyl chain of these bolaamphiphiles. In early stages of the DCL with **T3**, small amounts of both γ-CD and δ-CD were also detected. Both of these CDs form higher order *pseudo*rotaxanes with **T3**, but the high intrinsic stability of β-CD in combination with its strong binding to **T3** leads to the eventual high selectivity for β-CD. However, in the early stages of the reaction, where there is a deficiency of hosts for binding to **T3** (the starting material α-CD does not bind to **T3**), means that complexes that involve the threading of multiple templates, such δ-CD•**T3**$_2$, are favored.

In the DCLs with **T1** and **T2**, an amplification of linear α-1,4-glucans with 8–20 glucose units (**G8**–**G20**) was observed, and a significant amplification of δ-CD was also observed for **T1**. In a previous study in the Beeren group, it was shown that linear α-1,4-glucans with 5 to 12 glucose units (**G5**–**G12**) bind to an amphiphilic probe with a C_{16} alkyl chain with increasing binding constants as the degree of polymerization increases, due to the formation of left handed helices around the hydrophobic alkyl chain.[28] For **G10**, **G11** and **G12**, the determined binding constants were $K_a = 3.0 \times 10^3\ M^{-1}$, $K_a = 6.9 \times 10^3\ M^{-1}$ and $K_a = 1.6 \times 10^4\ M^{-1}$, respectively. Similarly strong binding affinities for bolaamphiphiles **T1** and **T2** to linear α-1,4-glucans can be expected. At the same time, **T1** and **T2** do not bind strongly to any of the small CDs with high intrinsic stabilities, α-, β-, and γ-CD. The systems then evidently favor the production of the long linear α-1,4-glucans with strong binding to the templates, and for **T1**, δ-CD is also produced due to the moderately strong affinity of **T1** for δ-CD.

3.5. Optimization of glucan and template concentrations to maximize δ-CD yield

To explore the use of **T1** for the preparative scale synthesis of δ-CD, a screening of concentrations to optimize for the yield of δ-CD was performed. The libraries were prepared with α-CD and **T1** at various concentrations in sodium phosphate buffer (0.2 M, pH 7.5). The distributions of glucans shown in Figure 3.19 are from the time points in the DCLs where the highest yield of δ-CD was achieved. The time-resolved data, showing the evolution of the DCLs over time is shown in Supporting Figures S3.55–S3.57. The glucan concentration is quoted in mg/ml, which with the molecular weight of α-CD (973 g/mol) corresponds to about the same numerical value in mM. Thus, a DCL with a glucan concentration of 10 mg/ml and a **T1** concentration of 10 mM corresponds roughly to a 1:1 mixture α-CD and **T1**. When the glucan concentration (5–15 mg/ml) was varied while keeping the equivalents of **T1** constant at ~1 (entries 1–3), a clear trend was found, where higher concentrations favor δ-CD formation as well as the formation of long linear α-1,4-glucans **G8**–**G20**. This is the consequence of a classical phenomenon observed in dynamic combinatorial chemistry, where longer oligomers are favored by higher building block concentrations.[29] As such, the size of the CD subsystem (which contain only 6 to 9 glucose monomers) decreases and the yield of **G8**–**G20** increases with higher glucan concentrations. This is counteracted by a larger fraction of δ-CD within the CD subsystem, leading to an overall increase in the total fraction of δ-CD with increasing glucan concentration.

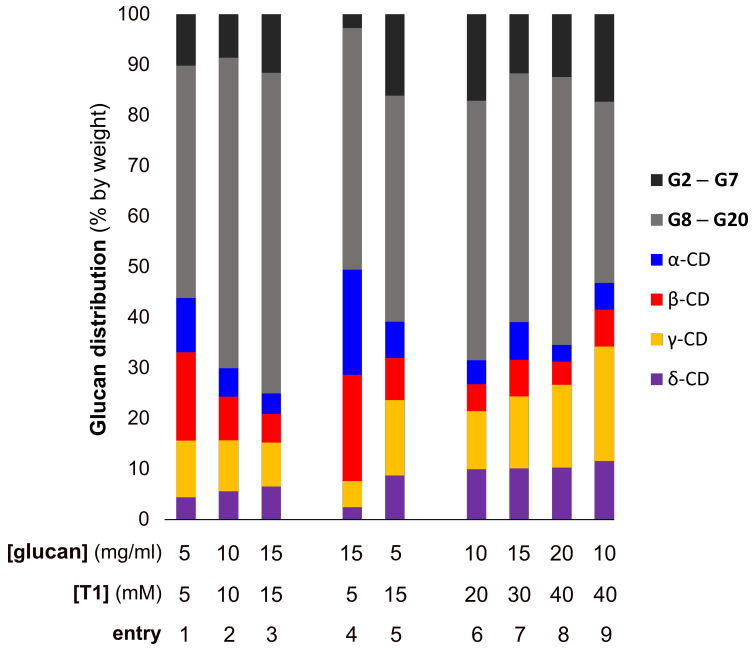


Figure 3.19. Optimization of the enzymatic synthesis of δ-CD using α-CD and **T1** at various concentrations in sodium phosphate buffer (0.2 M, pH 7.5). Distribution (% by weight) of CDs and linear α-1,4-glucans (minus **G1**) at the time point where the fraction of δ-CD was the highest.

Changing the concentrations to a ~3-fold excess of glucan (entry 4) diminished the yield of δ-CD. This is presumably due to a combination of two effects. With fewer templates available for binding, the dynamic system will favor the formation of the strongest binders (**G9**–**G20**) over a larger number of smaller oligomers. Also, the strong binding of **T1** to δ-CD is due to the formation of the δ-CD•**T1**$_2$ complex, meaning that a deficiency of **T1** is not favorable for the binding to δ-CD. Conversely, a 3-fold excess of **T1** (entry 5) was found to be favorable, leading to a significant increase in the fraction of δ-CD both within the CD subsystem and in the total system. Consequently, further screenings were carried out with various concentrations of glucan (10–20 mg/ml) and an excess of **T1** (20–40 mM) (entries 6–9). Increasing the excess of **T1** from 2-fold (entry 6) to 4-fold (entry 9), the CD subsystem increases in size at the expense of the fraction of **G9**–**G20**, but now primarily driven by an increase in the overall yield of γ-CD, with only a slight increase in the overall yield of δ-CD. As the NMR spectroscopy titrations show, **T1** also binds to γ-CD, albeit more weakly than to δ-CD. With a large excess of **T1**, the overall energy of the system is evidently minimized by forming a higher concentration of shorter oligomers with weaker binding to **T1**, rather than a lower concentration of longer oligomers (δ-CD and **G9**–**G20**) with stronger binding to **T1**.

3.6. Preparative-scale synthesis of δ-CD with recycling of T1

For the preparative scale synthesis of δ-CD, it was decided to proceed with the optimized conditions with 15 mg/ml glucan and 30 mM **T1**. Using α-CD as the starting material, the reaction mixture in sodium phosphate buffer (pH 7.5) was treated with commercial unfiltered CGTase. The use of the unfiltered solution of CGTase for the preparatory scale reactions led to the formation of extra peaks next to the peaks of **G8–G20** (presumably glyceryl glycosides

due to the presence of glycerol in the unfiltered stock solution of CGTase) as evident in the HPLC-chromatograms obtained after 18 hours reaction shown in Figure 3.20 a–c. The reaction was halted after 18 hours by acidification with trifluoroacetic acid (TFA), which led to the slow precipitation of template **T1**, which could be isolated and reused in another cycle of the same reaction. After five reaction cycles, δ-CD was isolated from the combined reaction mixtures by enzymatic treatment with α-glucosidase, to hydrolyze linear α-1,4-glucans (Figure 3.20d), and then precipitation and preparative HPLC, to give an unprecedented yield of 7.2% δ-CD in high purity (Figure 3.20e). **T1** was effectively recycled, with 83% of **T1** recovered again in high purity after five reaction cycles (Figure S3.54).

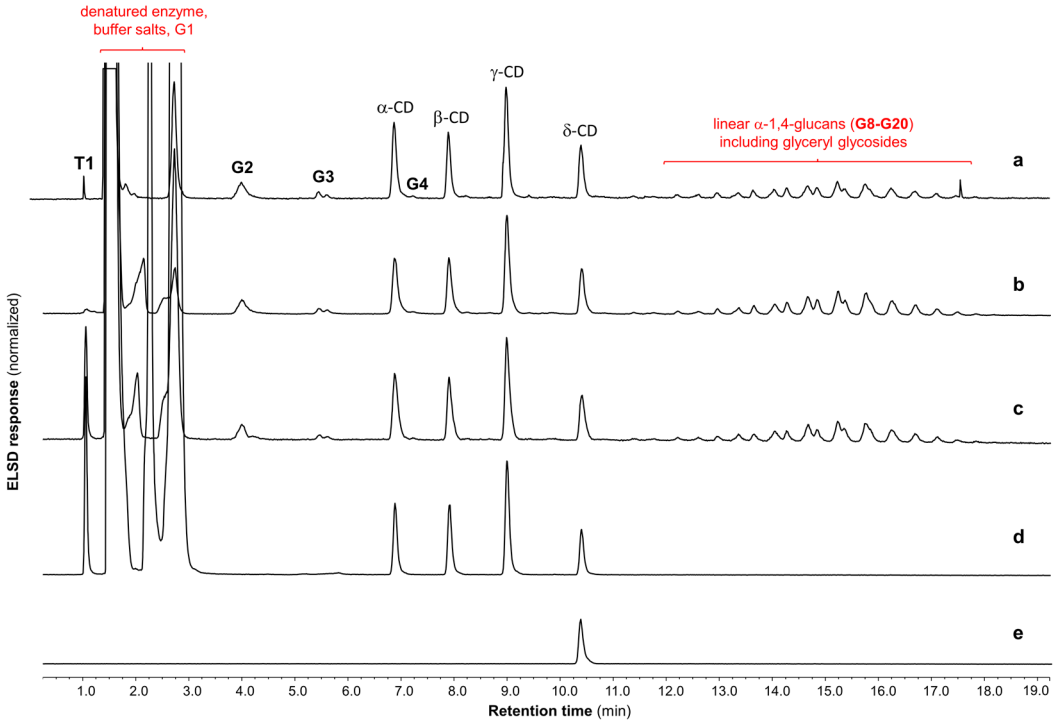


Figure 3.20. HPLC-ELSD chromatograms from the templated enzymatic synthesis of δ-CD using **T1**. (a) Supernatant of reaction cycle 1. (b) Supernatant of reaction cycle 5. (c) Combined supernatants of reaction cycles 1–5. (d) Reaction mixture after treatment with α-glucosidase. (e) Isolated δ-CD.

3.6.1. Starch as starting material for the synthesis of δ-CD

To further showcase the usefulness of **T1** for the enzymatic synthesis of δ-CD, starch was used as a starting material, as it is the glucan substrate used in the industrial synthesis of the native CDs. The reactions were set up with soluble starch (20 mg/ml) in the absence or presence of **T1** (20 mM) with CGTase in sodium phosphate buffer (0.2 M, pH 7.5) (Figure 3.21). The untemplated (Figure 3.21, a, b) reaction gave a CD distribution very similar to the untemplated reaction with α-CD (10 mg/ml) as starting material (Figure 3.3), with the exception that a small quantity of δ-CD (~1–2% of CDs by mass) was observed in the reaction with starch. In the templated reaction (Figure 3.21 c–e), the results were quite similar to the results obtained with 10 mg/ml α-CD as starting material and with 20 mM **T1** (Supporting Figure S3.57), which fits well with the observation that roughly 50% of the starch used in the reaction ends up in the

DCL of CDs and linear α-1,4-glucans within the initial hours (based on the calibrated peaks in the chromatograms).

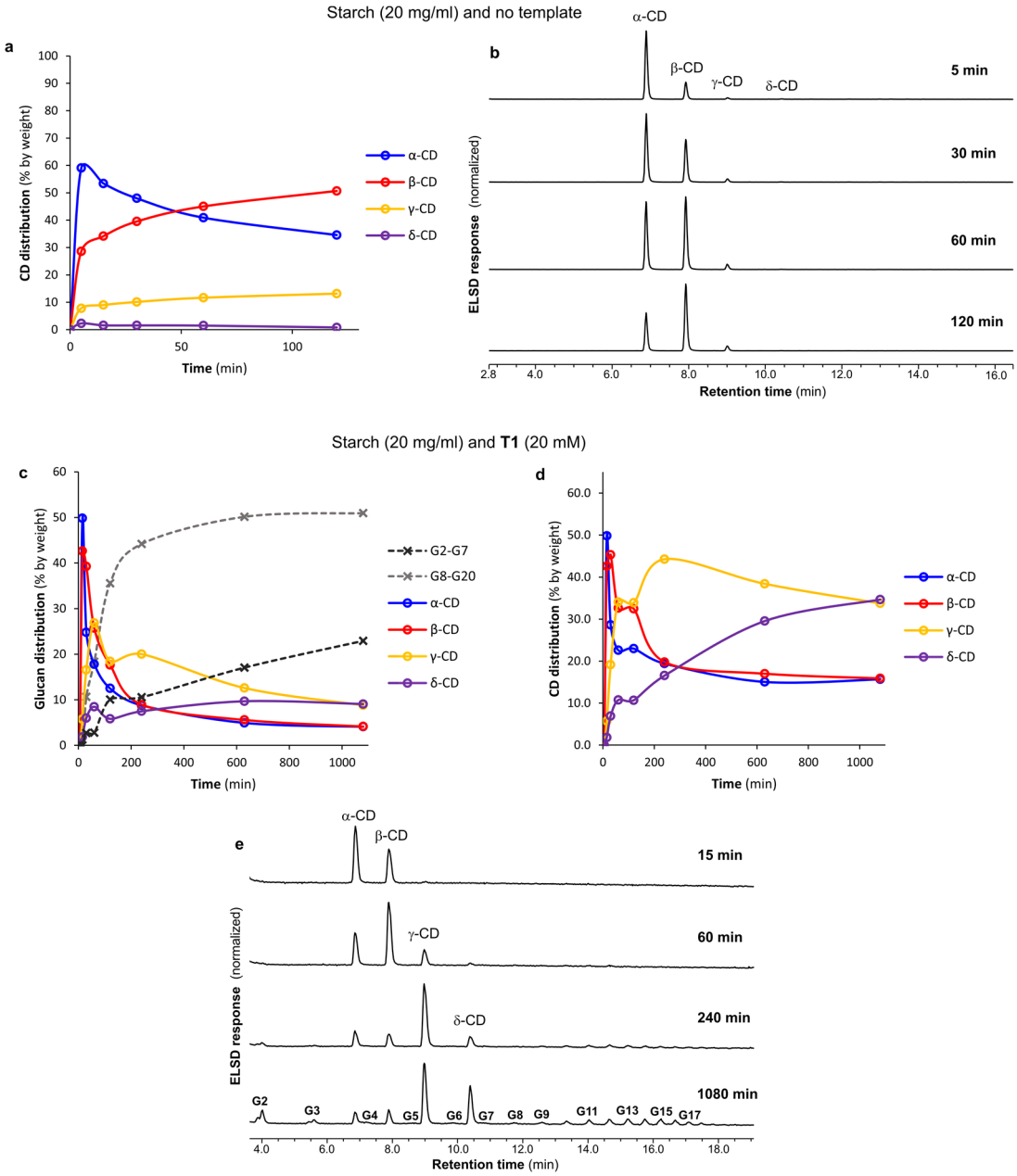


Figure 3.21. Enzymatic reactions with starch (20 mg/ml) in the absence or presence of template **T1** (20 mM). For the untemplated reaction, the reaction was performed in sodium phosphate buffer (50 mM, pH 7.5) and commercial stock CGTase (50 µl) was used, while the templated reaction was performed in sodium phosphate buffer (0.2 M, pH 7.5) and glycerol-free CGTase (98 µl per ml reaction mixture) was used. (a) and (d): CD distribution as a function of time. (c) Distribution of all glucans (minus **G1**), CDs (α, β, γ, δ-CD) and linear α-1,4-glucans (**G2**–**G7** and **G8**–**G20**), as a function of time (b) and (e): HPLC-ELSD chromatograms from the reactions.

3.6.2. Conclusion

In summary, a range of bolaamphiphiles were synthesized and explored for the templated enzymatic synthesis of the large-ring cyclodextrins. The templated enzymatic synthesis of δ-CD was achieved using α-CD and starch as starting materials and in unprecedented yields. The templates were designed to strategically block the binding of α-CD and β-CD, and to be recovered and reused repeatedly in enzymatic syntheses. The set of bolaamphiphile templates could be used to explore the supramolecular chemistry of δ-CD and the small CDs. By varying the size of the hydrophilic head groups and the length of the hydrophobic binding motifs, it was revealed how the interplay between these features leads to stark differences in the binding modes and binding affinities achieved. Moreover, the outcomes in templated CGTase-mediated dynamic combinatorial libraries of CDs and linear α-1,4-glucans could be rationalized on the basis of the binding affinities and stoichiometries of the various [n]-*pseudo*rotaxanes formed with β-, γ-, and δ-CD.

To more efficiently template the synthesis of LRCDs in future work, a bolaamphiphile template with a larger head group that blocks the binding of γ-CD might be useful. An even larger head group that blocks the binding of δ-CD (CD9) could be envisaged for the selective binding to CD10. However, there is a delicate balance here, as larger head groups might not allow the threading of more than one template and the formation of higher order *pseudo*rotaxanes, which was the comparatively strong binding event (formation of [3]-*pseudo*rotaxane) that led to the amplification of δ-CD in the work described here. To alleviate possible steric and/or electronic clashes between head groups of the bolaamphiphiles in higher order *pseudo*rotaxanes with LRCDs, a possibility would be to further lengthen the alkyl chains. For example, by extending the alkyl chain of **T3**, it could be explored whether the [4]-*pseudo*rotaxane that forms with **T3**, would be favored with the head groups further apart in the complex. Other possibilities for future work could involve the design of templates that contain two or more alkyl chains in a central binding motif, thus pre-organizing these alkyl chains for the binding to LRCDs such as δ-CD.

3.7. Experimental

3.7.1. Materials

All chemicals and solvents of HPLC (high performance liquid chromatography) grade or better were obtained from commercial suppliers and used as received for both synthesis and chromatographic analysis, unless stated otherwise. Soluble starch (product no. S9765) was purchased from commercial supplier Sigma Aldrich. The enzyme α-glucosidase (product code E-TSAGS, 1500 U/ml) from *Bacillus stearothermophilus* was purchased from commercial supplier Megazyme. Anhydrous solvents were obtained from an Innovative Technology PS-MS-7 Pure-solve solvent purification system. Flash column chromatography was performed using silica gel (40–63 μm) purchased from Merck while dry column vacuum chromatography (DCVC) was performed using silica gel (15–40 μm) purchased from Merck. High purity water used in both chromatographic analysis and in reactions was obtained using a Merck Milipore Synergy UV water purification system. Colorless Corning CoStar 0.65 ml centrifuge tubes were used for enzymatic reactions and sample preparation (dilution and centrifugation). Colorless 2 ml glass vials with PTFE-lined (polytetrafluoroethylene) screw-cap septa and 0.2 ml glass inserts were used for short-term sample storage and injection on HPLC equipment. NMR samples were analyzed in capped standard 5 mm borosilicate glass NMR tubes.

A stock solution of the enzyme CGTase derived from *Bacillus macerans* was received as a kind gift from Amano Enzyme, Inc., Nagoya, Japan and stored at 5 °C. According to specifications from the supplier, the stock solution contains approximately 20% glycerol. Previous work in the Beeren group has found that the presence of glycerol led to the formation of small amounts of glyceryl glycosides of linear α-1,4-glucans.[14] Glycerol was therefore removed by performing a 160-fold solvent exchange with water using a Pall MicroSep Advance Centrifugal Device (0.5–5 ml) with a 10 kDa Omega Membrane according to the procedure described by the manufacturer. The final volume after solvent exchange was kept constant. The activity of the enzyme stock solution was then determined as the production of β-CD (after 20 min) in a reaction starting from α-CD (10 mg/ml) in the presence of adamantane carboxylic acid (10 mM) (see previous publication for details of this reaction).[6] The activity was found to be 30% lower after solvent exchange. The obtained glycerol-free stock solution was stored at 5 °C and the activity of the glycerol-free enzyme stock solution was found to be unchanged after more than six months of storage. The commercial stock solution of CGTase and the glycerol-free stock solution of CGTase were used in the enzymatic reactions as indicated for each experiment.

3.7.2. Instruments and methods

Chromatographic analysis was performed on a Thermo Scientific Dionex Ultimate 3000 HPLC (ultra-high pressure) system equipped with a Waters Acquity BEH Amide 1.7 μm 2.1 × 150 mm column maintained at 30 °C and an autosampler maintained at 20 °C. Detection was carried out using an Agilent Technologies 1260 Infinity ELSD (evaporative light scattering detector), operating with the evaporator at 90° C, nebulizer at 70 °C and a N_2 gas flow of 1.0 L/min. For quantification from HPLC chromatograms, calibration curves for α, β, γ and δ-CD and linear

α-1,4-glucans (**Gn**) up to maltotriose (**G3**) were used to correct for differences in the ELS detector response for different oligosaccharides using a previously published method.[12] For linear α-1,4-glucans with more than three glucose units, the calibration parameters for **G3** were used. The gradient profile for HPLC was a linear gradient from 75% acetonitrile in water to 55% acetonitrile in water over 8 min with a flow rate of 0.6 ml/min (elutes α, β, γ and δ-CD and linear α-1,4-glucans up to maltooctaose). For separation of mixtures containing linear α-1,4-glucans with more than 9 glucose units, a linear gradient from 75% acetonitrile to 30% acetonitrile in water over 27 min with a flow rate of 0.4 ml/min was used. Both eluents contained 0.1% formic acid by volume.

NMR spectra were acquired on a Bruker Avance III 400 MHz NMR spectrometer equipped with a Prodigy broadband observe (BBO) probe, or a Bruker Avance III 400 MHz NMR spectrometer equipped with a BBO Smartprobe, or a 600 MHz Bruker Avance III spectrometer equipped with a 5 mm BBFO Probe or a Bruker Avance III 800 MHz NMR spectrometer equipped with a Bruker TCI Cryoprobe. NMR spectra were referenced to residual solvent peaks and measured at 298 K, unless stated otherwise. NMR spectra were processed in the software Topspin 3.6.2 or the software MestreNova 11.0. Data from NMR spectroscopy titrations was analyzed using the data analysis software OriginPro 2019. For 1:1 binding in fast exchange, the binding isotherms were fitted to a 1:1 binding model (equation 3.5).[4] For 1:1 binding in slow exchange, estimates of the binding constants were obtained from a linear plot of [HG] against [H][G]. The fitting procedures for mixed fast/slow exchange systems with 1:2 and 1:3 binding are described in detail in section S3.1. MALDI-TOF-MS experiments were carried out on a Bruker autoflex speed instrument. RP-UPLC-MS (ESI) analysis was performed on a S2 Waters Aquity RP-UPLC system with a diode array detector and equipped with a Thermo Accucore C18 column, 2.6 μm, 2.1 × 50 mm coupled to a SQD mass spectrometer. MALDI-TOF-HRMS and ESI-HRMS experiments were performed on a SolariX ESI/MALDI FTMS spectrometer. External calibration of the spectrometer was carried out using sodium trifluoroacetate cluster ions. Melting points were measured on a Stuart SMP 30 melting point apparatus. Preparative HPLC for the isolation of δ-CD was carried out with a Buchi Reveleris Prep Purification system equipped with an ELS detector and an XBridge BEH Amide OBD Prep column from Waters. Automated flash column chromatography was performed on a Buchi Reveleris Prep Purification system using pre-packed silica cartridges. A Heraeus Biofuge Pico centrifuge was used for the preparation of samples for HPLC analysis. An Eppendorf Centrifuge 5810R was used for preparative scale synthesis of δ-CD and during the synthesis of organic templates.

3.7.3. Computational details

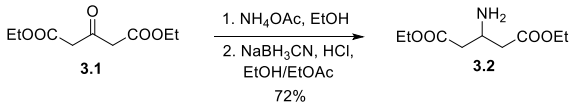
To investigate the binding mode of the 1:2 complex between δ-CD and template **T1** (δ-CD•**T1**$_2$), a molecular dynamics (MD) simulation was carried out. For the starting geometry of δ-CD a published crystal structure was used (CCDC deposition number: 186988),[30] while the starting geometry of **T1** as the octavalent carboxylate anion was constructed in the Chem3D molecular software. The starting geometry of the complex δ-CD•**T1**$_2$ was constructed in PyMOL by manually placing two **T1** templates in the cavity of δ-CD, with the alkyl chains threading the cavity and the hydrophilic head groups protruding from each side of δ-CD.

The MD simulation in explicit solvent was carried out using the Amber20 suite[23] employing the GLYCAM-06[24] force field for δ-CD and the general amber force field (GAFF)[25] for **T1**. Partial charges for **T1** were calculated using the antechamber program[31] implemented in Amber20 using the AM1-BCC method.[32] The LEaP program of Amber20 was used to solvate the complex using a truncated octahedron with a distance buffer of 20 Å between the complex and the edge of the periodic box. Water molecules were represented using the TIP3P[33] water model, and the system was neutralized using 8 sodium ions. The solvated and neutralized system was then minimized for 500 steps. The first 400 steps were performed using the steepest descent method followed by 100 steps of the conjugate gradient method. All MD simulations were carried out using a 8 Å cut-off for nonbonded interactions and the Particle Mesh Ewald method[34] for long range electrostatics. A time step of 2 fs was applied. Bonds to hydrogen atoms were constrained using the SHAKE algorithm.[35] The systems were then heated linearly from 10 K to 300 K within 0.3 ns followed by 0.7 ns of equilibration at 300 K in the *NVT* ensemble using the Langevin thermostat[36] with a collision frequency of 5 ps^{-1}. The systems were then equilibrated further for 4 ns using constant temperature (T = 300 K, using the Langevin thermostat with a collision frequency of 1 ps^{-1}) and constant pressure (p = 1 bar, using the Berendsen barostat[37] with isotropic position scaling). Finally, a 500 ns production run was performed using the same conditions as the equilibration protocol with constant pressure. Coordinates were saved every 2 ps for further analysis.

Analyses were performed using the CPPTRAJ module[38] in the Amber20 suite. Clustering analysis was performed to group similar structures from the trajectory into clusters. Clustering analysis was performed on the 500 ns production run based on average root-mean-square deviation (RMSD) of all non-hydrogen atom positions of δ-CD and **T1** using the "k-means" algorithm[27] to output 10 clusters. Cluster representatives were chosen based on the lowest distance to cluster centroid.

3.7.4. Synthetic protocols

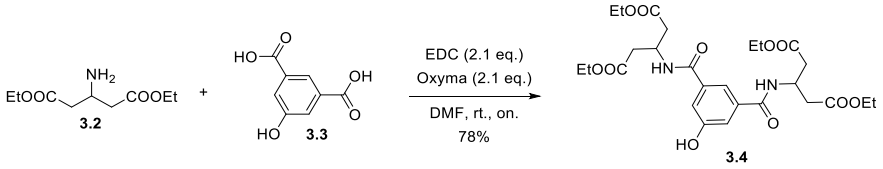
Synthesis of diethyl 3-aminoglutarate **3.2**



The procedure was adapted from ref.[8]
To a solution of diethyl 3-oxoglutarate **3.1** (20 g, 0.10 mol) in absolute ethanol (350 ml) with activated molecular sieves (12 g, 3Å) was added ammonium acetate (24 g, 0.38 mol). The suspension formed was stirred for 24 hours and then $NaBH_3CN$ (7.15 g, 0.11 mol) and HCl in EtOH/EtOAc (~1 M, 600 ml, made by the slow addition of acetyl chloride (60 ml) to a cooled (0 °C) solution of absolute ethanol (800 ml)) were added, resulting in a suspension with pH ~ 3 (by Litmus test). The suspension was stirred for one hour, then filtered and the resulting filtrate concentrated *in vacuo*. The resulting oil was dissolved in water (200 ml) and K_2CO_3 (40 g) added. Then aqueous NaOH (200 ml, 2 M) was added and the aqueous phase was extracted with ethyl acetate (3 × 200 ml) and dichloromethane (2 × 150 ml) until extraction of the product was completed as indicated by TLC ($KMnO_4$ stain). The combined organic extracts were washed with brine (3 × 80 ml) and dried over Na_2SO_4, filtered and evaporated *in vacuo* to leave a yellow oil. The crude oil was purified using flash column chromatography (5 % methanol in dichloromethane) to yield the product as a colorless oil. (14.5 g, 72%).

The ^{1}H NMR spectrum was in agreement with a reported spectrum.[39] **^{1}H NMR** (400 MHz, $CDCl_3$) δ 4.15 (q, *J* = 7.2 Hz, 4H), 3.63 (m, 1H), 2.50 (dd, *J* = 15.9, 4.7 Hz, 2H), 2.38 (dd, *J* = 15.9, 8.3 Hz, 2H), 1.26 (t, *J* = 7.2 Hz, 6H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 172.0, 60.7, 45.5, 42.0, 14.4. **HRMS** (ESI) *m/z*: $[M+H]^+$ calcd. for $C_9H_{18}NO_4$ 204.1231; found: 204.1235.

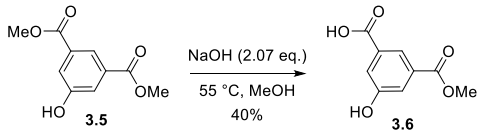
Synthesis of tetraethyl 3,3'-((5-hydroxyisophthaloyl)bis(azanediyl))diglutarate **3.4**



5-hydroxyisopthalic acid **3.3** (1.16 g, 6.37 mmol) and diethyl 3-aminoglutarate **3.2** (2.72 g, 13.4 mmol) were dissolved in DMF (60 ml) to form a colorless solution. Oxyma (1.90 g, 13.4 mmol) and EDC hydrochloride (2.60 g, 13.4 mmol) were then added, instantly forming a yellow solution which was stirred for 24 hours. Water (250 ml) was then added and the mixture was extracted with ethyl acetate (3 × 90 ml). The combined organic extracts were washed with hydrochloric acid (2 × 50 ml, 1 M), $NaHCO_3$ (sat., aq.) (3 × 80 ml) and brine (50 ml), dried over $MgSO_4$, filtered and evaporated *in vacuo*. The product was then isolated using dry column vacuum chromatography (*Ø*: 6 cm, length: 6 cm, crude product adsorbed on Celite, 100 ml fractions) eluting with a 3% ethyl acetate gradient in heptane starting from heptane : ethyl acetate, 57 : 43, to yield the product as a white solid (2.76 g, 78%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.54 (t, J = 1.5 Hz, 1H), 7.48 (d, J = 8.6 Hz, 2H), 7.39 (d, J = 1.5 Hz, 2H), 4.82 (dp, J = 8.6, 6.1 Hz, 2H), 4.15 (q, J = 7.1 Hz, 8H), 2.78 (dd, J = 16.1, 6.1 Hz, 8H), 1.24 (t, J = 7.1 Hz, 12H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 171.5, 166.4, 157.4, 135.8, 117.7, 116.8, 61.1, 44.0, 38.1, 14.3. **HRMS** (ESI) *m/z*: $[M+H]^+$ calcd. for $C_{26}H_{37}N_2O_{11}$ 553.2392; found: 553.2386. **M.P.**: 112.6–113.9 °C.

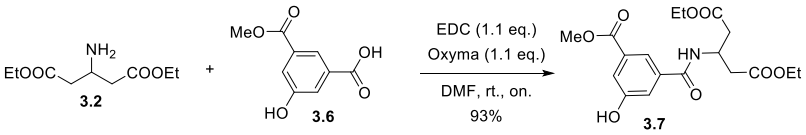
Synthesis of 3-hydroxy-5-(methoxycarbonyl)benzoic acid **3.6**



Dimethyl 5-hydroxyisophthalate **3.5** (2.03 g, 9.67 mmol) was dissolved in methanol (100 ml) and NaOH (aq., 20.13 ml, 1.00 2 M, 20.01 mmol) was added and the solution was stirred overnight. The flask was fitted with a Vigreux column and the reaction was heated to 55 °C for five hours after which the reaction mixture was left standing in the freezer (-18 °C) overnight. The solvent was evaporated *in vacuo* and the resulting solid was dissolved in water (50 mL). The aqueous solution was acidified to pH ~ 1–2 by the dropwise addition of hydrochloric acid (6 M) leading to precipitation of a white solid. The aqueous suspension was extracted with ethyl acetate (3 × 60 mL) and the combined organic phases were dried over $MgSO_4$, filtered and evaporated *in vacuo* to yield the crude product as a white solid. The crude product was purified by dry column vacuum chromatography (*Ø*: 4 cm, length: 7 cm, 30 ml fractions, crude product adsorbed on Celite). Eluent system: Dichloromethane (A) / methanol (B) / formic acid (C). Fraction 1: 100 % A, fraction 2–4: A/B 99/1, fraction 5: A/B/C 98/1/1, fraction 6-9: gradient elution with 1 % increments of B from A/B/C 97/2/1 to 94/5/1, fraction 10-15: A/B/C 94/5/1. The product elutes in fractions 11-14, which were collected and evaporated *in vacuo* to yield the product as a white solid (0.74 g, 40%).

The ^{1}H NMR spectrum was in agreement with a reported spectrum.[40] **^{1}H NMR** (400 MHz, DMSO-d_6) δ 13.17 (br, 1H), 10.24 (br, 1H), 7.98 – 7.90 (m, 1H), 7.54 (m, 2H), 3.86 (s, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 166.4, 165.5, 157.8, 132.6, 131.2, 120.5, 120.4, 119.7, 52.4. **MS** (ESI) *m/z*: $[M-H]^-$ calcd. for $C_9H_7O_5$ 195.03; found: 194.88.

Synthesis of diethyl 3-(3-hydroxy-5-(methoxycarbonyl)benzamido)pentanedioate **3.7**

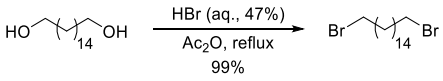


3-hydroxy-5-(methoxycarbonyl)benzoic acid **3.7** (0.73 g, 3.71 mmol) and diethyl 3-aminoglutarate **3.2** (0.99 g, 4.88 mmol) were dissolved in DMF (60 ml) to form a colorless solution. Oxyma (0.58 g, 4.10 mmol) and EDC hydrochloride (0.79 g, 4.11 mmol) were then added as a solution in DMF (30 ml) to the reaction mixture, instantly forming a yellow solution which was stirred overnight. Water (250 ml) was then added and the mixture was extracted with ethyl acetate (3 × 80 ml). The combined organic extracts were washed with hydrochloric acid (60 ml, 1 M), $NaHCO_3$ (sat., aq.) (4 × 40 ml, or until the organic phase was colorless) and

brine (60 ml), dried over $MgSO_4$, filtered and evaporated *in vacuo* to yield the product as a white solid (1.29 g, 92%), which was used in the next step without further purification.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.83 (t, J = 1.5 Hz, 1H), 7.77 (dd, J = 2.6, 1.5 Hz, 1H), 7.71 (br, 1H), 7.68 (dd, J = 2.6, 1.5 Hz, 1H), 7.46 (d, J = 8.9 Hz, 1H), 4.84 (dp, J = 8.9, 5.9 Hz, 1H), 4.16 (q, J = 7.1 Hz, 4H), 3.92 (s, 3H), 2.78 (dd, J = 16.2, 5.9 Hz, 4H), 1.26 (t, J = 7.1 Hz, 6H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 171.3, 166.3, 166.1, 157.2, 135.3, 131.9, 120.1, 119.5, 118.8, 61.1, 52.5, 43.8, 37.8, 14.1. **HRMS** (ESI) *m/z*: $[M+Na]^+$ calcd. for $C_{18}H_{23}NO_8Na$ 404.1316; found: 404.1322. **M.P**.: 128–131 °C.

Synthesis of 1,16-dibromohexadecane

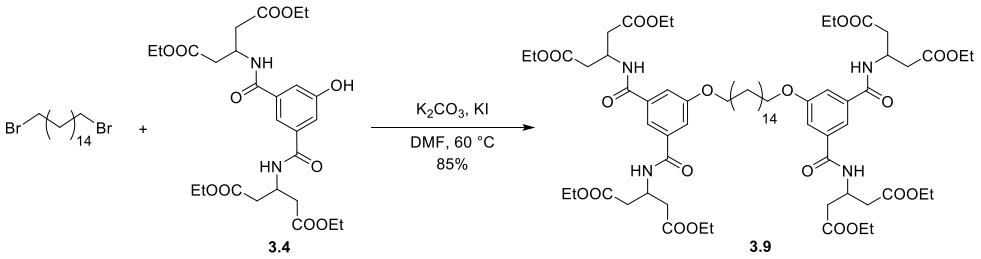


The procedure was adapted from ref.[9]

To acetic anhydride (17 ml) and 1,16-hexadecanediol (0.63 g, 2.4 mmol) cooled to 0 °C was added HBr (10 ml, aq., 48%) dropwise after which the solution was refluxed for 24 hours. The mixture was allowed to cool to room temperature, water (50 ml) was then added and the solution was extracted with hexanes (100 ml). The organic phase was washed with water (3 × 80 mL), dried over Na_2SO_4 and evaporated *in vacuo* to yield the product as a white solid (0.93 g, 99%).

The ^{1}H NMR spectrum was in agreement with a reported spectrum.[41] **^{1}H NMR** (400 MHz, $CDCl_3$) δ 3.41 (t, J = 6.9 Hz, 4H), 1.85 (p, J = 6.9 Hz, 4H), 1.41 (m, 4H), 1.27 (m, 20H).

Synthesis of octaester **3.9**

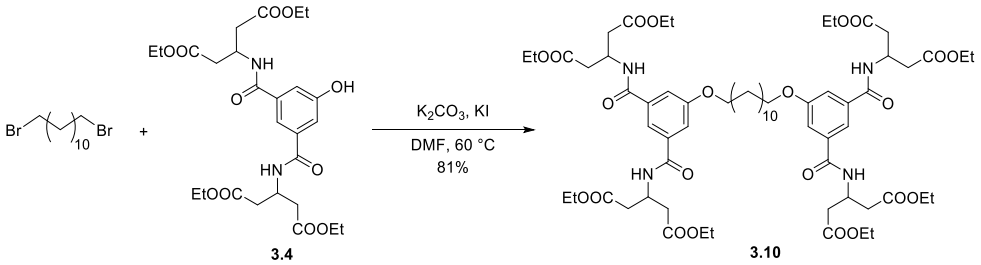


Tetraethyl 3,3'-((5-hydroxyisophthaloyl)bis(azanediyl))diglutarate **3.9** (1.14 g, 2.1 mmol), 1,16-dibromohexadecane (0.32 g, 0.81 mmol), oven-dried K_2CO_3 (0.59 g, 4.2 mmol) and KI (20 mg, 0.12 mmol) were suspended in dry DMF (70 ml) under a nitrogen atmosphere and stirred at 60 °C for two days. The solvent was removed *in vacuo* and the resulting mixture of oil and solids was partitioned between water (80 ml) and ethyl acetate (80 ml). The organic phase was isolated and the aqueous phase was extracted once more with ethyl acetate (80 ml). The combined organic phases were washed with hydrochloric acid (30 ml, 1 M), aqueous NaOH (70 ml, 1 M) and brine (60 ml), dried over Na_2SO_4, filtered, and evaporated *in vacuo* to yield the crude product as a yellow-white solid. The product was isolated using dry column vacuum chromatography (*Ø*: 3 cm, length: 5 cm, 50 ml fractions, gradient elution from 10%–50% acetonitrile in toluene in 5% increments). The product eluted in fractions 7-11, which

were collected and the solvent removed *in vacuo* to yield the product (0.95 g, 85%) as a white solid.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.67 (t, J = 1.4 Hz, 2H), 7.42 (d, J = 1.4 Hz, 4H), 7.24 (d, J = 8.8 Hz, 4H), 4.81 (dp, J = 8.8, 6.0 Hz, 4H), 4.16 (q, J = 7.2 Hz, 16H), 4.02 (t, J = 6.5 Hz, 4H), 2.86 – 2.65 (m, 16H), 1.79 (m, 4H), 1.45 (m, 4H), 1.40 – 1.20 (m, 44H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 171.2, 165.7, 159.5, 136.0, 117.1, 116.2, 68.6, 60.90, 43.6, 37.7, 29.7, 29.63, 29.59, 29.4, 29.1, 26.0, 14.2. (two aliphatic signals overlapping). **HRMS** (ESI) *m/z*: $[M+H]^+$ calcd. for $C_{68}H_{103}N_4O_{22}$ 1327.7059; found: 1327.7088 **M.P**.: 92–94 °C.

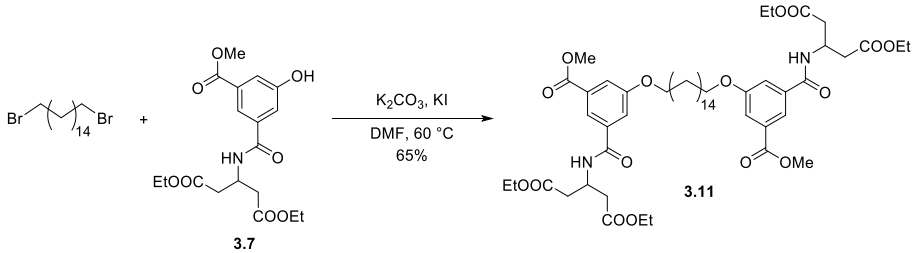
Synthesis of octaester **3.10**



Tetraethyl 3,3'-((5-hydroxyisophthaloyl)bis(azanediyl))diglutarate **3.4** (0.475 g, 0.86 mmol), 1,12-dibromododecane (0.105 g, 0.32 mmol), oven-dried K_2CO_3 (0.27 g, 1.95 mmol) and KI (26 mg, 0.157 mmol) were suspended in dry DMF (25 ml) and stirred at 60 °C for three days. The solvent was removed *in vacuo* and the resulting mixture of oil and solids was partitioned between water (60 ml) and ethyl acetate (70 ml). The organic phase was isolated and the aqueous phase was extracted once more with ethyl acetate (60 ml). The combined organic phases were washed with hydrochloric acid (50 ml, 1 M), aqueous NaOH (70 ml, 1 M) and brine (60 ml), dried over Na_2SO_4, filtered, and evaporated *in vacuo* to yield the crude product as a yellow-white solid. The product was then isolated using automated flash column chromatography (25 g silica cartridge, gradient elution from 0%–70% acetonitrile in toluene) and the solvent removed *in vacuo* to yield the product as a white solid (0.33 g, 81%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.66 (t, J = 1.4 Hz, 2H), 7.41 (d, J = 1.4 Hz, 4H), 7.24 (d, J = 8.8 Hz, 4H), 4.81 (dp, J = 8.8, 6.0 Hz, 4H), 4.16 (q, J = 7.1 Hz, 16H), 4.02 (t, J = 6.6 Hz, 4H), 2.92 – 2.63 (m, 16H), 1.79 (p, J = 6.6 Hz, 4H), 1.45 (m, 4H), 1.26 (m, 36H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 171.2, 165.7, 159.5, 136.0, 117.1, 116.2, 68.6, 60.9, 43.6, 37.7, 29.55, 29.54, 29.4, 29.1, 26.0, 14.2. **HRMS** (ESI) *m/z:* $[M+H]^+$ calcd. for $C_{64}H_{95}N_4O_{22}$ 1271.6433; found: 1271.6446 **M.P**.: 104–107 °C.

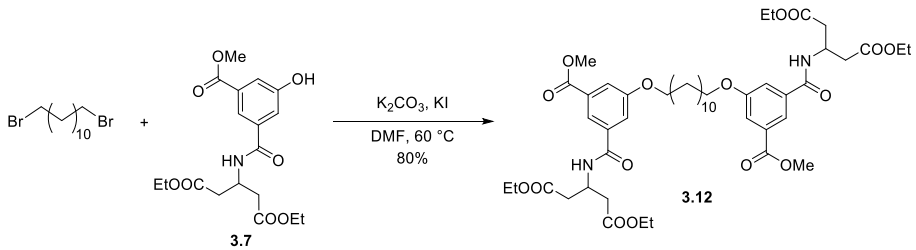
Synthesis of hexaester **3.11**



Diethyl 3-(3-hydroxy-5-(methoxycarbonyl)benzamido)pentanedioate **3.7** (229 mg, 0.60 mmol), 1,16-dibromohexadecane (119 mg, 0.31 mmol), K_2CO_3 (230 mg, 1.66 mmol) and KI (75 mg, 0.45 mmol) were suspended in DMF (20 mL) and stirred at 60 °C for three days. Water (140 mL) was added and the resulting solution was extracted with ethyl acetate (2 × 100 ml, then 50 ml). The combined organic phases were washed with hydrochloric acid (25 ml, 1 M), aqueous NaOH (75 ml, 2 M) and brine (50 ml), dried over $MgSO_4$ and evaporated *in vacuo* to yield the crude product as a yellow-white solid. The product was then isolated using automated flash column chromatography (12 g silica cartridge, gradient elution from 10%–100% ethyl acetate in *n*-heptane) and the solvent removed *in vacuo* to yield the product as a white solid (274 mg, 65%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.90 (m, 2H), 7.67 (m, 2H), 7.55 (m, 2H), 7.24 (d, J = 8.8 Hz, 2H), 4.82 (dp, J = 8.8, 5.9 Hz, 2H), 4.17 (q, J = 7.1 Hz, 8H), 4.03 (t, J = 6.6 Hz, 4H), 3.93 (s, 6H), 2.89 – 2.64 (m, 8H), 1.79 (p, J = 6.6 Hz, 4H), 1.51 – 1.40 (m, 4H), 1.40 – 1.20 (m, 32H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 171.4, 166.4, 165.7, 159.6, 136.0, 131.9, 119.6, 118.6, 118.3, 68.8, 61.1, 52.6, 43.7, 37.8, 29.83, 29.77, 29.7, 29.5, 29.3, 26.1, 14.3 (two aliphatic signals overlapping). **HRMS** (ESI) *m/z*: $[M+H]^+$ calcd. for $C_{52}H_{77}N_2O_{16}$ 985.5268; found: 985.5330. **M.P**.: 109–111 °C.

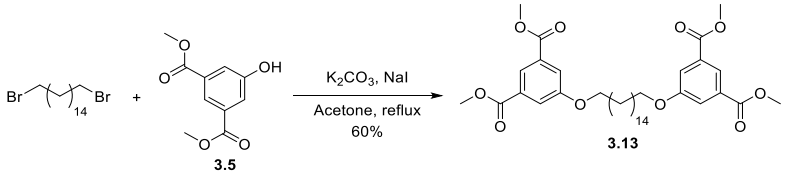
Synthesis of hexaester **3.12**



Diethyl 3-(3-hydroxy-5-(methoxycarbonyl)benzamido)pentanedioate **3.7** (404 mg, 1.06 mmol), 1,12-dibromododecane (149 mg, 0.45 mmol), K_2CO_3 (316 mg, 2.29 mmol) and KI (140 mg, 0.84 mmol) were suspended in DMF (25 ml) and stirred at 60 °C for three days. Water (100 ml) was added and the resulting solution was extracted twice with ethyl acetate (100 ml, then 80 ml). The combined organic phases were washed with hydrochloric acid (25 ml, 1 M), aqueous NaOH (75 ml, 2 M) and brine (50 ml), dried over $MgSO_4$ and evaporated *in vacuo* to yield the crude product as an off-white solid. The product was then isolated using automated flash column chromatography (12 g silica cartridge, gradient elution from 10%–100% ethyl acetate in *n*-heptane) and the solvent removed *in vacuo* to yield the product as a white solid (292 mg, 80%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.91 (m, 2H), 7.67 (m, 2H), 7.55 (t, J = 2.1 Hz, 2H), 7.24 (d, J = 8.8 Hz, 2H), 4.82 (dp, J = 8.8, 5.9 Hz, 2H), 4.17 (q, J = 7.1 Hz, 8H), 4.03 (t, J = 6.6 Hz, 4H), 3.93 (s, 6H), 2.90 – 2.64 (m, 8H), 1.80 (p, J = 6.6 Hz, 4H), 1.46 (m, 4H), 1.40 – 1.20 (m, 24H). **^{13}C NMR** (101 MHz, $CDCl_3$) 171.4, 166.4, 165.7, 159.6, 136.0, 131.9, 119.6, 118.5, 118.3, 68.8, 61.1, 52.6, 43.7, 37.8, 29.7, 29.7, 29.5, 29.3, 26.1, 14.3. **HRMS** (ESI) *m/z*: $[M+Na]^+$ calcd. for $C_{48}H_{68}N_2O_{16}Na$ 951.4461; found: 951.4492. **M.P**.: 107–109 °C.

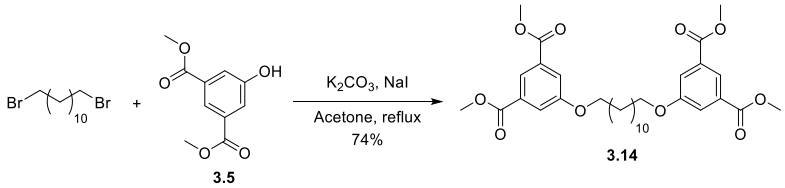
Synthesis of tetraester **3.13**



Dimethyl 5-hydroxyisophthalate **3.5** (154 mg, 0.73 mmol) and 1,16-dibromohexadecane (122 mg, 0.32 mmol) were suspended in acetone (45 ml). K_2CO_3 (108 mg, 0.78 mmol) and NaI (a few crystals) were added and the reaction mixture was refluxed for three days. The solvent was removed by boiling at atmospheric pressure and the resulting solid residue was suspended in a mixture of aqueous K_2CO_3 (50 ml, 0.5 M), CH_2Cl_2 (40 ml) and EtOAc (20 ml) by vigorous stirring. The mixture was transferred to a separatory funnel and the organic phase was collected and dried over $MgSO_4$. The crude product was purified by dry column vacuum chromatography (silica, *Ø*: 3 cm, length: 6 cm, 25 mL fractions, crude product adsorbed on celite, gradient elution with 10 % increments of ethyl acetate in heptane) to yield the product as a white solid (122 mg, 60%).

^{1}H NMR (400 MHz, $CDCl_3$) δ ^{1}H NMR (400 MHz, Chloroform-*d*) δ 8.25 (t, *J* = 1.4 Hz, 2H), 7.74 (d, *J* = 1.4 Hz, 4H), 4.03 (t, *J* = 6.6 Hz, 4H), 3.94 (s, 12H), 1.80 (p, *J* = 6.6 Hz, 4H), 1.44 (m, 4H), 1.39 – 1.22 (m, 20H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ ^{13}C NMR (101 MHz, Chloroform-*d*) δ 166.4, 159.4, 131.8, 122.9, 120.0, 68.8, 52.5, 29.8, 29.8, 29.7, 29.5, 29.3, 26.1 (two aliphatic signals overlapping) **HRMS** (ESI+) *m/z*: $[C_{36}H_{50}O_{10}+H]^+$; calc.: 643.3477 ; found: 643.3483. **M.P.**: 105 – 107 °C.

Synthesis of tetraester **3.14**

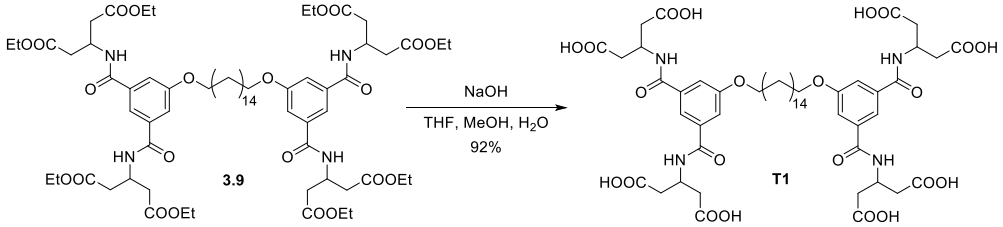


Dimethyl 5-hydroxyisophthalate **3.5** (295 mg, 1.40 mmol) and 1,12-dibromododecane (200 mg, 0.60 mmol) were added to acetone (90 ml) to form a suspension. K_2CO_3 (210 mg, 1.52 mmol) and NaI (a few crystals) were added and the reaction mixture was refluxed for three days. The solvent was removed by boiling at atmospheric pressure and the resulting solid residue was dispersed in a mixture of water (50 mL) and CH_2Cl_2 (100 ml), forming two phases; an aqueous suspension and an organic suspension. The organic suspension was collected and washed with NaOH (aq.) (30 ml, 2 M) and evaporated *in vacuo* to yield the crude product as a white solid. The crude product was purified by automated flash column chromatography (12 g silica cartridge, gradient elution from 0% to 100% ethyl acetate in *n*-heptane) to yield the product as a white solid (266 mg, 74%).

The ^{1}H NMR spectrum was in agreement with a reported spectrum.[42] **^{1}H NMR** (400 MHz, $CDCl_3$) δ 8.25 (t, *J* = 1.5 Hz, 2H), 7.74 (d, *J* = 1.5 Hz, 4H), 4.03 (t, *J* = 6.5 Hz, 4H), 3.93 (s,

12H), 1.80 (m, 4H), 1.47 (m, 4H), 1.41 – 1.27 (m, 12H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 166.4, 159.4, 131.8, 122.9, 120.0, 68.8, 52.5, 29.70, 29.69, 29.5, 29.3. **HRMS** (ESI+) *m/z*: $[C_{32}H_{42}O_{10}+H]^+$; calc.: 587.2851; found: 587.2870. **M.P.**: 102 – 104 °C.

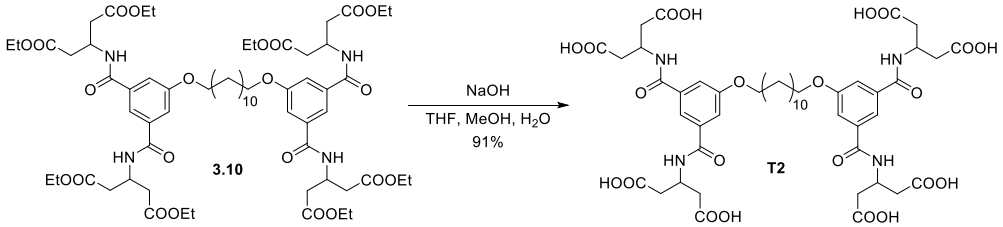
Synthesis of template **T1**



Octaester **3.9** (0.67 g, 0.50 mmol) was dissolved in a mixture of tetrahydrofuran (15 ml) and methanol (15 ml). Then aqueous NaOH (17 ml, 1 M) was added and the reaction was stirred at room temperature for 24 hours. The solvent was then removed *in vacuo* until ~12 ml of solvent was left. Hydrochloric acid (4 ml, 6 M) was then added and the mixture left at room temperature overnight, leading to the precipitation of the product, which was isolated by vacuum filtration, washed with water and dried *in vacuo* to yield the product as a white solid (0.54 g, 92%).

^{1}H NMR (800 MHz, D_2O phosphate buffer, 50 mM, pH 7.5) δ 7.59 (t, J = 1.5 Hz, 2H), 7.39 (d, J = 1.5 Hz, 4H), 4.60 (p, J = 7.0 Hz, 4H), 4.09 (t, J = 6.6 Hz, 4H), 2.46 (m, 16H), 1.73 (m, 4H), 1.39 (m, 4H), 1.28 (m, 4H), 1.26 – 1.17 (m, 16H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 172.2, 165.0, 158.3, 136.0, 118.7, 115.6, 67.9, 44.0, 38.4, 29.15, 29.13, 29.10, 29.09, 28.9, 28.7, 25.6. **HRMS** (ESI) *m/z*: $[M-H]^-$ calcd. for $C_{52}H_{69}N_4O_{22}$ 1101.4409; found: 1101.4475. **M.P**.: 142–145 °C.

Synthesis of template **T2**

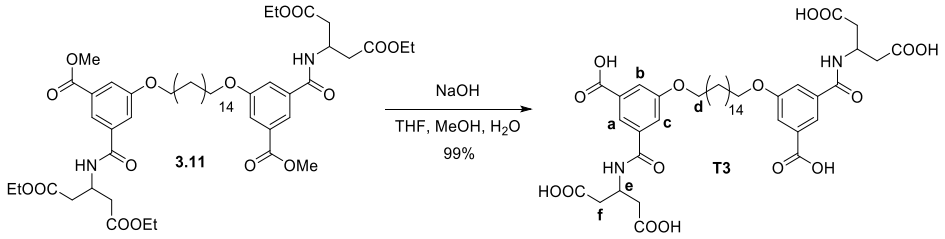


Octaester **3.10** (306 mg, 0.24 mmol) was dissolved in a mixture of tetrahydrofuran (15 ml) and methanol (15 ml). Then aqueous NaOH (15 ml, 1 M) was added and the reaction was stirred at room temperature for 24 hours. The solvents were then removed *in vacuo* until ~10 ml of solvent was left. Hydrochloric acid (4 ml, 6 M) was then added and the mixture left at room temperature overnight, leading to the precipitation of the product, which was isolated by centrifugation, washed twice with water and dried *in vacuo* to yield the product as a white solid (230 mg, 91%).

^{1}H NMR (400 MHz, D_2O phosphate buffer, 50 mM, pH 7.5) δ 7.68 (t, J = 1.5 Hz, 2H), 7.49 (d, J = 1.5 Hz, 4H), 4.69 (p, J = 7.0 Hz, 4H), 4.19 (t, J = 6.7 Hz, 4H), 2.67 – 2.48 (m, 16H), 1.84 (p, J = 6.7 Hz, 4H), 1.49 (m, 4H), 1.44 – 1.30 (m, 12H). **^{13}C NMR** (101 MHz, DMSO-d_6)

δ 172.7, 165.4, 158.8, 136.4, 119.1, 116.1, 68.4, 44.5, 38.9, 29.6, 29.4, 29.2, 26.1 (two aliphatic signals overlapping). **HRMS** (ESI) *m/z*: $[M\text{-}H]^-$ calcd. for $C_{48}H_{61}N_4O_{22}$ 1045.3783; found: 1045.3798. **M.P**.: 155–158 °C.

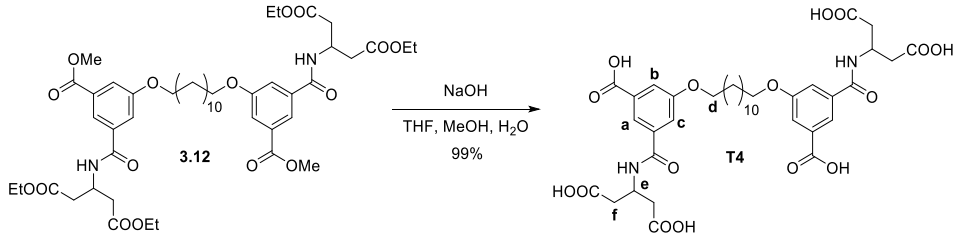
Synthesis of template **T3**



Hexaester **3.11** (187 mg, 0.19 mmol) was dissolved in a mixture of tetrahydrofuran (10 ml) and methanol (20 ml). Then aqueous NaOH (10 ml, 1 M) was added and the reaction was stirred at room temperature for 24 hours. The solvents were then removed *in vacuo* until ~8 ml of solvent was left. Hydrochloric acid (6 ml, 6 M) was then added, leading to the precipitation of the product, which was isolated by centrifugation, washed twice with water and dried *in vacuo* to yield the product as a white solid (160 mg, 99%).

^{1}H NMR (400 MHz, D_2O phosphate buffer, 50 mM, pH 7.5) δ 7.80 (s, 2H, Ha), 7.63 (s, 2H, Hb), 7.44 (s, 2H, Hc), 4.68 (p, J = 7.0 Hz, 2H, He), 4.18 (t, J = 6.4 Hz, 4H, Hd), 2.56 (m, 8H, Hf), 1.82 (p, J = 6.4 Hz, 4H), 1.47 (m, 4H), 1.43 – 1.16 (m, 20H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 172.2, 166.7, 164.5, 158.6, 136.1, 132.3, 120.3, 117.9, 117.0, 68.0, 44.0, 38.4, 29.1, 29.1, 29.0, 28.8, 28.6, 25.5 (two aliphatic signals overlapping). **HRMS** (ESI) *m/z*: $[M\text{-}H]^-$ calcd. for $C_{42}H_{55}N_2O_{16}$ 843.3557; found: 843.3576. **M.P**.: 176–178 °C.

Synthesis of template **T4**

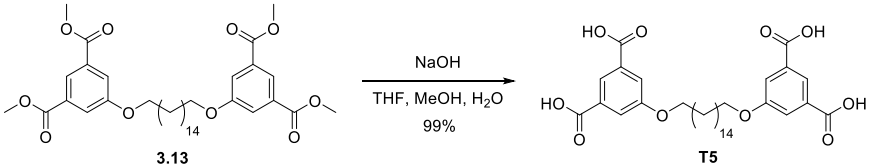


Hexaester **3.12** (0.28 g, 0.30 mmol) was dissolved in a mixture of tetrahydrofuran (15 ml) and methanol (30 ml). Then aqueous NaOH (15 ml, 1 M) was added and the reaction was stirred at room temperature for 24 hours. The solvents were then removed *in vacuo* until ~10 ml of solvent was left. Hydrochloric acid (6 ml, 6 M) was then added, leading to the precipitation of the product, which was isolated by centrifugation, washed twice with water and dried *in vacuo* to yield the product as a white solid (0.24 g, 99%).

^{1}H NMR (400 MHz, D_2O phosphate buffer, 50 mM, pH 7.5) δ 7.79 (t, J = 1.5 Hz, 2H, Ha), 7.63 (dd, J = 2.5, 1.5 Hz, 2H, Hb), 7.44 (dd, J = 2.5, 1.5 Hz, 2H, Hc), 4.68 (p, J = 7.0 Hz, 2H, He), 4.19 (t, J = 6.6 Hz, 4H, Hd), 2.65 – 2.48 (m, 8H, Hf), 1.84 (p, J = 6.6 Hz, 4H), 1.49 (m, 4H), 1.44 – 1.30 (m, 12H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 172.2, 166.8, 164.5, 158.6,

136.1, 132.3, 120.3, 117.9, 117.0, 68.0, 44.0, 38.5, 29.0, 28.8, 28.6, 25.5 (two aliphatic signals overlapping). **HRMS** (ESI) *m/z*: $[M-H]^-$ calcd. for $C_{38}H_{48}N_2O_{16}$ 787.2931; found: 787.2933. **M.P.**: 179 – 182 °C.

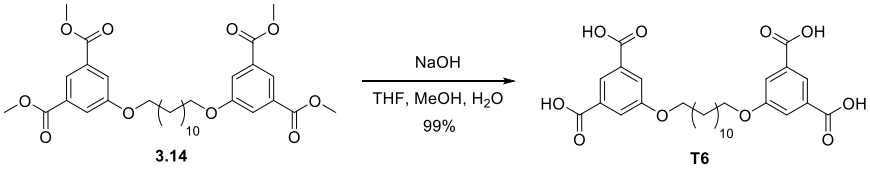
Synthesis of template **T5**



Tetraester **3.13** (103 mg, 0.16 mmol) was dissolved in a mixture of tetrahydrofuran (20 ml) and methanol (4 ml). Then aqueous NaOH (4 ml, 1 M) was added and the reaction was stirred at room temperature for 3 days. The solvents were then removed *in vacuo* until ~3 ml of solvent was left. Hydrochloric acid (2 ml, 6 M) was then added, leading to the precipitation of the product, which was isolated by centrifugation, washed twice with water and dried *in vacuo* to yield the product as a white solid (93 mg, 99%).

^{1}H NMR (400 MHz, D_2O phosphate buffer, 50 mM, pH 7.5) δ 7.90 (t, J = 1.4 Hz, 2H), 7.56 (d, J = 1.4 Hz, 4H), 4.14 (t, J = 6.6 Hz, 4H), 1.78 (p, J = 6.6 Hz, 4H), 1.44 (m, 4H), 1.38 – 1.20 (m, 20H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 166.4, 158.8, 132.6, 122.1, 119.0, 68.1, 29.0, 29.0, 29.0, 28.7, 28.5, 25.4 (two aliphatic signals overlapping). **HRMS** (ESI-) *m/z*: $[C_{32}H_{42}O_{10}-H]^-$; calc.: 585.2705; found: 585.2709. **M.P.**: 243 – 247 °C.

Synthesis of template **T6**



Tetraester **3.14** (218 mg, 0.37 mmol) was dissolved in a mixture of tetrahydrofuran (40 ml) and methanol (8 ml). Then aqueous NaOH (8 ml, 1 M) was added and the reaction was stirred at room temperature for 3 days. The solvents were then removed *in vacuo* until ~6 ml of solvent was left. Hydrochloric acid (6 ml, 6 M) was then added, leading to the precipitation of the product, which was isolated by centrifugation, washed twice with water and dried *in vacuo* to yield the product as a white solid (194 mg, 99%).

^{1}H NMR (400 MHz, D_2O phosphate buffer, 50 mM, pH 7.5) δ 7.89 (t, J = 1.4 Hz, 2H), 7.56 (d, J = 1.4 Hz, 4H), 4.15 (t, J = 6.6 Hz, 4H), 1.79 (p, J = 6.6 Hz, 4H), 1.45 (m, 4H), 1.40 – 1.22 (m, 12H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 166.4, 158.8, 132.6, 122.1, 119.0, 68.1, 28.98, 28.96, 28.7, 28.5, 25.4 (two aliphatic signals overlapping). **HRMS** (ESI-) *m/z*: $[C_{28}H_{34}O_{10}-H]^-$; calc.: 529.2079; found: 529.2080. **M.P.**: 272 – 274 °C.

3.7.5. Preparative scale templated enzymatic synthesis of δ-CD

T1 (121 mg, 0.11 mmol) and α-CD (54 mg, 0.057 mmol) were suspended in water (2 ml) in a 15 ml centrifuge tube. Aqueous solutions of Na_2HPO_4 (73 µl, 0.5 M) and NaOH (0.21 ml, 4 M) were then added and **T1** was dissolved using agitation. The pH was then adjusted using NaOH (aq., ~20 µl, 4 M) to give a final pH of 7.5. Water (1.2 ml) was then added, and the reaction was started by the addition of unfiltered commercial CGTase stock solution (0.157 ml) to give a final phosphate concentration of 10 mM, template concentration of 30 mM and α-CD concentration of 15 mg/ml and a total reaction volume of 3.65 ml. The reaction was kept at room temperature for 18 hours, after which the reaction was stopped by the addition of water with 2% TFA (7.3 ml). The reaction mixture was left at room temperature overnight, leading to the precipitation of **T1** as a white solid. The reaction mixture was centrifuged (10,000 rpm for 20 min) and the supernatant removed. The precipitate was then washed with water (6 ml), centrifuged again (10,000 rpm for 20 min), the solids collected, and the combined supernatants retained and stored at 5° C.

The isolated template was then used again in another cycle of the same reaction. A total of 5 reaction cycles were carried out this way. **T1** was recovered again after the 5 reaction cycles (100 mg, 83% recovery, high purity confirmed by ^{1}H-NMR, Figure S6). The combined supernatants from all 5 reaction cycles were concentrated to dryness *in vacuo*, and the residue then dissolved in water (~18 ml) and the pH adjusted to 7.0 using phosphate buffer (pH 7.5, 6 ml, 0.2 M) and NaOH (~0.2 ml, 4 M) for a total reaction volume of 25 ml. Then α-glucosidase stock solution (Megazyme, 18 µl, 27 U) was added and the reaction kept at 40 °C. When all linear α-1,4-glucans had been hydrolyzed to glucose (4 days, determined using HPLC-ELSD, Figure 3.20), the reaction was heated to 95 °C for 45 min to stop the reaction. The reaction mixture was centrifuged (10,000 rpm for 15 min), and the supernatant was collected and concentrated to 2.7 ml by blowing nitrogen over the surface. Acetone (15 ml) was then added to precipitate a white solid. The suspension was centrifuged (1900 rpm for 20 min), decanted and the solids dried *in vacuo*. The solid was then dissolved in water (2.1 ml) and filtered through a syringe filter. The filtrate was split into three portions and each portion injected on a preparative HILIC type HPLC column (XBridge BEH Amide OBD Prep column from Waters, 130 Å, 5 µm, 19 × 150 mm) using a Buchi Reveleris Prep Purification system equipped with an ELS detector. Gradient elution, 25% water in acetonitrile to 43% water in acetonitrile over 50 minutes. δ-CD eluted from 45 to 48 minutes. All fractions containing δ-CD were combined, concentrated and lyophilized to yield pure δ-CD (19.4 mg, 7.2% yield).

^{1}H NMR (400 MHz, D_2O) δ 5.27 (d, J = 3.9 Hz, 9H, **H1**), 3.97 (t, J = 9.6 Hz, 9H, **H3**), 3.94 – 3.81 (m, 27H, **H5**, **H6**), 3.65 (t, J = 9.6 Hz, 9H, **H4**), 3.62 (dd, J = 9.6, 3.9 Hz, **H2**). **^{13}C NMR** (101 MHz, D_2O, referenced to residual solvent in DMSO-*d6* locktube) δ 98.9 (**C1**), 77.0 (**C4**), 71.9 (**C3**), 71.2 (**C2**), 70.3 (**C5**), 59.3 (**C6**). **HRMS** (ESI) *m/z*: $[M+Na]^+$ calcd. for $C_{54}H_{90}O_{45}Na$ 1481.465; found 1481.467.

3.8. Bibliography

[1] K. L. Larsen, *J. Incl. Phenom.* **2002**, *43*, 1–13.

[2] T. Endo, *Trends Glycosci. Glycotechnol.* **2011**, *23*, 79–92.

[3] E. M. M. Del Valle, *Process Biochem.* **2004**, *39*, 1033–1046.

[4] A. R. Hedges, *Chem. Rev.* **1998**, *98*, 2035–2044.

[5] N. Sharma, A. Baldi, *Drug Deliv.* **2016**, *23*, 739–757.

[6] D. Larsen, S. R. Beeren, *Chem. Sci.* **2019**, *10*, 9981–9987.

[7] K. I. Assaf, D. Gabel, W. Zimmermann, W. M. Nau, *Org. Biomol. Chem.* **2016**, *14*, 7702–7706.

[8] M. J. Crossley, M. L. Fisher, J. J. Potter, P. W. Kuchel, M. J. York, *J. Chem. Soc. Perkin Trans* **1990**, 2365.

[9] N. S. Curvey, S. E. Luderer, J. K. Walker, G. W. Gokel, *Synth.* **2014**, *46*, 2771–2779.

[10] D. Larsen, P. M. Bjerre, S. R. Beeren, *Chem. Commun.* **2019**, *55*, 15037–15040.

[11] D. Larsen, S. R. Beeren, *Chem. Eur. J.* **2020**, *26*, 11032–11038.

[12] D. Larsen, S. R. Beeren, *Chem. Commun.* **2021**, *57*, 2503–2506.

[13] S. Yang, D. Larsen, M. Pellegrini, S. Meier, D. F. Mierke, S. R. Beeren, I. Aprahamian, *Chem* **2021**, *7*, 2190–2200.

[14] D. Larsen, M. Ferreira, S. Tilloy, E. Monflier, S. R. Beeren, *Chem. Commun.* **2022**, *58*, 2287–2290.

[15] G. Wenz, B. H. Han, A. Müller, *Chem. Rev.* **2006**, *106*, 782–817.

[16] M. Nilsson, A. J. M. Valente, G. Olofsson, O. Söderman, M. Bonini, *J. Phys. Chem. B* **2008**, 11310–11316.

[17] J. G. Gao, Y. J. Ding, H. W. Chen, Q. P. Song, Q. J. Zhang, *Chinese J. Chem. Phys.* **2008**, *21*, 387–392.

[18] M. V. Rekharsky, Y. Inoue, *Chem. Rev.* **1998**, *98*, 1875–1917.

[19] S. Makedonopoulou, I. M. Mavridis, *Carbohydr. Res.* **2001**, *335*, 213–220.

[20] J. Czescik, Y. Lyu, S. Neuberg, P. M. Scrimin, F. Mancin, *J. Am. Chem. Soc.* **2020**, *142*, 6837–6841.

[21] S. A. Nepogodiev, J. F. Stoddart, *Chem. Rev.* **1998**, *98*, 1959–1976.

[22] P. Thordarson, *Chem. Soc. Rev.* **2011**, *40*, 1305–1323.

[23] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, et al., *Amber 2020*, University of California, San Francisco, **2020**.

[24] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.

[25] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[26] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem.*

Phys. **1983**, *79*, 926–935.

[27] J. A. Hartigan, M. A. Wong, *J. R. Stat. Soc.* **1979**, *28*, 100–108.

[28] S. Meier, S. R. Beeren, *J. Am. Chem. Soc.* **2014**, *136*, 11284–11287.

[29] P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J. L. Wietor, J. K. M. Sanders, S. Otto, *Chem. Rev.* **2006**, *106*, 3652–3711.

[30] K. Harata, H. Akasaka, T. Endo, H. Nagase, H. Ueda, *Chem. Commun.* **2002**, 1968–1969.

[31] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

[32] A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623–1641.

[33] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[34] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[35] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.

[36] R. W. Pastor, B. R. Brooks, A. Szabo, *Mol. Phys.* **2006**, *65*, 1409–1419.

[37] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[38] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

[39] Y. Wu, C. J. Aquino, D. J. Cowan, D. L. Anderson, J. L. Ambroso, M. J. Bishop, E. E. Boros, L. Chen, A. Cunningham, R. L. Dobbins, et al., *J. Med. Chem.* **2013**, *56*, 5094–5114.

[40] Galderma Res & Dev, B. Jean-Michel, *Vitamin D Analogues*, **2004**, US6689922, B1.

[41] M. Wu, A. M. Vartanian, G. Chong, A. K. Pandiakumar, R. J. Hamers, R. Hernandez, C. J. Murphy, *J. Am. Chem. Soc.* **2019**, *141*, 4316–4327.

[42] W. H. Binder, M. J. Kunz, C. Kluger, G. Hayn, R. Saf, *Macromolecules* **2004**, *37*, 1749–1759.

Chapter 4. Probing the properties of figure-of-eight large-ring cyclodextrins

Abstract

In this chapter, the properties and host-guest chemistry of large-ring cyclodextrins (LRCDs) with 12, 14 and 16 glucose units (CD12, CD14 and CD16) are investigated. Molecular dynamics (MD) simulations of these LRCDs reveal that CD12 adheres to a bent-boat like macrocyclic structure, while CD14 and CD16 show larger flexibilities with figure-of-eight type geometries. MD simulations with CD12, CD14 and CD16 in the presence of adamantane carboxylate (ACA) showed that CD16 was able to bind two ACAs in a figure-of-eight type conformation with one ACA bound to each loop of the figure-of-eight. CD12 and CD14 were found to be too small to accommodate two ACAs in the same type of binding mode. A range of LRCDs were isolated from a mixture of LRCDs, and CD16 was used in an NMR titration with ACA that confirmed the 1:2 binding stoichiometry, and showed that binding exhibited positive cooperativity. A double adamantane bolaamphiphile was synthesized and titrated with CD16, where it was found that the bolaamphiphile bound to CD16 with a 1:1 stoichiometry and negative chelate cooperativity.

4.1. Introduction

In 2014, the author's supervisor Sophie Beeren and co-supervisor Sebastian Meier developed a method for the simultaneous determination of binding constants for multiple carbohydrate hosts in complex mixtures.[1] In the study, high-resolution ^{1}H-^{13}C HSQC experiments were used to resolve the complexes of amphiphiles and more than 10 different maltooligosaccharides (Figure 4.1). From this data, relative binding constants between each component of the mixture were extracted, which could then be converted into absolute binding constants using the known absolute binding parameters of one of the components in the mixture.

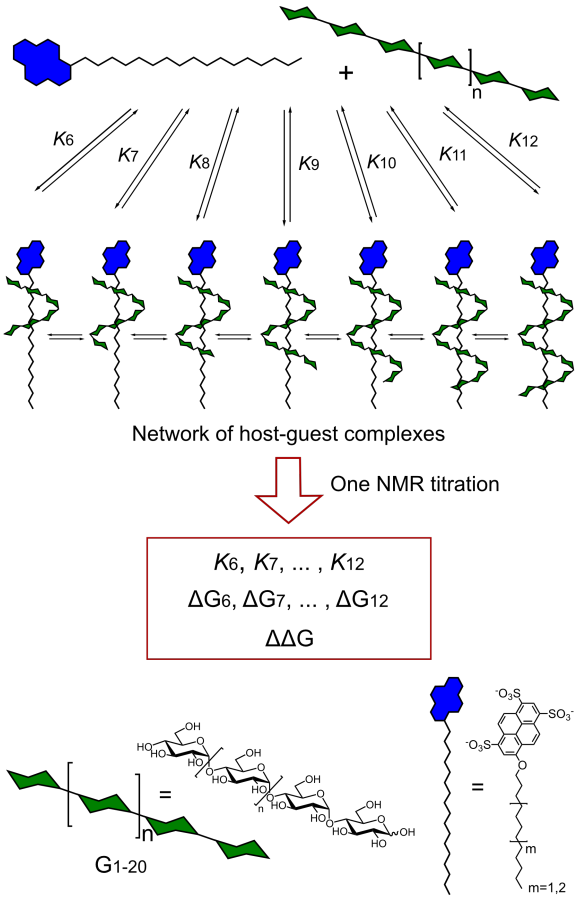


Figure 4.1. The examination of the interaction of maltooligosaccharides (G_1–G_{20}) with amphiphiles in mixtures to extract binding parameters for each complex from a single titration. From ref[1] with permission from publisher.

In the Beeren group, this method was later utilized with mixtures of cyclodextrins as well by Senior Researcher Dennis Larsen (unpublished work). The LRCDs with 9 to about 20 glucose units were of particular interest for binding studies. As described in Chapter 1, the smaller LRCDs (CD9, CD10), while more flexible than the native CDs (CD6–CD8, α-, β-, and γ-CD), still exhibit a semi-rigid, annular shape.[2–4] Going from CD9 and up to CD18, the LRCDs become progressively more flexible and start exhibiting band flips, loops and figure-of-eight type geometries.[5–7] As such, this series of CDs should exhibit a variety of different binding modes and binding affinities to hydrophobic guest molecules, which would be interesting to explore, like the different binding profiles observed for the binding of I_2/I_3^- to CD21–CD32.[8] The treatment of a commercially available mixture of very large CDs called cycloamylose

(CD24 to ~CD60) (10 mg/ml) with CGTase (1 µl per ml reaction mixture) in water at room temperature for 8 hours yielded a mixture of CDs with 6 to ~30 glucose units as well as linear α-1,4-glucans. The linear α-1,4-glucans were then hydrolyzed to glucose by the action of the enzyme α-glucosidase, which hydrolyzes linear α-1,4-glucans while leaving the CDs intact. Figure 4.2 shows an HPLC-ELSD chromatogram and a high resolution ^{1}H-^{13}C HSQC spectrum of the LRCD mixture.

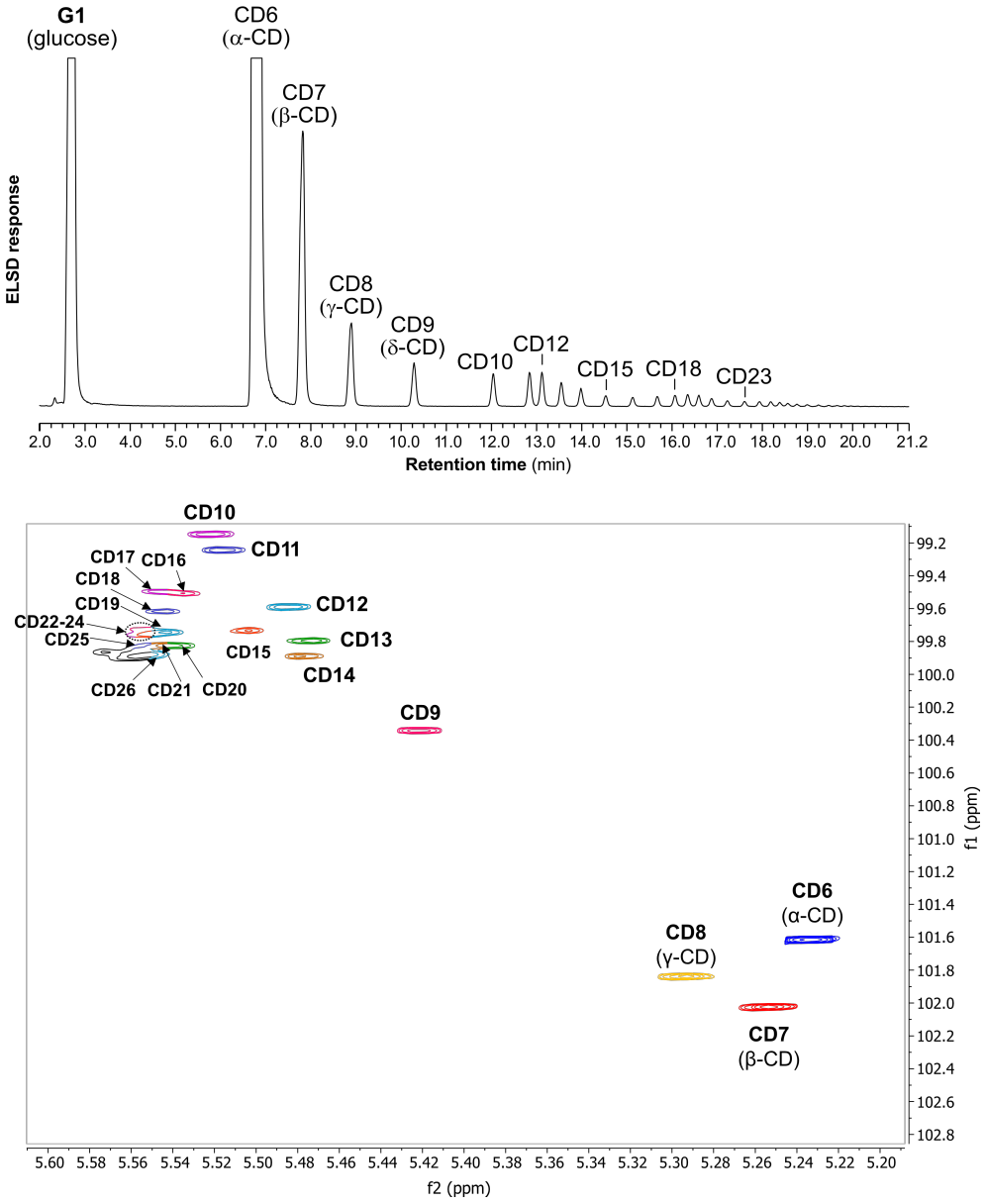


Figure 4.2. LRCD mixture with CD6 to ~CD30 produced by the action of the enzyme CGTase on 'Cycloamylose' (CD24 to ~CD60) followed by treatment with the enzyme α-glucosidase to hydrolyze linear α-1,4-glucans. *Top:* HPLC-ELSD chromatogram showing the distribution of CDs and glucose in the mixture. *Bottom:* Partial high resolution ^{1}H-^{13}C HSQC (800 MHz) spectrum of the LRCD mixture (10 mg/ml) in phosphate buffered D_2O (100 mM, pH 8.0) at 50 °C with assignment of the anomeric protons of CDs up to CD26. Noise removed for clarity.

This LRCD mixture was then used in an NMR titration at 50 °C with adamantane carboxylate (ACA) as a hydrophobic probe, to investigate the binding properties of these LRCDs (Figure 4.3). Using the high resolution HSQC technique developed previously,[1] the changes in chemical shifts for CD6 to CD26 were followed (Figure 4.3a). The shape of the resulting binding isotherms showed clear evidence of higher order binding in some cases, presumably 1:2 binding. The method developed in 2014 only dealt with 1:1 binding, which meant that new fitting models needed to be developed. The development of this new model and the accompanying fitting functions for use in Origin was carried out by Associate Professor Wei Yan (unpublished work). The binding constant for the binding of ACA to CD6 ($K_a = 96 \pm 5$ M^{-1}) in phosphate buffered D_2O (0.1 M, pH 7.5) at 50 °C was determined from an NMR titration using isolated CD6, (unpublished results). Using this value as a reference, the binding constants (K_{a1} and K_{a2}) for the binding of ACA to CD6–CD26 were determined (Figure 4.3b). Some trends are quite clear. ACA binds very strongly to CD7 due to good size complementarity between ACA and the cavity of CD7.[9] Going from CD7 to CD10, the binding constants dramatically decrease with roughly an order of magnitude for each additional glucose unit in the CD. Assuming an annular shape of these CDs, their cavities are becoming increasingly too large for ACA. Moving on from CD10 to CD11 and CD12, the binding becomes slightly stronger again. This can be explained with the increased flexibility of these CDs as they become larger, leading to an increased ability to adopt conformations that are complementary to the guest, ACA.

For CD9–CD12, no evidence of 1:2 binding was observed. Moving on to CD13 and especially CD14 and CD15, the fits of the data to a 1:1 binding model become less ideal and the 1:2 binding model shows decent fits to the data, yielding first and second binding constants of similar magnitude. Moving on to CD16–CD20, the data only fits well to a 1:2 binding model, and the second binding becomes increasingly larger and more cooperative with increasing CD size. The proposed binding mode responsible for this trend from CD13–CD20 is a 'figure-of-eight' type conformation of the CDs (Figure 4.3b), where two binding pockets are available for binding ACA. As mentioned previously, CD7 has the ideal geometry for the binding to ACA. As such, 14 glucose units (2 × 7) are, as a minimum, required for the optimal binding of two ACAs to a CD in a figure-of-eight conformation, not taking into account the extra glucose units needed to connect the two loops/binding pockets of the figure-of-eight. With CD13–CD15, the CDs are thus not large enough to form two separate binding pockets of the right size for binding ACA, while CD16–CD20 on the other hand are large enough to do this. The cooperativity of the 1:2 binding of ACA to these figure-of-eight LRCDs can be explained by the first ACA binding event paying the bulk of the entropic cost (and possibly enthalpic cost due to non-optimal glycosidic bond angles) of forming the figure-of-eight conformation, thus pre-organizing the CD for the binding of the second ACA.

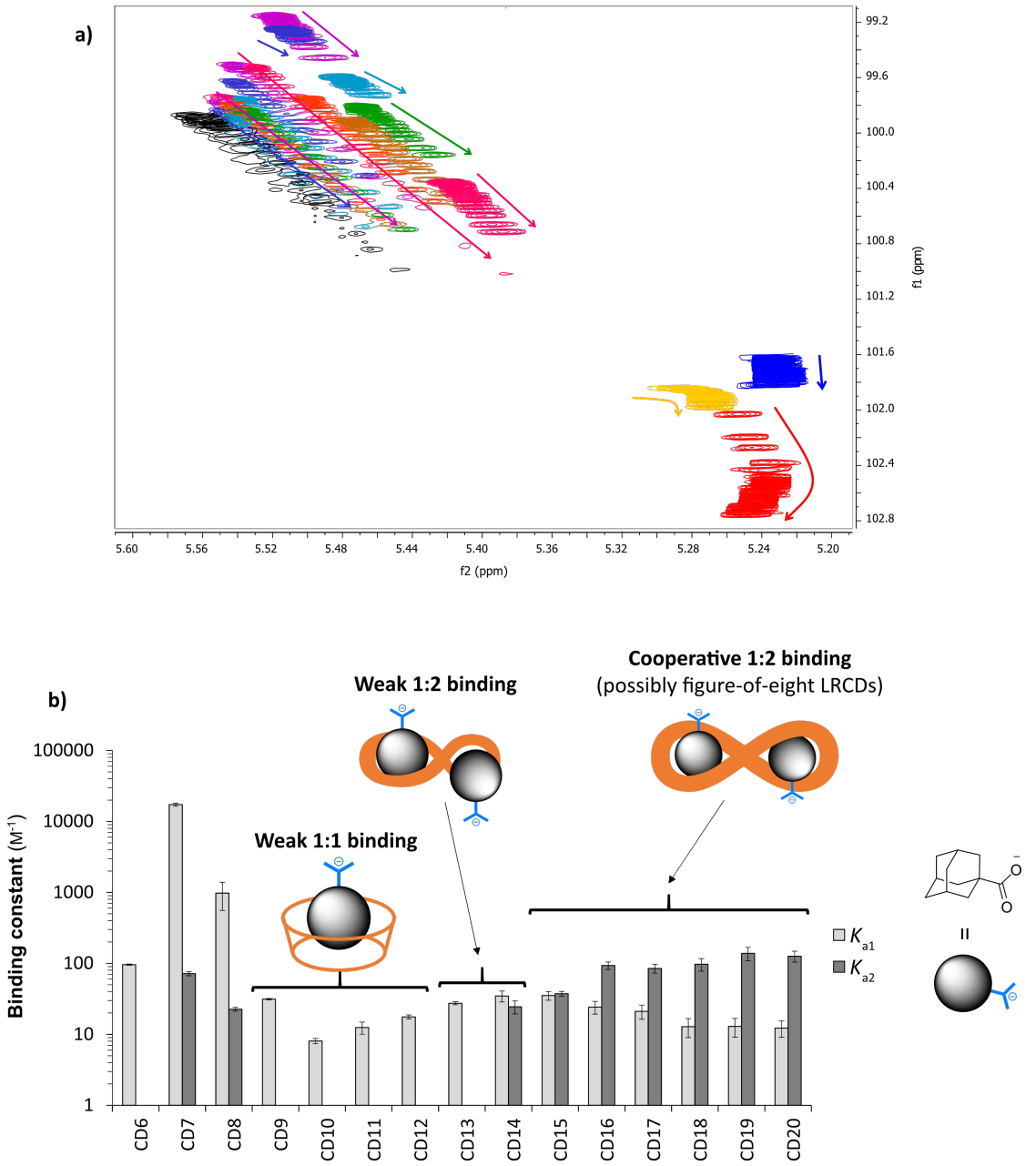


Figure 4.3. Results from the high resolution ^{1}H-^{13}C HSQC NMR titration with 'LRCD mixture' (10 mg/ml) and adamantane carboxylic acid (0–50 mM) in sodium phosphate buffer (100 mM, pH 8.0) at 50 °C. (a) Superimposed spectra from the titration. The colors of the peaks correspond to the colors of the peaks assigned to CDs shown in Figure 4.2. Arrows indicate movement of peaks during titration. (b) Determined binding constants and proposed binding modes.

As described in detail in Chapter 1, Ivanov, Jaime and co-workers have previously carried out molecular dynamics (MD) simulation studies of a wide range of LRCDs.[6,7,10–14] Most of these simulations were carried out on the isolated LRCDs. One study, however, dealt with the inclusion complexes of LRCDs and adamantane or 1-hydroxyadamantane.[14] Snapshots from these simulations carried out in explicit water at 300 K are shown in Figure 4.4. The simulation with CD13 and adamantane (Figure 4.4a) shows how CD13 mostly adopts bent-boat like macrocyclic conformations within the 100 ns simulation. At the 40 ns snapshot however, the geometry of CD13 adapts to the bound adamantane, forming a binding pocket that fits the size

of adamantane, with the geometry of CD13 slightly resembling a figure-of-eight. In the case of CD14 and adamantane (Figure 4.4b), a figure-of-eight type conformation is observed within the relatively short 40 ns simulation. On the other hand, in the case of CD14 and 1-hydroxyadamanane (Figure 4.4c), CD14 only exhibited geometries similar to the bent-boat like macrocyclic conformation of the crystal structure of CD14.[5] These resuflts by Ivanov, Jaime and co-workers seem to somewhat agree with the binding modes and trends proposed from the NMR titration discussed above. However, more studies with other LRCDs and longer simulations are required to further substantiate this.

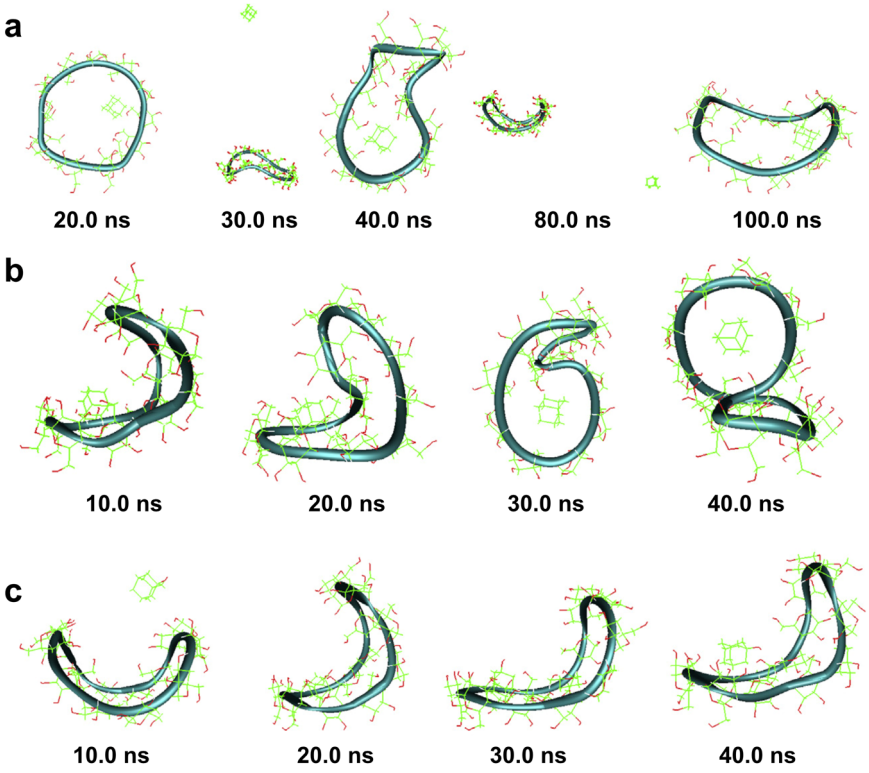


Figure 4.4. Snapshots of optimized geometries of complexes from MD simulations (300 K, TIP4P water[15]) carried out by Jaime and co-workers. (a) CD13 with adamantane. (b) CD14 with adamantane. (c) CD14 with 1-hydroxyadamantane. Adopted with permission from ref.[14]

The work described in this introduction served as motivation for the work carried out by the author, which is described in the rest of this chapter. The aim was to further explore the host-guest chemistry of figure-of-eight LRCDs with adamantane carboxylate using a variety of methods, such as MD simulations and NMR titrations. Another aim was to synthesize bola-amphiphiles with two linked adamantane moieties as hydrophobic recognition units, which could be used to bind figure-of-eight LRCDs, as shown in Figure 4.5. A simple double-adamantane bolaamphiphile (Figure 4.5a) could be used as a probe for the figure-of-eight binding mode, and potentially lead to increased binding affinities due to chelate cooperativity.[16,17] Installing bulky hydrophilic head groups on both sides of the double-adamantane binding motif could block out the binding of the small CDs, making such bolaamphiphiles (Figure 4.5b) potential templates for CGTase-mediated dynamic

combinatorial libraries targeting LRCDs, where selective binding to LRCDs over the small CDs would be required.

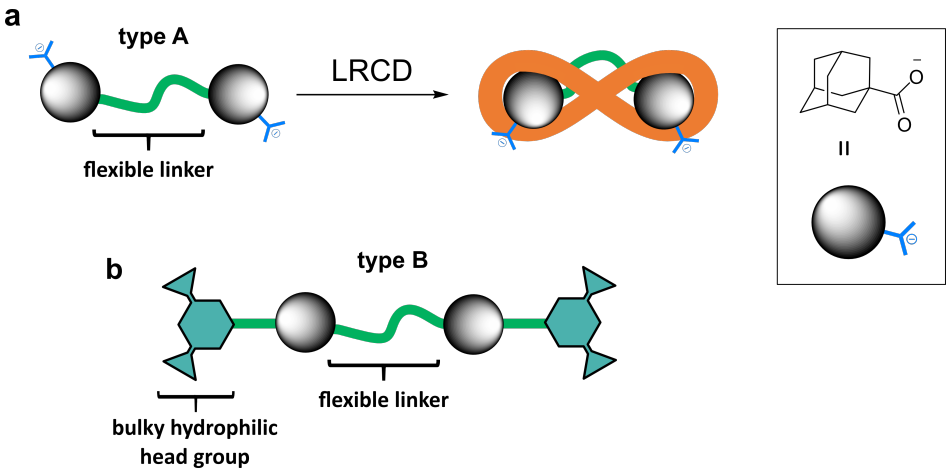


Figure 4.5. Concept of double-adamantane bolaamphiphiles. (a) Binding of a simple 'type A' double-adamantane bolaamphiphile to figure-of-eight LRCDs. (b) 'Type B' template for selective binding to figure-of-eight LRCDs, with bulky hydrophilic head groups to block out the binding of small CDs.

4.2. Computational studies with adamantane carboxylate and CD12, CD14 and CD16

In order to further investigate and substantiate the trends in the conformations and proposed binding modes of LRCDs to adamantane carboxylate described in the introduction to this chapter, a computational study was carried out. MD simulations were chosen as the computational tool, as it enables the studies of dynamics and various conformations of these flexible LRCDs and their complexes in explicit aqueous solvent. CD12, CD14 and CD16 were chosen as the LRCDs to be studied, as these three LRCDs are closely related with a difference of only 2 and 4 glucose units to CD12, while still covering the main trends in binding modes proposed in the introduction to this chapter: CD12 falls in the category of 'weak 1:1 binding' to adamantane carboxylate, while CD14 and CD16 both fall in the general category of 1:2 binding in a figure-of-eight conformation, but with CD16 showing stronger cooperativity for the second binding.

4.2.1. MD simulations with CD12, CD14 and CD16 without guests

Similar to the approach by Ivanov,[7] the crystal structure of CD14[5] (Cambridge Crystallographic Data Centre (CCDC) deposition number: 124917) was chosen as a starting point for generating the starting geometries of CD12 and CD16. The structures of CD12 and CD16 were then generated by manually deleting or adding extra glucose units and then reconnecting the α-1,4-glycosidic bonds using the molecular software PyMOL. The two 'band flips' (*anti* oriented glucose units) in the crystal structure of CD14 were kept in the generated structures of CD12 and CD16. Two band flips are seen in the crystal structures of both CD10 and CD14,[5] so keeping these structural motifs in the starting geometries of CD12 and CD16 was deemed sensible. The highly strained geometries (due to distorted bond lengths and angles) generated for CD12 and CD16 were then minimized in the Amber20 suite[18] in vacuum using the carbohydrate force field GLYCAM-06.[19] Computational details can be found in the experimental section of this chapter. The minimized CD12 and CD16 structures and the crystal structure of CD14 were then solvated with TIP3P water,[15] and the solvated systems were then minimized, heated to 300 K and equilibrated for 4 ns. MD production runs (1200 ns) were then carried out in the *NPT* ensemble at 300 K (snapshots from the simulations are shown in Figure 4.6). The full simulation trajectories were then subjected to clustering analysis using the 'k-means' algorithm[20] in the CPPTRAJ module[21] of Amber20 to output clusters based on root-mean square deviation (RMSD) of non-hydrogen atoms. The results from the clustering analyses are shown in Figures 4.7 and 4.8.

The snapshots from the MD production runs (Figure 4.6) show the CDs both with traditional 'sticks' or 'licorice' structures, as well as with the 'twister' algorithm developed for the visualization of cyclic and multi-branched polysaccharides[22] in the Visual Molecular Dynamics (VMD) package.[23] The 'twister' algorithm traces glycosidic bonds with a ribbon that twists according to the relative orientation of successive sugar residues. This 'twister' ribbon thus assists in the visualization and interpretation of the structures of the LRCDs from the MD simulations.

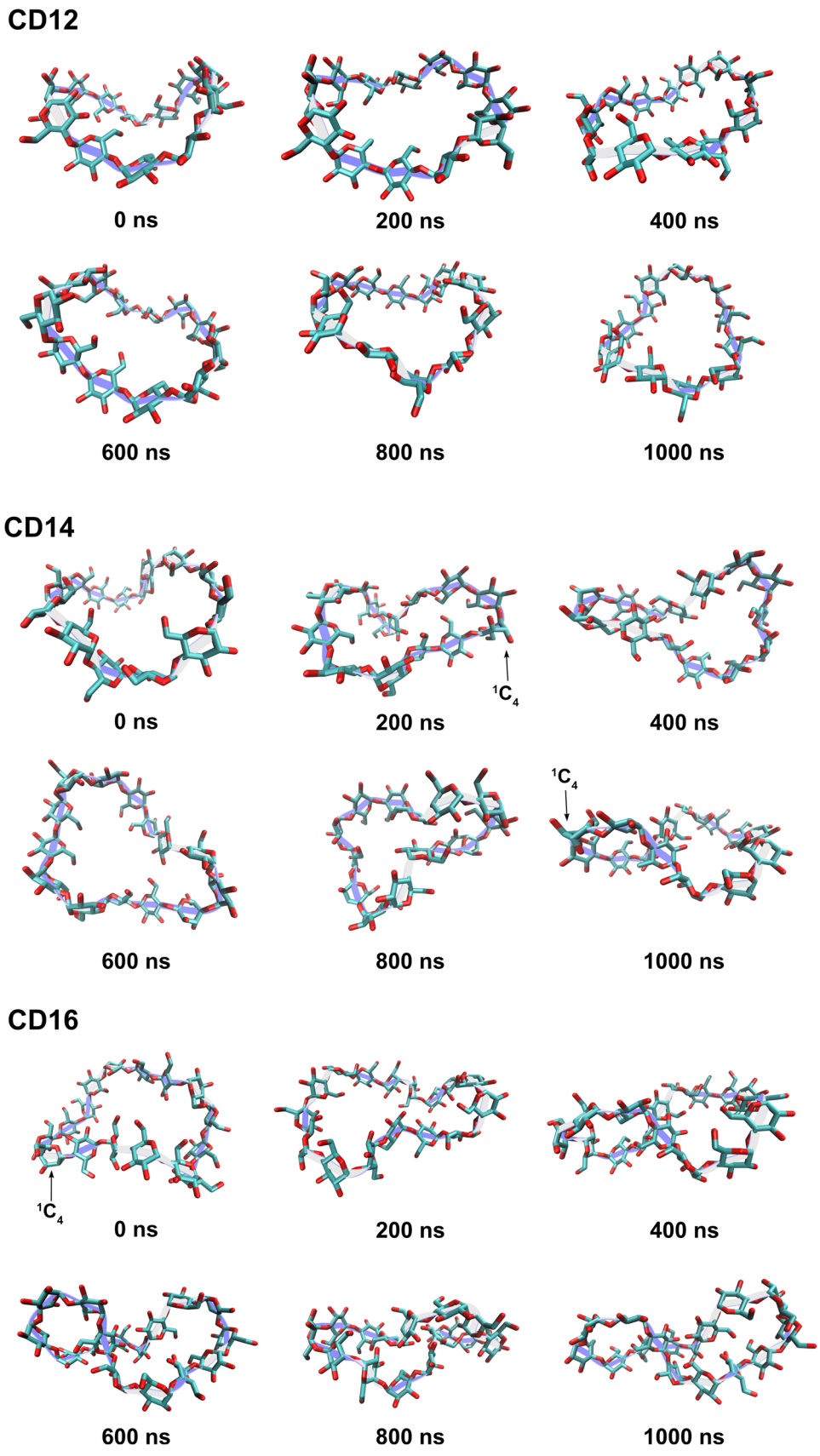


Figure 4.6. Snapshots from the MD simulations (1200 ns) with CD12, CD14 and CD16. The CDs are depicted both with traditional 'sticks' or 'licorice' structures, as well as with the 'twister' algorithm developed for the visualization of cyclic and multi-branched polysaccharides[22] in the Visual Molecular Dynamics (VMD) package.[23] All $^1C_4$ chair conformations are indicated on the figure.

From the snapshots of the simulation with CD12, it is apparent how CD12 adheres to the open bent-boat like macrocyclic structures proposed previously. This is also apparent in the representative structures extracted from the three most dominant clusters of the clustering analysis of the trajectory (Figure 4.7). These results are in agreement with the previously published results on CD12.[7]

From the simulations with CD14 and CD16, the larger flexibilities of these LRCDs relative to CD12 are quite clear. The open bent-boat like macrocyclic structures of the first snapshots (0 ns), which are closely related to the starting geometries of CD14 and CD16 (and thus the crystal structure of CD14), converts into different variations of the proposed figure-of-eight like geometry during the MD simulations (Figure 4.6). The cluster population graphs show that structures that are present in early parts of the simulation (**C2**, third most dominant cluster for both CD14 and CD16) relate to the more open starting geometries (Figure 4.7 and 4.8). The structures representing the two most dominant clusters (**C0** and **C1**) for CD14 and CD16, which are present in later parts of the simulation, adhere to the figure-of-eight like motif. For CD14, the **C0** and **C1** structures contain two loops, where one of the loops is slightly smaller than the other. The loops can be considered as representing approximately two 'CD6 + CD7' pseudo-cavities, as observed in previous studies.[6] CD16 generally forms a more 'folded in' structure where one of the loops of the figure-of-eight has 'collapsed', as apparent in the **C0** cluster as well as snapshots (400, 800, 1000 ns) from the simulation. This type of structure, while possibly predominantly an energy minimum for the glycosidic angles in CD16, would also lead to exposure of fewer of the hydrophobic surfaces of the glucose units to the aqueous solvent, which would also be thermodynamically favorable due to hydrophobic effects.

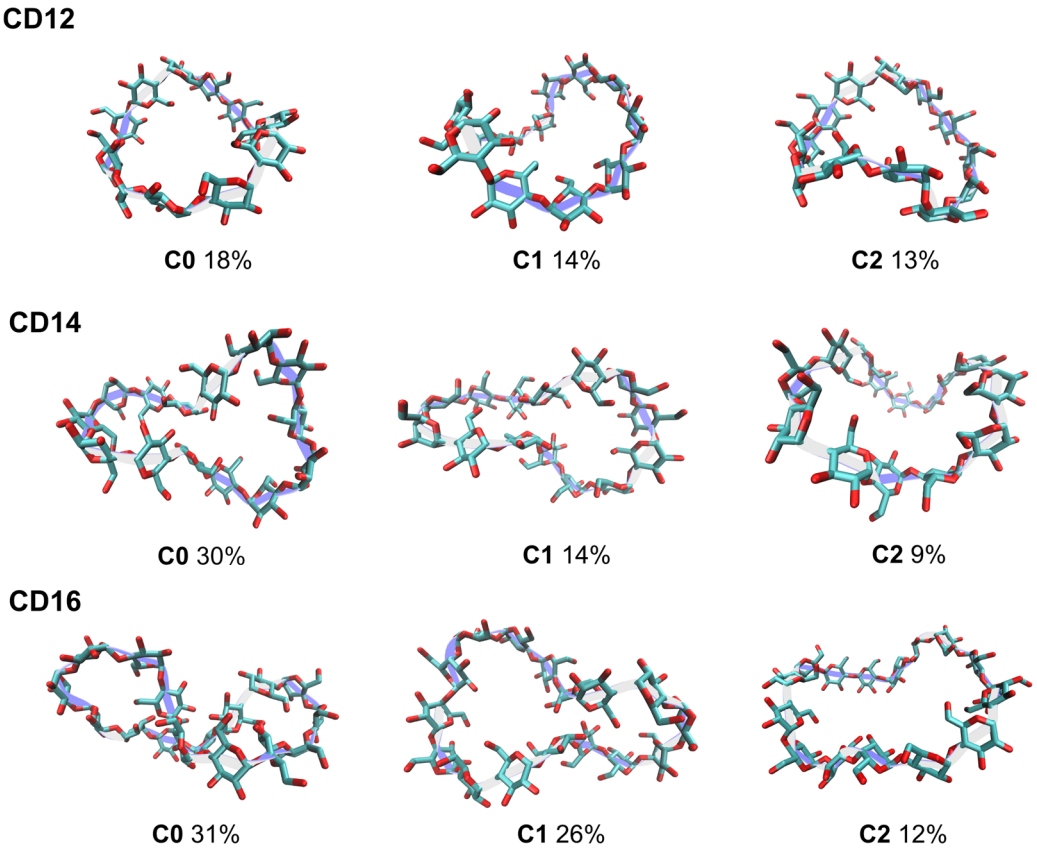


Figure 4.7. Three most dominant structures (**C0**–**C2**) of CD12, CD14 and CD16 extracted from the clustering analysis of the MD simulations (1200 ns, 300 K) with these CDs. Number of frames assigned to each cluster is shown as % of total number of frames.

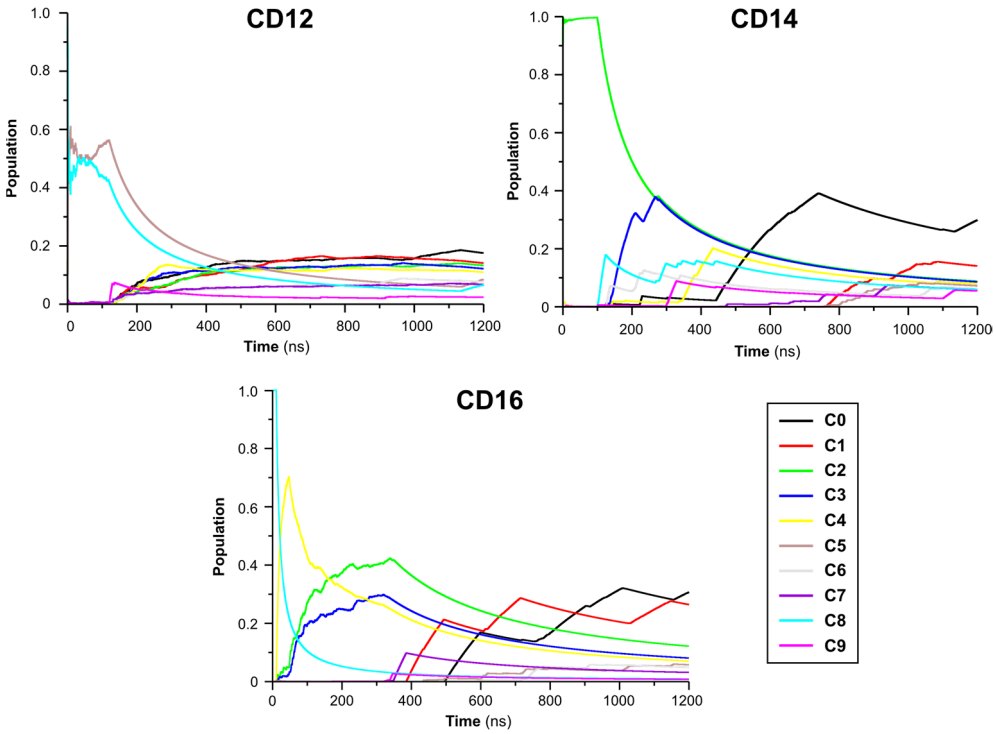


Figure 4.8. Cluster population (number of frames in each cluster normalized by the total number of frames at each time point) vs simulation time from the clustering analysis of the MD simulations with CD12, CD14 and CD16.

A previous study and evaluation of carbohydrate force fields in relation to 'amylose' found that the *syn* ⇌ *anti* transition rate ('flips') in maltododecaose (linear α-1,4-glucan with 12 glucose units) was in the range of 1–20 μs^{-1} for the GLYCAM-06 force field at 298 K.[24] As such, *syn* ⇌ *anti* transitions are to be expected within the simulation time frame (1.2 μs) carried out here if the LRCDs are flexible enough to allow these transitions without incurring excessive strain. The ψ torsional glycosidic angle, defined with standard formalism as C_1–O–C_4–C_3, can be used to assess whether successive glucose residues are *syn* ($\psi \approx 90°$) or *anti* ($\psi \approx -70°$). The time evolution of ψ values of one glycosidic linkage from each of the full trajectories with CD12, CD14 and CD16 are shown in Figure 4.9.

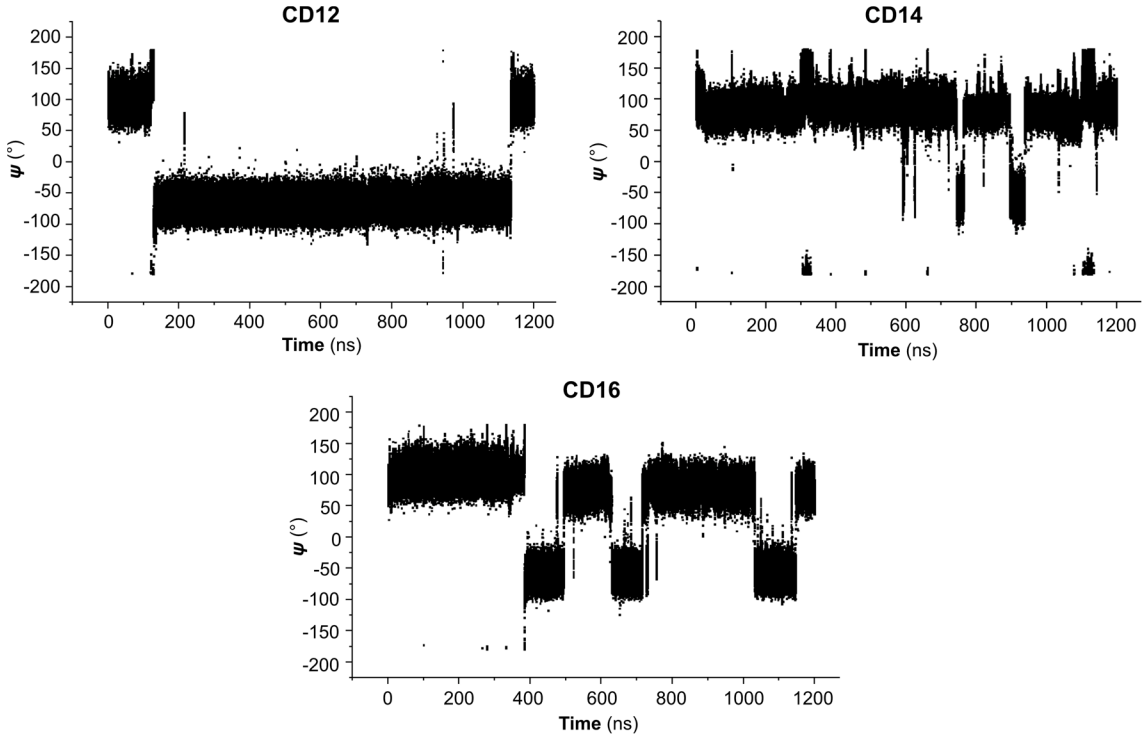


Figure 4.9. Time evolution of the torsional angle ψ (C_1–O-C_4-C_3) for one glycosidic linkage from each of the simulations with CD12, CD14 and CD16. Glycosidic linkages were randomly chosen for CD14 and CD16, while for CD12, the glycosidic linkage with the most *syn* ⇌ *anti* transitions was selected.

These results, while just one example for each CD, show the time frame of the transitions. A transition of a linkage from *syn* to *anti* that occurs in CD12 after ~100 ns remains primarily as *anti* for about 1000 ns before reverting back to *syn*. These two transitions are concurrent with complementary *anti* to *syn* to *anti* transitions in another linkage (not shown in figure), and these four transitions are the only stable transitions ones that occur, (i.e. where the transition does not quickly revert back within a couple of frames) for the duration of the simulation with CD12. These transitions are also apparent in the cluster population graph (Figure 4.8), where the transitions are concurrent with the switching between two groups of clusters (**C6**+**C8** vs the rest). In the case of CD14 and CD16, *syn* ⇌ *anti* transitions are seen with higher frequencies in the simulations, as apparent in the torsional angles for a randomly chosen linkage of CD14 and CD16 shown in Figure 4.9. Ring flips from the $^4C_1$ to the $^1C_4$ chair conformation were also

observed for CD14 ($^1C_4$ chairs shown in snapshots at 200 and 1000 ns) and for CD16 (shown in snapshot at 0 ns). These transitions occur at a much slower rate (0.2–3.5 μs^{-1}) in maltododecaose,[24] and are thus relatively rare events in a 1.2 μs simulation.

The results from these simulations with CD12, CD14 and CD16 are generally in agreement with the previously published results[6,7] and confirm the general trends proposed for the structures of these LRCDs. These results also show the rates at which the transitions that are the key drivers for the conformational changes in these CDs,[6,7,10,13] such as the *syn* ⇌ *anti* flip, occur under the conditions employed in these MD simulations.

4.2.2. MD simulations with CD12, CD14 and CD16 and adamantane carboxylate

To investigate the binding mode of adamantane carboxylate (ACA) to CD12, CD14 and CD16, simulations with both one and two ACAs present were carried out. For detailed descriptions of the computational protocols, see experimental section 4.6.3. As starting geometries of CD12 and CD16, the equilibrated structures from the previous MD simulations with these CDs (corresponding to the 0 ns snapshots) were chosen, while for CD14, the crystal structure was used again. The structure of ACA was based on a published crystal structure[5] of adamantane carboxylic acid (CCDC deposition number: 1283437).[25] For ACA, the general amber force field (GAFF)[26] was employed and partial charges were calculated using the AM1-BCC method.[27] The starting geometries of the complexes were generated by manually placing one or two ACAs with a minimum distance of 5 Å to the CDs (and to each other, when two ACAs were used). The complexes were then solvated with TIP3P water with a distance buffer of 20 Å from the CDs and ACAs to the edges of the periodic box, and the system was neutralized using one or two sodium ions (corresponding to the one or two adamantane carboxylates). The complexes were then minimized, heated to 300 K and equilibrated for 4 ns. MD production runs (1200 ns and 2000 ns for the simulations with one and two ACAs, respectively) were then carried out in the *NPT* ensemble at 300 K.

The resulting trajectories of the simulations with one ACA contained significant parts where ACA was not bound to the CDs. The trajectories were then analyzed according to the distance between the center of mass (COM) of ACA and the center of mass of the CDs. The results showed that for the parts of the simulation where ACA was bound to the CDs, the distance between the COM of ACA and the COM of CDs was generally no greater than 10 Å (CD12 and CD14) and 12 Å (CD16). Since the structure of the complexes was the topic of interest, all frames of the trajectory with distances between the COM of ACA and the COM of CDs greater than these cutoffs where then filtered out, resulting in the removal of 12%, 7% and 14% of the trajectories for CD12, CD14 and CD16, respectively. The resulting distance-filtered trajectories were then subjected to clustering analysis using the same protocol as previously described, with both the RMSDs of the CDs and the ACAs used for the clustering. The most dominant structures extracted from the clustering analysis are shown in Figure 4.10.

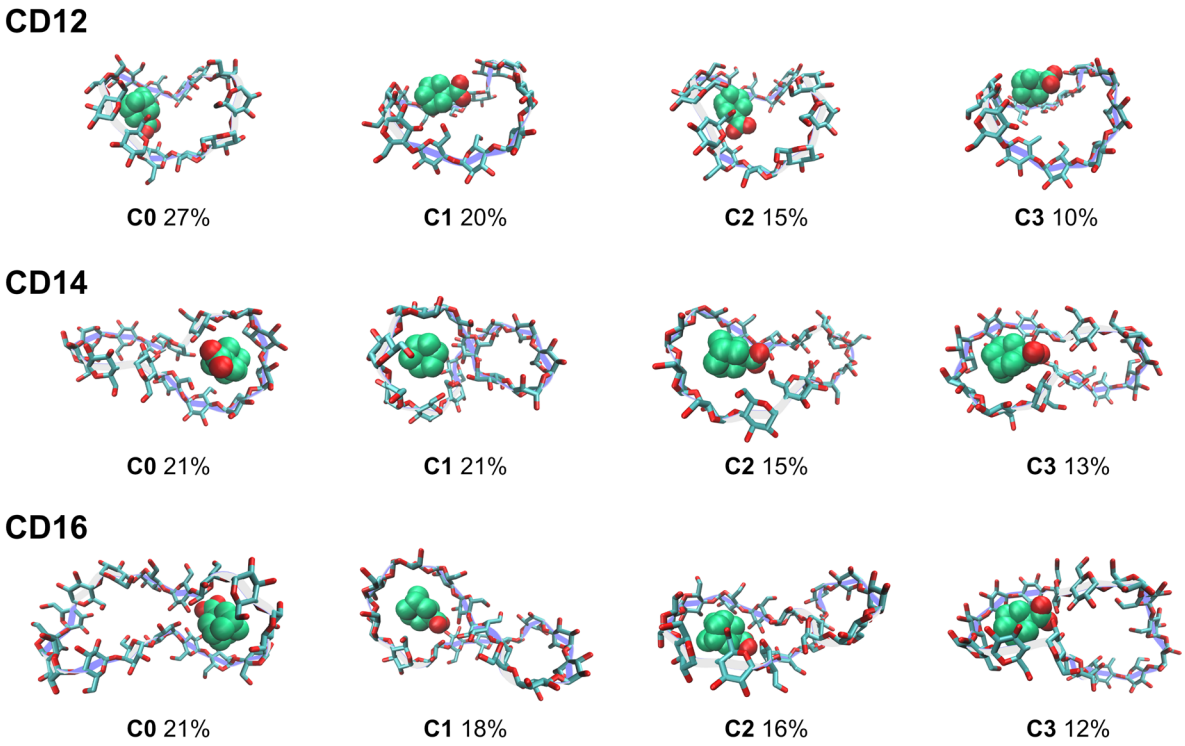


Figure 4.10. Four most dominant structures (**C0**–**C3**) of the complexes of CD12, CD14 or CD16 with ACA extracted from the clustering analysis of the distance-filtered MD simulations at 300 K of these complexes. Number of frames assigned to each cluster is shown as % of total number of distance-filtered frames.

The structures of CD12+ACA clearly show how CD12 still adheres to variations of the open bent-boat conformations. It seems that CD12 is able to somewhat adapt its cavity to the ACA guest, using about 5 glucose units to form an incomplete CD7-like cavity. For CD14 and CD16, variations of the figure-of-eight like conformation are seen. In both cases, about 7 glucose units are used to form a CD7-like cavity that binds ACA, which leaves about 5–6 glucose units for the second loop of CD14, and about 7–8 glucose units for the second loop of CD16.

To investigate the structures of 1:2 complexes between CD12, CD14 or CD16 with ACA, the trajectories with two ACAs were distance-filtered according to the same criteria established above, but now requiring that both ACAs were within the cutoffs, which filtered out 15%, 22% and 53% of the trajectories for CD12, CD14 and CD16, respectively. Large parts of the simulation with CD16 were thus spent with zero or only one ACA bound. The resulting filtered trajectories were subjected to clustering analysis, using the RMSDs of the CDs and both of the ACAs. The four most dominant structures extracted from the clustering analysis are shown in Figure 4.11.

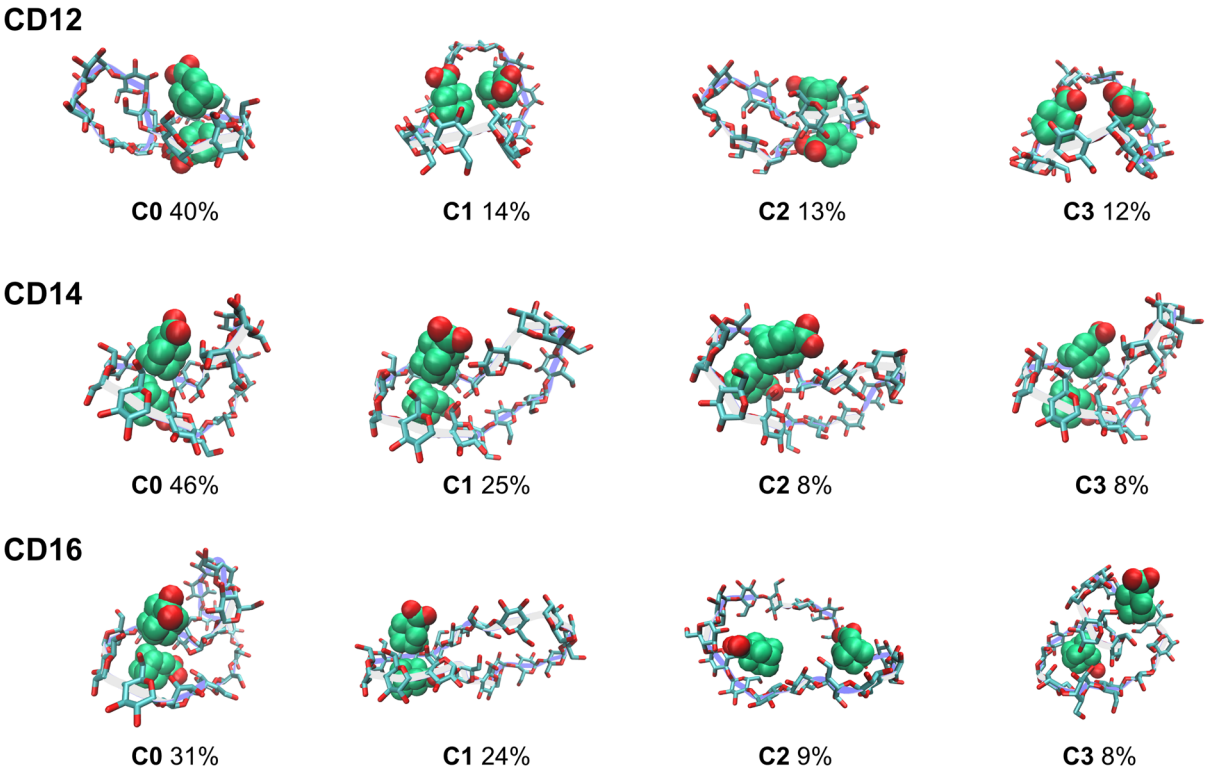


Figure 4.11. Four most dominant structures (**C0**–**C3**) of the complexes of CD12, CD14 or CD16 with two ACAs extracted from the clustering analysis of the distance-filtered MD simulations at 300 K of these complexes. Number of frames assigned to each cluster is shown as % of total number of distance-filtered frames.

In the case of all three CDs, the 1:2 complexes formed generally has two ACA molecules bound side-by-side to the same hydrophobic loop of the CD. For CD12, the loop again contains about 5 glucose units, and as apparent in the cluster representatives **C0** (40%) and **C2** (13%), one of the ACAs does not have any significant contacts with hydrophobic parts of CD12, which could be in agreement with the NMR titrations that show 1:1 binding. While other structures with CD12 and two ACA bound are also seen, like **C1** (14%) and **C3** (12%), where two ACAs are actually bound inside the bent-boat macrocyclic structure of CD12, these will not be discussed further here, as the NMR titrations show no evidence of 1:2 binding for CD12.

The most dominant structures of the 1:2 complexes of CD14 and CD16 and ACA (**C0**) adhere to the same type of structure. One loop forms a hydrophobic binding pocket with 7–8 glucose units in one V-amylose-like helical turn,[28] where both ACAs have contacts with the hydrophobic surfaces of the CDs. The rest of the glucose units of CD14 and CD16 are then 'used' to twist and loop around to connect the two ends of the helical hydrophobic binding pocket. As such, these structures could potentially explain the stronger cooperativity seen in the 1:2 binding of ACA to CD16 compared to CD14, as the extra glucose units and flexibility of CD16 could alleviate strain related to the formation of the helical hydrophobic binding pocket required for binding two ACAs. The proposed figure-of-eight binding mode with one ACA in each loop of the figure-of-eight is generally not observed for CD14 and CD16, which could be due to it simply not being a favoured structure, or perhaps that the conformational space of the 1:2 complexes has not been sufficiently sampled in these 2000 ns simulations. As described previously in this chapter, the dynamics of the *syn* ⇌ *anti* 'flip' that is important for

conformational changes in α-1,4-glucans is relatively slow. Also, half of the simulation with CD16 and two ACAs is spent where one or more ACAs are not in the vicinity of CD16. The chance that a collision between an ACA and CD16 happens at a time point where CD16 has the right geometry to accommodate two ACAs in a figure-of-eight binding mode with an ACA in each loop, could be too low to occur in the time span of the simulation (2000 ns).

4.2.3. MD simulations with CD14 and CD16 and adamantane carboxylate at 323 K

To test this hypothesis, new MD simulations with CD14 and CD16 and two ACAs were designed to speed up both internal dynamics in the CDs and collisions between the CDs and ACA. Previous simulations were carried out with a 20 Å buffer of water around the initial starting geometry, which was cut down to a 10 Å buffer for these simulations to increase the number of collisions. The temperature was increased from 300 K in the previous simulations to 50 °C (323 K), which is the same temperature used in the NMR titration described in the introduction to this chapter. For the starting geometry of the CDs, a more equilibrated structure than previous simulations was employed, by using the structure corresponding to the most dominant clusters (**C0**) of the simulations of the CDs with no guest (Figure 4.7).

Three simulations for each CD were run in parallel, using slightly different starting geometries, which were obtained by carrying out the initial heating to either 328, 318 or 313 K before equilibrating at 323 K. MD production runs (800 ns) were carried out as previously, but at 323 K instead of 300 K. The resulting trajectories where filtered according to the distance between the ACAs and the CDs using the same criteria as previously, which led to the filtering out of 19%, 18% and 10% of the frames in the three trajectories with CD14, and the filtering out of 52%, 8% and 44% of the frames in the three trajectories with CD16. Clustering analysis was then carried out as described previously, and the resulting four most dominant clusters for each simulation are shown in Figure 4.12 (CD14) and Figure 4.13 (CD16).

The new simulation conditions gave results for CD14 that were relatively similar to the previously obtained results at 300 K. The dominant structures adhere to the binding mode with both ACAs bound to the same helical hydrophobic binding pocket consisting of 7–8 glucose units. Only in one of the three parallel simulations with CD14 does one of the four most dominant structures correspond to the figure-of-eight binding motif with one ACA in each loop (Figure 4.12b, **C3**), and this structure only corresponds to 7% of the frames in the filtered trajectory. A closer look at the structure helps explain why this binding motif is not generally observed — each loop consists of only 6 glucose units (a CD6-like binding pocket), which is not the optimal size (CD7-like) for the binding of ACA. This can also be seen in the fact that one of the ACAs is protruding from the (too small) binding pocket (Figure 4.12b, **C3**).

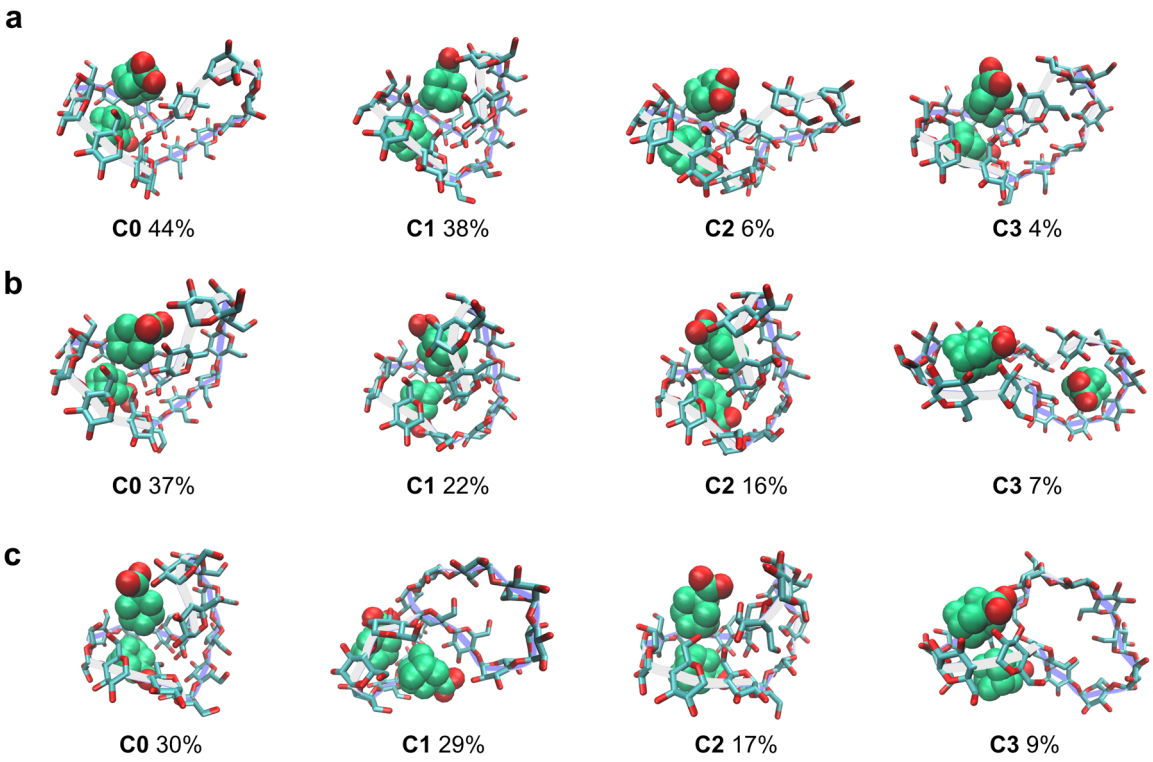


Figure 4.12. Four most dominant structures (**C0**–**C3**) of the complexes of CD14 and two ACAs extracted from the clustering analysis of three distance-filtered MD simulations at 323 K of these complexes with initial heating to (a) 328 K, (b) 318 K and (c) 313 K. Number of frames assigned to each cluster is shown as % of total number of distance-filtered frames.

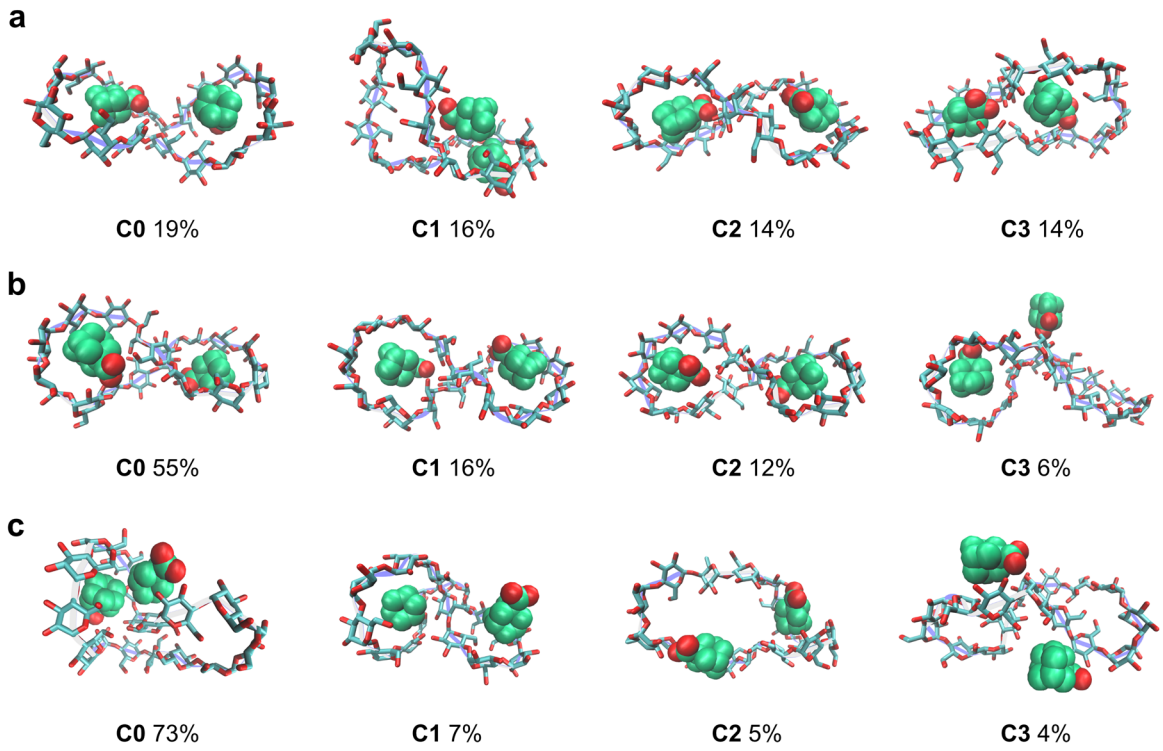


Figure 4.13. Four most dominant structures (**C0**–**C3**) of the complexes of CD16 with two ACAs extracted from the clustering analysis of three distance-filtered MD simulations at 323 K of these complexes with initial heating to (a) 328 K, (b) 318 K and (c) 313 K. Number of frames assigned to each cluster is shown as % of total number of distance-filtered frames.

With these new simulation conditions, the proposed binding motif with one ACA in each loop of a figure-of-eight was observed for large portions of the simulations with CD16. For two of the three parallel simulations, variations of this binding mode was dominant (Figure 4.13a and 4.13b), with roughly 7 glucose units per loop or binding pocket. As hypothesized previously, this structure could explain the cooperativity of the 1:2 binding of ACA to CD16, where the first ACA binding event pays the primary energetic cost (entropic and enthalpic) of forming figure-of-eight conformations with two binding pockets that can accommodate two ACA guests. In the third simulation (Figure 4.13c), the previously observed binding mode, with one helical loop binding both ACAs, was dominant. These results highlight the importance of running several parallel simulations (or very long simulations) to more sufficiently sample the conformational space of these α-1,4-glucans with relatively slow dynamics, especially when binding events also need to be simulated.

The results obtained here support both the proposed figure-of-eight 1:2 binding of ACA to CD16 and the proposal that CD14 is not sufficiently large and/or flexible to efficiently form the same binding motif. The true 1:2 binding mode(s) of ACA to CD16 in aqueous solution might be both of the two main binding modes observed here, but regardless, significant differences between CD14 and CD16 were observed, which is in agreement with the experimental NMR results.

4.3. Synthesis of double adamantane bolaamphiphiles for binding to figure-of-eight LRCDs

The design of the synthetic double adamantane bolaamphiphiles (introduced with schematic representations previously in this chapter) of interest for the binding to figure-of-eight LRCDs are shown in the Figure 4.14. For the flexible linker, a simple alkyl chain with ether linkages to the adamantane moieties was selected, while for the bulky hydrophilic head groups, the '5-hydroxy isophthalic acid'-based head groups (small and large) introduced in chapter 2 were chosen.

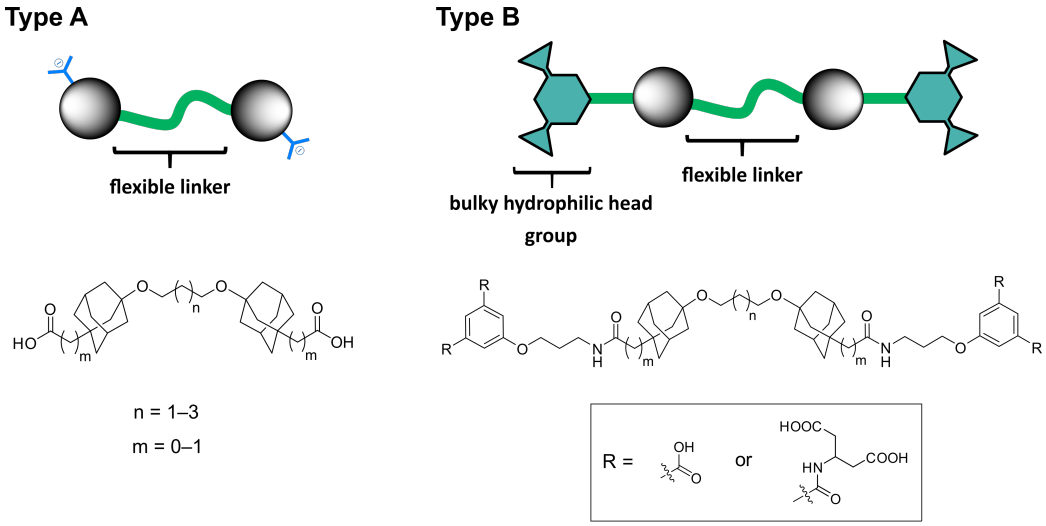
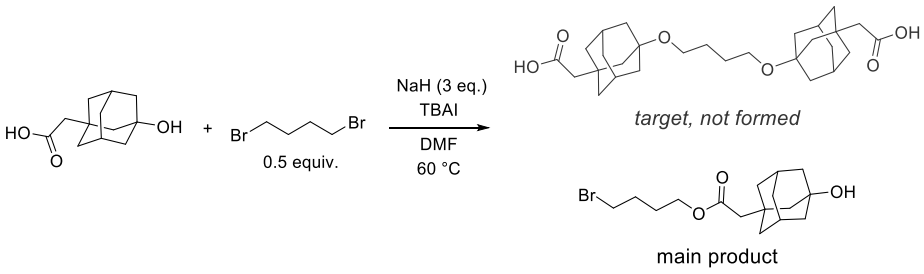


Figure 4.14. Double adamantane bolaamphiphiles of interest.

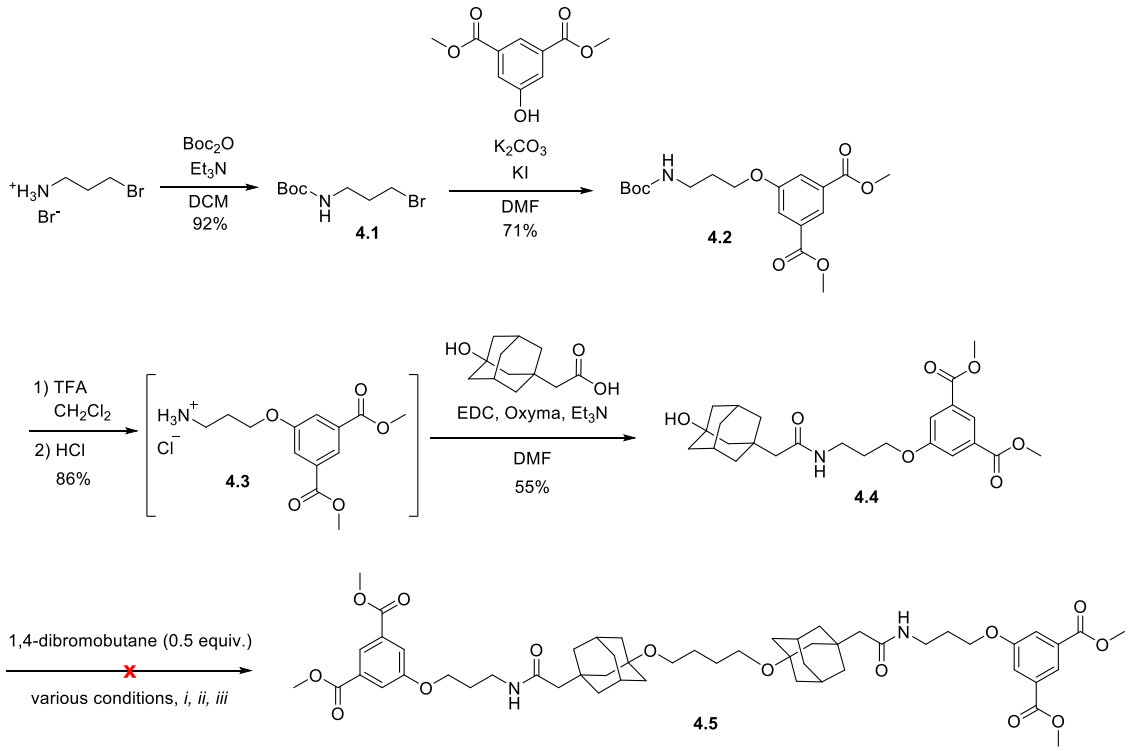
There are numerous reports of the alkylation of 1-hydroxyadamatanes with alkyl bromides (examples[29–31]). To test whether a 'type A' double adamantane bolaamphiphile could be accessed in one step using a alkyl dibromide, a reaction with 3-hydroxy-1-adamantaneacetic acid and 1,4-dibromobutane with sodium hydride was carried out (Scheme 4.1). The general higher nucleophilicity of alkoxides in comparison to carboxylates did not materialize in this case however, and the product formed in the reaction was the alkylation product of the carboxylate, likely due to the relatively low reactivity of the sterically hindered, tertiary alkoxide employed in the reaction.



Scheme 4.1. Alkylation of 3-hydroxy-adamantane-1-acetic acid does not yield the desired product.

To synthesize double adamantane bolaamphiphiles of the design shown in Figure 4.14 by alkylating the tertiary hydroxyl group of adamantane with alkyl dibromides, protection of the carboxylic acid was deemed necessary. The synthesis of 'type B' bolaamphiphile **4.5** with

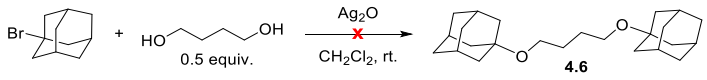
small hydrophilic head groups was thus attempted in an 'inwards' approach, where the 'head groups + linker' were first coupled to the carboxylic acid moiety of 3-hydroxy-1-adamantaneacetic acid before attempting the alkylation of the tertiary alkoxide (Scheme 4.2). Standard Boc-protection of the amine group of 3-bromo-1-propylamine using a published procedure[32] yielded compound **4.1**, which was used in an alkylation reaction with dimethyl 5-hydroxyisophthalate to yield compound **4.2**. Deprotection of the Boc group followed by a peptide coupling with 3-hydroxy-1-adamantaneacetic acid using EDC and Oxyma as coupling reagents yielded **4.4**. Several attempts of the double reaction of 1,4-dibromobutane with **4.4** to yield tetramethyl ester **4.5** of the desired 'type B' bolaamphiphile were unsuccessful. Variations of bases (Cs_2CO_3 or NaH), solvents (THF, DMF) and sources of nucleophilic catalyst iodide (potassium iodide or tetrabutylammonium iodide) and long reaction times (>48 h) did not lead to any detectable amount of compound **4.5** by LC-MS or TLC analysis. It seems that the tertiary adamantane alkoxide is quite unreactive. While most of the conditions (temperature, solvent, reagents) employed in these reaction attempts were similar to the previously published alkylations of the adamantane alkoxide,[29–31] one key difference was the use of an excess of alkyl bromide in the literature protocols. Using an excess of 1,4-dibromobutane would be counterproductive, and in this case and not lead to desired compound **4.5**, but instead to the singly alkylated product.



Scheme 4.2. Synthesis of precursors **4.1–4.4** and attempted syntheses of **4.5**. Conditions: *i)* Cs_2CO_3, DMF, KI, 60 °C. *ii)* NaH, DMF, KI, 60 °C. *iii)* NaH, TBAI, THF, reflux.

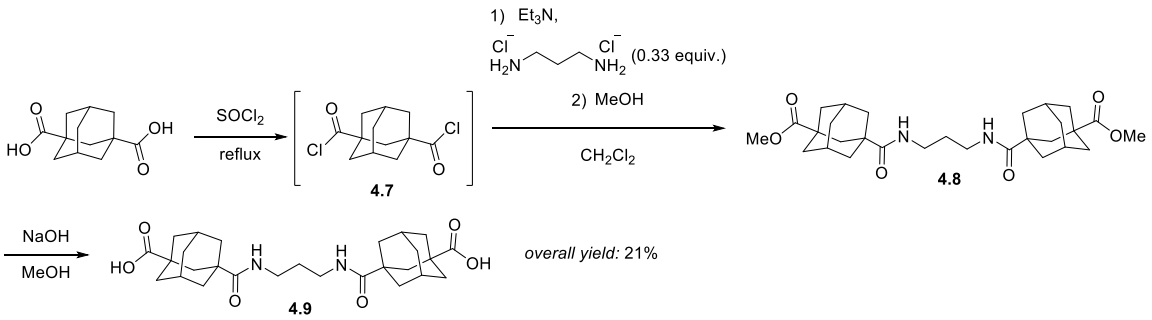
Due to the apparent very low reactivity of the tertiary adamantane alkoxide, a different approach was needed. A type of reaction commonly reported for substituted adamantanes is the S_N1-type reaction of 1-adamantane bromides with alcohols in the presence of base (examples[30,33,34]). These reactions use an excess of the nucleophile, usually by using the

alcohol as the solvent (solvolysis). In one approach using methanol as the nucleophile, $AgNO_3$ was added to the reaction to push the reaction to completion by precipitating bromide as AgBr.[35] To test whether the addition of silver might allow an analogous reaction to occur without excess alcohol, the reaction between 1-bromoadamatane and 1,4-butanediol with Ag_2O as both the silver source and the base was set up (Scheme 4.3). The reaction did not yield the desired product **4.6**, however, and a complex mixture was observed. The approach of using either a tertiary alcohol or tertiary bromide of adamantane in reactions without an excess of nucleophile was thus abandoned at this point.



Scheme 4.3. The reaction of 1-bromoadamantane with 1,4-butanediol (0.5 equiv.) in the presence of Ag_2O does not yield the desired product.

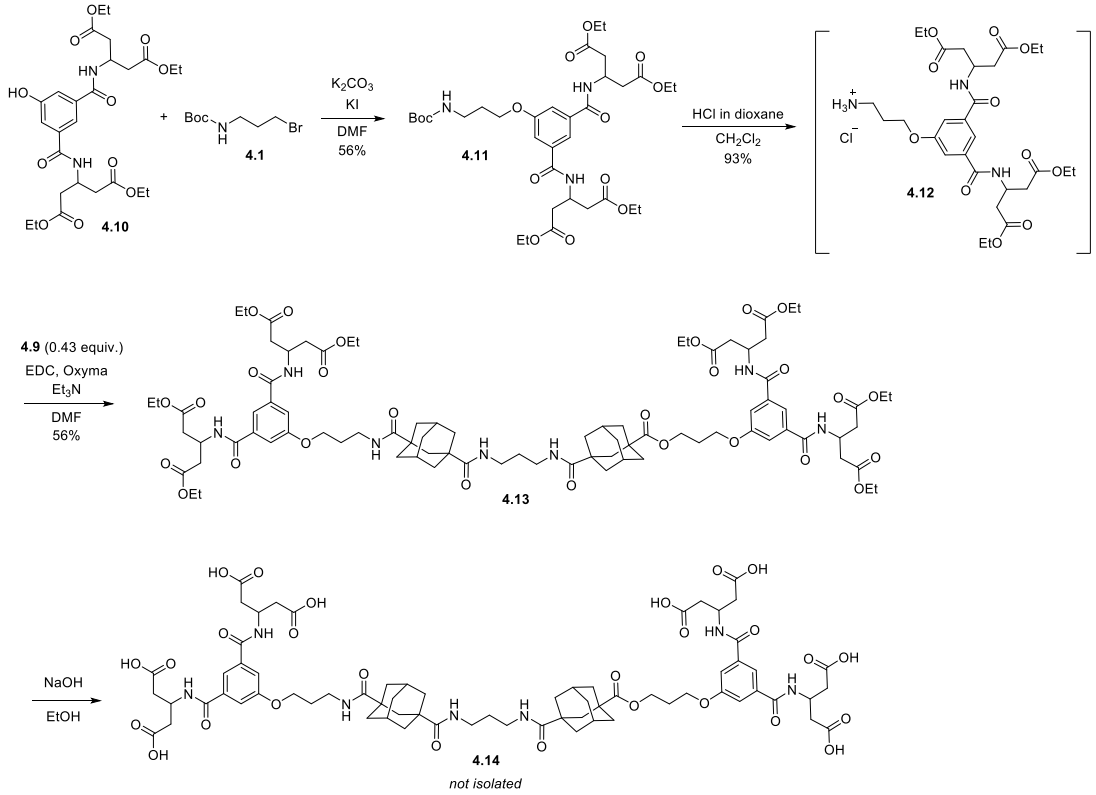
To access type A and B double adamantane bolaamphiphiles, a slightly different synthetic strategy was needed. 1,3-adamantanedicarboxylic acid was chosen as the adamantane source and amide functionalities were selected to join the adamantane moieties to the flexible linker, despite the fact that these amide bonds would provide less flexibility compared to the previously targeted ether functionalities. To synthesize type A bolaamphiphile **4.9** with a propyl linker (Scheme 4.4), 1,3-adamantanedicarboxylic acid was first refluxed in thionyl chloride to provide the corresponding bis-(acyl chloride) **4.7**, which was immediately reacted with 1,3-diaminopropane (with the bis-(acyl chloride) **4.7** in excess to favor the product where each 1,3-diaminopropane had reacted with two equivalents of bis-(acyl chloride) **4.7**). To obtain a product that would be isolable using flash column chromatography on silica gel with standard organic solvents, the reaction was quenched with methanol to yield dimethyl ester **4.8**. Purification of **4.8**, followed by hydrolysis of the methyl esters using sodium hydroxide, gave type A bolaamphiphile **4.9**.



Scheme 4.4. Synthesis of type A bolaamphiphile **4.9**.

The large head group introduced in Chapter 3 that blocks out the binding of α-CD and β-CD was chosen as the bulky hydrophilic head groups for the synthesis of a type B bolaamphiphile based on type A bolaamphiphile **4.9**. Consequently, tetraethyl ester **4.10** (synthesized as described in Chapter 3) was alkylated with Boc-protected 3-bromopropylamine **4.1** to give compound **4.11** (this reaction step was carried out by former MSc student Mie Avnegaard Larsen). Deprotection of the Boc group with HCl in dioxane gave hydrochloride salt **4.12**

which was used in slight excess in the next step — a peptide coupling with dicarboxylic acid **4.9**. The subsequent hydrolysis of octaethyl ester **4.13** yielded the corresponding type B bolaamphiphile as the octa-carboxylic acid, but the compound had significant impurities. At this point, results from an NMR titration had revealed that type A bolaamphiphile **4.9** did not have desirable binding properties to a large-ring cyclodextrin (CD16, see section 4.4.3 later in this chapter). As compounds **4.14** and **4.9** contain the same binding motif, further attempts at the synthesis and isolation of type B bolaamphiphile **4.14** were not made.



Scheme 4.5. Synthesis of compounds **4.11–4.14**.

4.4. Binding studies between adamantane-based guests and CD16

4.4.1. Isolation of large-ring cyclodextrins

To investigate the binding between CD16 and adamantane carboxylate (ACA) as well as binding between the synthesized type A bolaamphiphile **4.9** and CD16, an isolation procedure for the LRCDs was developed in collaboration with PhD student Giorgia Masciotta. A batch of an LRCD mixture with CD6 to ~CD30 was produced (as described in section 4.1) by treating 'Cycloamylose' with the enzymes CGTase and α-glucosidase (CD distribution obtained shown in Figure 4.15, top trace).

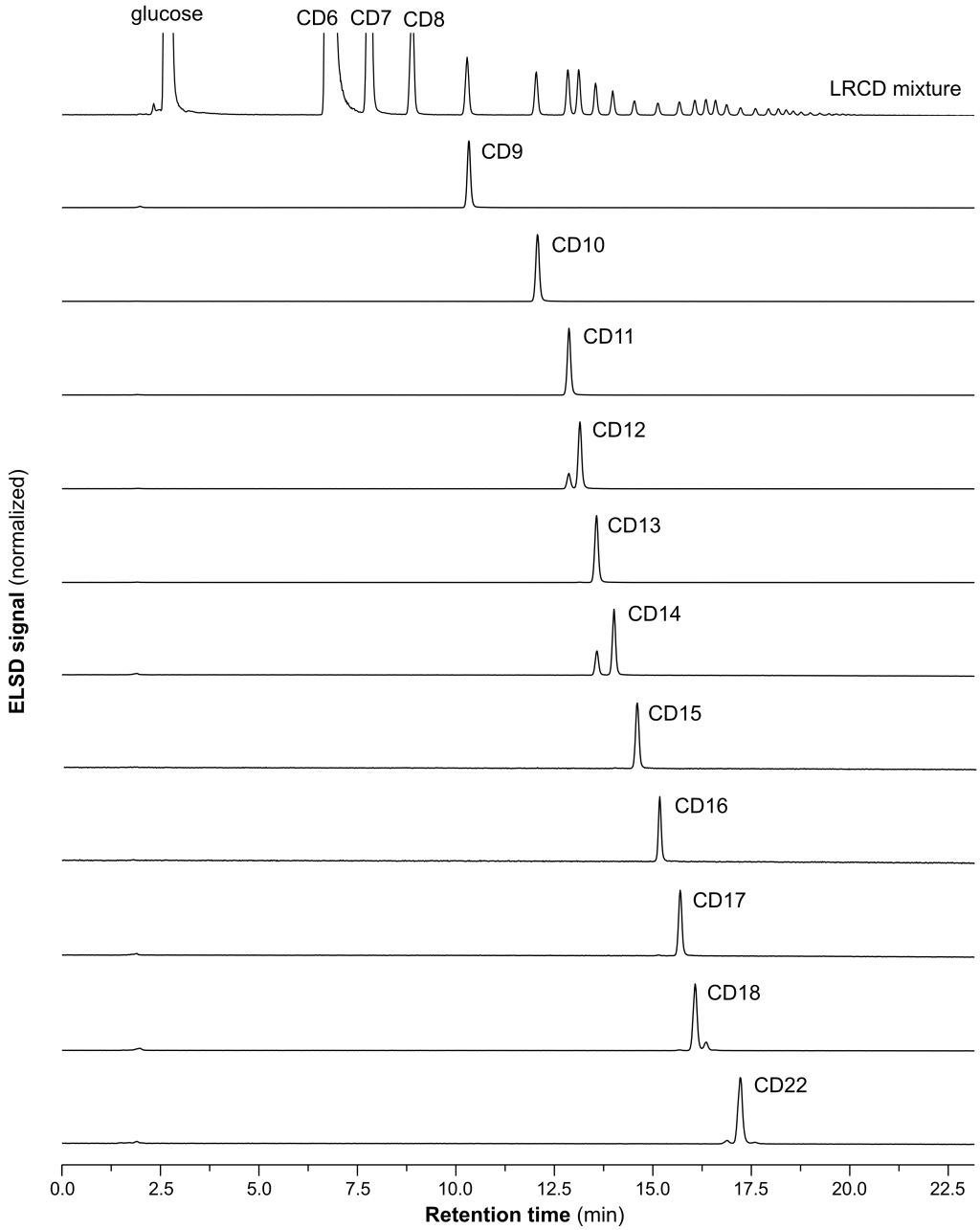


Figure 4.15. HPLC-ELSD chromatograms of LRCD mixture and isolated LRCDs.

The mixture was then purified using preparative HPLC with a HILIC column and a gradient eluent system (water/acetonitrile). A single chromatographic run yielded CD9–CD18 and CD22 in various but generally quite good purities as shown with HPLC-ELSD chromatograms in Figure 4.15. Some of the LRCDs, such as CD9–CD11, CD13 and CD15–CD17 were obtained in high purities judging by the HPLC-ELSD chromatograms. However, due to the non-linearity of the ELS detector,[36–38] small peaks in the chromatograms correspond to larger analyte quantities than are suggested by the size of the peaks. ^{1}H NMR analysis of the CD16 fraction revealed that there was in fact still about 10–15% of CD15 + CD17 present. To obtain CD16 in high purity for NMR titrations, CD16 was purified three more times using the same preparative HPLC system.

4.4.2. HSQC titration with CD16 and adamantane carboxylate

A high resolution HSQC titration with the isolated CD16 and adamantane carboxylate was then carried out. The titration was carried out at 50 °C (like the titration with the LRCD mixture) in phosphate buffered D_2O (0.1 M, pH 7.5) with CD16 at a constant concentration of 0.05 mM and ACA concentrations up to 37.5 mM (Figure 4.16).

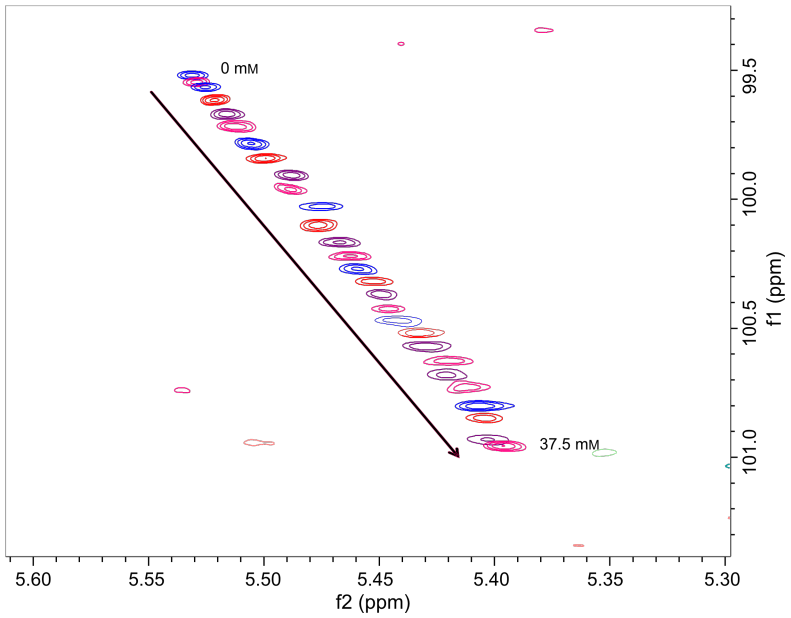


Figure 4.16. ^{1}H-^{13}C HSQC (800/202 MHz) spectra showing the anomeric proton of CD16 in the titration with CD16 (0.05 mM) and adamantanecarboxylic acid (0–37.5 mM) in phosphate buffered D_2O (0.1 M, pH 7.5) at 50 °C. Arrow indicates movement of peak during titration.

The change in chemical shift in the ^{13}C dimension was then fitted in the software Origin to 1:1 and 1:2 binding models, as outlined by Anslyn and co-workers[39] and modified for use with NMR data (Supporting section 4.1). Like the 1:1 fitting procedure outlined in Chapter 3, the 1:2 binding model also fits two types of parameters: The binding constants (K_{a1} and K_{a2}) and the difference in chemical shifts between the unbound host and the fully bound host ($\Delta\delta_{max1}$ and $\Delta\delta_{max2}$). The fit to a 1:1 binding model was poor, showing systematic error, a clear sinusoidal distribution[40,41] in the residuals (Figure 4.17) and a suspiciously large fitted $\Delta\delta_{max}$ (6.6 ppm). The fit to a 1:2 binding model was clearly better, with a more stochastic distribution

of residuals with lower numerical values and $\Delta\delta_{max}$ values in a seemingly more realistic range ($\Delta\delta_{max1}$ = 0.7 ppm and $\Delta\delta_{max2}$ = 2.6 ppm). As such, the titration provides further evidence of the 1:2 binding of ACA to CD16 with fitted binding constants of K_{a1} = 42 ± 22 M^{-1} and K_{a2} = 38 ± 11 M^{-1}. While these values are numerically slightly different than the values obtained with CD16 from the NMR titration of the LRCD mixture (K_{a1} = 24 ± 5 M^{-1} and K_{a2} = 94 ± 11 M^{-1}), they lead to the same conclusion, namely that the 1:2 binding exhibits positive cooperativity. As described in more detail in the Chapter 1, positive allosteric cooperativity exists where $K_2 > K_1/4$, which is the case for both sets of binding constants obtained from the two different titrations.

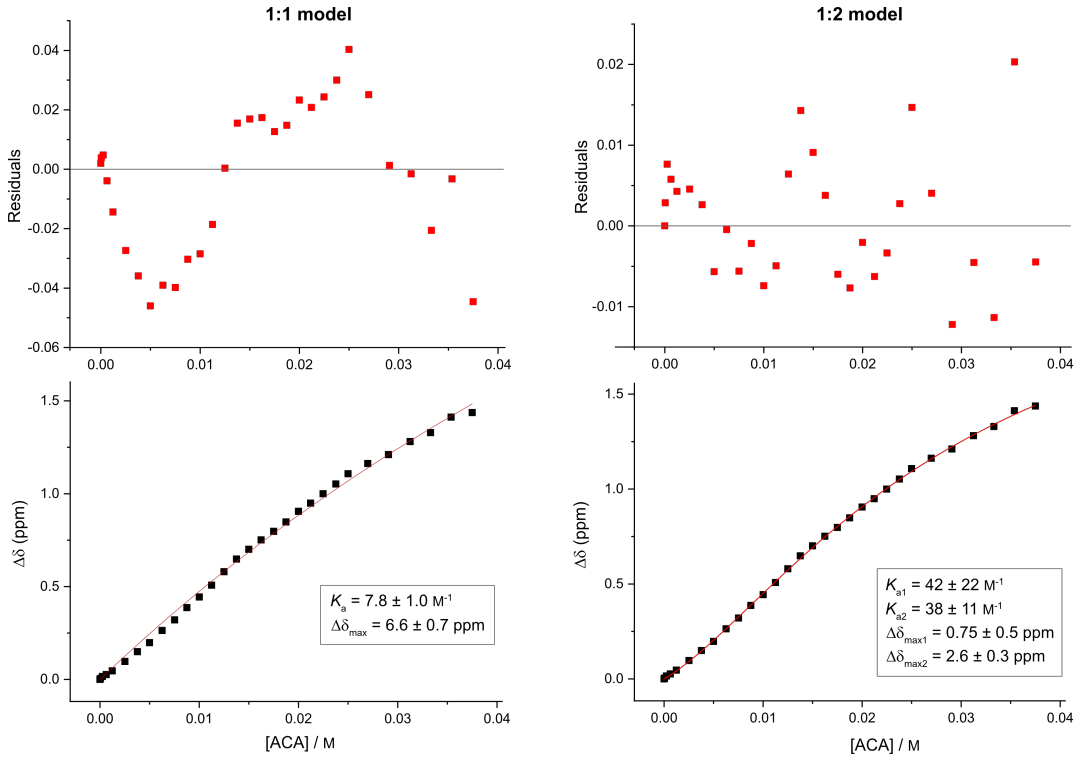


Figure 4.17. Fits to 1:1 (poor fit) and 1:2 (good fit) binding models of the ^{13}C observed chemical shift changes of the anomeric proton of CD16 in the presence of ACA.

4.4.3. ^{1}H NMR titration with CD16 and bolaamphiphile 4.9

To investigate whether the design of double adamantane bolaamphiphile **4.9** leads to improved binding to CD16 compared to ACA due to possible chelate cooperativity, a ^{1}H NMR titration with CD16 (0.023 mM) and **4.9** (0–11.6 mM) was carried out in phosphate buffered D_2O (0.1 M, pH 7.5). Partial ^{1}H NMR spectra from the titration are shown in Figure 4.18, while the fit of the change in chemical shift for the anomeric proton to a 1:1 binding model is shown in Figure 4.19. The fit to a 1:1 binding model was good, albeit with a slight systematic trend in the residuals. Attempted fits to a 1:2 binding model gave nonsensical values of binding parameters. As such, it seems that bolaamphiphile **4.9** binds to CD16 primarily with a 1:1 stoichiometry, as expected due to the presence of the two adamantyl moieties in **4.9** and the two binding pockets of CD16. However, the fitted binding constant of $K_{a(\mathbf{4.9})}$ = 61 ± 6 does not indicate that the binding occurs with positive chelate cooperativity.

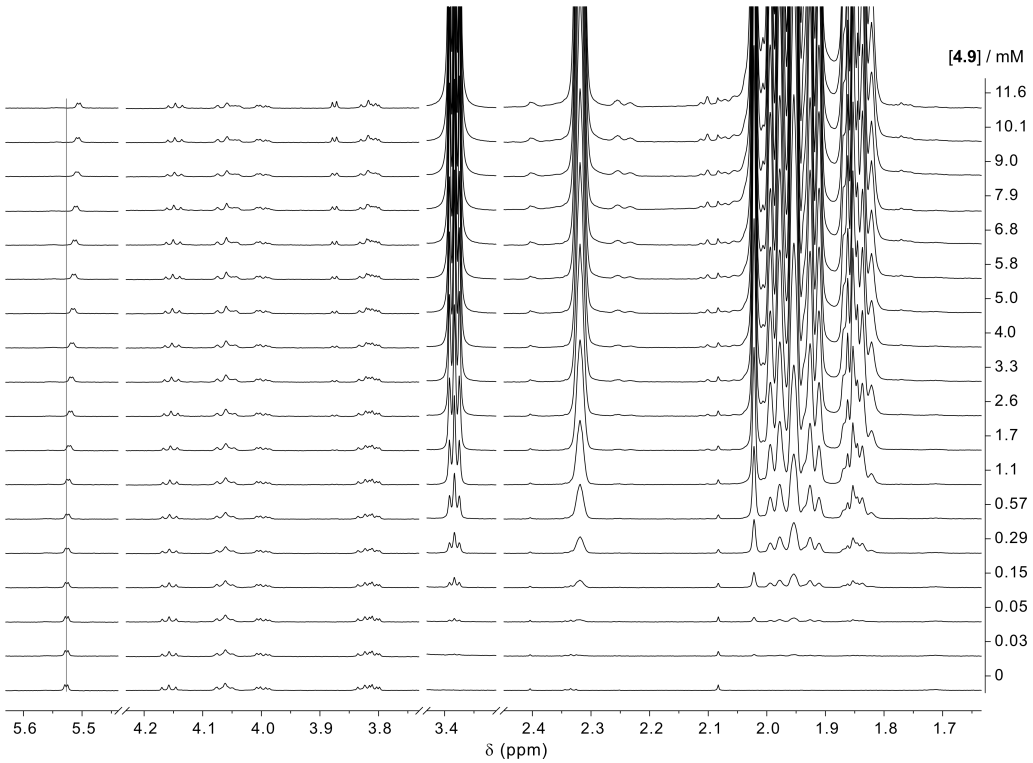


Figure 4.18. Partial ^{1}H NMR (800 MHz) spectra of CD16 (0.023 mM) in phosphate buffered D_2O (0.1 M, pH 7.5) with increasing concentrations of bolaamphiphile **4.9**.

In a case with no net positive or negative chelate cooperativity, the binding affinity of the double adamantane guest **4.9** can be predicted to be the square of first binding constant of ACA to CD16, $K_{a(\mathbf{4.9}\_pred)} = K_{a1(ACA)}^2$, assuming two identical binding sites in CD16.[16] This would give a predicted binding constant of **4.9** to CD16 of $K_{a(\mathbf{4.9}\_pred)} = 1.8 \times 10^3$ M^{-1}, using the previously obtained value $K_{a1(ACA)} = 42$ M^{-1}. The actual obtained fitted binding constant is more than an order of magnitude lower than this predicted value, and the binding of **4.9** to CD16 can thus be described as exhibiting negative chelate cooperativity.

An explanation for this negative cooperativity could be that a considerable amount of unfavorable strain is incurred in either CD16 or **4.9** to facilitate the effective binding of the second adamantyl moiety of **4.9** to CD16. Another closely related explanation could be that when the first adamantyl moiety of **4.9** is bound to the one of the binding pockets of CD16, the second adamantyl moiety can only partly reach the second binding pocket of CD16, leading to a suboptimal or even absent second binding event. In both cases, a longer and/or more flexible linker between the adamantyl moieties of **4.9** might help alleviate these issues.

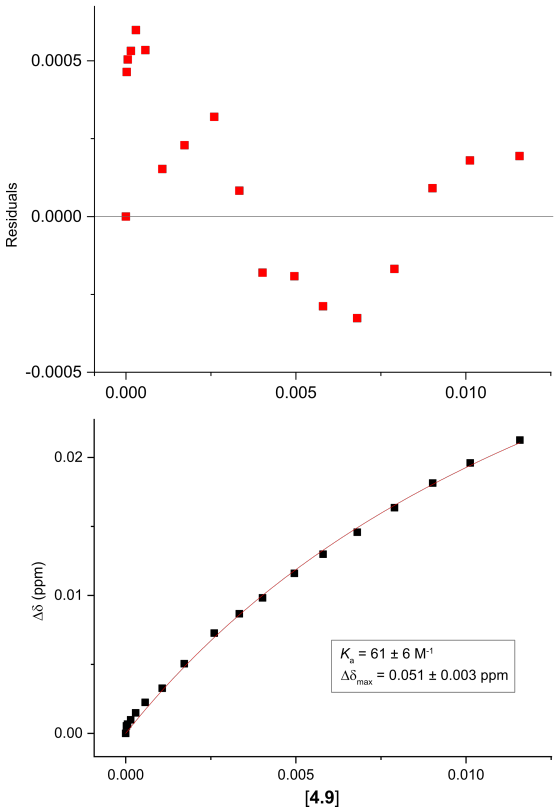
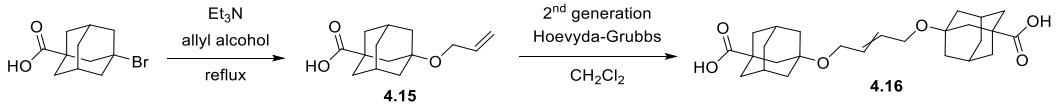


Figure 4.19. Fit to a 1:1 binding model of the observed chemical shift change for the anomeric proton of CD16 with increasing concentrations of bolaamphiphile **4.9.**

4.5. Conclusion

In this chapter, the properties of large-ring cyclodextrins CD12, CD14 and CD16 was explored. MD simulations of CD12, CD14 and CD16 substantiated the expected geometric trends in large-ring cyclodextrins, where CD12 forms mainly bent-boat like macrocyclic structures, while CD14 and CD16 are more flexible and form figure-of-eight-like geometries. MD simulations with CD12, CD14 and CD16 and adamantane carboxylate confirmed the trends observed in previous studies in the group, where CD16 is able to effectively bind two adamantane carboxylate in two loops of a figure-of-eight, while CD14 and CD12 are not able to do this. An NMR titration with isolated CD16 confirmed the cooperative 1:2 binding of adamantane carboxylate to CD16. Synthetic double adamantane bolaamphiphile **4.9** was also titrated with CD16, where it was found that the binding exhibited negative chelate cooperativity, presumably due to a sub-optimal linker design.

For new iterations of the design of a double adamantane bolaamphiphile, a longer and/or more flexible linker would be of interest. One simple possibility would be to increase the length of the alkyl chain linking the two adamantyl moieties. Another approach could be to avoid the inflexible amide bonds, which was also the initial design principle employed in this chapter. To do so, a different synthetic approach would be needed. One idea would be to use the commonly employed reaction of a tertiary adamantane bromide with an excess of alcohol in the presence of base, as previously described. By using allyl alcohol (Scheme 4.6), compound **4.15** with a terminal alkene could be produced, which could then be homo-coupled in a subsequent alkene metathesis reaction to yield **4.16**. These two steps were carried out by the author on a test scale, where crude **4.15** was obtained by refluxing 3-bromo-1-adamantanecarboxylic acid in allyl alcohol with triethylamine overnight. The addition of 2nd generation Hoveyda-Grubbs catalyst to crude **4.15** in dichloromethane showed good conversion to the desired product by LC-MS analysis. In possible subsequent steps, the internal alkene of **4.16** could then be hydrogenated to yield a bolaamphiphile with a flexible alkyl linker. A similar approach could be to use an alcohol with a terminal alkyne such as propargyl alcohol for the reaction with 3-bromoadamantanecarboxylic acid. The resulting product with terminal alkynes could then be connected using a Glaser-type homo-coupling.



Scheme 4.6. Test scale synthesis towards double adamantane bole-amphiphiles with a more flexible linker.

4.6. Experimental

4.6.1. Materials

For materials, see section 3.7.1.

The commercially available mixture of large-ring cyclodextrins 'Cycloamylose' was purchased from commercial supplier Ezaki Glico Co., LTD. (product number 302-32406, average molecular weight: 7300 g/mol).

4.6.2. Instrumentation and Methods

For Instrumentation and Methods, see section 3.7.2.

High-resolution ^{1}H-^{13}C HSQC titration with CD16 and adamantane carboxylate

Sensitivity enhanced ^{1}H-^{13}C HSQC spectra were recorded on an 800 MHz Bruker Avance III spectrometer equipped with a TCI Z-gradient CryoProbe and an 18.7 T magnet (Oxford Magnet Technology, Oxford, UK) using the hsqcetgpsisp2.2 pulse sequence with shaped pulses for all 180 degree pulses on ^{13}C. The spectra were acquired as matrices of 512×100 complex data points sampling 80 milliseconds in the direct (^{1}H) and 110 milliseconds in the indirect (^{13}C) dimension. The spectra were acquired with a spectral width of 8.0 ppm in the direct dimension with a transmitter frequency offset of 4.7 ppm and with a spectral with of 4.5 ppm in the indirect dimension with a transmitter frequency offset of 101.5 ppm. The spectra were acquired at 50 °C with non-uniform sampling of 30% of the data points in the indirect dimension. With concentrations of adamantane carboxylate up to 10 mM, 8 scans were collected. With higher concentrations of adamantane carboxylate, 16 scans were collected to acquire sufficient signal due to broadening of the signal corresponding to the anomeric proton of CD16.

4.6.3. Computational details

General details

All minimizations and MD simulations were carried out using the Amber20 suite.[18] For large-ring cyclodextrins CD12, CD14 and CD16, the GLYCAM-06 force field[19] was employed, while for adamantane carboxylate (ACA), the general Amber force field (GAFF)[26] was employed. Water molecules were represented using the TIP3P model.[15] All manual manipulations of starting geometries were performed in the molecular software PyMOL.[42] Snapshots and cluster representatives from MD simulations were visualized using the Visual Molecular Dynamics (VMD) package.[23] All MD simulations were carried out using an 8 Å cut-off for nonbonded interactions and the Particle Mesh Ewald method[43] for long range electrostatics. Bonds to hydrogen atoms were constrained using the SHAKE algorithm[44] and a time step of 2 fs was applied in all cases. MD simulations were performed at constant temperature and constant pressure. Constant temperature (T = 300 K or 323 K) was applied using the Langevin thermostat with a collision frequency of 1 ps^{-1} while constant pressure (p = 1 bar) was applied using the Berendsen barostat[45] with isotropic position scaling.

All analyses of MD trajectories were performed in the CPPTRAJ module[21] of Amber20. Clustering analysis was performed using the 'k-means' algorithm[20] to output 10 clusters based on root-mean square deviation (RMSD) of non-hydrogen atoms. Cluster representatives were chosen based on the lowest distance to cluster centroid.

Generation of CD12, CD14 and CD16 starting geometries and simulations without ACA

The crystal structure of CD14[5] (CCDC deposition number: 124917) was chosen as a starting point for generating the starting geometries of CD12 and CD16, similar to the work by Ivanov.[7] The structures of CD12 and CD16 were then generated by manually deleting or adding extra glucose units and then reconnecting the α-1,4-glycosidic bonds. The two *anti* oriented glucose units in the crystal structure of CD14 were retained in the structures of CD12 and CD16. The generated structures for CD12 and CD16 were then minimized for 5000 steps in vacuum. The first 4800 steps were performed using the steepest descent method, while the remaining 200 steps were performed using the conjugate gradient method. The minimized CD12 and CD16 structures and the crystal structure of CD14 were then solvated with water in the LEaP program of Amber20 using a truncated octahedron with a distance buffer of 10 Å between the CDs and the edges of the periodic box. The systems were then heated linearly from 10 K to 300 K within 0.3 ns followed by 0.7 ns of equilibration at 300 K in the *NVT* ensemble using the Langevin thermostat[46] with a collision frequency of 5 ps^{-1}. Equilibration (4 ns) and production runs using constant temperature (300 K) and pressure were then carried out using the same settings as described previously.

Simulations with CD12, CD14 and CD16 and one or two ACAs at 300 K

As starting geometries of CD12 and CD16, the structures corresponding to the final structure obtained in the 4 ns equilibration with constant pressure of each of these CDs (see above) were used. For CD14, the crystal structure was used. The structure of ACA was based on a published crystal structure of adamantane carboxylic acid (CCDC deposition number: 1283437).[25] The starting geometries of the complexes were generated by manually placing one or two ACAs with a minimum distance of 5 Å to the CDs (and to each other, when two ACAs were used). The complexes were then solvated with water in the LEaP program of Amber20 using a truncated octahedron with a distance buffer of 20 Å between the CDs and the ACAs and the edges of the periodic box. The systems were neutralized using one or two sodium ions (corresponding to the one or two adamantane carboxylates). The solvated and neutralized systems were then minimized for 500 steps. The first 400 steps were performed using the steepest descent method followed by 100 steps of the conjugate gradient method. The systems were then heated linearly from 10 K to 300 K within 0.3 ns followed by 0.7 ns of equilibration at 300 K in the *NVT* ensemble using the Langevin thermostat with a collision frequency of 5 ps^{-1}. Equilibration (4 ns) and production runs using constant temperature (300 K) and pressure (1 bar) were then carried out using the same settings as described previously.

Simulations with CD14 and CD16 and two ACAs at 323 K

For the starting geometry of the CD14 and CD16, the structure corresponding to the most dominant clusters (**C0**) of the simulations of these CDs with no guest was used. The starting geometry of the complexes were generated by manually placing two ACAs with a minimum distance of 5 Å to the CDs and to each other. The complexes were then solvated with water in the LEaP program of Amber20 using a truncated octahedron with a distance buffer of 10 Å between the CDs and the ACAs and the edges of the periodic box. The systems were then neutralized using two sodium ions, and the solvated and neutralized systems were then minimized for 500 steps. The first 400 steps were performed using the steepest descent method followed by 100 steps of the conjugate gradient method.

Three parallel simulations for each CD with slightly different starting conditions were then set up. To generate the different starting conditions, the systems were initially heated to three different temperatures. The minimized systems were thus heated linearly from 10 K to either 313 K, 318 K or 328 K within 0.3 ns followed by 0.7 ns of equilibration at 323 K in the *NVT* ensemble using the Langevin thermostat with a collision frequency of 5 ps^{-1}. Equilibration (4 ns) and MD production runs using constant temperature (323 K) and pressure (1 bar) were then carried out using the same settings as described previously.

4.6.4. Production of LRCD mixture

Carried out by PhD Student Giorgia Masciotta

Cycloamylose (1 g) was dissolved in water (100 ml) and treated with a commercial stock solution of CGTase (100 μl). The resulting solution was mixed by gentle agitation and then kept at room temperature for 8 hours. The reaction was then heated to 100 °C (to denature the enzyme CGTase) for 15 minutes and then allowed to cool to 60 °C. α-glucosidase stock solution (100 μl, 150 U) was then added to the reaction mixture, which was mixed by gentle agitation and then kept at 60 °C overnight. The reaction was then heated to 100 °C for 30 minutes (to denature the enzyme α-glucosidase) and the solvent removed *in vacuo*. The resulting solids were then dissolved in water (50 ml), filtered, and the filtrate freeze dried to yield the LRCD mixture as a white solid (0.83 g). For product distribution, see Figure 4.15.

4.6.5. General procedure for the isolation of LRCDs

LRCD mixture (0.1 g) was dissolved in a 1:1 mixture of water and acetonitrile (2 ml), leading to an almost clear solution with a small amount of precipitate. The mixture was then filtered through a syringe filter and injected on a preparative HILIC type HPLC column (XBridge BEH Amide OBD Prep Column from Waters, 130 Å, 5 μm, 19 × 150 mm) using a Buchi Reveleris Prep Purification system equipped with an ELS detector. Gradient elution: 25% water in acetonitrile to 59% water in acetonitrile over 94 minutes. Peaks corresponding to different CDs or groups of CDs were split into separate fractions during the run, which were then analyzed by HPLC to assess purity. This separation process was repeated three more times to accumulate more material. About 2–4 mg each of CD9–CD18 in various purities (see Figure 4.15) were obtained.

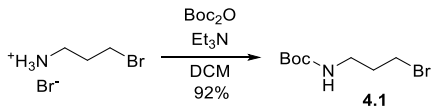
4.6.6. Isolation of CD16

Fractions with mainly CD16 (10–15% of CD15 and CD17 were also present) from two separate runs of the separation of LRCDs were combined, the solvents evaporated *in vacuo* and the resulting solid dissolved in a 1:1 mixture of water and acetonitrile (1 ml) and injected on the same preparative chromatographic equipment used for the separation of LRCDs using gradient elution, 40% water in acetonitrile to 53% water in acetonitrile over 33 minutes. The fraction containing mainly CD16 was then collected and concentrated *in vacuo* to yield a white solid, which was used in two more rounds of the same purification procedure to yield CD16 in high purity (0.9 mg).

^{1}H NMR (800 MHz, D_2O) δ 5.38 (d, J = 3.8 Hz, 16H, **H1**), 3.99 (t, J = 9.5 Hz, 16H, **H3**), 3.89 (m, 32H, **H5**, **H6**), 3.83 (dd, J = 12.5, 5.1 Hz, 16H, **H6'**), 3.66 (t, J = 9.5 Hz, 16H, **H4**), 3.65–3.60 (dd, J = 9.8, 3.8 Hz, 16H, **H2**). **^{13}C NMR** (201 MHz, D_2O, from ^{1}H-^{13}C HSQC spectrum) δ 99.1 (**C1**), 76.8 (**C4**), 73.0 (**C3**), 71.6 (**C2**), 71.0 (**C5**), 60.4 (**C6**). **MS** (MALDI) *m/z*: $[M+Na]^+$ calcd. for $C_{96}H_{160}O_{80}Na$ 2616.838; found: 2616.858.

4.6.7. Synthetic protocols

<u>*t*-Butyl (3-bromopropyl)carbamate **4.1**</u>

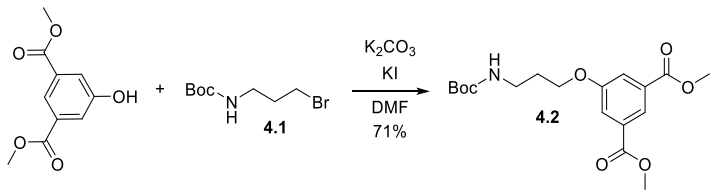


t-Butyl (3-bromopropyl)carbamate was synthesized according to a published procedure.[32]

To an ice-cooled suspension of 3-bromopropan-1-amine hydrobromide (1.0 g, 4.6 mmol) in dry dichloromethane (5 ml) was added di-*tert*-butyl dicarbonate (1.0 g, 4.6 mmol). Triethylamine (1.27 ml, 9.2 mmol) was then added dropwise over the course of 10 minutes. The reaction was allowed to slowly heat to room temperature and stirred overnight. The reaction mixture was diluted with dichloromethane (15 ml) and washed successively with hydrochloric acid (2 × 8 ml, 1 M) and water (2 × 8 ml), dried over Na_2SO_4, filtered and the solvent removed *in vacuo* to yield the product **4.1** as a yellow oil (1.0 g, 92%), which was used in the next step without further purification.

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.64 (br, 1H), 3.44 (t, J = 6.5 Hz, 2H), 3.27 (m, 2H), 2.05 (p, J = 6.5 Hz, 2H), 1.44 (s, 9H).

<u>Compound **4.2**</u>



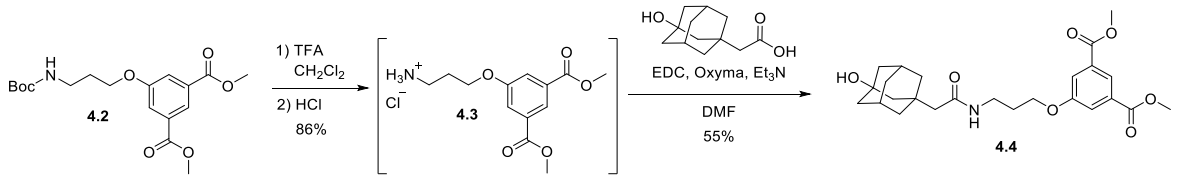
To a solution of dimethyl 5-hydroxyisophthalate (1.02 g, 4.83 mmol) and *t*-butyl (3-bromopropyl)carbamate **4.1** (1.0 g, 4.2 mmol) in DMF (55 ml) was added potassium carbonate

(1.45 g, 10.5 mmol) and potassium iodide (0.21 g, 4.2 mmol). The resulting mixture was stirred for 24 hours at 60 °C, after which the solvent was removed *in vacuo*. The resulting residue was partitioned between ethyl acetate (120 ml) and water (70 ml). The aqueous phase was then extracted once more with ethyl acetate (50 ml) and the combined organic phases were washed successively with aqueous NaOH (2 × 40 ml, 1 M), hydrochloric acid (2 × 50 ml, 0.2 M) and brine (40 ml), dried over Na_2SO_4, filtered and evaporated *in vacuo*. The product was isolated using automated flash column chromatography (25 g silica cartridge, gradient elution from 10%–70% ethyl acetate in *n*-heptane) and the solvent was removed *in vacuo* to yield the product **4.2** as a white solid (1.1 g, 71%).

The NMR data was in agreement with a reported spectrum.[47]

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.27 (br, 1H), 7.73 (br, 2H), 4.72 (br, 1H), 4.10 (t, J = 6.1 Hz, 2H), 3.93 (s, 6H), 3.33 (q, J = 6.4 Hz, 2H), 2.01 (p, J = 6.1 Hz, 2H), 1.44 (s, 9H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 166.3, 159.0, 156.1, 131.9, 123.2, 119.9, 52.6, 32.0, 28.5, 22.8, 14.3. **M.P.** 70–72 °C.

Compound **4.4**



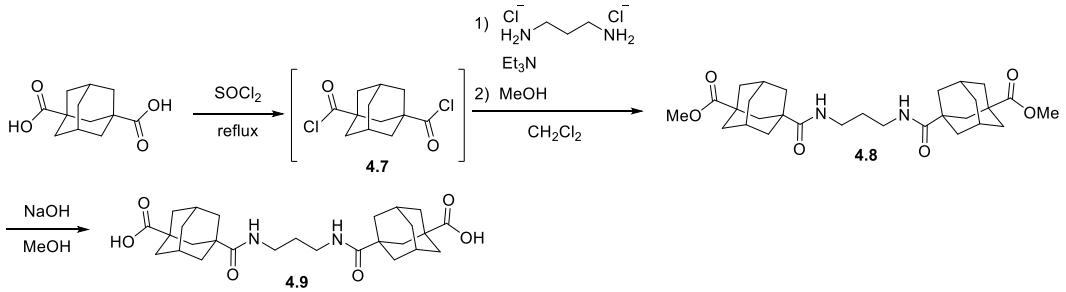
Compound **4.2** (0.93 g, 2.5 mmol) was dissolved in a mixture of dichloromethane (8 ml) and trifluoroacetic acid (3 ml) and the resulting solution was stirred at room temperature for one hour. The solvents were then removed *in vacuo* and the residue suspended in hydrochloric acid (20 ml). The solvent was then removed *in vacuo* and the resulting residue suspended in hydrochloric acid (20 ml) again. The solvent was then removed *in vacuo* to yield HCl salt **4.3** as a white solid (0.67 g, 86%), which was used in the next step without further purification.

HCl salt **4.3** (0.19 g, 0.63 mmol), 3-hydroxyadamantaneacetic acid (0.14 g, 0.67 mmol), EDC hydrochloride (0.13 g, 0.66 mmol), Oxyma (94 mg, 0.66 mmol) and triethylamine (0.1 ml, 0.72 mmol) were dissolved in DMF (10 ml), forming a yellow solution which was stirred at room temperature for 24 hours. The mixture was then transferred to a separatory funnel and water (120 ml) was added. The resulting solution was extracted with ethyl acetate (100 ml, then 2 × 50 ml). The combined organic phases were then washed with hydrochloric acid (60 ml, 1 M), saturated $NaHCO_3$ (2 × 60 ml) and brine (50 ml), dried over Na_2SO_4, filtered and evaporated *in vacuo*. The product was isolated using dry column vacuum chromatography (*Ø*: 2 cm, length: 7 cm, 50 ml fractions, crude product adsorbed on Celite) eluting with a 7.5% gradient of acetonitrile in toluene (starting from 30% acetonitrile in toluene) to yield the product **4.4** as a waxy white solid (0.16 g, 55%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.28 (t, J = 1.3 Hz, 1H), 7.74 (d, J = 1.3 Hz, 2H), 5.66 (t, 6.3 Hz, 1H), 4.13 (t, J = 5.8 Hz, 2H), 3.94 (s, 6H), 3.47 (q, J = 6.3 Hz, 2H), 2.19 (s, 2H), 2.10–1.98

(m, 4H), 1.72–1.49 (m, 12H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ171.1, 166.6, 132.3, 123.7, 120.2, 69.2, 67.4, 53.0, 51.2, 50.7, 44.9, 41.7, 37.6, 36.6, 35.7, 31.0, 29.5. **HRMS** (MALDI) *m/z*: $[M+Na]^+$ calcd. for $C_{25}H_{33}NO_7Na$ 482.2149; found: 482.2147.

Compound **4.9**

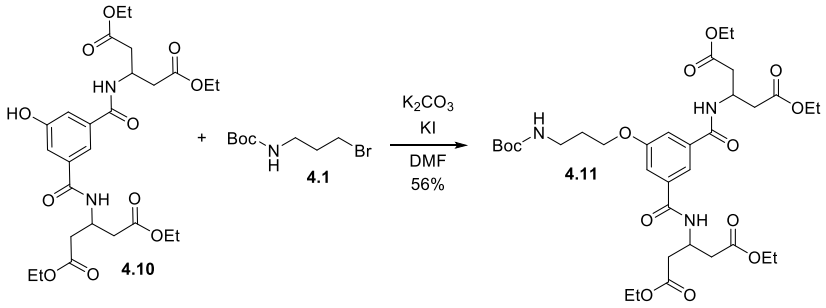


Adamantane-1,3-dicarboxylic acid (0.47 g, 2.1 mmol) was dissolved in thionyl chloride (8 ml) and the resulting solution was stirred at reflux overnight. The solvent was then removed *in vacuo* to yield acyl chloride **4.7**. 1,3-diaminopropane dihydrochloride (102 mg, 0.69 mmol) was then suspended in dry CH_2Cl_2 (8 ml) with triethylamine (0.95 ml, 6.8 mmol) in a flame-dried addition funnel. Acyl chloride **4.7** was dissolved in dry CH_2Cl_2 (10 ml) and the solution was cooled to 0 °C. The suspension of 1,3-diaminopropane dihydrochloride and triethylamine was then added dropwise over the course of 20 minutes to the solution of acyl chloride **4.7** under a nitrogen atmosphere. Residue of 1,3-diaminopropane dihydrochloride left behind in the addition funnel was rinsed down and added to the reaction using dry CH_2Cl_2 (10 ml). The reaction was then stirred for 1.5 hours, after which it was quenched by the addition of methanol (30 ml) at 0 °C. The quenched reaction was left at room temperature overnight. The solvents were removed *in vacuo* and the resulting residue was suspended in CH_2Cl_2 and washed with hydrochloric acid (2 × 30 ml, 0.2 M) and brine (30 ml), dried over Na_2SO_4, filtered and the solvent removed *in vacuo*. Dimethyl ester **4.8** was then isolated as a pale yellow solid (87 mg, 26% yield based on 1,3-diaminopropane dihydrochloride) using automated flash column chromatography (12 g silica cartridge, gradient elution from 10%–100% ethyl acetate in *n*-heptane). Dimethyl ester **4.8** (85 mg, 0.17 mmol) was dissolved in methanol (8 ml) and then a solution of NaOH in water (2 ml, 1 M) was added and the reaction was stirred at room temperature for 24 hours. The solvents were then removed *in vacuo* and the resulting residue dissolved in water (14 ml). The reaction was acidified with a few drops of hydrochloric acid (6 M) leading to precipitation of the product **4.9** a white solid, which was isolated by vacuum filtration and dried *in vacuo* (63 mg, 79% yield based on dimethyl ester **4.8.** Overall yield for all steps 21%).

^{1}H NMR (400 MHz, D_2O phosphate buffer, 0.1 M, pH 7.5) δ 3.20 (t, J = 6.7 Hz, 4H), 2.22–2.05 (m, 4H), 1.88–1.60 (m, 26H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 187.0, 181.6, 42.7, 41.5, 50.9, 38.7, 38.0, 36.7, 35.1, 28.5, 28.2 **HRMS** (MALDI) *m/z*: $[M+H]^+$ calcd. for $C_{27}H_{39}N_2O_6$ 487.2803; found: 487.2805. **M.P**.: 132–135 °C.

Compound **4.11**

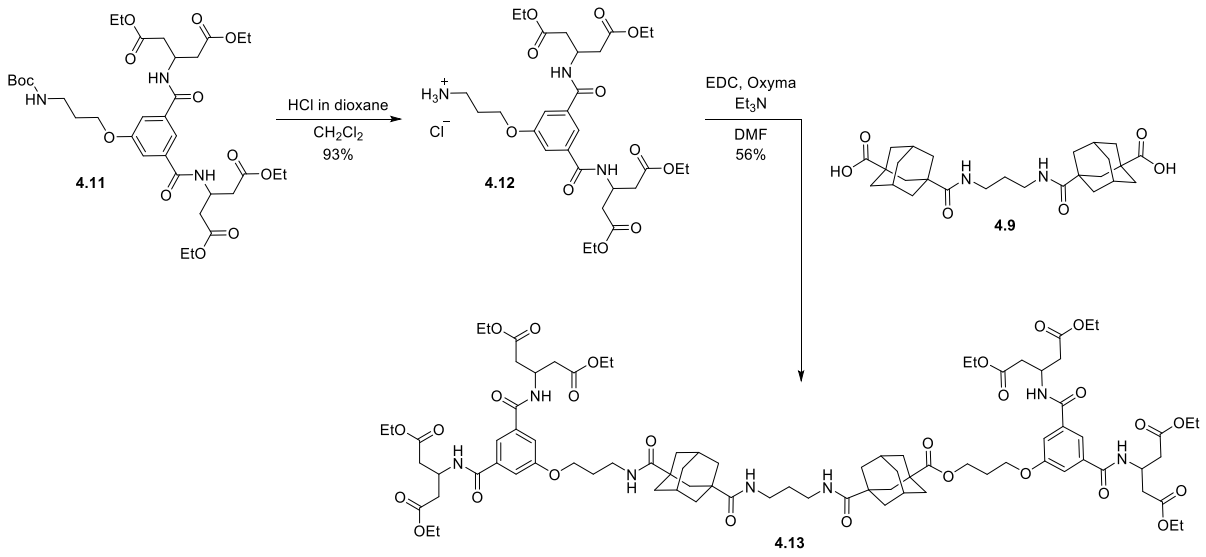
Synthesized by (MSc student at the time) Mie Avnegaard Larsen.



To a solution of compound **4.10** (1.05 g, 1.90 mmol) and *t*-butyl (3-bromopropyl)carbamate **4.1** (0.51 g, 2.1 mmol) in DMF (55 ml) was added K_2CO_3 (0.53 g, 3.8 mmol) and KI (0.15 g, 0.93 mmol). The solution was stirred at 60 °C for 24 h and then concentrated *in vacuo* until approximately 5 ml of solvent remained. The solution was then diluted with water (75 ml) and extracted with ethyl acetate (3 × 75), and the combined organic extracts were washed with aqueous NaOH (2 × 40 ml, 2 M), hydrochloric acid (2 × 50 ml, 0.1 M), brine (50 ml), dried over $MgSO_4$, filtered and concentrated *in vacuo*. The product was isolated using dry column vacuum chromatography (Ø: 4 cm, length: 7 cm, 50 ml fractions, crude product adsorbed on celite, gradient eluting with 6.6% increments of ethyl acetate in *n*-heptane), yielding the product **4.11** as a colorless oil (0.75 g, 56%).

^{1}H NMR (400 MHz,) δ 7.69 (s, 1H), 7.43 (s, 2H), 7.24 (d, J = 8.6 Hz, 2H), 4.82 (dp, J = 8.6, 5.8 Hz, 2H), 4.71 (s, 1H), 4.17 (q, J = 7.1 Hz, 8H), 4.10 (t, J = 5.8 Hz, 2H), 3.33 (m, 2H), 2.97–2.54 (m, 8H), 2.00 (p, J = 6.3 Hz, 2H), 1.44 (s, 12H), 1.26 (t, J = 7.1 Hz, 12H). (^{13}C NMR not collected). **HRMS** (MALDI) *m/z*: $[M+Na]^+$ calcd. for $C_{34}H_{51}N_3O_{13}Na$ 732.3314; found: 732.3307.

Compound **4.14**



Compound **4.11** (0.19 g, 0.27 mmol) was dissolved in dry dichloromethane (2 ml). HCl in dioxane (3.7–4.3 M, 4 ml) was then added and the reaction was stirred at room temperature for two hours. The solvents were then removed *in vacuo* to yield **4.12** as a waxy pale yellow solid (0.16 g, 56%) which was used in the next step without further purification.

To a solution of **4.9** (38 mg, 0.078 mmol) and **4.12** (110 mg, 0.18 mmol) in DMF (3 ml) was added triethylamine (24 µl, 0.17 mmol), EDC hydrochloride (31 mg, 0.16 mmol) and Oxyma (23 mg, 0.16 mmol). The resulting yellow solution was then stirred overnight at room temperature. Ethyl acetate (40 ml) was added, and the mixture was washed successively with hydrochloric acid (10 ml, 1 M), aqueous NaOH (2 × 10 ml, 1 M) and brine (10 ml), then dried over $MgSO_4$, filtered and evaporated *in vacuo*. The product was then isolated using dry column vacuum chromatography (*Ø*: 2 cm, length: 8 cm, 50 ml fractions, crude product adsorbed on Celite) eluting with a 0.7% gradient of methanol in dichloromethane to yield the product **4.13** as a white solid (73 mg, 56%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.65 (t, J = 1.4 Hz, 2H), 7.52 (d, J = 8.6 Hz, 4H), 7.37 (d, J = 1.4 Hz, 4H), 6.64 (t, J = 6.2 Hz, 2H), 6.40 (t, J = 5.5 Hz, 2H), 4.82 (dp, J = 8.6, 6.2 Hz, 4H), 4.15 (q, J = 7.1 Hz, 16H), 4.06 (t, J = 5.6 Hz, 4H), 3.45 (q, J = 6.1 Hz, 4H), 3.20 (q, J = 6.1 Hz, 4H), 2.94–2.68 (m, 16H), 2.18 (s, 4H), 2.07–1.61 (m, 30H), 1.53 (p, 5.6 Hz, 2H), 1.25 (t, J = 7.1 Hz, 24H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 178.0, 177.4, 171.4, 165.8, 159.0, 136.1, 117.9, 116.2, 67.0, 61.0, 44.0, 41.3, 41.2, 40.9, 38.5, 38.4, 38.1, 37.3, 35.5, 35.3, 29.7, 29.0, 28.3, 14.3. **HRMS** (MALDI) *m/z*: $[M+H]^+$ calcd. for $C_{85}H_{121}N_8O_{26}$ 1669.8387; found: 1669.8331. **M.P**.: 71–74 °C.

4.7. Bibliography

[1] S. Meier, S. R. Beeren, *J. Am. Chem. Soc.* **2014**, *136*, 11284–11287.

[2] T. Fujiwara, N. Tanaka, S. Kobayashi, *Chem. Lett.* **1990**, *19*, 739–742.

[3] K. Harata, H. Akasaka, T. Endo, H. Nagase, H. Ueda, *Chem. Commun.* **2002**, 1968–1969.

[4] T. Endo, *Trends Glycosci. Glycotechnol.* **2011**, *23*, 79–92.

[5] J. Jacob, K. Gessler, D. Hoffmann, H. Sanbe, K. Koizumi, S. M. Smith, T. Takaha, W. Saenger, *Angew. Chem. Int. Ed.* **1998**, *37*, 605–609.

[6] M. G. Gotsev, P. M. Ivanov, *J. Phys. Chem. B* **2009**, *113*, 5752–5759.

[7] P. Ivanov, *J. Mol. Struct.* **2012**, *1009*, 3–10.

[8] S. Kitamura, K. Nakatani, T. Takaha, S. Okada, *Macromol. Rapid Commun.* **1999**, *20*, 612–615.

[9] M. V. Rekharsky, Y. Inoue, *Chem. Rev.* **1998**, *98*, 1875–1917.

[10] P. M. Ivanov, C. Jaime, *J. Phys. Chem. B* **2004**, *108*, 6261–6274.

[11] M. G. Gotsev, P. M. Ivanov, *Int. J. Quantum Chem.* **2007**, *107*, 1657–1672.

[12] I. Maestre, I. Beà, P. M. Ivanov, C. Jaime, *Theor. Chem. Acc.* **2007**, *117*, 85–97.

[13] P. M. Ivanov, *J. Phys. Chem. B* **2010**, *114*, 2650–2659.

[14] P. Ivanov, E. Atanassov, C. Jaime, *J. Mol. Struct.* **2014**, *1056–1057*, 238–245.

[15] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[16] E. V. Anslyn, D. A. Dougherty, *Modern Physical Organic Chemistry*, University Science Books, **2006**.

[17] C. A. Hunter, H. L. Anderson, *Angew. Chem. Int. Ed.* **2009**, *48*, 7488–7499.

[18] E. Sugawara, H. Nikaido, *Antimicrob. Agents Chemother.* **2014**, *58*, 7250–7257.

[19] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.

[20] J. A. Hartigan, M. A. Wong, *J. R. Stat. Soc.* **1979**, *28*, 100–108.

[21] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

[22] S. Cross, M. M. Kuttel, J. E. Stone, J. E. Gain, *J. Mol. Graph. Model.* **2009**, *28*, 131–139.

[23] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38.

[24] J. K. Koneru, X. Zhu, J. Mondal, *J. Chem. Theory Comput.* **2019**, *15*, 6203–6212.

[25] F. Bélanger-Gariépy, F. Brisse, P. D. Harvey, D. F. R. Gilson, I. S. Butler, *Can. J. Chem.* **1990**, *68*, 1163–1169.

[26] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[27] A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623–1641.

[28] G. Fittolani, P. H. Seeberger, M. Delbianco, *Pept. Sci.* **2019**, e24124.

[29] P. J. Kropp, G. E. Fryxell, M. W. Tubergen, M. W. Hager, G. D. Harris, T. P. McDermott, R. Tornero-Velez, *J. Am. Chem. Soc.* **1991**, *113*, 7300–7310.

[30] S. Inomata, Y. Harada, S. I. Matsuoka, T. Ishizone, *Tetrahedron* **2013**, *69*, 3238–3248.

[31] N. Kandoth, E. Vittorino, M. T. Sciortino, T. Parisi, I. Colao, A. Mazzaglia, S. Sortino, *Chem. Eur. J.* **2012**, *18*, 1684–1690.

[32] A. Barnard, K. Long, D. J. Yeo, J. A. Miles, V. Azzarito, G. M. Burslem, P. Prabhakaran, T. A. Edwards, A. J. Wilson, *Org. Biomol. Chem.* **2014**, *12*, 6794–6799.

[33] K. J. C. van Bommel, G. A. Metselaar, W. Verboom, D. N. Reinhoudt, *J. Org. Chem.* **2001**, *66*, 5405–5412.

[34] A. N. Reznikov, M. Y. Skomorokhov, Y. N. Klimochkin, *Russ. J. Org. Chem.* **2010**, *46*, 1741–1742.

[35] P. Šilhár, N. R. Silvaggi, S. Pellett, K. Čapková, E. A. Johnson, K. N. Allen, K. D. Janda, *Bioorg. Med. Chem.* **2013**, *21*, 1344–1348.

[36] A. Stolyhwo, H. Colin, G. Guiochon, *J. Chromatogr. A* **1983**, *265*, 1–18.

[37] B. A. Kimball, W. M. Arjo, J. J. Johnston, *J. Liq. Chromatogr. Relat. Technol.* **2004**, *27*, 1835–1848.

[38] L. E. Magnusson, D. S. Risley, J. A. Koropchak, *J. Chromatogr. A* **2015**, *1421*, 68–81.

[39] A. E. Hargrove, Z. Zhong, J. L. Sessler, E. V. Anslyn, *New J. Chem.* **2010**, *34*, 348–354.

[40] F. Ulatowski, K. Dabrowa, T. Bałakier, J. Jurczak, *J. Org. Chem.* **2016**, *81*, 1746–1756.

[41] D. B. Hibbert, P. Thordarson, *Chem. Commun.* **2016**, *52*, 12792–12805.

[42] **N.d.**

[43] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[44] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.

[45] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[46] R. W. Pastor, B. R. Brooks, A. Szabo, *Mol. Phys.* **2006**, *65*, 1409–1419.

[47] R. P. Megens, T. A. van den Berg, A. D. de Bruijn, B. L. Feringa, G. Roelfes, *Chem. Eur. J.* **2009**, *15*, 1723–1733.

Chapter 5. Templated enzymatic synthesis of large-ring cyclodextrins

Abstract

The following chapter is divided into two subchapters that both deal with the templated enzymatic synthesis of large-ring cyclodextrins.

The first subchapter describes the endeavours towards the templated enzymatic synthesis of very large-ring cyclodextrins using (bola)amphiphilic templates. The experimental work showed how CD26, CD27, CD38 and other large-ring cyclodextrins could be kinetically trapped in CGTase-mediated libraries using the commercially available mixture of very large cyclodextrins called 'Cycloamylose' as starting material. A range of different templates were employed, which, along with MD simulations, revealed how the length of the aliphatic chain of (bola)-amphiphiles, as well as the nature of the hydrophilic head groups, play important roles in determining the binding to CD26 and the other large-ring cyclodextrins.

The second subchapter is a brief description of the high-yielding, chromatography-free templated enzymatic synthesis of δ-CD. By employing the superchaotropic anion $B_{12}Cl_{12}^{2-}$ as template and α-CD as starting material, an unprecedented yield of 50% δ-CD was achieved. The template could be re-isolated and re-used in subsequent reactions.

*In subchapter 5.1, experimental work with templates **5.1** and **5.2** was carried out by PhD Student Giorgia Masciotta. The rest of the work described was carried out by the author.*

In subchapter 5.2, the synthesis of $Na_2B_{12}Cl_{12}$ and the synthesis and isolation of δ-CD was carried out in collaboration with PhD student Kasper Hornstrup Hansen. Optimizations with α-CD as starting material (only described briefly here) was carried out by Kasper Hansen. Reactions with starch were carried out by the author.

5.1. Introduction

The templated enzymatic synthesis of large-ring cyclodextrins (LRCDs) is one of the main topics of this thesis. LRCDs are still relatively unexplored due to synthetic inaccessibility, and Chapters 3 and 4 dealt with the design, development and exploration of new binding motifs for LRCDs. This chapter describes the exploration of the templated enzymatic synthesis of LRCDs using binding motifs that are more well-established. The first part of this chapter (subchapter 5.1) deals with targeting the synthesis of CD26 and other very large LRCDs, where the binding of alkyl-based amphiphiles in a 1:2 stoichiometry (CD/amphiphile) has been established for CD26 through crystal structures with undecanoic acid and dodecanol[1] (Figure 5.1) and through isothermal titration calorimetry (ITC) with sodium dodecyl sulfate (SDS).[2] The second part of this chapter (subchapter 5.2) deals with the templated enzymatic synthesis of δ-CD using superchaotropic dodecaborate clusters that were reported as high-affinity binders to CD9–CD11.[3]

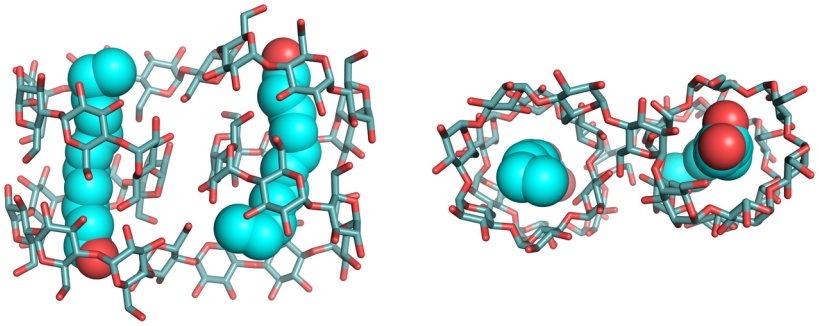


Figure 5.1. Crystal structure[1] of CD26 in complex with undecanoic acid showing the binding of the alkyl chains of two undecanoic acids to the two V-amylose-like helical segments of CD26. *Left*: Side view. *Right:* Top view.

5.2. CD26 as a target for templated enzymatic synthesis

5.2.1. Previous work

Templates **5.1** and **5.2** (Figure 5.2) were designed according to the same principles as the design of bolaamphiphile templates **T1**–**T6** introduced in Chapter 3: bulky hydrophilic head groups, (in the case of **5.1** and **5.2** based on trisodium 8-hydroxypyrene-1,3,6-trisulfonate (HTPS)), serve to give water solubility and block the binding of small CDs to the central alkyl chain. The syntheses and analyses of bolaamphiphile templates **5.1** and **5.2** were carried out by PhD Student Giorgia Masciotta. To investigate the possible templated enzymatic synthesis of LRCDs using these templates, DCLs starting from a commercially available mixture of LRCDs called 'Cycloamylose' (consisting of CD24 to CD60+) were set up. The HPLC-ELSD chromatograms obtained from the reaction are depicted in Figure 5.3. In an untemplated library starting from Cycloamylose (10 mg/ml) in sodium phosphate buffer (50 mM, pH 7.5) treated with CGTase (10 μl per ml reaction mixture), the LRCDs were quickly converted into mainly α-, β-, and γ-CD. In the initial parts of the reaction (1–8 hours), δ-CD and CD10–CD23 were also observed. After 18 hours reaction time, all LRCDs were converted into α-, β-, and γ-CD. In contrast, in the templated DCL with template **5.1** (10 mM) and with otherwise identical reaction conditions to the untemplated reaction, the concentration of CD26 was observed to slightly build up initially and remain in the library until after 18 hours of reaction time. After 72 hours of reaction time, all LRCDs (except for δ-CD) had been converted into α-, β-, and γ-CD. CD38 was also observed to remain in the library with template **5.1**. In contrast, in the DCL in the presence of template **5.2** (10 mM), which has a shorter central alkyl chain compared to template **5.1**, no build-up of CD26 was seen, and all LRCDs (except δ-CD) were removed from the library at approximately the same rate as in the untemplated library. In a different set of experiments, α-CD or γ-CD were used as starting materials (10 mg/ml) in enzymatic reactions with template **5.1** (10 mM). Under these conditions, no CD26 (or CD38) was observed.

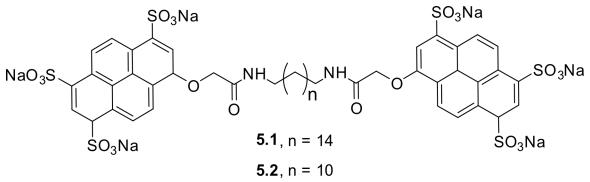


Figure 5.2. Structure of bolaamphiphile templates **5.1** and **5.2**.

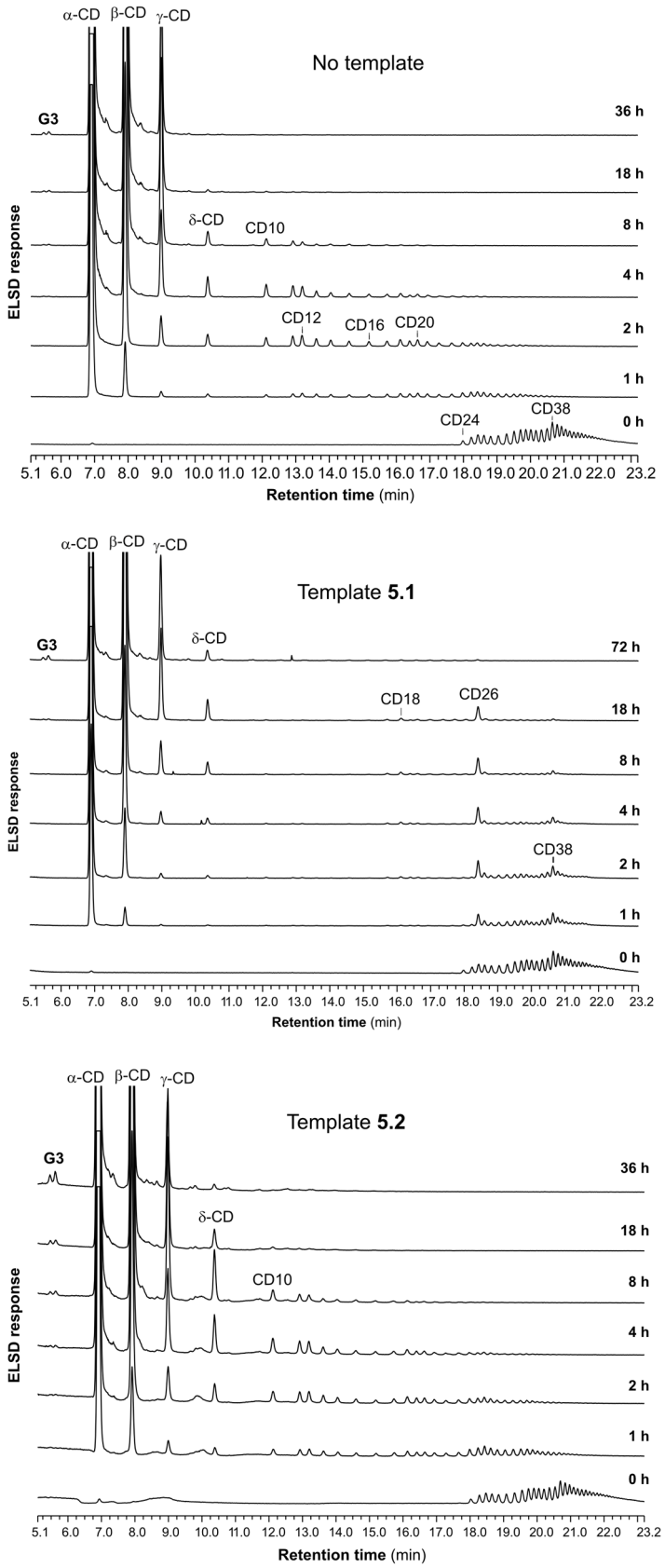


Figure 5.3. HPLC-ELSD chromatograms showing the product distributions in reactions with Cycloamylose (10 mg/ml) in the absence or presence of templates **5.1** and **5.2** (10 mM) treated with CGTase (10 µl per ml reaction mixture) in sodium phosphate buffer (50 mM, pH 7.5).

5.2.2. Kinetic trapping of CD26 as the 1:2 complex with template 5.1

The fact that CD26 remains for a while in the DCL in the presence of template **5.1**, after which CD26 is converted into the small CDs, seems to indicate that this effect is mainly due to the kinetic trapping of CD26 with template **5.1**. If the subsystem of CDs (small CDs and CD26) was primarily under *pseudo*-thermodynamic control,[4] one would except a (somewhat) stable distribution of these CDs before the eventual conversion of the CD subsystem into small linear α-1,4-glucans and ultimately glucose as the true thermodynamic product. The fact that no CD26 is seen in reactions starting from α-CD or γ-CD supports this idea that CD26 is kinetically trapped in reactions starting from Cycloamylose with template **5.1**. If the small CDs (α-, β-, and γ-CD) and CD26 were truly interconverting in a dynamic system under *pseudo*-thermodynamic control, the identity of the α-1,4-glucan starting material should be largely inconsequential.[4]

The kinetic trapping of CD26 with template **5.1** is proposed to be caused by the formation of a 1:2 complex between CD26 and template **5.1**, where the alkyl chain of the template threads and binds to two V-amylose-like helical channels of CD26. This type of binding mode for CD26 was observed in the crystal structure of CD26 in complex with undecanol and undecanoic acid,[1] which is depicted in Figure 5.1, where each V-amylose segment consist of two helical turns with 6 glucoses in each turn of the helix. The fact that template **5.1**, which has a longer alkyl chain than template **5.2**, leads to the kinetic trapping of CD26, and template **5.2** does not, shows that there is a requirement for a specific length of the alkyl chain for effective complexation to occur. The crystal structure with undecanoic acid shows that an alkyl chain with 11 carbons is long enough to thread the full length of the channel, but only just long enough. As template **5.2** has an alkyl chain with 12 carbons (and then short amide linkers connecting the alkyl chain to the HPTS groups), the lack of kinetic trapping with **5.2** could be due to the steric bulk of the HPTS head groups leading to steric and/or electronic clashes between CD26 and the head groups. Another possibility for the lack of binding of **5.2** to CD26 could be a kinetic effect. The proposed threading of templates **5.1** and **5.2** through the V-amylose-like channels would require a certain amount of flexibility; both from the template and CD26, and the shorter linker of **5.2** would provide less flexibility.

The fact that CD38 also remained in the DCL with template **5.1** can be explained with a similar binding mode as the proposed binding mode for **5.1** and CD26. CD38, with 12 additional glucose units, should be able to form a structure analogous to CD26, but with one additional helical turn (using 6 glucose units) for each helical V-amylose segment, giving 3 helical turns per segment.[5,6]

5.2.3. Computational study of the binding mode between template 5.1 and CD26

To get a better understanding of this possible binding mode between bolaamphiphilic template **5.1** and CD26, a molecular dynamics (MD) simulation of the complex was carried out. The simulation was carried out using the AMBER20 suite[7] and employing the GLYCAM-06 force field[8] for CD26 and the general amber force field (GAFF)[9] for template **5.1** (for computational details see experimental section 5.5.6). The starting geometry was manually

generated with two **5.1** templates pre-threaded through the V-amylose-like helical channels of CD26. The resulting structure was solvated with TIP3P water,[10] energy minimized, heated to 300K and equilibrated. An MD production run (300 ns) was then carried out in the *NPT* ensemble. The resulting trajectory was then subjected to clustering analysis using the 'kmeans' algorithm.[11] The representative structures of the complex extracted from the most dominant cluster is displayed in Figure 5.4. Here, the structure shows how the linker between the HPTS head groups of template **5.1** is long enough to effectively fill out the cavity of the V-amylose-like segments of CD26, while also providing enough flexibility for the head groups to explore the most stable positions and geometries. All of the aromatic HPTS moieties of the templates are effectively stacked against the hydrophobic phase of one or more glucose units of CD26, thus minimizing the hydrophobic areas exposed to the water. This effect, in addition to the possible other mechanisms mentioned previously, could be contributing to the lack of kinetic trapping of CD26 with template **5.2**, which has the shorter linker between the HPTS groups and thus less flexibility to stack the HPTS groups against the hydrophobic phases of the glucose units.

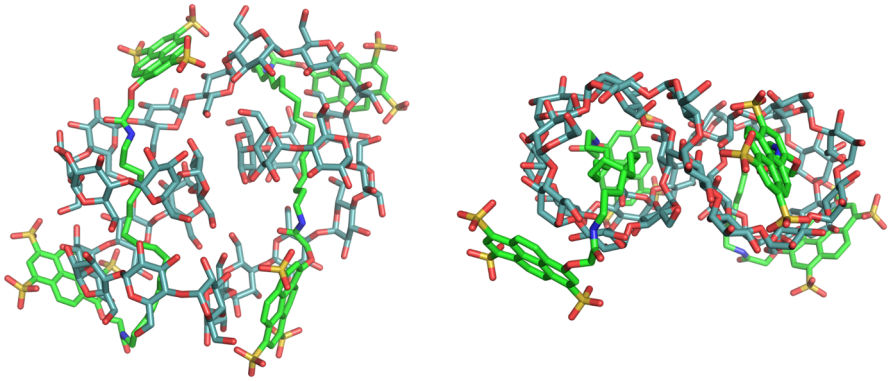


Figure 5.4. The representative structure of the most dominant cluster from clustering analysis of the MD simulation (300 ns) of the 1:2 complex between CD26 and two templates **5.1**.

5.2.4. Exploring other templates for the kinetic trapping and synthesis of CD26

Based on these experimental and computational results with template **5.1** and CD26, the next goal to explore whether the kinetic trapping of CD26 was specific to template **5.1**, or if other templates could provide similar results. Template **5.1** had also proven tedious to synthesize in large quantities, so to expand the experimental possibilities and to be able to move to a preparatory scale for the templated enzymatic synthesis of CD26, different templates that were easier to acquire or synthesize were of interest. As such, bolaamphiphile templates **T1** and **T2** (Figure 5.5) introduced in Chapter 3, were explored for the synthesis of CD26.

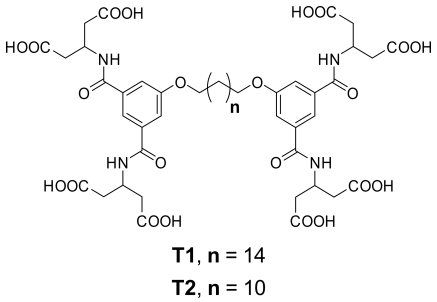


Figure 5.5. Structure of bolaamphiphile templates **T1** and **T2** introduced in Chapter 3.

A series of reactions were set up with 'Cycloamylose' (10 mg/ml) and glycerol-free CGTase (10 μl per ml reaction mixture) in sodium phosphate buffer (0.2 M, pH 7.5) in the presence of templates **T1** and **T2** (10 mM). The HPLC-ELSD chromatograms obtained from the reactions are shown in Figure 5.6. Remarkably, both in the reactions with **T1** and **T2**, CD26 was kinetically trapped and remained in the libraries while the other LRCDs were gradually removed. In the library with **T2**, CD26 was also clearly observed to build up in the library. Another intriguing result is that CD27 remained in both libraries as well. One noteworthy difference here, is that CD27 remained for a longer time in the library with **T1**. CD38 also remained in the library with template **T1**, which has the longer C_{16} alkyl chain, while CD38 did not remain in the library with **T2**. The later stages of the libraries closely resemble the libraries with **T1** and **T2** described in Chapter 3, where a distribution of α-, β-, γ-CD (and δ-CD) and linear α-1,4-glucans is achieved. The long linear α-1,4-glucans (**G9–G20**) and the LRCDs can be clearly distinguished by treating an aliquot of the reaction mixture (where CGTase has been denatured) with the enzyme α-glucosidase, which hydrolyzes all the linear α-1,4-glucans down to glucose while leaving all CDs untouched (Figure 5.6 (a)).

The kinetic trapping of CD26 and CD27 with **T1** and **T2** (and CD38 with **T1**) and the difference between the libraries with templates **T1**, **T2**, **5.1** and **5.2** can be rationalized by proposing the same binding mode for **T1** and **T2** to these LRCDs as for **5.1**. The less bulky head groups of **T1** and **T2** (compared to **5.1** and **5.2**) seems to enable the binding of **T2**, which has a "short" C_{12} alkyl chain (+ an ether linkage), to CD26. This is in contrast with template **5.2**, where a C_{12} alkyl chain (+ an amide linker) is seemingly not long enough — presumably due to the bulky HPTS head groups. The fact that CD38 is kinetically trapped with **T1** and not **T2** can be explained by the shorter C_{12} alkyl chain of **T2**, which, by comparison with the crystal structure of CD26 with undecanoic acid (C_{11} alkyl chain, Figure 5.1), does not seem long enough to accommodate a third turn of the helix required for the binding to CD38. On the other hand, the C_{16} alkyl chains and linkers of **T1** and **5.1** are evidently long enough to enable the binding to CD38, even for **5.1**, which has the larger HPTS head groups.

To explain the kinetic trapping of CD27 with **T1** and **T2**, a similar binding mode of these bolaamphiphiles as for CD26 is proposed. It seems plausible that CD27 could exhibit a similar geometry to the crystal structure of CD26, albeit with more distortion and less symmetry than CD26, as CD27 has to accommodate an extra glucose unit somewhere in the structure. Molecular dynamics simulations in water have also shown that the conformations exhibited by

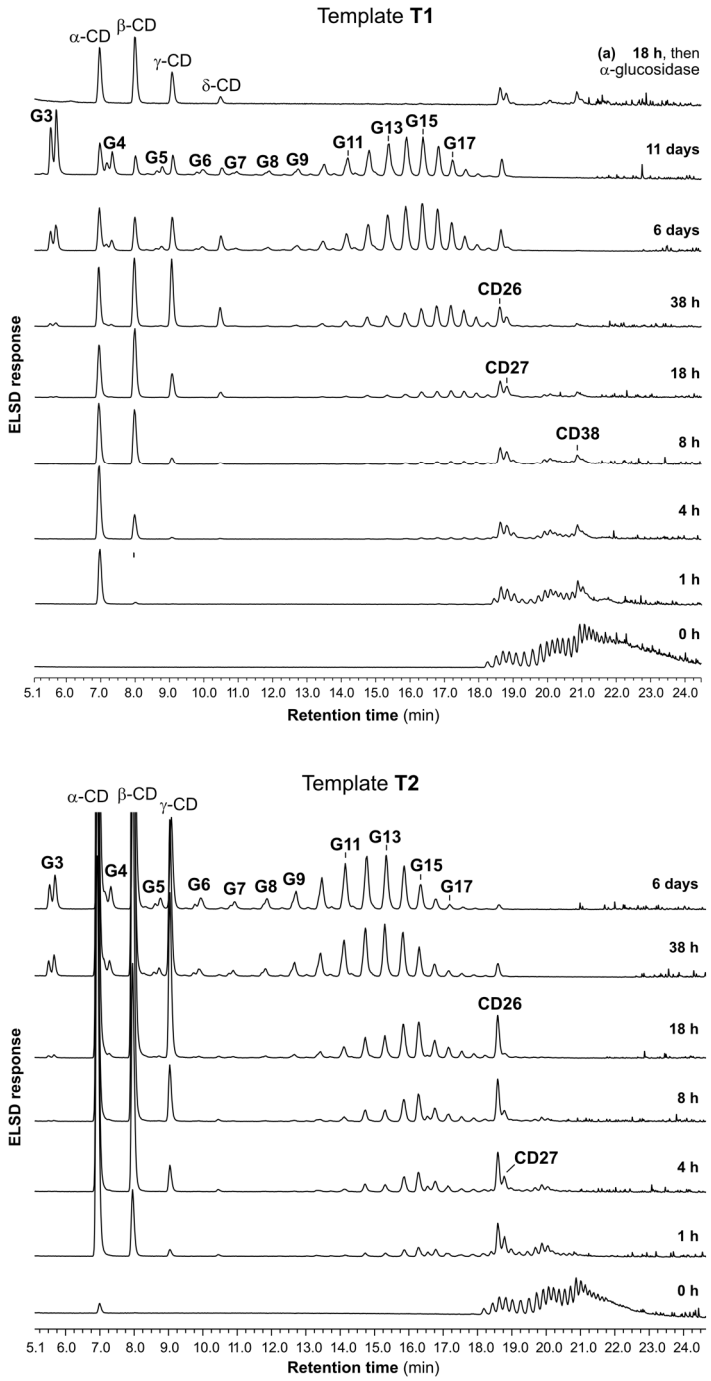


Figure 5.6. HPLC-ELSD chromatograms showing product distributions from the enzymatic reactions with Cycloamylose (10 mg/ml) in the presence of templates **T1** or **T2** (10 mM) treated with glycerol-free CGTase (10 µl per ml reaction mixture) in sodium phosphate buffer (50 mM, pH 7.5). (a) Reaction mixture after 18 hours of reaction with CGTase followed by treatment with α-glucosidase to hydrolyze linear α-1,4-glucans. *Note:* Very sharp peaks ('spikes') in the chromatograms at retention times of 22–25 minutes are due to issues with the ELS detector.

CD27 resembles the conformations exhibited by CD26.[12,13] The fact that CD27 is trapped by **T1** and **T2** and not by **5.1** could be due to the proposed distortions present for CD27 (compared to CD26), which are presumably more easily accommodated by the less bulky and more flexible templates **T1** and **T2**. This could also explain why CD27 remains for longer in the reaction with **T1** compared to the reaction with **T2**, as the longer alkyl chain of **T1** could be

better suited for accommodating these distortions. It is also possible that the templates only bind to CD27 with a 1:1 stoichiometry, where CD27 forms a longer helical V-amylose segment instead of two parallel segments. ITC analysis has shown that CD27 binds sodium dodecyl sulfate (SDS) in a 1:1 stoichiometry,[2] in contrast to the cooperative 1:2 binding reported for SDS to CD26. The binding of the templates **T1** and **T2** to one (longer) V-amylose segment in CD27 would also explain why the template with the longer C_{16} alkyl chain of **T1** leads to the more prominent kinetic trapping of CD27.

From these studies, it is quite evident that both the length of the alkyl chain and the structure and identity of head groups and extra linkers play important roles in the binding to LRCDs with V-amylose-like channels. However, it also seems that as long as an alkyl chain with a sufficient length and low steric demands of the hydrophilic head group is present, the kinetic trapping of these LRCDs is plausible. So to further expand the scope of templates that could be used, simple, commercially available amphiphiles with alkyl chains of different length were explored. For this, sodium dodecyl sulfate (SDS) and sodium hexadecyl sulfate (SHS) were selected. However, the mixing of Cycloamylose and SHS in sodium phosphate buffer led to the formation of large amounts of precipitates, so only SDS was used.

A library was set up with Cycloamylose (10 mg/ml) in sodium phosphate buffer (50 mM, pH 7.5) in the presence of sodium dodecyl sulfate (10 mM) and with glycerol-free CGTase (10 µl per ml reaction mixture). A small amount of precipitate was also seen upon the mixing of Cycloamylose and SDS in sodium phosphate buffer, which was removed by centrifugation before the addition of enzyme. The HPLC-ELSD chromatograms obtained from the reaction are shown in Figure 5.7. From the chromatogram of the zero hour time point, it is evident that the observed and removed precipitate corresponds to the LRCDs in the Cycloamylose starting material with about 42–60+ glucose units, as the peaks corresponding to these LRCDs are diminished here in comparison with the zero hour time point of previously discussed libraries (Figure 5.3 and 5.6). These observations are in agreement with the precipitation studies carried out by Suzuki and co-workers, where amphiphiles/surfactants with C_{12} and C_{14} alkyl chains led to the precipitation of LRCDs of similar size.[14] In the enzymatic library (Figure 5.7), CD26 was again observed to build up and remain in the library while all other LRCDs were removed, an effect seen clearly after just one hour of reaction. CD26 then remains in the library for several more hours, but is removed mostly within 7 hours and completely within 19 hours. CD27 and CD38 were not kinetically trapped to any significant degree in this reaction. Instead, some of the LRCDs with about 42 to 60+ glucose units, which were also the LRCDs that precipitated upon addition of SDS, were observed to remain for much longer here, compared to an untemplated library (Figure 5.3). Evidently, some interaction between SDS and these LRCDs is taking place, presumably due to the formation of V-amylose segments in the LRCDs as well. Unsurprisingly, the main product of the reaction is α-CD, as it binds strongly to SDS and is also the main product in CGTase-mediated DCLs in the presence of SDS with other α-1,4-glucan starting materials.[4,15]

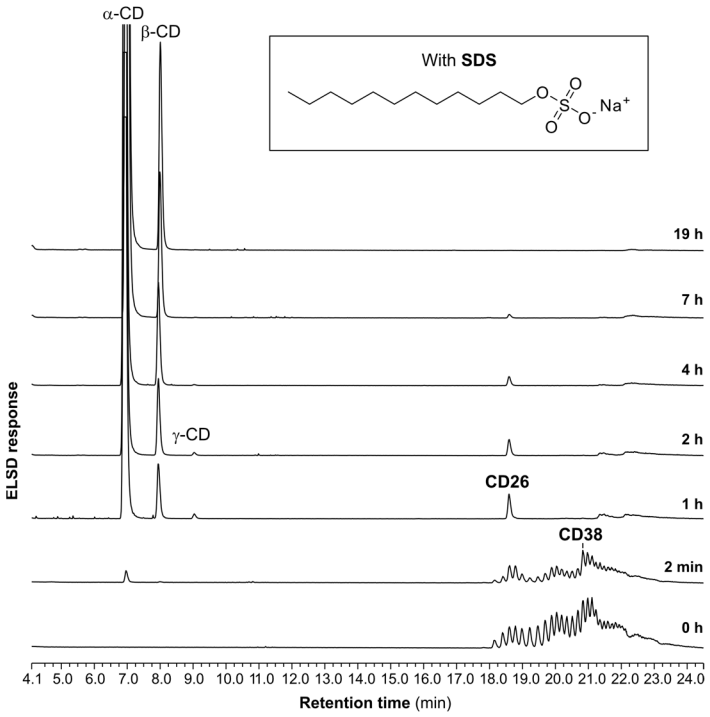


Figure 5.7. HPLC-ELSD chromatograms showing the product distribution from the enzymatic reaction with Cycloamylose (10 mg/ml) and template SDS (10 mM) treated with glycerol-free CGTase (10 µl per ml reaction mixture) in sodium phosphate buffer (50 mM, pH 7.5).

5.2.5. Polymeric α-glucan sources for CD26 synthesis

The possibility of using other α-glucan polymers for the synthesis of CD26 was also desirable. It has been shown that the major products of the initial action of CGTase on the substrate synthetic amylose is in fact not the small CDs, but instead a wide range of LRCDs from δ-CD and up to at least CD60, which were then converted to the small CDs α-, β-, and γ-CD with longer reaction times.[16] With these LRCDs produced in the initial reaction, it was hypothesized that the addition of any template that could bind to and kinetically trap CD26 would lead to the templated synthesis of CD26. As such, starch was chosen as a cheap, polymeric, primarily α-1,4-glucan starting material that contains both amylose and amylopectin. Reactions with 'soluble starch' (20 mg/ml) and CGTase in buffer and with templates **T1, T2** or SDS (10 mM) were set up and monitored by HPLC-ELSD. Unfortunately, no CD26 or other LRCDs with more than 10 glucose units were observed in the reactions, both after short reaction times (1–5 minutes) and after several hours. A possible explanation for this could be that amylopectin, which does not consist of very long strands of linear α-1,4-glucan segments and is the more soluble component of starch, is the primary component of commercial 'soluble starch'. To achieve the templated synthesis of CD26 using a polymeric α-glucan source other than 'Cycloamylose', (synthetic) amylose might thus be the best option.

5.3. High-yielding, chromatography-free templated enzymatic synthesis of δ-CD

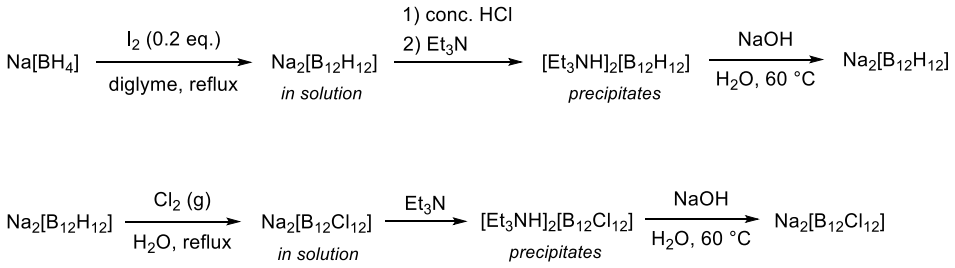
In 2016, Werner Nau and co-workers discovered the high-affinity binding of superchaotropic dodecaborate clusters ($B_{12}Cl_{12}^{2-}$, $B_{12}Br_{12}^{2-}$ and $B_{12}I_{12}^{2-}$) to γ-CD and to the LRCDs with 9 to 11 glucose units.[17] The results from their studies are summarized in Table 5.1.

Table 5.1. Binding constants K_a of dodecaborate cluster dianions determined using isothermal titration calorimetry (ITC) by Nau and co-workers.[17]

	K_a / 10^3 M^{-1}		
Host	$B_{12}Cl_{12}^{2-}$	$B_{12}Br_{12}^{2-}$	$B_{12}I_{12}^{2-}$
CD8 (γ-CD)	14	960	67
CD9 (δ-CD)	2500	2600	680
CD10	29	140	2100
CD11	2	6	8

In a previous study in the Beeren group, $B_{12}I_{12}^{2-}$ was synthesized and used as a template to target the enzymatic synthesis of CD10 due to the very high affinity of this dodecaborate cluster towards CD10. Despite this high affinity, only small amounts of CD9 (δ-CD) and CD10 were produced.[4] Instead, γ-CD formed as the primary product due to the intrinsically higher stability of γ-CD. To achieve the selective templated enzymatic synthesis of any of these LRCDs, an even more selective binding over γ-CD was required. This was found for $B_{12}Cl_{12}^{2-}$, which binds to δ-CD with a more than two orders of magnitude higher affinity than to γ-CD.[17]

To explore the templated enzymatic synthesis of δ-CD using $B_{12}Cl_{12}^{2-}$, the water soluble sodium salt $Na_2B_{12}Cl_{12}$ was synthesized. The synthesis was performed in accordance to a published procedure,[18] as shown in Scheme 5.1. First, the hydride cluster $Na_2B_{12}H_{12}$ was synthesized from sodium borohydride by oxidation with iodine. Then $Na_2B_{12}Cl_{12}$ was synthesized by the oxidation of $Na_2B_{12}H_{12}$ using chlorine gas. The chlorine gas was produced using a chlorine generator with hydrochloric acid and calcium hypochlorite. Both the synthesis of $Na_2B_{12}H_{12}$ and $Na_2B_{12}Cl_{12}$ takes advantage of the fact that the triethylammonium salts of these dodecaborate clusters are insoluble in water, while the sodium salts are water soluble.



Scheme 5.1. Synthesis of $Na_2B_{12}Cl_{12}$ according to a published procedure.[18]

5.3.1. Analytical scale templated enzymatic reactions with $Na_2B_{12}Cl_{12}$

To explore the use of the template $Na_2B_{12}Cl_{12}$ in CGTase-mediated enzymatic libraries, reactions with α-CD as starting material in the presence of $Na_2B_{12}Cl_{12}$ were set up. HPLC-ELSD chromatograms obtained from the reaction with α-CD (10 mg/ml), $Na_2B_{12}Cl_{12}$ (10 mM) and CGTase (50 µl per ml reaction mixture) in sodium phosphate buffer (50 mM, pH 7.5) are depicted in Figure 5.8. Remarkably, δ-CD was produced in the library with very high selectivity over α-, β-, and γ-CD after 72 hours of reaction. After one hour of reaction, β-CD was the main component of the library, which was then converted into mainly γ-CD after 4–8 hours. After 24 hours, δ-CD had become the main component of the library with the selectivity increasing after 48 and 72 hours. The slow conversion of γ-CD to δ-CD is presumably due to the kinetic trapping of γ-CD in the library, as the template $B_{12}Cl_{12}^{2-}$ also binds strongly to γ-CD, thereby making γ-CD largely unavailable for enzymatic reactions.

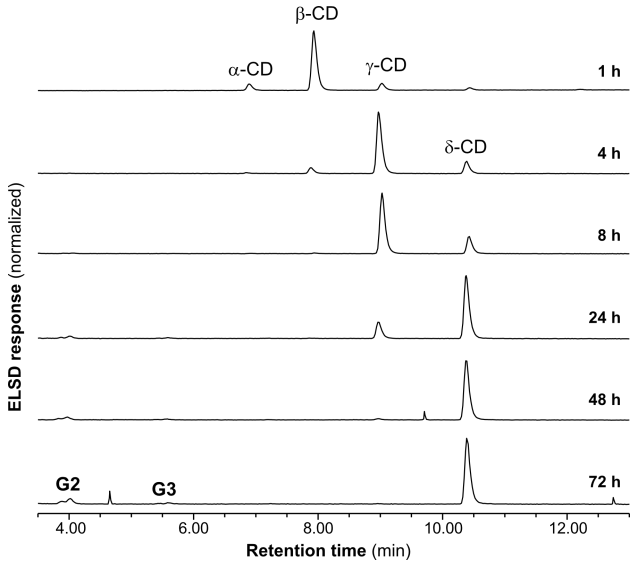


Figure 5.8. HPLC-ELSD chromatograms from the enzymatic reaction with α-CD (10 mg/ml), $Na_2B_{12}Cl_{12}$ (10 mM) and CGTase (50 µl per ml reaction mixture) in sodium phosphate buffer (50 mM, pH 7.5).

Optimizations with regards to template concentration and temperature were carried out, but will not be described in detail in this thesis. To briefly summarize, a template concentration of 5 mM at room temperature was found to give an optimal balance of yield and selectivity for δ-CD over the small CDs. Long reaction times (more than 48 hours) were found to be favourable for the selectivity for δ-CD, while at the same time diminishing the yield of δ-CD slightly. This is shown in Figure 5.9, where the yields of all glucans (determined from the HPLC-ELSD chromatograms) in a library with α-CD (10 mg/ml) and $Na_2B_{12}Cl_{12}$ (5 mM) are plotted as a function of time. With long reaction times, all CDs including δ-CD are slowly hydrolysed to the small linear α-1,4-glucans and ultimately glucose. However, δ-CD, with its strong binding to the template $B_{12}Cl_{12}^{2-}$, is kinetically trapped to a larger extent than the other CDs and less accessible to the enzyme for hydrolysis. As such, the selectivity for δ-CD improves with long reaction times at the cost of the yield of δ-CD.

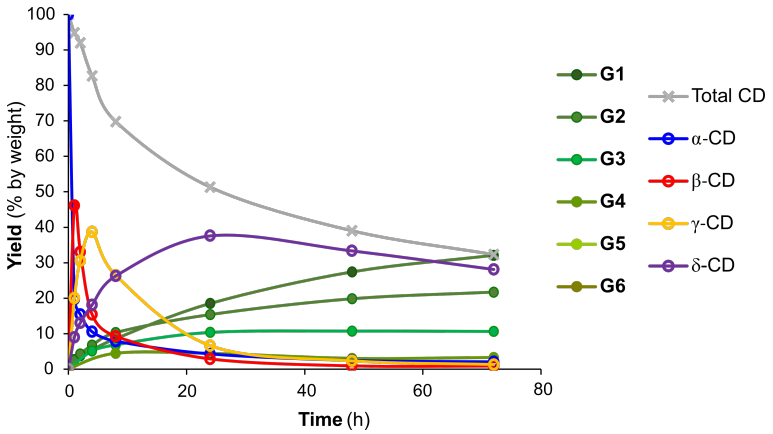


Figure 5.9. Glucan yields (% by weight, based on HPLC-ELSD chromatograms) as a function of time in the reaction with α-CD (10 mg/ml) and $Na_2B_{12}Cl_{12}$ (5 mM) with CGTase (50 µl per ml reaction mixture) in sodium phosphate buffer (50 mM, pH 7.5).

Other α-1,4-glucan substrates were also of interest for the templated enzymatic synthesis of δ-CD using $Na_2B_{12}Cl_{12}$. Starch, the glucan substrate used in the industry for the synthesis of small CDs, was thus chosen as a starting material. Initial studies with 'soluble starch' had shown that roughly half of the starch used in CGTase-mediated DCLs is converted into CDs within the first hours of the reaction. The rest of the starch is presumably removed upon centrifugation before injection on the HPLC equipment. As such, a reaction with soluble starch (20 mg/ml, twice the amount used in the reactions with α-CD) and $Na_2B_{12}Cl_{12}$ (5 mM) and glycerol-free CGTase (65 µl per ml reaction mixture) in sodium phosphate buffer (50 mM, pH 7.5) was set up. Figure 5.10 depicts HPLC-ELSD chromatograms obtained from this reaction. The results closely resemble the results with α-CD as starting material. α-CD and β-CD are the main components of the library in the initial hour of the reaction. Then γ-CD becomes the main component around 4 hours, after which the CDs are then slowly converted into δ-CD with high selectivity.

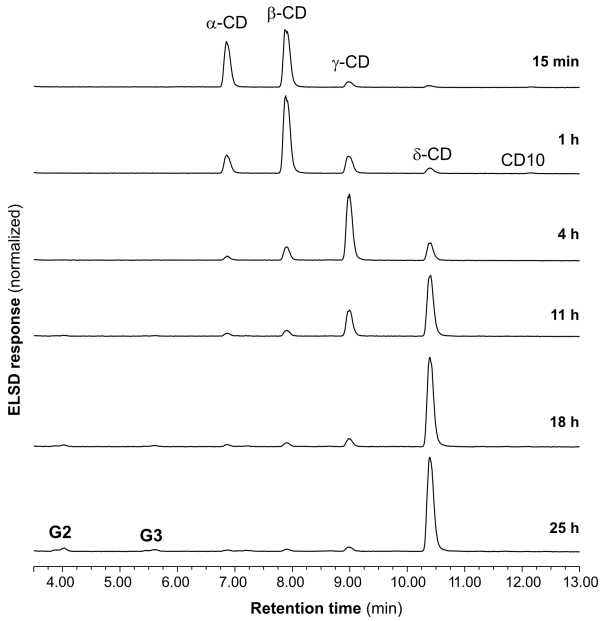


Figure 5.10. HPLC-ELSD chromatograms with time points as indicated on figure from the enzymatic reaction with soluble starch (20 mg/ml), $Na_2B_{12}Cl_{12}$ (5 mM) and glycerol-free CGTase (65 µl per ml reaction mixture) in sodium phosphate buffer (50 mM, pH 7.5).

5.3.2. Preparative scale synthesis and isolation of δ-CD

The ultimate goal of this project was to efficiently synthesize and isolate δ-CD on a preparatory scale. Chapter 3 describes how δ-CD can be isolated from enzymatic libraries using preparatory HPLC, but as this method requires an expensive HILIC column with a maximum loading capacity of ~100 mg, a chromatography-free method for the isolation of δ-CD on a large scale was desired. Scaling up from analytical scale (0.1 ml reaction mixture, 1 mg α-CD starting material) to preparatory scale (100 ml reaction mixture, 1 g α-CD starting material) turned out to be unproblematic, and reaction mixtures with the same composition as the analytical scale reactions with α-CD (Figure 5.8) were achieved. Work by Zimmerman and co-workers[19] and the work described in Chapter 3 had shown that the addition of five volumes of acetone to CGTase-mediated reactions with roughly 50 mg/ml α-glucan selectively precipitates CDs while leaving linear α-1,4-glucans in solution. The addition of acetone to the reaction mixture of the preparatory scale reaction (which had been concentrated five times) led to selective precipitation of the CDs, while leaving the template and most of the linear α-1,4-glucans in solution. By repeatedly dissolving, precipitating and filtering the CDs (~five times), all traces of template and linear α-1,4-glucans were removed, yielding δ-CD with >99% purity (Figure 5.11). Further optimizations of the large-scale reaction showed that a reaction temperature of 30 °C led to reaction completion within 2 days and a yield of 50% on a 10 g scale.

The re-isolation and reuse of the template in subsequent syntheses of δ-CD was also highly desirable. Using the same principles as the precipitation protocols from the synthesis of the template, it was discovered that the template could be isolated from the filtrate left behind from the precipitation of δ-CD. By evaporating the acetone, followed by acidification with

hydrochloric acid and addition of an excess of triethylamine, the selective precipitation of the triethylammonium salt of $B_{12}Cl_{12}^{2-}$ was achieved. The triethylammonium salt could then be converted into the sodium salt using the method from the synthesis of $Na_2B_{12}Cl_{12}$ and reused in subsequent syntheses of δ-CD.

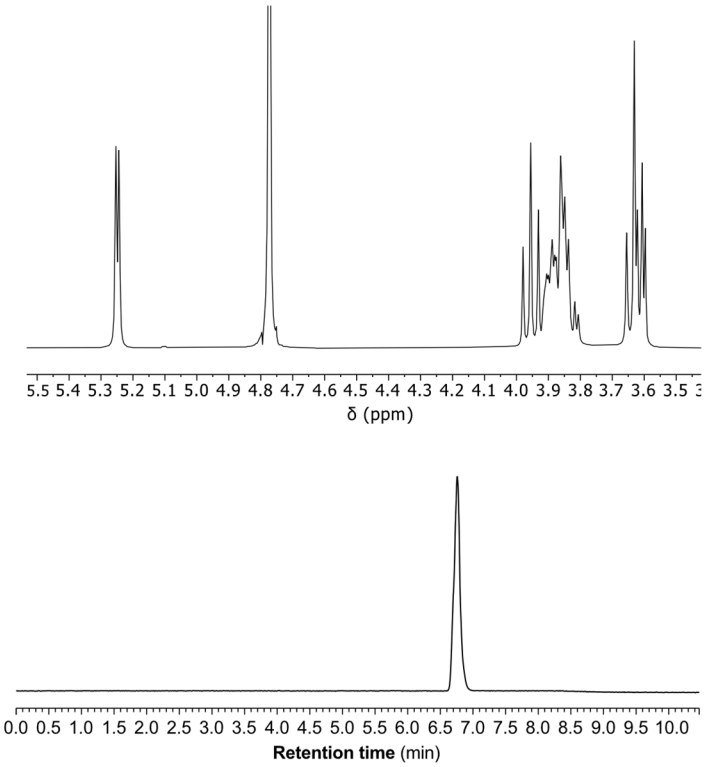


Figure 5.11. Isolated δ-CD. *Top*: Partial ^{1}H NMR spectrum of isolated δ-CD. *Bottom:* HPLC-ELSD chromatogram of isolated δ-CD.

5.4. Conclusion

In this chapter, two different approaches to the templated enzymatic synthesis of large-ring cyclodextrins was described.

In subchapter 5.1, the endeavours towards the templated enzymatic synthesis of CD26 were described. It was found that several different (bola)amphiphiles with alkyl chains were able to template the synthesis of CD26 by kinetically trapping CD26 in an enzymatic reaction with CGTase starting from 'Cycloamylose'. CD27 and CD38 and even larger LRCDs were also found to be kinetically trapped and remain in some of the libraries.

For further studies towards the enzymatic synthesis of CD26 and other LRCDs with more than 26 glucose units, the use of synthetic amylose would interesting to explore. For initial studies, commercially available synthetic amylose could be used. For later studies, synthetic amylose could be produced, either by utilizing glucan phosphorylase or by using a coupled enzyme system with sucrose phosphorylase and glucan phosphorylase.[14] Another enzyme to explore for the synthesis of LRCDs is the 'disproportionating enzyme' (D-enzyme, EC 2.4.1.25), which catalyses inter- and intramolecular transglycosylation reactions, just like CGTase. D-enzyme however, produces only CDs with 17 to several hundred glucose units, and not the small CDs. The thermodynamically stable products of α-, β- and γ-CD, which are usually the main components in the CGTase-mediated dynamic combinatorial libraries, could thus be avoided.

Other (bola)amphiphiles for use as templates in the synthesis of CD26 and other LRCDs could also be explored. Simple, commercially available amphiphiles with alkyl chains longer than 12 carbons that do not lead to excessive precipitation of the α-glucan substrates would be desirable. Amphiphiles that contain two alkyl chains, such as phospholipids, which could then bind to both V-amylose channels of CD26 (or CD38) would also be interesting to explore.

In subchapter 5.2, the high-yielding, chromatography-free templated enzymatic synthesis of δ-CD was described. The yield of δ-CD achieved (50%) is more than two orders of magnitude better than the highest reported yield (0.26%),[20] and also quite an improvement over the results with the synthesis of δ-CD described in Chapter 3 (7.2%).

The work described here could be of significant scientific value, as it makes δ-CD available to the scientific community in quantities where routine use of δ-CD could, potentially, become the norm in laboratories that routinely work with α-, β-, and γ-CD. As such, the methods developed could also be of commercial value.

5.5. Experimental

5.5.1. Materials

For materials, see section 3.7.1 and section 4.6.1.

5.5.2. Instrumentation and Methods

For Instrumentation and Methods, see section 3.7.2.

5.5.3. Enzymatic reactions

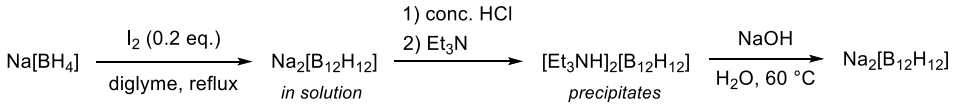
Enzymatic reactions (200–240 µl total volume) starting from 'Cycloamylose' were set up by the addition of a CGTase stock solution (commercial or glycerol-free, 10 µl per ml reaction mixture) to a reaction mixture consisting of Cycloamylose (10 mg/ml) and in the absence or presence of template (10 mM) in sodium phosphate buffer (50–200 mM, pH 7.5). Reactions starting from Cycloamylose were analysed by injecting the reaction mixture directly on the HPLC instrument. For reactions with starch, soluble starch was dissolved in water by heating in the microwave (5-8 heating cycles with vigorous mixing between cycles). Enzymatic reactions (100–200 µl total volume) with α-CD or starch as starting materials were set up by the addition of a CGTase stock solution (commercial or glycerol-free, 50-65 µl per ml reaction mixture) and in the absence or presence of templates as indicated for each experiment. Reactions with α-CD as starting material were monitored by taking out aliquots (3–6 µl) and diluting these (21–41-fold) with a denaturing mixture (1% TFA in 3:1 acetonitrile/water with 10 mM NH_4Cl), which stops the enzymatic reaction and preserves the reaction mixture composition for later analysis. Reactions with starch as starting material were monitored by taking out aliquots (3–10 µl) and diluting these (6–9-fold) with a denaturing mixture (1–2% TFA in water). All diluted samples were then centrifuged at 10,000 rpm for a minimum of 4 minutes and the top fractions (leaving behind 9–10 µl) were then transferred to 2 ml glass vials with 0.2 ml glass inserts and kept at 20° C until injection on the HPLC instrument. Injection volumes at the HPLC instrument were 8–15 µl when the acetonitrile/water denaturing mixture was used and 1–2 µl when the aqueous denaturing mixture or direct injection was used. Peaks in the chromatograms corresponding to α-, β-, γ-, and δ-CD and linear α-1,4-glucans up to maltooctaose were identified by comparison with authentic samples while peaks in the chromatograms corresponding to linear α-1,4-glucans with more than 8 glucose units and LRCDs were identified by comparison with characterized reference mixtures. When direct injection or the aqueous denaturing mixture was used, the glucose (**G1**) peak in the chromatograms co-eluted with the peak from the buffer, and it could therefore not be integrated reliably.

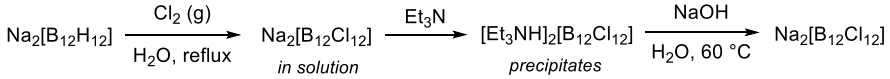
Hydrolysis of linear α-1,4-glucans to glucose was carried out using the enzyme α-glucosidase. Reaction mixtures with the enzyme CGTase (not denatured with TFA-based solutions) were heated to 95 °C for 15 min to denature the enzyme CGTase. The reaction mixtures were then cooled to 40 °C and α-glucosidase (1.0–1.5 U per ml reaction mixture) was added and the reactions were kept at 40 °C for up to 48 hours, until the reaction was complete (monitored by HPLC-ELSD). The reaction mixture was then heated to 95 °C for 30 min to denature the

enzyme α-glucosidase. The samples were then diluted (3–6 fold) in water, centrifuged at 10,000 rpm for a minimum of 4 minutes and the top fractions (leaving behind 10 µl) were transferred to 2 ml glass vials with 0.2 ml glass inserts and kept at 20° C until injection on the HPLC instrument.

5.5.4. Synthesis of $Na_2B_{12}Cl_{12}$

The synthesis of $Na_2B_{12}Cl_{12}$ was performed according to a published procedure.[18]





$[NHEt_3][B_{12}H_{12}]$. A suspension of $NaBH_4$ (47 g, 1.24 mol) in diglyme (200 ml) was heated to 100 °C under a nitrogen atmosphere. To this, a solution of iodine (102.8 g, 0.41 mol) in diglyme (175 ml) was then added dropwise. The dropwise addition was carried out over 24 hours. During the addition, the amount of insoluble $NaBH_4$ gradually decreased while the color of the reaction mixture became yellow. The reaction mixture was then stirred at 100 °C overnight, after which it was refluxed for another 24 hours. At this point, the reaction mixture had become a colorless suspension with white precipitate. The solvent was then removed by vacuum distillation and the resulting white solid was then dissolved by the slow addition of water (300 ml). Concentrated hydrochloric acid (140 ml) was then added slowly, leading to the evolution of hydrogen gas. The solution was then kept at 5 °C overnight, leading to the precipitation of boric acid as a white solid, which was removed by gravity filtration. The filtrate was then treated with triethylamine (200 ml), leading to the precipitation of a white solid, which was collected by vacuum filtration, re-suspended in water (125 ml) and stirred for two hours to dissolve any remaining boric acid. The resulting suspension was then filtered by vacuum filtration and dried *in vacuo* to yield $[NHEt_3][B_{12}H_{12}]$ as a white solid (13.5 g, 39 mmol, 48%, based on iodine)

$Na_2[B_{12}H_{12}]$. $[NHEt_3][B_{12}H_{12}]$ (11.5 g, 33 mmol) was added to a solution of NaOH (2.78 g, 69 mmol) in water (120 ml) in a polypropylene beaker. The suspension formed was heated on a water bath (60 °C) until a clear solution was formed. The solvents were then removed *in vacuo* to yield $Na_2[B_{12}H_{12}]$ as a white solid (6.2 g, 33 mmol, quantitative). **^{1}H NMR** (400 MHz, D_2O) δ 0.50–1.90 (broad multiplet). **^{11}B NMR** (128 MHz, D_2O) δ -15.4 (d, $^1J_{BH}$ = 123 Hz).

$[Et_3NH]_2[B_{12}Cl_{12}]$. $Na_2[B_{12}H_{12}]$ (3 g, 16 mmol) was dissolved in water (30 ml) in a three-necked round-bottom flask fitted with a reflux condenser. This flask was then then connected to a 'chlorine generator' flask and a 'chlorine trap' system. The original paper[18] contains tips on how to set up the apparatus for this synthesis. The chlorine trap contained an aqueous solution of NaOH (1 M). The chlorine generator consisted of a three-necked round-bottom flask containing solid $Ca(ClO)_2$ (100 g, 0.70 mol) equipped with an addition funnel with

concentrated hydrochloric acid (100 ml). Chlorine was then generated and bubbled through the solution of $Na_2[B_{12}H_{12}]$ at room temperature for 5 hours by the slow, dropwise addition of hydrochloric acid to the $Ca(ClO)_2$ solids (2 × 100 ml hydrochloric acid used). The reaction was then refluxed overnight (no chlorine generation). Another chlorine generator (same quantities of $Ca(ClO)_2$ and hydrochloric acid) was then set up, and chlorine was bubbled through the solution again for 6 hours at room temperature. The reaction was then refluxed overnight (no chlorine generation). The process of chlorine generation and bubbling through the reaction followed by reflux overnight was repeated once more. Completion of chlorination was checked by ^{11}B NMR spectroscopy over the course of the reaction. After complete reaction, the resulting acidic, clear solution was treated with Et_3N (100 ml) leading to the immediate precipitation of a white solid. The solid was then isolated by vacuum filtration, washed repeatedly with water and dried *in vacuo* to yield $[Et_3NH]_2[B_{12}Cl_{12}]$ as a white solid (7.3 g, 9.65 mmol, 60%). **^{1}H NMR** (400 MHz, CD_3CN) δ 3.14 (q, J = 7.3 Hz, 12H) 1.24 (t, J = 7.3 Hz, 18H). **^{11}B NMR** (128 MHz, CD_3CN) δ -15.8 (s).

$Na_2[B_{12}Cl_{12}]$. $[Et_3NH]_2[B_{12}Cl_{12}]$ (7.2 g, 9.5 mmol) was added to a solution of NaOH (0.80 g, 19.9 mmol) in water (80 ml) in a polypropylene beaker. The suspension formed was heated on a water bath (60 °C) until a clear solution was formed. The solvents were then removed *in vacuo* to yield $Na_2[B_{12}Cl_{12}]$ as a white solid (5.7 g, 9.5 mmol, quantitative). **^{11}B NMR** (128 MHz, CD_3CN) δ -13.0 (s).

5.5.5. Large scale synthesis of δ-CD

The synthesis of δ-CD at a 10 g scale, as described here, was carried out by PhD student Kasper Hornstrup Hansen

α-CD (10.00 g, 10.03 mmol) and $Na_2B_{12}Cl_{12}$ (3.00 g, 5 mmol) were dissolved in MilliQ water (900 ml) adjusted to pH 7.5 with 1 M HCl. The mixture was transferred to a 1-liter volumetric flask. A solution of CGTase (25 ml) was then added to start the reaction and MilliQ water was added until the 1-liter mark. The reaction mixture was transferred to a 2-liter round bottomed flask and placed in a water bath at 30 °C for 42 hours, after which the reaction was stopped by heating to boiling for 15 minutes. The reaction mixture was concentrated to 0.2 l *in vacuo*, then centrifuged and the supernatant decanted, leaving behind precipitated enzyme. δ-CD was then precipitated by the addition of acetone (1.0 l), and isolated by filtration. The white solids collected were then dissolved in water (100 ml) and precipitated with acetone (500 ml), followed by filtration to isolate the solids. This process was repeated 2 times, after which the white solids were dried *in vacuo*.

The NMR data and MS data of δ-CD were consistent with the results obtained and described in Chapter 3.

5.5.6. Computational details

To investigate the binding mode of the 1:2 complex between CD26 and template **5.1**, a molecular dynamics (MD) simulation was carried out. For the starting geometry of CD26 a published crystal structure of CD26 in complex with undecanoic acid (undecanoic acids were removed) was used (CCDC deposition number: 197411)[1] while the starting geometry of template **5.1** as the hexavalent sulfonate ion was constructed in the Chem3D molecular software. The starting geometry of the complex was constructed in the software PyMOL by manually placing one **5.1** template in each of the helical V-amylose-like channels of CD26, with the alkyl chains threading the channels and the hydrophilic head groups protruding from each side.

The MD simulation in explicit solvent was carried out using the Amber20 suite[7] employing the GLYCAM-06[8] force field for δ-CD and the general amber force field (GAFF)[9] for template **5.1**. Partial charges for **5.1** were calculated using the antechamber program[21] implemented in Amber20 using the AM1-BCC method.[22] The LEaP program of Amber20 was used to solvate the complex using a truncated octahedron with a distance buffer of 20 Å between the complex and the edges of the periodic box. Water molecules were represented using the TIP3P[10] water model, and the system was neutralized using 6 sodium ions. The solvated and neutralized system was then minimized for 500 steps. The first 400 steps were performed using the steepest descent method followed by 100 steps of the conjugate gradient method. The MD simulation was carried out using a 8 Å cut-off for nonbonded interactions and the Particle Mesh Ewald method[23] for long range electrostatics. A time step of 2 fs was applied. Bonds to hydrogen atoms were constrained using the SHAKE algorithm.[24] The systems were then heated linearly from 10 K to 300 K within 0.3 ns followed by 0.7 ns of equilibration at 300 K in the *NVT* ensemble using the Langevin thermostat[25] with a collision frequency of 5 ps^{-1}. The systems were then equilibrated further for 4 ns using constant temperature (T = 300 K, using the Langevin thermostat with a collision frequency of 1 ps^{-1}) and constant pressure (p = 1 bar, using the Berendsen barostat[26] with isotropic position scaling). Finally, a 300 ns production run was performed using the same conditions as the equilibration protocol with constant pressure. Coordinates were saved every 2 ps for further analysis.

Analyses were performed using the CPPTRAJ module[27] in the Amber20 suite. Clustering analysis was performed to group similar structures from the trajectory into clusters. Clustering analysis was performed on the 300 ns production run based on average root-mean-square deviation (RMSD) of all non-hydrogen atom positions of CD26 and **5.1** using the "k-means" algorithm[11] to output 10 clusters. Cluster representatives were chosen based on the lowest distance to cluster centroid.

5.6. Bibliography

[1] O. Nimz, K. Gessler, I. Usón, G. M. Sheldrick, W. Saenger, *Carbohydr. Res.* **2004**, *339*, 1427–1437.

[2] S. Kitamura, *Japanese Soc. Appl. Glycosci.* **2003**, *50*, 321–325.

[3] K. I. Assaf, D. Gabel, W. Zimmermann, W. M. Nau, *Org. Biomol. Chem.* **2016**, *14*, 7702–7706.

[4] D. Larsen, S. R. Beeren, *Chem. Sci.* **2019**, *10*, 9981–9987.

[5] K. Gessler, I. Usón, T. Takaha, N. Krauss, S. M. Smith, S. Okada, G. M. Sheldrick, W. Saenger, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 4246–4251.

[6] P. Ivanov, E. Atanassov, C. Jaime, *J. Mol. Struct.* **2014**, *1056–1057*, 238–245.

[7] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, et al., *Amber 2020*, University of California, San Francisco, **2020**.

[8] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.

[9] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[10] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[11] J. A. Hartigan, M. A. Wong, *J. R. Stat. Soc.* **1979**, *28*, 100–108.

[12] I. Maestre, I. Beà, P. M. Ivanov, C. Jaime, *Theor. Chem. Acc.* **2007**, *117*, 85–97.

[13] P. Ivanov, *J. Mol. Struct.* **2012**, *1009*, 3–10.

[14] S. Kitamura, S. Suzuki, *Linear and Cyclic Amyloses: Beyond Natural*, Elsevier Inc., **2020**.

[15] D. Larsen, S. R. Beeren, *Chem. Eur. J.* **2020**, *26*, 11032–11038.

[16] Y. Terada, M. Yanase, H. Takata, T. Takaha, S. Okada, *J. Biol. Chem.* **1997**, *272*, 15729–15733.

[17] K. I. Assaf, D. Gabel, W. Zimmermann, W. M. Nau, *Org. Biomol. Chem.* **2016**, *14*, 7702–7706.

[18] V. Geis, K. Guttsche, C. Knapp, H. Scherer, R. Uzun, *Dalt. Trans.* **2009**, 2687–2694.

[19] C. Sonnendecker, S. Melzer, W. Zimmermann, *Microbiologyopen* **2019**, *8*, 1–8.

[20] T. Endo, H. Ueda, S. Kobayashi, T. Nagai, *Carbohydr. Res.* **1995**, *269*, 369–373.

[21] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

[22] A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623–1641.

[23] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[24] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.

[25] R. W. Pastor, B. R. Brooks, A. Szabo, *Mol. Phys.* **2006**, *65*, 1409–1419.

[26] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[27] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

Conclusions and outlook

In this work, enzyme-mediated dynamic combinatorial chemistry was used to explore cyclodextrins (CDs), including the elusive and inaccessible large-ring cyclodextrins (LRCDs). The templated dynamic enzymatic synthesis of δ-CD, the large-ring cyclodextrin with 9 glucose units, was achieved in two different ways. In one approach, a bolaamphiphile template design was used to achieve selective binding to δ-CD over the native CDs, and δ-CD was isolated in unprecedented yields using preparative HPLC. The host-guest chemistry of δ-CD was explored using these bolaamphiphiles, and thorough NMR spectroscopy titrations revealed that, depending on the structure of the bolaamphiphile, δ-CD can form [2]-, [3]-, and [4]-*pseudo*rotaxanes. In a different approach, the superchaotropic dodecaborate cluster anion $B_{12}Cl_{12}^{2-}$ was used to template the enzymatic synthesis of δ-CD, leading to the scalable, chromatography-free production and isolation of δ-CD in 50% yield. Molecular dynamics (MD) simulations and NMR spectroscopy studies of the host-guest chemistry of LRCDs with 12–16 glucose units revealed that some of these LRCDs form 1:2 complexes with adamantane carboxylate in a figure-of-eight-like geometry. Additionally, it was shown how very large LRCDs (with 26 or more glucose units) were kinetically trapped in enzyme-mediated DCLs with (bola)amphiphile templates starting from 'cycloamylose', potentially paving the way for the preparative scale synthesis of these very large LRCDs.

The initial 'golden years' of research into LRCDs came following the successful isolation of a range of LRCDs in the 1990s with the seminal work of Takaha and co-workers as well as Ueda and co-workers, among others. Despite the significant discoveries made in these years, as well as strides made in recent years, such as the remarkable use of engineered CGTases to favor LRCD production by Zimmermann and co-workers, LRCDs have remained mostly an academic curiosity. This thesis has led to the development of synthetic procedures for δ-CD, which could make it available to the scientific community in substantial quantities. Potentially, this could lead to the use of δ-CD becoming routine, with bottles of δ-CD finding their place next to α-, β-, and γ-CD on shelves in laboratories around the world.

It is the hope of this author that the work presented here, including the synthesis of δ-CD and the exploration of the host-guest chemistry of LRCDs, might inspire further investigations into the properties and applications of these exciting, non-toxic, water-soluble macrocycles. The knowledge of the host-guest chemistry of LRCDs obtained could potentially lead to the design of new and better templates for the templated enzymatic synthesis of other LRCDs. For instance, the figure-of-eight-like geometries of the LRCDs with more than 15 glucose units that were explored using MD simulations could lead to the improved design of guests with two recognition motifs linked together, such as the double-adamantane design that was briefly explored in this thesis.

With the knowledge that δ-CD binds several alkyl chains and forms [2]-, [3]-, and [4]-*pseudo*rotaxanes with suitable bolaamphiphiles, it is now also easier to envisage which

guests would bind effectively to δ-CD. This could both lead to the use of δ-CD in various supramolecular systems, such as the synthesis of mechanically interlocked molecules with δ-CD or the use of δ-CD with suitable polymers to form materials (like hydrogels). The availability and knowledge of δ-CD might also help guide the use of δ-CD in industrial applications, such as drug delivery, where it for instance could help in the solubilization of large hydrophobic drugs. Modified δ-CDs are of course also of interest, and could be interesting both in terms of classical applications (drug delivery etc.), but might also find use in catalysis. Both the larger cavity of δ-CD as well as its increased flexibility could open up new possibilities here in comparison with the classical CDs.

Additionally, this thesis has further showcased the usefulness of enzyme-mediated dynamic combinatorial chemistry for the synthesis of rare carbohydrates such as LRCDs, and it serves as proof-of-concept for future endeavors in this regard.

List of publications

- A. Erichsen, G. H. J. Peters, S. R. Beeren. Templated dynamic enzymatic synthesis of δ-cyclodextrin by [n]-*pseudo*rotaxane formation. (Will be submitted January **2023**, to *J. Am. Chem. Soc.*)

- K. Hansen, A. Erichsen, D. Larsen, S. R. Beeren. Method for the production of delta-cyclodextrin. EP application no. 22192440.0. **2022**.

- A. Erichsen, D. Larsen, S. R. Beeren. Chaotropic and Kosmotropic Anions Regulate the Outcome of Enzyme-Mediated Dynamic Combinatorial Libraries of Cyclodextrins in Two Different Ways. *Front. Chem.* **2021**, *9*, 1–8.

Supporting information

SI.2: Supporting information for Chapter 2

The supporting information for chapter 2 is a copy of the "Supplementary Information" of the article "Chaotropic and Kosmotropic Anions Regulate the Outcome of Enzyme-Mediated Dynamic Combinatorial Libraries of Cyclodextrins in Two Different Ways" which was published in Frontiers in Chemistry in August 2021.[1] The contents have been formatted to fit the style of this thesis.

S2.1. Enzyme-mediated Dynamic Combinatorial Libraries

A series of reactions were set up by treating α-CD (10 mg/mL) with CGTase at room temperature in sodium phosphate buffer in the presence of different sodium salts at concentrations up to 4 M and in the absence or presence of cyclohexanol or cyclohexane carboxylate. The reactions were monitored by HPLC with an ELS detector and the distributions of α-CD, β-CD and γ-CD as a function of time are plotted in Figures S2.1–S2.7.

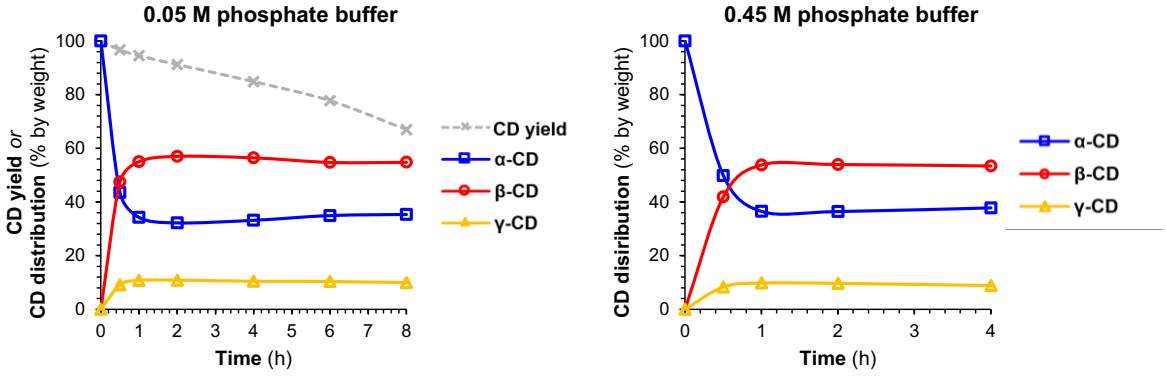


Figure S2.1. Distribution of α-CD, β-CD and γ-CD as a function of time in reactions started from α-CD (10 mg/mL) in sodium phosphate buffer (50 mM or 450 mM, as indicated on graph) at pH 7.5 treated with CGTase at room temperature. (Lines connecting data points are only to guide the eye.)

Note that the total CD yield (grey line, Supplementary Figure 1, left) decreases overtime, due to background hydrolysis and the build-up of short linear α-1,4-glucan and glucose. The same was observed for the DCLs prepared in the presence of salt. However, quantification was not possible as the salt peaks obscured relevant glucan peaks in the chromatograms. In the subsequent figures, only the changes in the distribution of CDs are plotted.

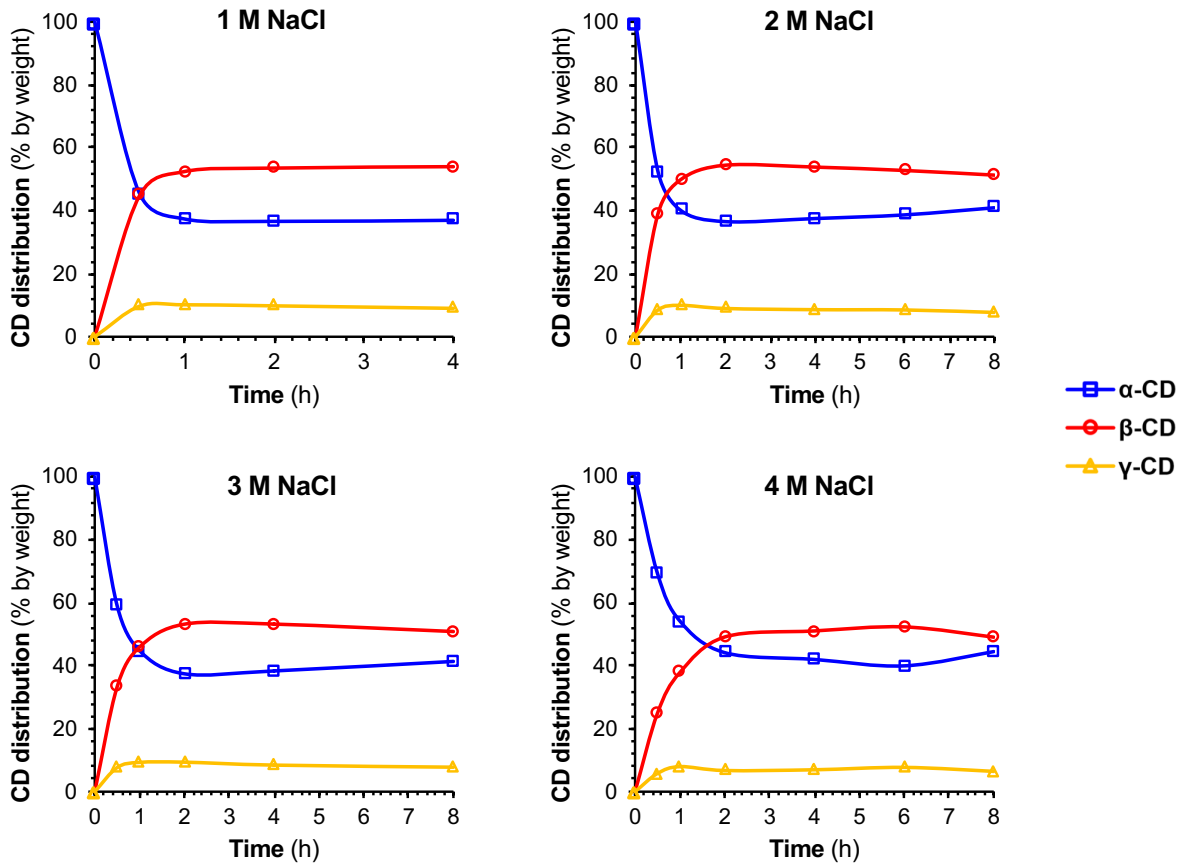


Figure S2.2. Distribution of α-CD, β-CD and γ-CD as a function of time in reactions started from α-CD (10 mg/mL) in sodium phosphate buffer (50 mM, pH 7.5) with the indicated concentration of NaCl treated with CGTase at room temperature. (Lines connecting data points are only to guide the eye.)

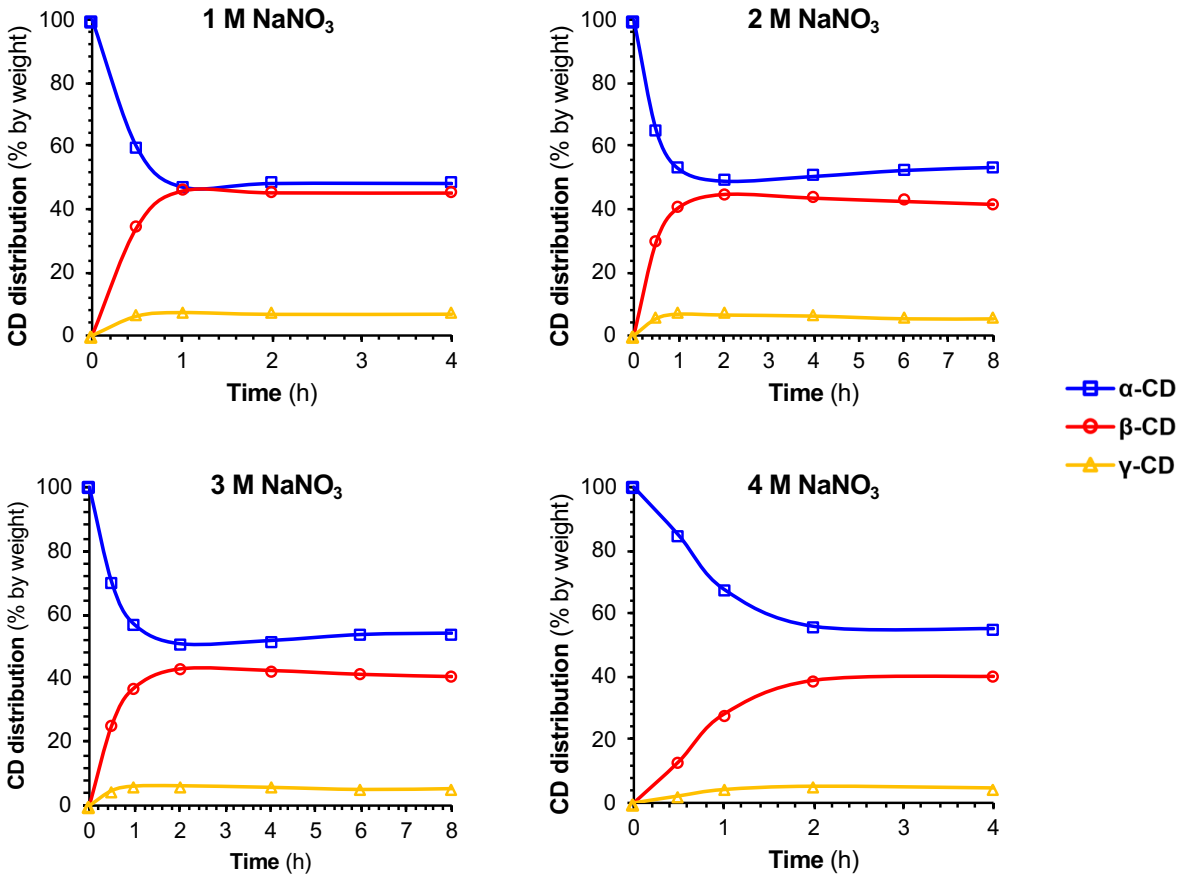


Figure S2.3. Distribution of α-CD, β-CD and γ-CD as a function of time in reactions started from α-CD (10 mg/mL) in sodium phosphate buffer (50 mM, pH 7.5) with the indicated concentration of $NaNO_3$ treated with CGTase at room temperature. (Lines connecting data points are only to guide the eye.)

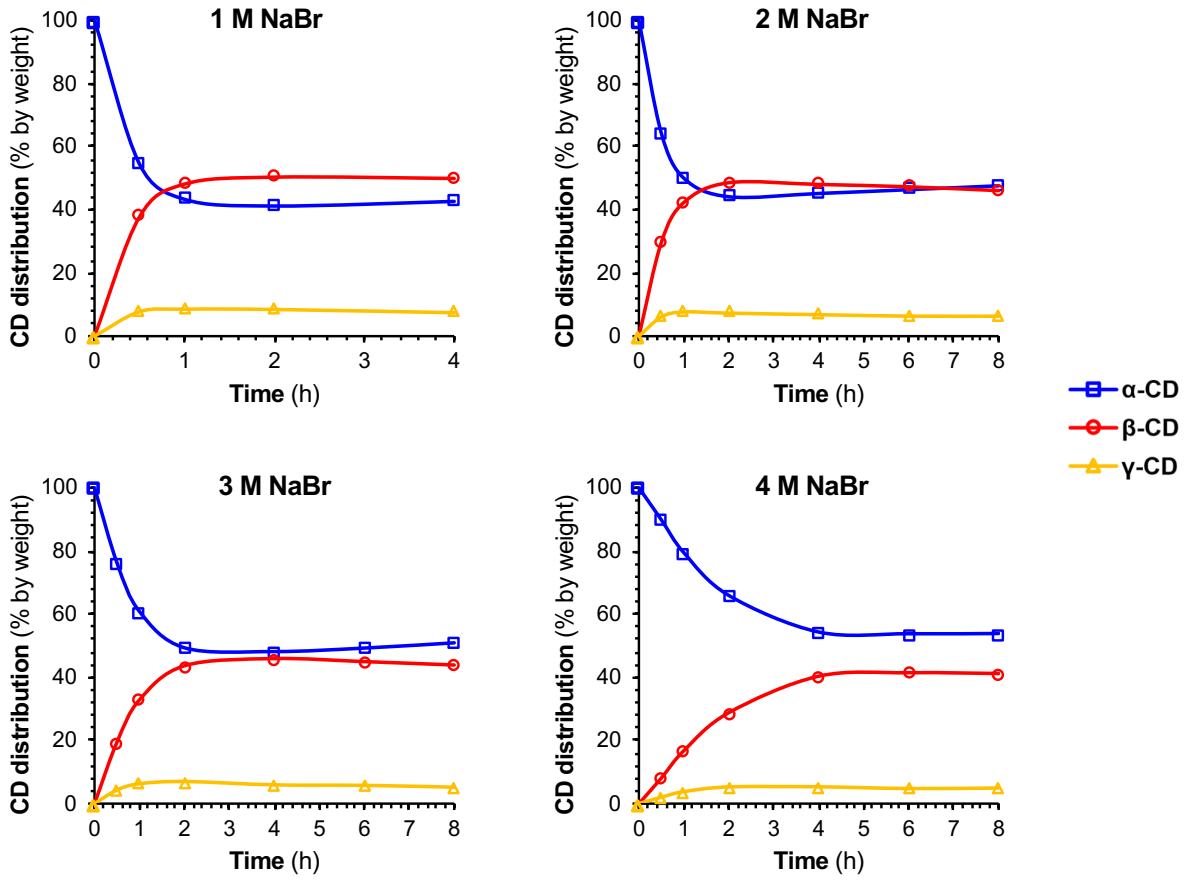


Figure S2.4. Distribution of α-CD, β-CD and γ-CD as a function of time in reactions started from α-CD (10 mg/mL) in sodium phosphate buffer (50 mM, pH 7.5) with the indicated concentration of NaBr treated with CGTase at room temperature. (Lines connecting data points are only to guide the eye.)

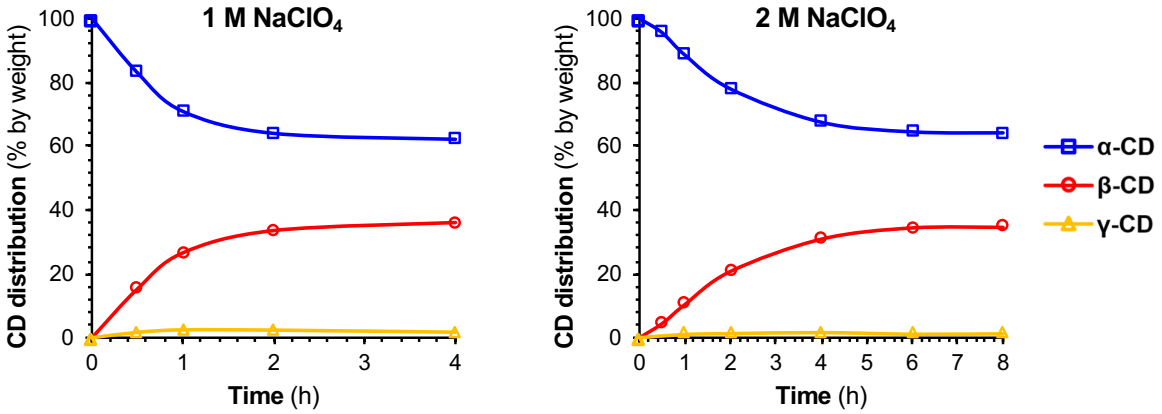


Figure S2.5. Distribution of α-CD, β-CD and γ-CD as a function of time in reactions started from α-CD (10 mg/mL) in sodium phosphate buffer (50 mM, pH 7.5) with the indicated concentration of $NaClO_4$ treated with CGTase at room temperature. (Lines connecting data points are only to guide the eye.)

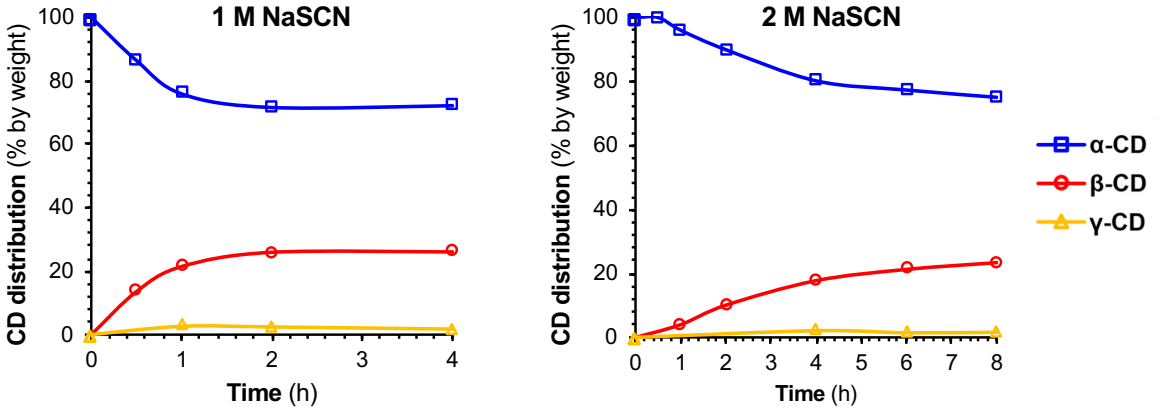


Figure S2.6. Distribution of α-CD, β-CD and γ-CD as a function of time in reactions started from α-CD (10 mg/mL) in sodium phosphate buffer (50 mM, pH 7.5) with the indicated concentration of NaSCN treated with CGTase at room temperature. (Lines connecting data points are only to guide the eye.)

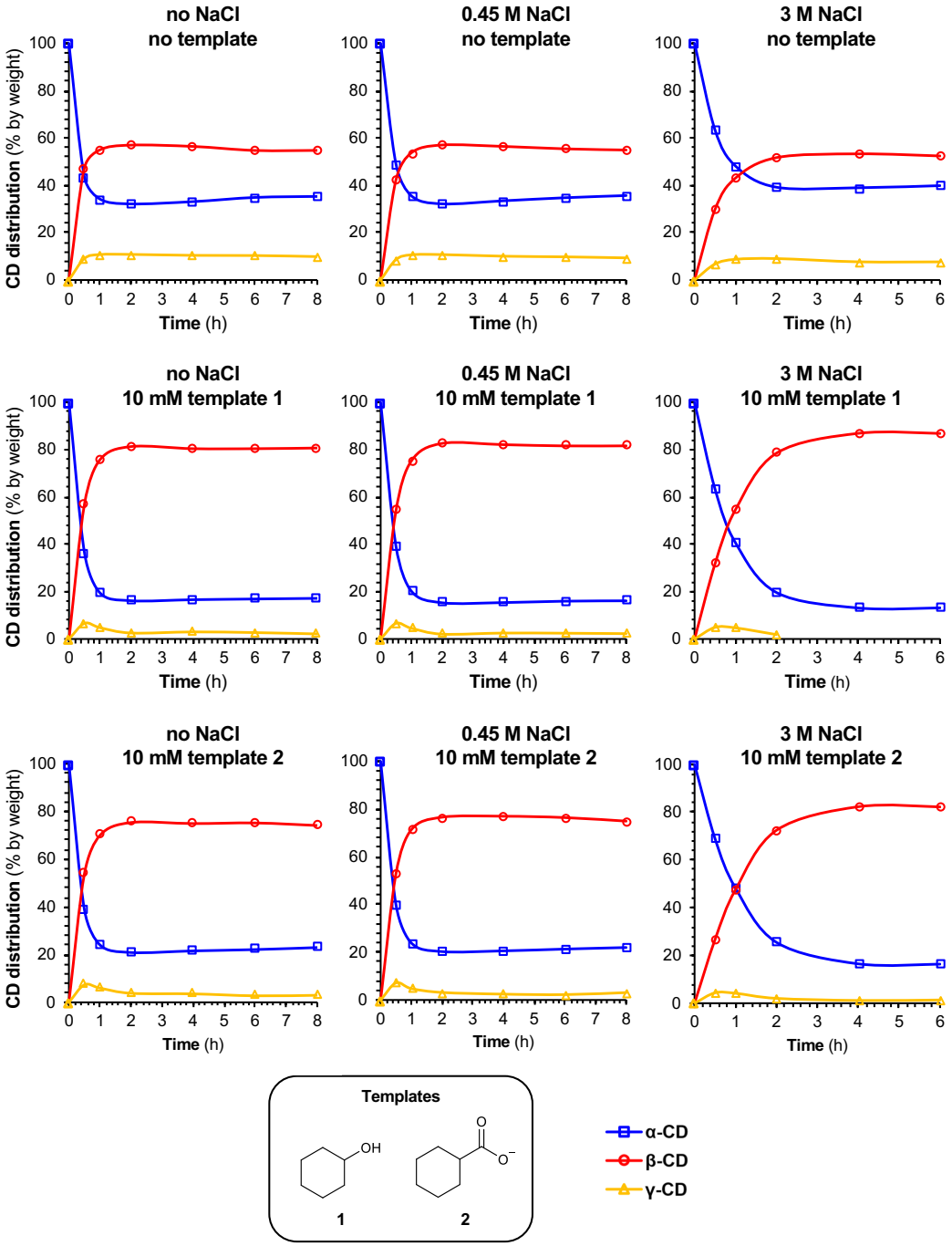


Figure S2.7. Distribution of α-CD, β-CD and γ-CD as a function of time in reactions started from α-CD (10 mg/mL) and treated with CGTase at room temperature in sodium phosphate buffer (50 mM, pH 7.5) in the presence of the indicated concentrations of cyclohexanol, cyclohexane carboxylate and NaCl. (Lines connecting data points are only to guide the eye.)

S2.2. Enzyme activity in the presence of denaturing salts NaSCN and $NaClO_4$

A series of solutions of CGTase in sodium phosphate buffer (50 mM, pH 7.5) with different sodium salts in concentrations up to 4 M were prepared. After 30 minutes of incubation of CGTase in salt solutions at room temperature, maltohexaose (**G6**) (10 mg/mL) was added. After 5 minutes, the reactions were quenched and the concentrations of **G6** remaining were determined by HPLC with an ELS detector. The consumption of **G6** relative to a reference (no salt added) are plotted in Figure S2.8.

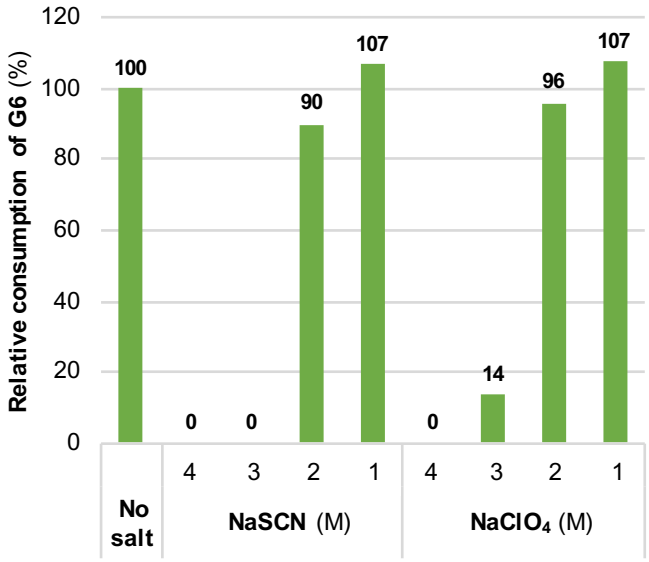


Figure S2.8. Consumption of G6 relative to a reference with no salt added after 30 minutes of exposing the enzyme to the salt solutions, followed by 5 minutes reaction time. With 3–4 M NaSCN a dramatic loss of enzyme activity was observed. With $NaClO_4$ NaSCN (1 M) and $NaClO_4$ (1 M) there was a perceived increase in activity, indicated by the increased consumption of G6. This is presumably due to the kinetic trapping of the α-CD / anion complex formed, leading to an apparent higher consumption of G6 due to a slower rate of the reverse ring-opening reaction.

S2.3. Simulations of dynamic combinatorial libraries

Simulations of dynamic combinatorial libraries using the program *DCLSim* requires the input of the relative formation constants K_f of the library members. K_f for α-, β-, and γ-CD were calculated according to Equation S2.1:

$$K_f(\text{CD-}n) = \frac{[\text{CD-n}]_{eq}}{[G_0]^n} \quad \textbf{(S2.1)}$$

Where CD-*n* is a cyclodextrin with a degree of polymerization *n*, K_f(CD-*n*) is the relative formation constant of the cyclodextrin, $[\text{CD-}n]_{eq}$ is the concentration (M) of the cyclodextrin at *pseudo*-equilibrium and $[G_0]$ is the total concentration of glucose units (M) in the cyclodextrin library at *pseudo*-equilibrium. The calculated K_f values are 9.5×10^4 M^{-5}, 2.6×10^6 M^{-6} and 7.6×10^6 M^{-7} for α-CD, β-CD, and γ-CD, respectively.

SI.3: Supporting information for Chapter 3

S3.1. NMR spectroscopy titrations

S3.1.1. Derivations of functions for fitting of 1:2 binding in mixed fast/slow exchange

The following equations were derived by Associate Professor Sophie R. Beeren

For the situation where the guest (G) can bind twice to the host (H) to give HG and HG_2, we have three equilibria and two association constants, K_{a1} and K_{a2}.

$$H + G \rightleftharpoons HG$$

$$HG + G \rightleftharpoons HG_2$$

The association constants are expressed in equations (S3.2) and (S3.3).

$$K_{a1} = \frac{[HG]}{[H][G]} \quad \text{(S3.2)} \qquad K_{a2} = \frac{[HG_2]}{[HG][G]} \quad \text{(S3.3)}$$

Equations (S3.2) and (S3.3) can be rearranged to give equations (S3.4) and (S3.5), which define the concentrations of the different complexes.

$$[HG] = K_{a1}[H][G] \quad \text{(S3.4)}$$

$$[HG_2] = K_{a1}K_{a2}[H][G]^2 \quad \text{(S3.5)}$$

The total concentrations of H and G are defined as $[H]_t$ and $[G]_t$ and give the mass balance equations (S3.6) and (S3.7).

$$[H]_t = [H] + [HG] + [HG_2] \quad \text{(S3.6)}$$

$$[G]_t = [G] + [HG] + 2[HG_2] \quad \text{(S3.7)}$$

Substitution of equations (S3.4) and (S3.5) into equation (S3.6) gives equation (S3.8).

$$[H]_t = [H] + K_{a1}[H][G] + K_{a1}K_{a2}[H][G]^2 \quad \text{(S3.8)}$$

Solving equation (S3.8) for [H] gives equation (S3.9). By substituting equation (S3.9) into equations (S3.4) and (S3.5) we obtain equations for [H], [HG] and $[HG_2]$ based on only one unknown concentration ([G]) (equations (S3.9), (S3.10), (S3.11)).

$$[H] = \frac{[H]_t}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2} \quad \text{(S3.9)}$$

$$[HG] = \frac{K_{a1}[G][H]_t}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2} \quad \text{(S3.10)}$$

$$[HG_2] = \frac{K_{a1}K_{a2}[G]^2[H]_t}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2} \quad \text{(S3.11)}$$

Substituting equations (S3.9), (S3.10) and (S3.11) into mass balance equation (S3.7) gives equation (S3.12).

$$[\mathrm{G}]_\mathrm{t} = [\mathrm{G}] + \frac{K_{\mathrm{a}1}[\mathrm{G}][\mathrm{H}]_\mathrm{t} + 2K_{\mathrm{a}1}K_{\mathrm{a}2}[\mathrm{G}]^2[\mathrm{H}]_\mathrm{t}}{1 + K_{\mathrm{a}1}[\mathrm{G}] + K_{\mathrm{a}1}K_{\mathrm{a}2}[\mathrm{G}]^2} \quad \text{(S3.12)}$$

Rearranging equation (S3.12) results in a quartic equation for [G].

$$0 = K_{\mathrm{a}1}K_{\mathrm{a}2}[\mathrm{G}]^3 + K_{\mathrm{a}1}[\mathrm{G}]^2 + 2K_{\mathrm{a}1}K_{\mathrm{a}2}[\mathrm{H}]_\mathrm{t}[\mathrm{G}]^2 - K_{\mathrm{a}1}K_{\mathrm{a}2}[\mathrm{G}]_\mathrm{t}[\mathrm{G}]^2 + K_{\mathrm{a}1}[\mathrm{H}]_\mathrm{t}[\mathrm{G}] - K_{\mathrm{a}1}[\mathrm{G}]_\mathrm{t}[\mathrm{G}] + [\mathrm{G}] - [\mathrm{G}]_\mathrm{t}$$

For convenience, we can express the cubic equation as equation (S3.13).

$$\mathrm{A}[\mathrm{G}]^3 + \mathrm{B}[\mathrm{G}]^2 + \mathrm{C}[\mathrm{G}] + \mathrm{D} = 0 \quad \text{(S3.13)}$$

wherein

$$\mathrm{A} = K_{\mathrm{a}1}K_{\mathrm{a}2}$$

$$\mathrm{B} = K_{\mathrm{a}1} + 2K_{\mathrm{a}1}K_{\mathrm{a}2}[\mathrm{H}]_\mathrm{t} - K_{\mathrm{a}1}K_{\mathrm{a}2}[\mathrm{G}]_\mathrm{t}$$

$$\mathrm{C} = 1 + K_{\mathrm{a}1}[\mathrm{H}]_\mathrm{t} - K_{\mathrm{a}1}[\mathrm{G}]_\mathrm{t}$$

$$\mathrm{D} = -[\mathrm{G}]_\mathrm{t}$$

With the [G] dependent equation in hand, we must now derive equations that relate the NMR observable parameters to K_{a1}, K_{a2}, [G] and known concentrations $[H]_t$ and $[G]_t$.

In the situation where the first binding is seen in fast exchange and the second binding is seen in slow exchange, we can directly determine the concentrations $[HG_2]$, and ([H]+[HG]) by integration of the two sets of peaks.

In the NMR spectrum we can detect the formation of HG from H by the change in chemical shift of the peak corresponding to these two species in fast exchange. The observed chemical shift (δ_{obs}) is determined by the weighted average of the chemicals shifts of H + HG (δ_H, δ_{HG}) according to equation (S3.14).

$$\delta_{\mathrm{obs}} = \delta_\mathrm{H}\frac{[\mathrm{H}]}{[\mathrm{H}]+[\mathrm{HG}]} + \delta_{\mathrm{HG}}\frac{[\mathrm{HG}]}{[\mathrm{H}]+[\mathrm{HG}]} \quad \text{(S3.14)}$$

Rearranging equation (S3.14) gives equation (S3.15) where $\Delta\delta_{HG}$ is the chemical shift difference between δ_{HG} and δ_H, and $\Delta\delta_{obs}$ is the observed chemical shift change during the titration.

$$\delta_{\mathrm{obs}} = \left(1 - \frac{[\mathrm{HG}]}{[\mathrm{H}]+[\mathrm{HG}]}\right)\delta_\mathrm{H} + \frac{[\mathrm{HG}]}{[\mathrm{H}]+[\mathrm{HG}]}\delta_{\mathrm{HG}}$$

$$\delta_{\mathrm{obs}} - \delta_\mathrm{H} = (\delta_{\mathrm{HG}} - \delta_\mathrm{H})\frac{[\mathrm{HG}]}{[\mathrm{H}]+[\mathrm{HG}]}$$

$$\Delta\delta_{\mathrm{obs}} = \frac{[\mathrm{HG}]}{[\mathrm{H}]+[\mathrm{HG}]}\Delta\delta_{\mathrm{HG}} \quad \text{(S3.15)}$$

Rearranging equation (S3.15), and dividing both side by $[H]_t$, gives equation (S3.16).

$$([H] + [HG])\Delta\delta_{obs} = [HG]\Delta\delta_{HG}$$

$$\frac{[H]+[HG]}{[H]_t}\Delta\delta_{obs} = \frac{[HG]}{[H]_t}\Delta\delta_{HG} \quad \text{(S3.16)}$$

Substitution of equation (S3.10) into equation (S3.16) gives equation (S3.17).

$$\frac{[H]+[HG]}{[H]_t}\Delta\delta_{obs} = \frac{K_{a1}[G]}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2}\Delta\delta_{HG} \quad \text{(S3.17)}$$

Furthermore, from equation (S3.14) we can obtain relationship (S3.18).

$$\frac{[HG_2]}{[H]_t} = \frac{K_{a1}K_{a2}[G]^2}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2} \quad \text{(S3.18)}$$

Together, equations (S3.17) and (S3.18) describe the relationships between the observed chemical shift change, the observed molar fractions of H+HG, and HG_2, and the concentration of guest ([G]). Experimentally generated isotherms are generated by plotting $\frac{[H]+[HG]}{[H]_t}\Delta\delta_{obs}$ and $\frac{[HG_2]}{[H]_t}$ against $[G]_t$. The two isotherms, respectively, are simultaneously fitted to equations (S3.17) and (S3.18), as described below.

Although the value of [G] in equations (S3.17) and (S3.18) varies depending on $[G]_t$ (as described by equation (S3.13)), the solution can be readily approached through an iterative data fitting procedure. We have adapted the method described by Hargrove *et al.*[25] for fitting of 1:2 binding in optical binding isotherms using the non-linear curve fitting program Origin. The script we wrote for the fitting function in Origin is shown below. The code is written in Labscript. Briefly, equation (S3.13) is solved numerically for [G] using Newton's method while fitting the experimental data to equations (S3.17) and (S3.18). Given a set of estimated initial values for K_{a1}, K_{a2} and $\Delta\delta_{HG}$, the program iteratively determines the parameter values that best fit the experimental data in terms of both the integrals and the chemical shift changes.

```
A=K1*K2;
B=K1+2*K2*K1*Ht-K1*K2*x;
C=1+K1*Ht-K1*x;
D=-x;
for (G=x, step=1; abs(step)>1e-15; G=G-
step){step=(A*G*G*G+B*G*G+C*G+D)/(3*A*G*G+2*B*G+C);};
y1=(D1*K1*G)/(1+K1*G+K1*K2*G*G);
y2=(K1*K2*G*G)/(1+K1*G+K1*K2*G*G)
```

Where K1, K2, D1, Ht, x, y1, and y2 represent K_{a1}, K_{a2}, $\Delta\delta_{HG}$, $[H]_t$, $[G]_t$, $\frac{[H]+[HG]}{[H]_t}\Delta\delta_{obs}$ and $\frac{[HG_2]}{[H]_t}$, respectively.

S3.1.2. Derivations of functions for fitting of 1:3 binding in mixed fast/slow exchange

The following equations were derived by Associate Professor Sophie R. Beeren

For the situation where the guest (G) can bind three times to the host (H) to give HG, HG_2 and HG_3, we have three equilibria and three association constants, K_{a1}, K_{a2}, and K_{a3}.

$$H + G \rightleftharpoons HG$$

$$HG + G \rightleftharpoons HG_2$$

$$HG_2 + G \rightleftharpoons HG_3$$

The association constants are expressed in equations (S3.19), (S.20) and (S3.21).

$$K_{a1} = \frac{[HG]}{[H][G]} \quad \text{(S3.19)} \qquad K_{a2} = \frac{[HG_2]}{[HG][G]} \quad \text{(S3.20)} \qquad K_{a3} = \frac{[HG_3]}{[HG_2][G]} \quad \text{(S3.21)}$$

Equations (S3.19), (S3.20) and (S3.21) can be rearranged to give equations (S3.22), (S3.23) and (S3.24), which define the concentrations of the different complexes.

$$[HG] = K_{a1}[H][G] \quad \text{(S3.22)}$$

$$[HG_2] = K_{a1}K_{a2}[H][G]^2 \quad \text{(S3.23)}$$

$$[HG_3] = K_{a1}K_{a2}K_{a3}[H][G]^3 \quad \text{(S3.24)}$$

The total concentrations of H and G are defined as $[H]_t$ and $[G]_t$ and give the mass balance equations (S3.25) and (S3.26).

$$[H]_t = [H] + [HG] + [HG_2] + [HG_3] \quad \text{(S3.25)}$$

$$[G]_t = [G] + [HG] + 2[HG_2] + 3[HG_3] \quad \text{(S3.26)}$$

Substitution of equations (S3.22), (S3.23) and (S3.24) into equation (S3.25) gives equation (S3.27).

$$[H]_t = [H] + K_{a1}[H][G] + K_{a1}K_{a2}[H][G]^2 + K_{a1}K_{a2}K_{a3}[H][G]^3 \quad \text{(S3.27)}$$

Solving equation (S3.27) for [H] gives equation (S3.28). By substituting equation (S3.28) into equations (S3.22), (S3.23) and (S3.24) we obtain equations for [H], [HG], $[HG_2]$ and $[HG_3]$ based on only one unknown concentration ([G]) (equations (S3.29), (S3.30), (S3.31)).

$$[H] = \frac{[H]_t}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2+K_{a1}K_{a2}K_{a3}[G]^3} \quad \text{(S3.28)}$$

$$[HG] = \frac{K_{a1}[G][H]_t}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2+K_{a1}K_{a2}K_{a3}[G]^3} \quad \text{(S3.29)}$$

$$[\mathrm{HG_2}] = \frac{K_{a1}K_{a2}[\mathrm{G}]^2[\mathrm{H}]_t}{1+K_{a1}[\mathrm{G}]+K_{a1}K_{a2}[\mathrm{G}]^2+K_{a1}K_{a2}K_{a3}[\mathrm{G}]^3} \quad \text{(S3.30)}$$

$$[\mathrm{HG_3}] = \frac{K_{a1}K_{a2}K_{a3}[\mathrm{G}]^3[\mathrm{H}]_t}{1+K_{a1}[\mathrm{G}]+K_{a1}K_{a2}[\mathrm{G}]^2+K_{a1}K_{a2}K_{a3}[\mathrm{G}]^3} \quad \text{(S3.31)}$$

Substituting equations (S3.28), (S3.29), (S3.30), and (S3.31) into mass balance equation (S3.26) gives equation (S3.32).

$$[\mathrm{G}]_t = [\mathrm{G}] + \frac{K_{a1}[\mathrm{G}][\mathrm{H}]_t+2K_{a1}K_{a2}[\mathrm{G}]^2[\mathrm{H}]_t+3K_{a1}K_{a2}K_{a3}[\mathrm{G}]^3[\mathrm{H}]_t}{1+K_{a1}[\mathrm{G}]+K_{a1}K_{a2}[\mathrm{G}]^2+K_{a1}K_{a2}K_{a3}[\mathrm{G}]^3} \quad \text{(S3.32)}$$

Rearranging equation (S3.32) results in a quartic equation for [G].

$$0 = K_{a1}K_{a2}K_{a3}[\mathrm{G}]^4 + K_{a1}K_{a2}[\mathrm{G}]^3 + 3K_{a1}K_{a2}K_{a3}[\mathrm{G}]^3[\mathrm{H}]_t - K_{a1}K_{a2}K_{a3}[\mathrm{G}]^3[\mathrm{G}]_t + K_{a1}[\mathrm{G}]^2 + 2K_{a1}K_{a2}[\mathrm{G}]^2[\mathrm{H}]_t - K_{a1}K_{a2}[\mathrm{G}]^2[\mathrm{G}]_t + K_{a1}[\mathrm{G}][\mathrm{H}]_t - K_{a1}[\mathrm{G}][\mathrm{G}]_t + [\mathrm{G}] - [\mathrm{G}]_t$$

For convenience, we can express the quartic equation as equation (S3.33).

$$\mathrm{A}[\mathrm{G}]^4 + \mathrm{B}[\mathrm{G}]^3 + \mathrm{C}[\mathrm{G}]^2 + \mathrm{D}[\mathrm{G}] + \mathrm{E} = 0 \quad \text{(S3.33)}$$

wherein

$$\mathrm{A} = K_{a1}K_{a2}K_{a3}$$

$$\mathrm{B} = K_{a1}K_{a2} + 3K_{a1}K_{a2}K_{a3}[\mathrm{H}]_t - K_{a1}K_{a2}K_{a3}[\mathrm{G}]_t$$

$$\mathrm{C} = K_{a1} + 2K_{a1}K_{a2}[\mathrm{H}]_t - K_{a1}K_{a2}[\mathrm{G}]_t$$

$$\mathrm{D} = 1 + K_{a1}K_{a2} - K_{a1}[\mathrm{G}]_t$$

$$\mathrm{E} = -[\mathrm{G}]_t$$

With the [G] dependent equation in hand, we must now derive equations that relate the NMR observable parameters to K_{a1}, K_{a2}, K_{a3}, [G] and known concentrations $[H]_t$ and $[G]_t$.

In the situation where the first binding is seen in fast exchange and the second and third bindings are seen in slow exchange, we can directly determine the concentrations $[HG_2]$, $[HG_3]$ and ([H]+[HG]) by integration of the three sets of peaks.

In the NMR spectrum we can detect the formation of HG from H by the change in chemical shift of the peak corresponding to these two species in fast exchange. As described for the 1:2 binding case (where the first binding is in fast exchange and the second binding is in slow exchange) we can derive equation (S3.16)

$$\frac{[\mathrm{H}]+[\mathrm{HG}]}{[\mathrm{H}]_t}\Delta\delta_{obs} = \frac{[\mathrm{HG}]}{[\mathrm{H}]_t}\Delta\delta_{HG} \quad \text{(S3.16)}$$

Substitution of equation (S3.29) into equation (S3.16) gives equation (S3.36).

$$\frac{[\mathrm{H}]+[\mathrm{HG}]}{[H]_t}\Delta\delta_{obs} = \frac{K_{a1}[\mathrm{G}]}{1+K_{a1}[\mathrm{G}]+K_{a1}K_{a2}[\mathrm{G}]^2+K_{a1}K_{a2}K_{a3}[\mathrm{G}]^3}\Delta\delta_{HG} \quad \text{(S3.36)}$$

Furthermore, from equation (S3.31) we can obtain relationship (S3.37).

$$\frac{[HG_3]}{[H]_t} = \frac{K_{a1}K_{a2}K_{a3}[G]^3}{1+K_{a1}[G]+K_{a1}K_{a2}[G]^2+K_{a1}K_{a2}K_{a3}[G]^3} \quad \text{(S3.37)}$$

Together, equations (S3.36) and (S3.37) describe the relationships between the observed chemical shift change, the observed molar fractions of H+HG, HG_2 and HG_3, and the concentration of guest ([G]). Experimentally generated isotherms are generated by plotting $\frac{[H]+[HG]}{[H]_t}\Delta\delta_{obs}$ and $\frac{[HG_3]}{[H]_t}$ against $[G]_t$. The two isotherms, respectively, are simultaneously fitted to equations (S3.36) and (S3.37), as described below.

We have further adapted the method described by Hargrove *et al.*[25] in order to fit this 1:3 binding data using the non-linear curve fitting program Origin. The script we wrote for the fitting function in Origin is shown below. The code is written in Labscript. Briefly, equation (S3.33) is solved numerically for [G] using Newton's method while fitting the experimental data to equations (S3.36) and (S3.37). Given a set of estimated initial values for K_{a1}, K_{a2}, K_{a3} and $\Delta\delta_{HG}$, the program iteratively determines the parameter values that best fit the experimental data in terms of both the integrals and the chemical shift changes.

```
A=K1*K2*K3;

B=K1*K2+3*K1*K2*K3*Ht-K1*K2*K3*x;

C=K1+2*K1*K2*Ht-K1*K2*x;

D=1+K1*Ht-K1*x;

E=-x;

For (G=x, step=1; abs(step)>1e-15; G=G-
step){step=(A*G*G*G*G+B*G*G*G+C*G*G+D*G+E)/(4*A*G*G*G+3*B*
G*G+2*C*G+D);};

y1=(D1*K1*G)/(1+K1*G+K1*K2*G*G+K1*K2*K3*G*G*G);

y2=(K1*K2*K3*G*G*G)/(1+K1*G+K1*K2*G*G+K1*K2*K3*G*G*G)
```

Where K1, K2, K3, D1, Ht, x, y1, and y2 represent K_{a1}, K_{a2}, K_{a3}. $\Delta\delta_{HG}$, $[H]_t$, $[G]_t$, $\frac{[H]+[HG]}{[H]_t}\Delta\delta_{obs}$ and $\frac{[HG_3]}{[H]_t}$, respectively.

S3.1.3. NMR titrations with α-CD

α-CD and **T3**

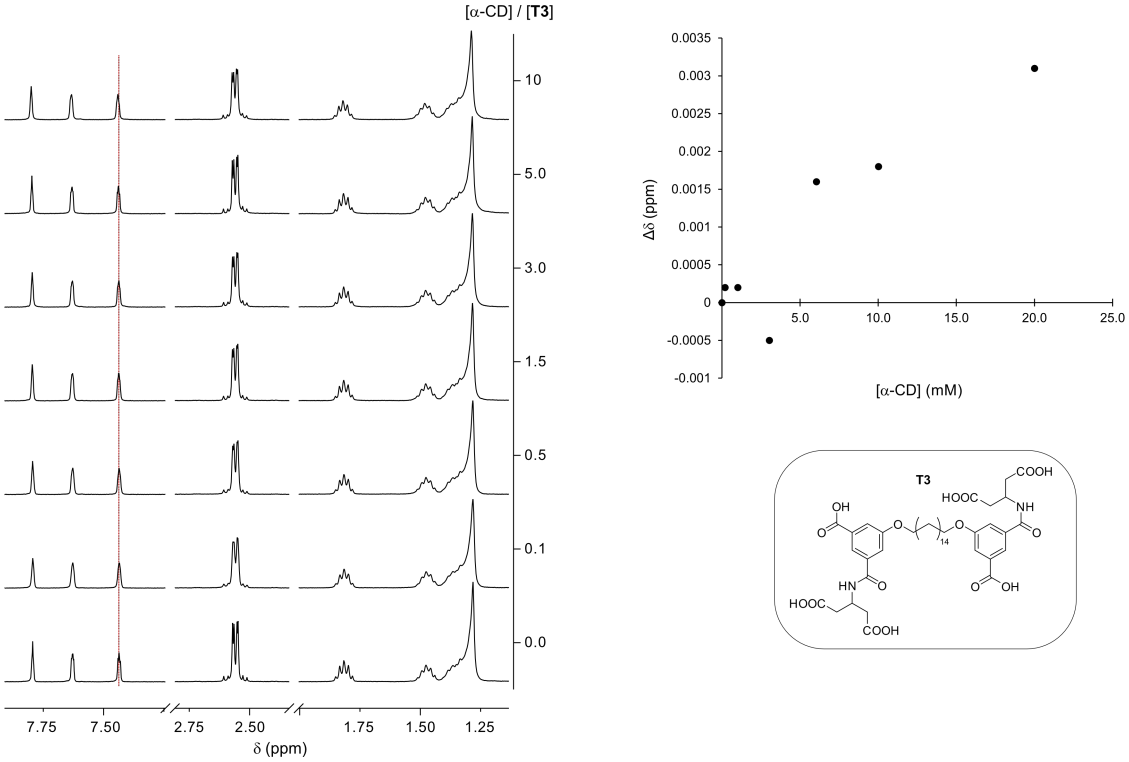


Figure S3.1. ^{1}H NMR spectroscopy titration of **T3** with α-CD shows no binding *or* very weak, nonspecific binding. *Left*: Partial ^{1}H NMR spectra (400 MHz) of **T3** with increasing concentrations of α-CD. The titration was performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T3** at 2.0 mM. *Right:* Change in chemical shift for the aromatic proton of **T3** at 7.44 ppm**.**

α-CD and **T4**

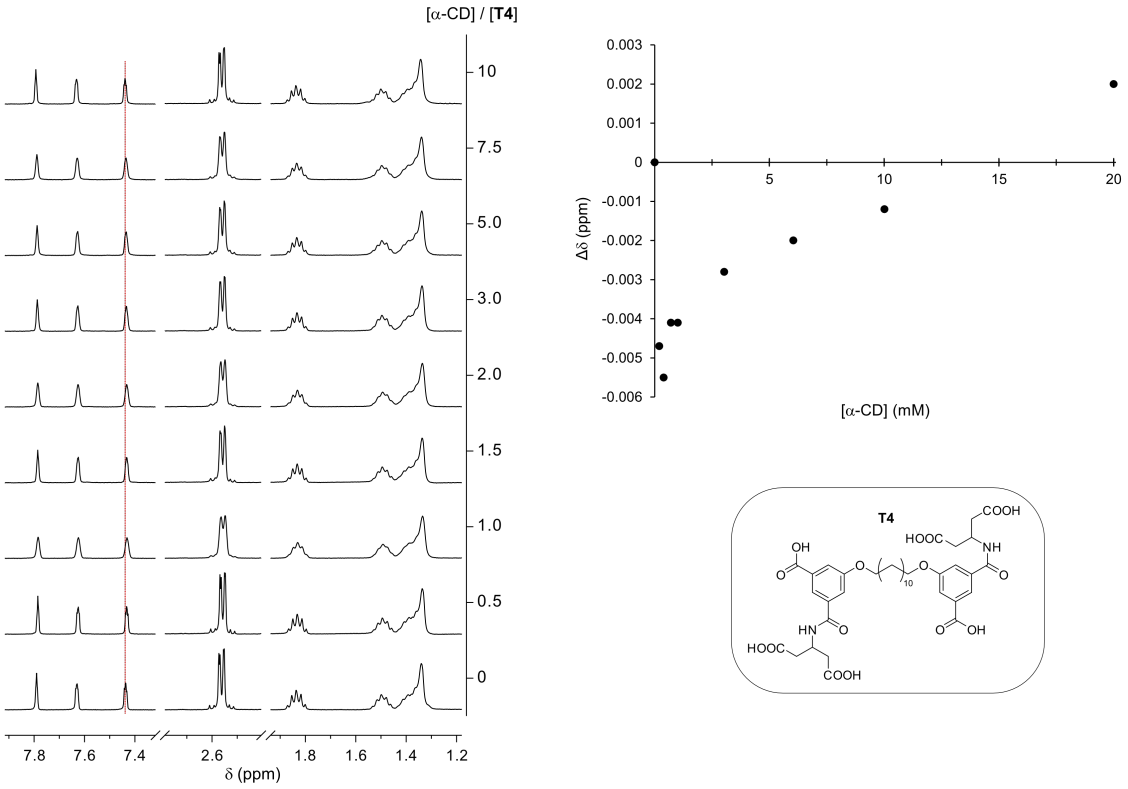


Figure S3.2. ^{1}H NMR spectroscopy titration of **T4** with α-CD shows no binding *or* very weak, nonspecific binding. The titration was performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T4** at 2.0 mM. *Left*: Partial ^{1}H NMR spectra (400 MHz) of **T4** with increasing concentrations of α-CD. *Right:* Change in chemical shift for the aromatic proton of **T4** at 7.44 ppm.

S3.1.4. NMR titrations with β-CD

β-CD and **T2**

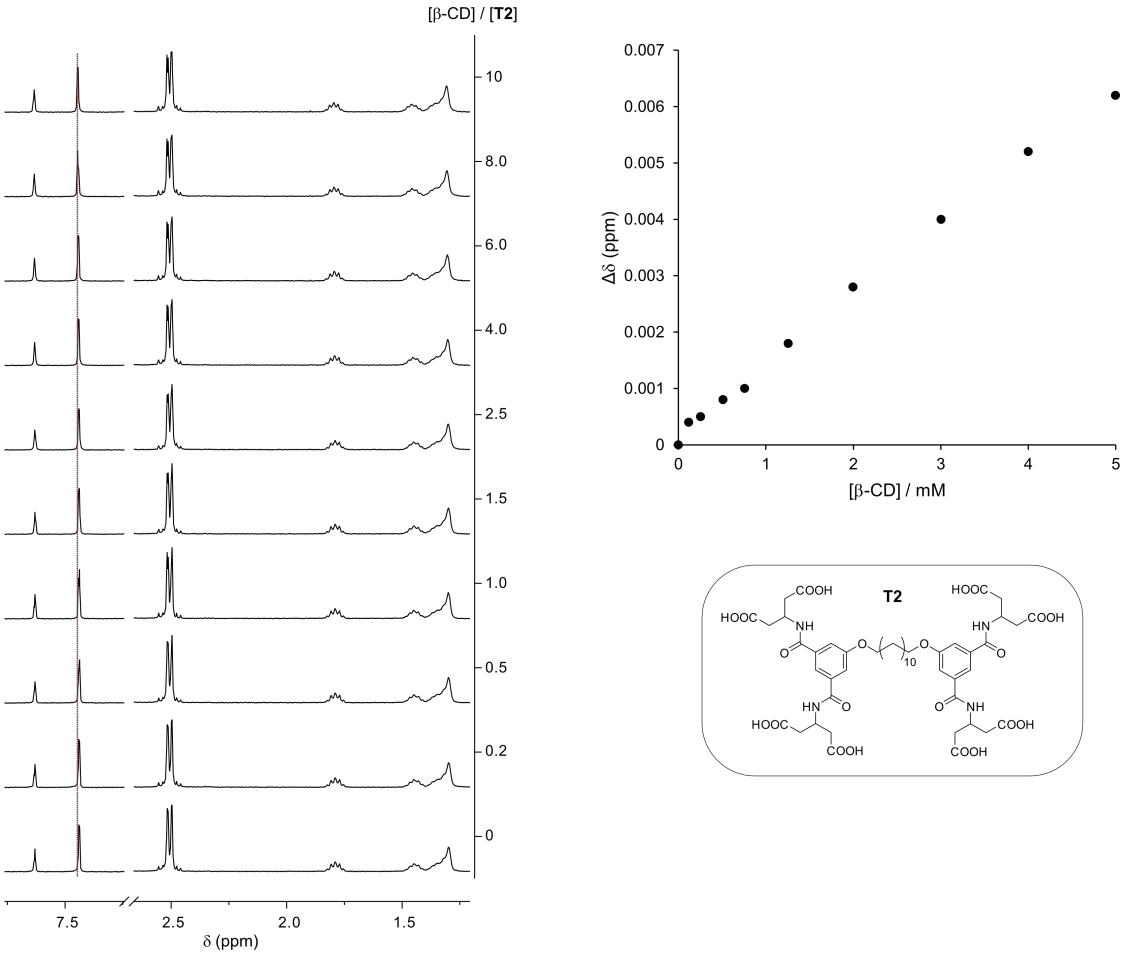


Figure S3.3. ^{1}H NMR spectroscopy titration of **T2** with β-CD shows no binding *or* weak, nonspecific binding. The titration was performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T2** at 2.0 mM. *Left*: Partial ^{1}H NMR spectra (400 MHz) of **T2** with increasing concentrations of β-CD. *Right:* Change in chemical shift for the aromatic proton of **T2** at 7.44 ppm.

β-CD and **T4**

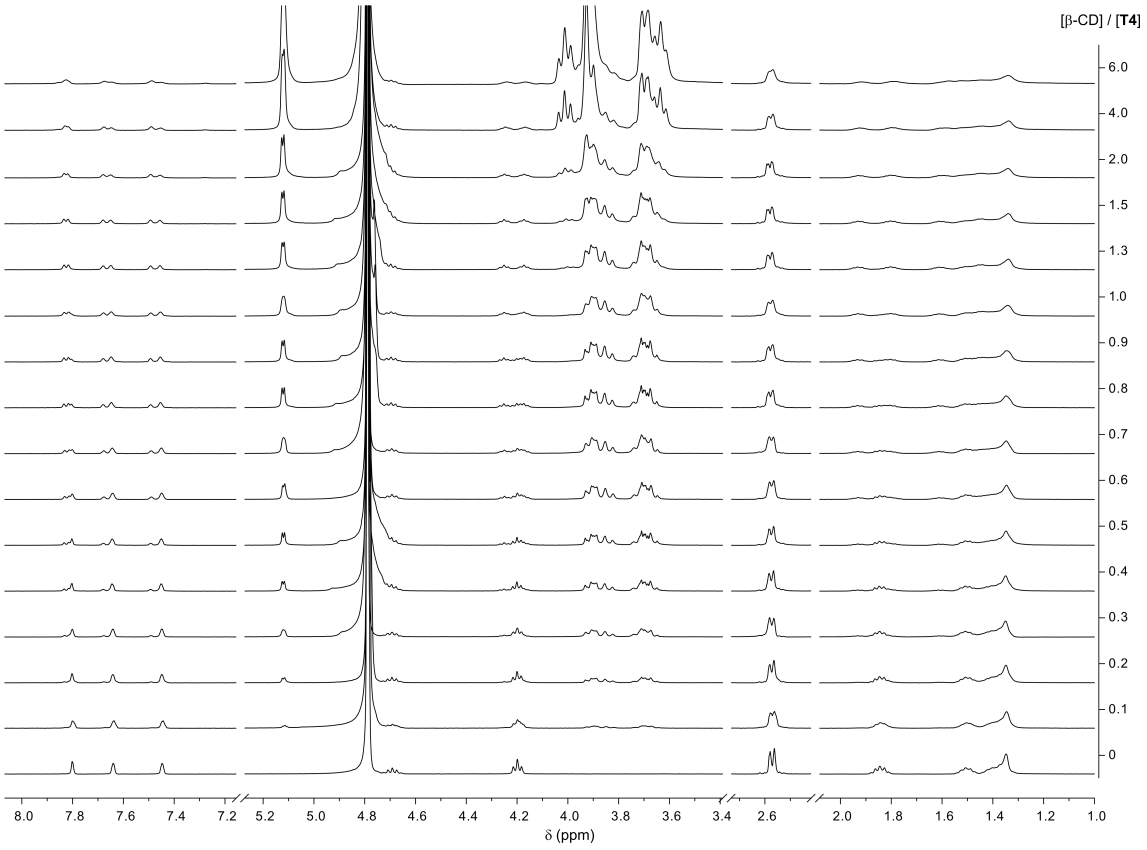


Figure S3.4. ^{1}H NMR spectra (400 MHz) of the titration of **T4** with β-CD performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T4** at 2.0 mM.

S3.1.5. NMR titrations with γ-CD

γ-CD and **T1**

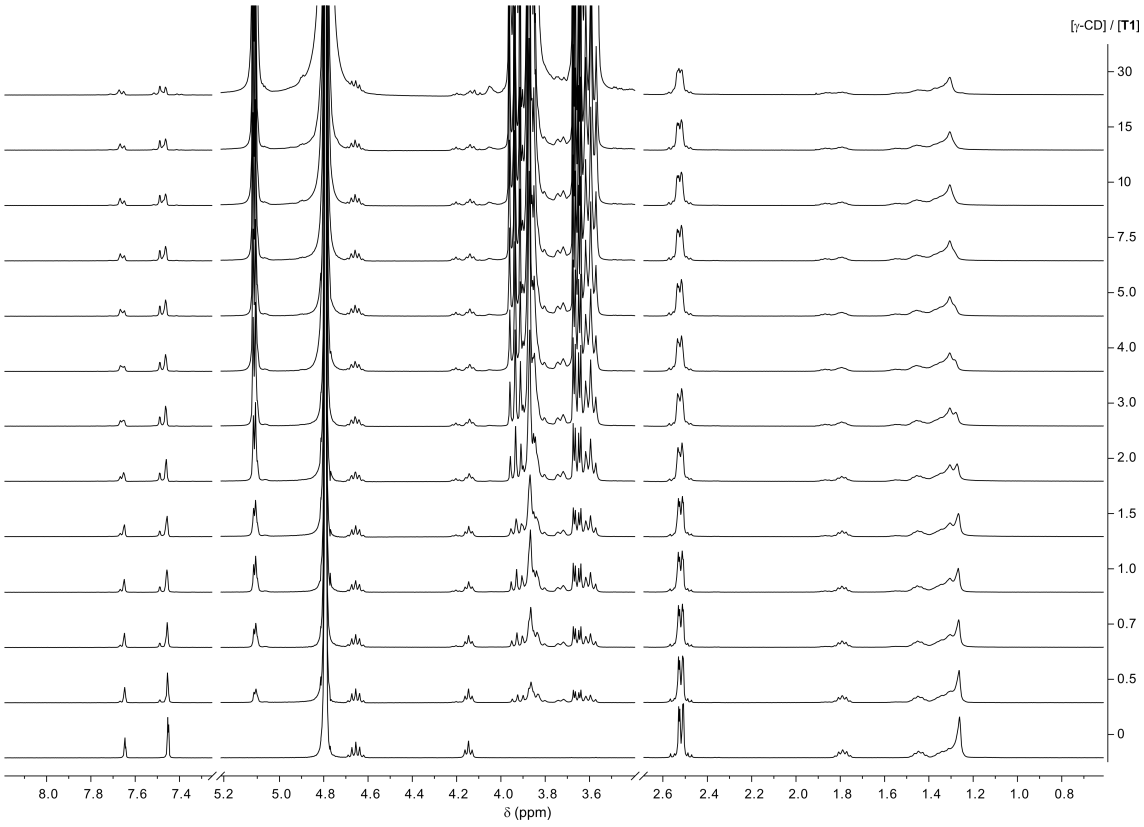


Figure S3.5. ^{1}H NMR spectra (400 MHz) of the titration of γ-CD with **T1**. The titration was performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T1** at 1.0 mM.

γ-CD and **T2**

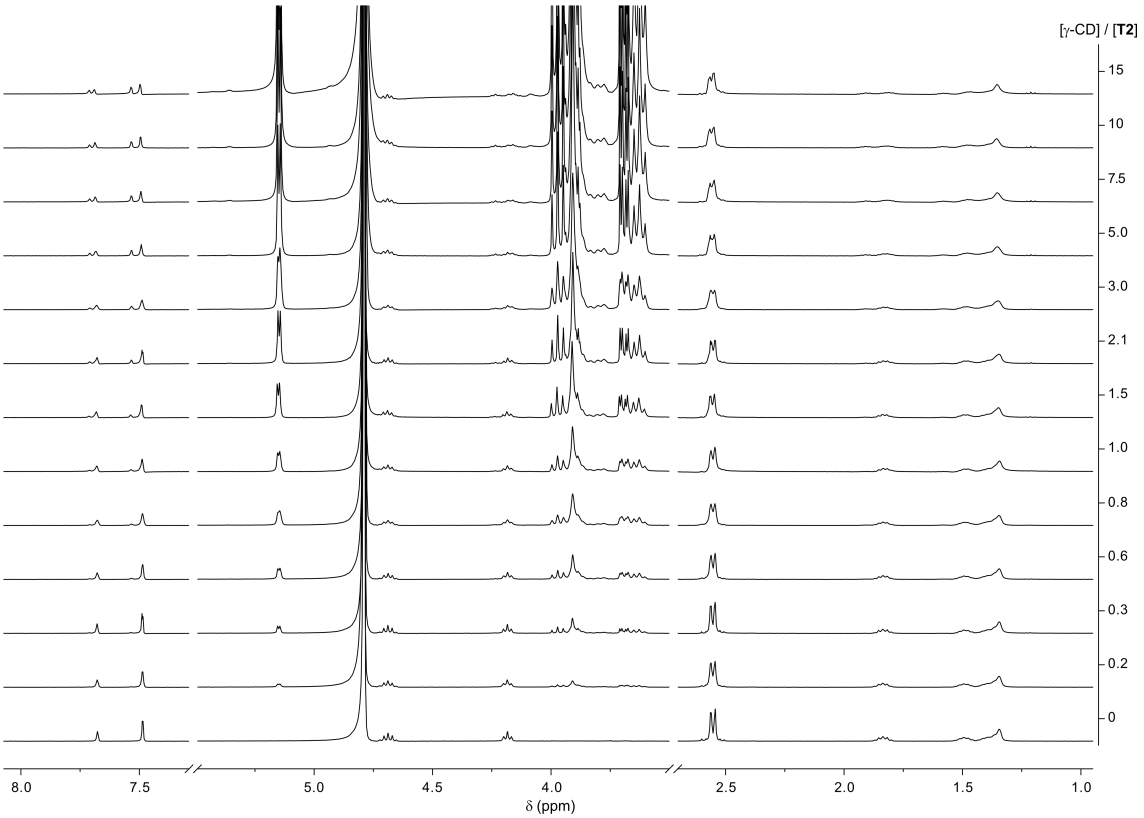


Figure S3.6. NMR spectra (400 MHz) of the titration of **T2** with γ-CD performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T2** at 1.0 mM.

γ-CD and **T3**

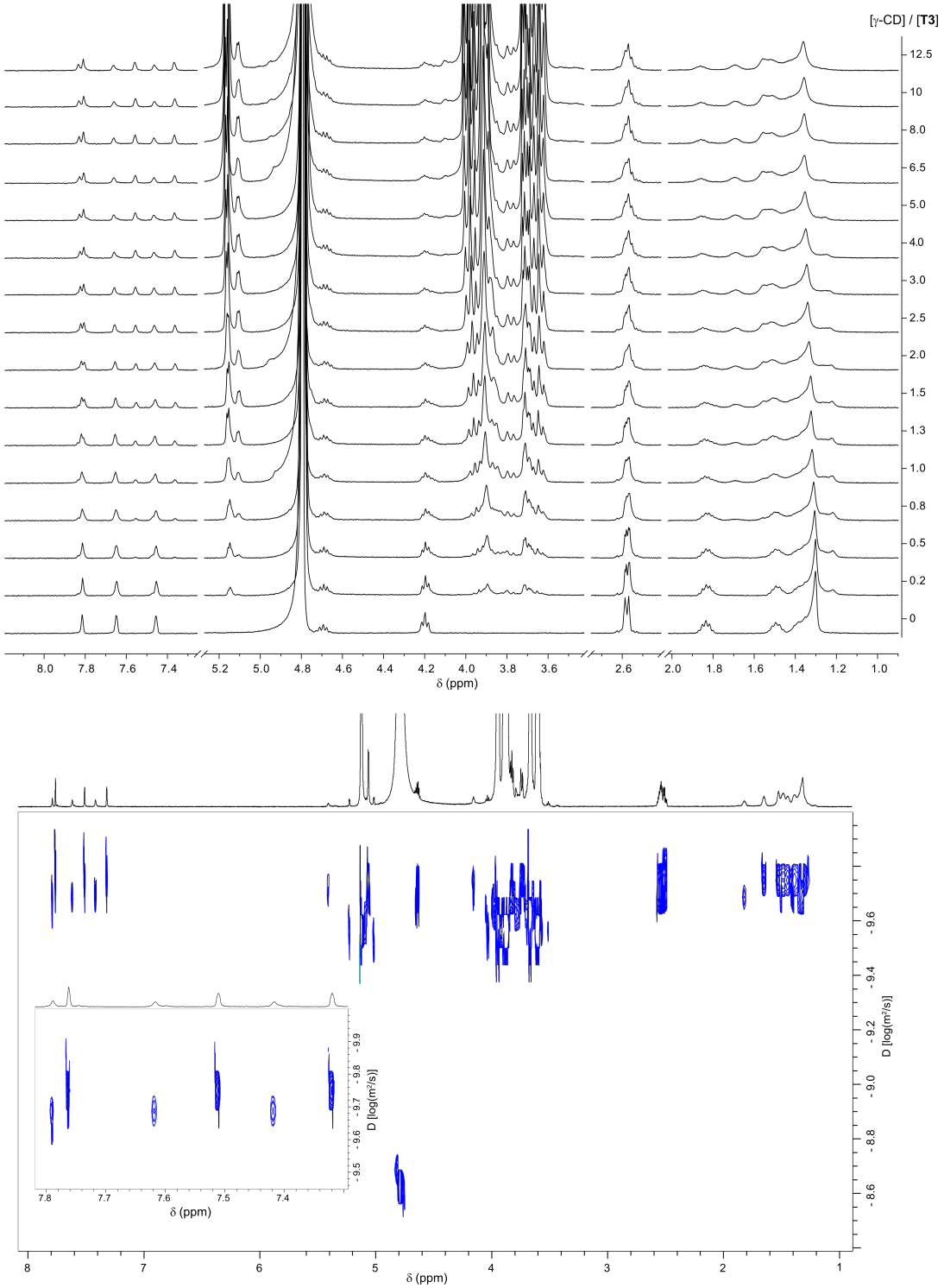


Figure S3.7. ^{1}H NMR spectroscopy titration of **T3** with γ-CD performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T3** at 1.0 mM. *Top:* ^{1}H NMR spectra (400 MHz) of the titration of **T3** with γ-CD. *Bottom:* DOSY spectrum (800 MHz) of the data point with 12.5 mM γ-CD.

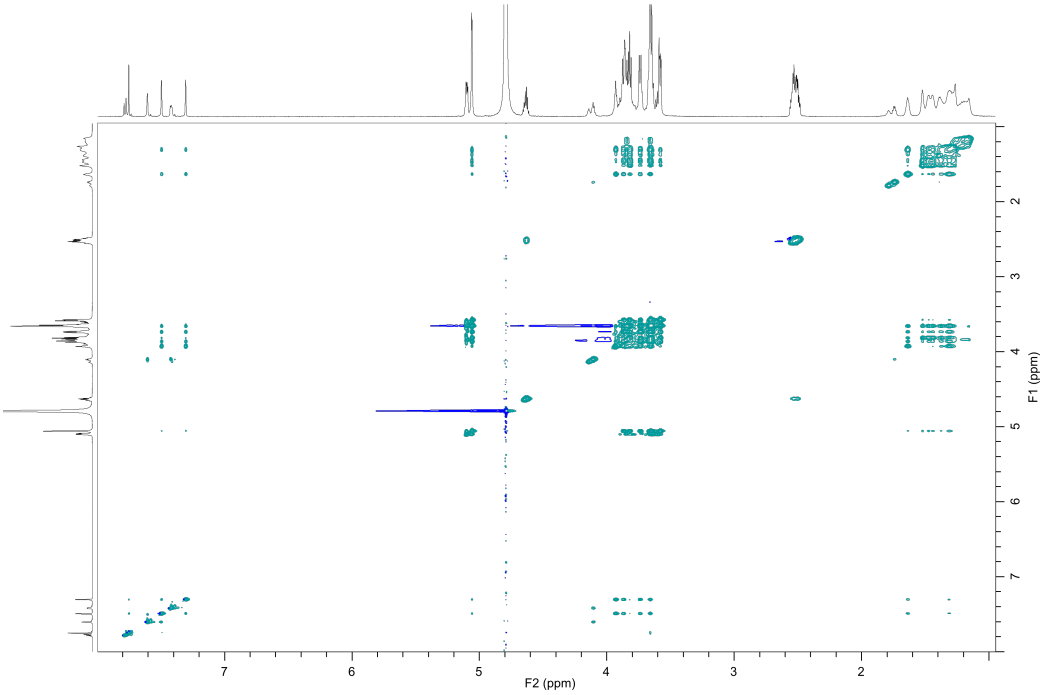


Figure S3.8. ^{1}H-^{1}H NOESY spectrum (800 MHz) of **T3** (5 mM) and γ-CD (5 mM) in D_2O phosphate buffer (pH 7.5, 100 mM).

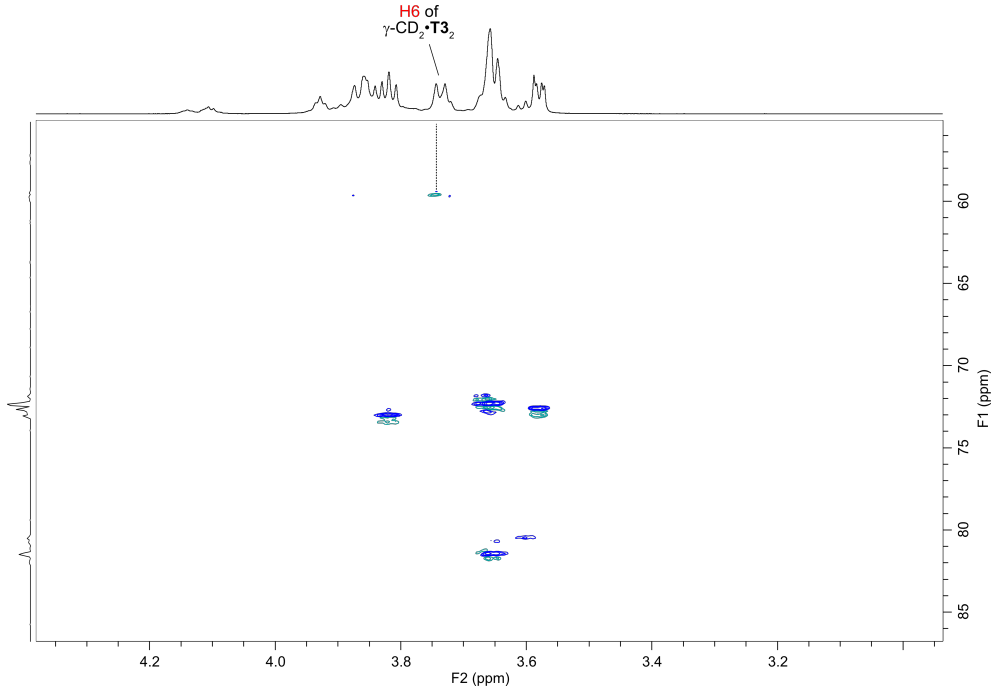


Figure S3.9. Partial multiplicity-edited ^{1}H-^{13}C HSQC spectrum (800 MHz) of **T3** (5 mM) and γ-CD (5 mM) in D_2O phosphate buffer (pH 7.5, 100 mM) with assignment of the H6 protons of γ-CD in the γ-CD_2•**T3**$_2$ complex.

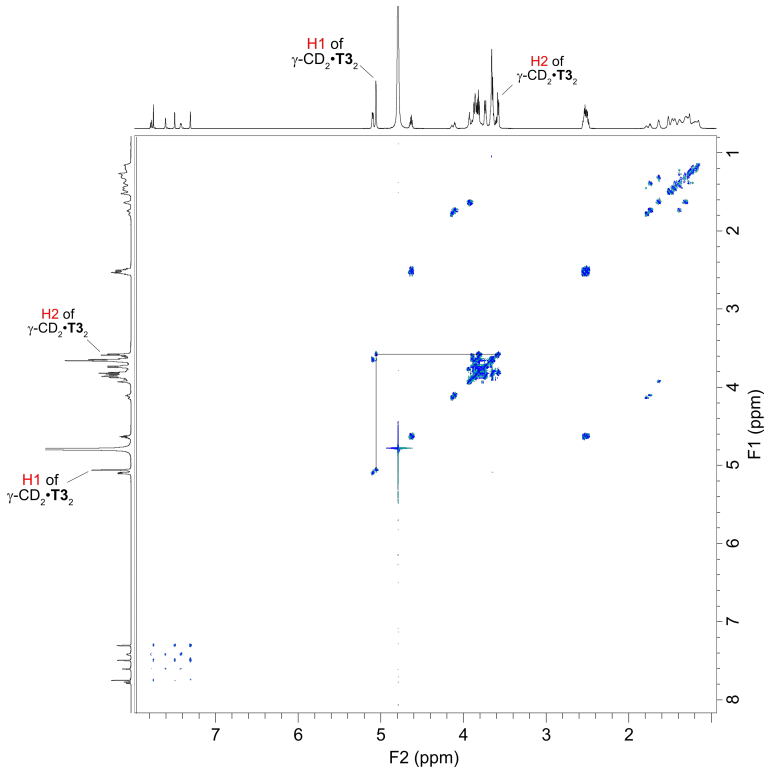


Figure 3.10. ^{1}H-^{1}H DQF-COSY spectrum (800 MHz) of **T3** (5 mM) and γ-CD (5 mM) in D_2O phosphate buffer (pH 7.5, 100 mM) with assignment of protons (H1, H2) of γ-CD in the γ-CD_2•**T3**$_2$ complex.

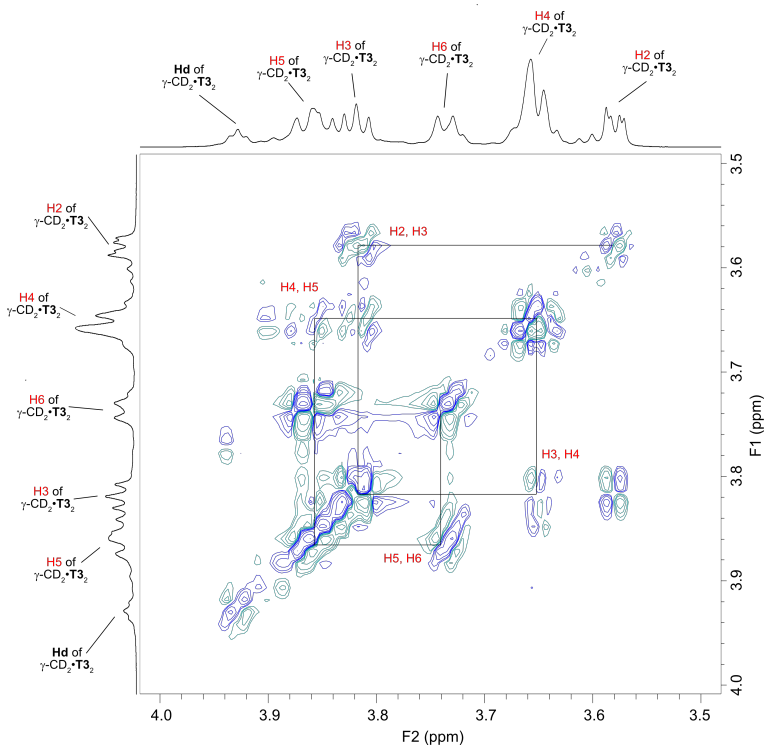


Figure 3.11. Partial ^{1}H-^{1}H DQF-COSY spectrum (800 MHz) of **T3** (5 mM) and γ-CD (5 mM) in D_2O phosphate buffer (pH 7.5, 100 mM) with assignment of protons (H1–H6) of γ-CD in the γ-CD_2•**T3**$_2$ complex.

γ-CD and **T4**

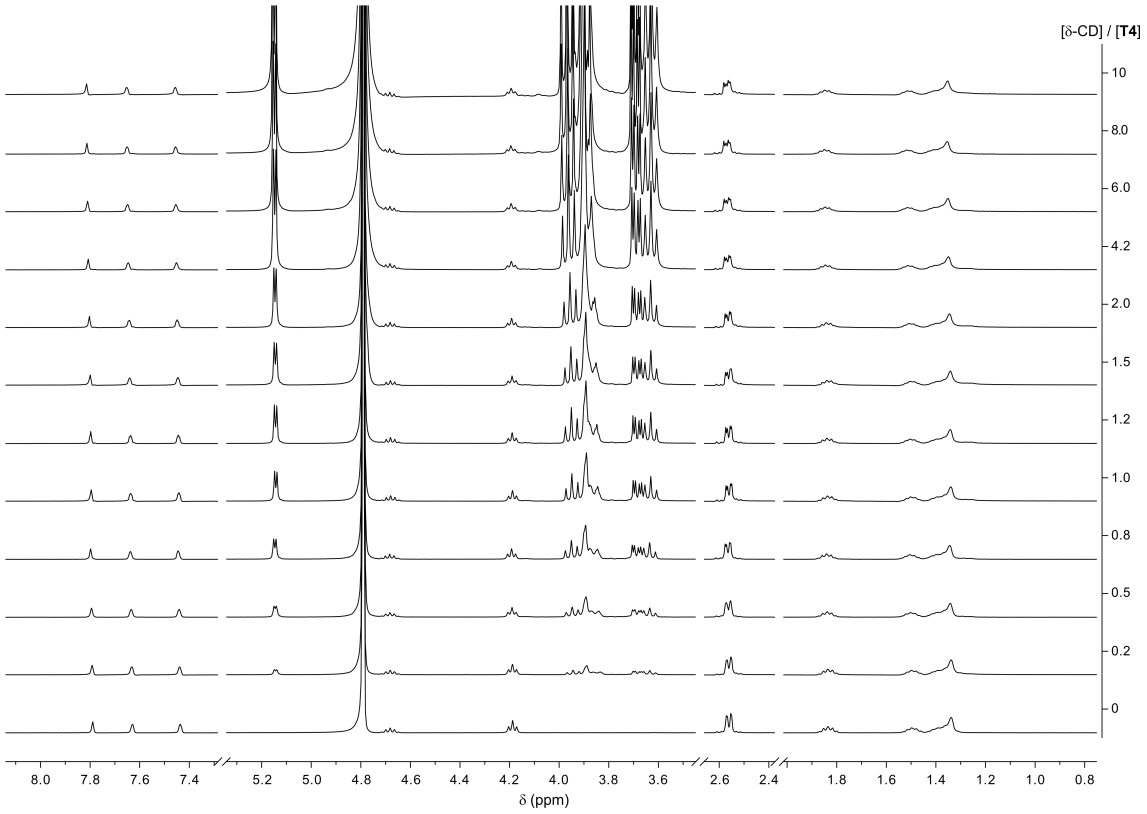


Figure S3.12. ^{1}H NMR spectra (400 MHz) of the titration of **T4** with γ-CD performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T4** at 1.0 mM.

S3.1.6. NMR titrations with δ-CD

δ-CD and **T1**

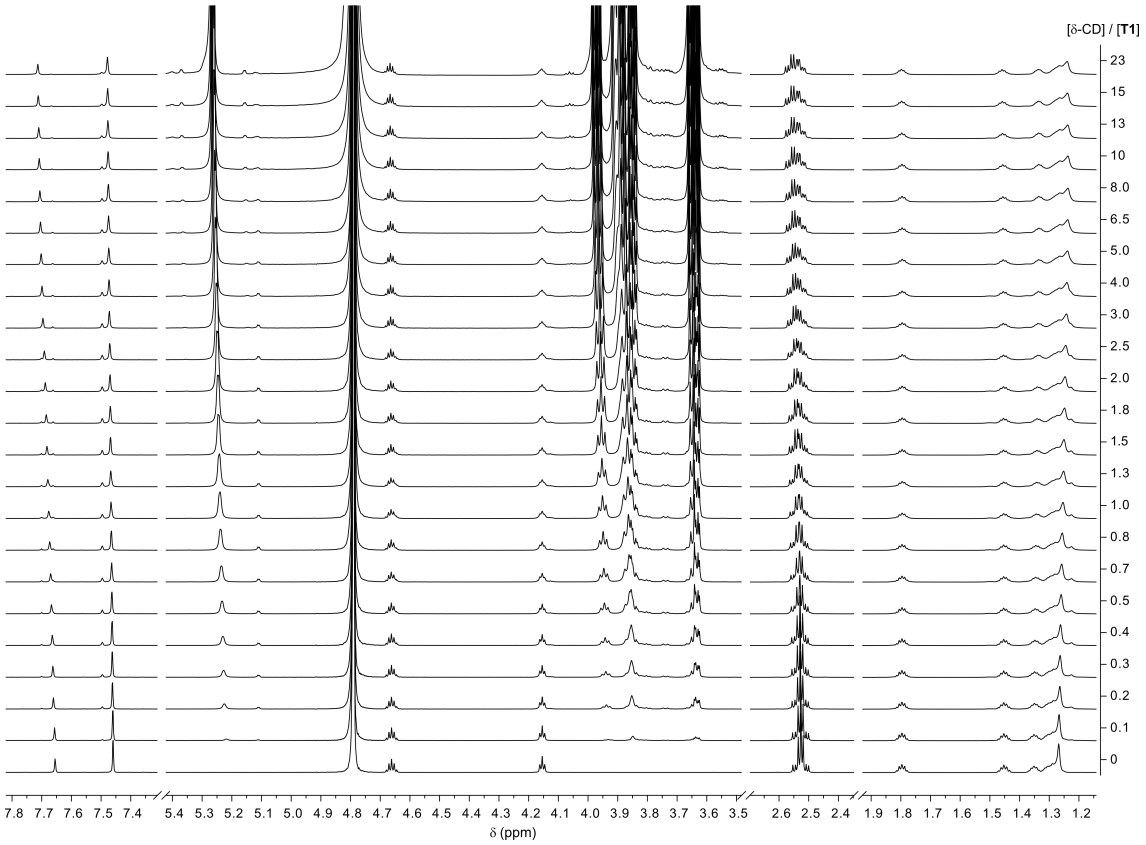


Figure S3.13. ^{1}H NMR spectroscopy (800 MHz) titration with **T1** and δ-CD performed in sodium phosphate buffer (100 mM) at pH 7.5 in D_2O with a constant concentration of **T1** at 1.0 mM.

δ-CD and **T1**

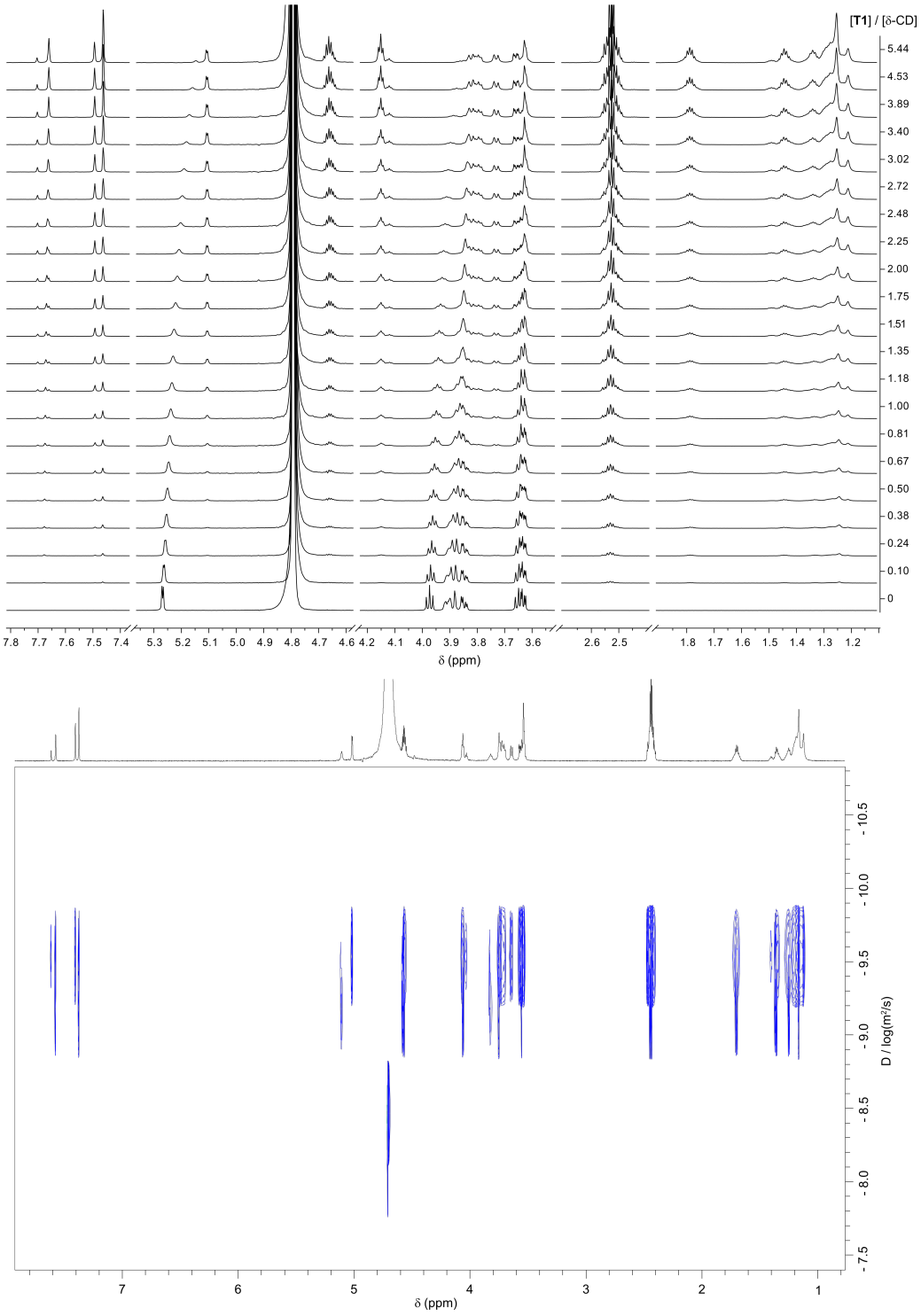


Figure S3.14. ^{1}H NMR spectroscopy titration of δ-CD with **T1** performed in sodium phosphate buffer (100 mM) at pH 7.0 in D_2O with a constant concentration of δ-CD at 1.0 mM. *Top:* ^{1}H NMR spectra (800 MHz) of the titration of δ-CD with **T1**. *Bottom:* DOSY spectrum (800 MHz) of the data point with 2.72 mM **T1**.

δ-CD and **T2**

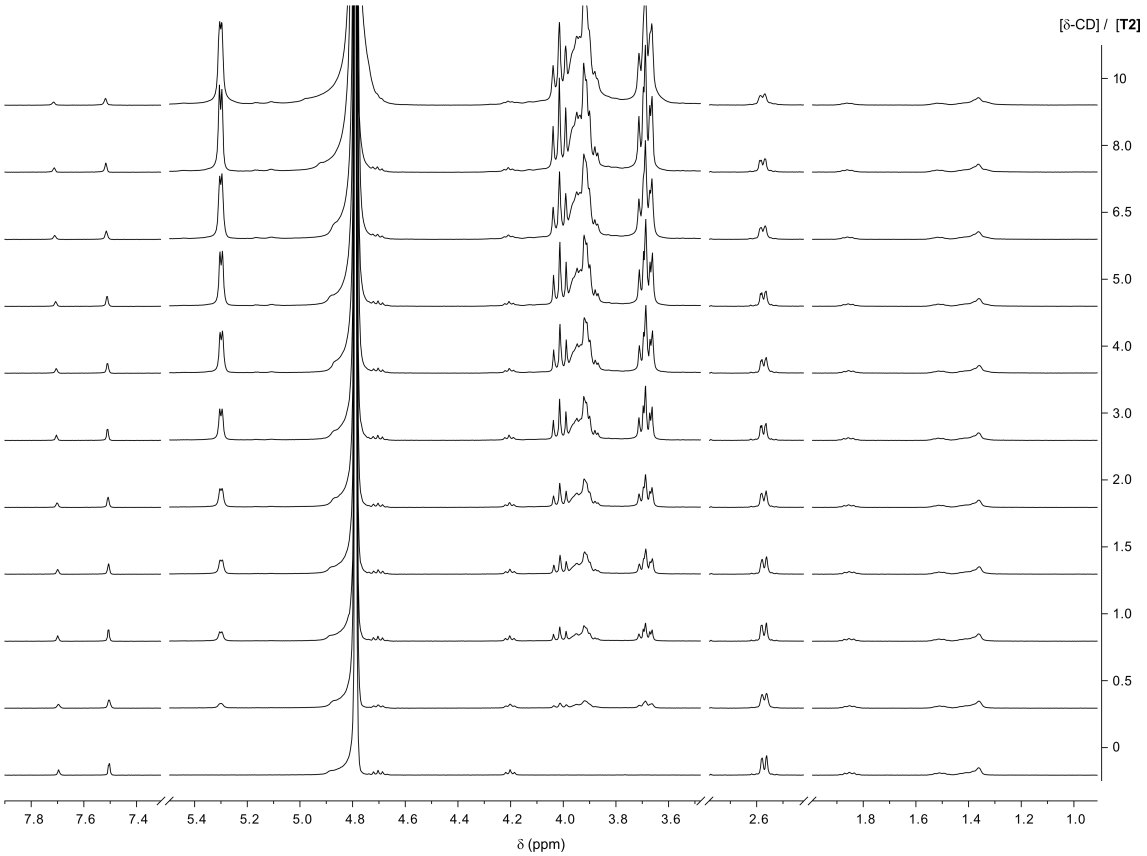


Figure S3.15. ^{1}H NMR spectra (400 MHz) of the titration of **T2** with δ-CD performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T2** at 1.0 mM.

δ-CD and **T2**

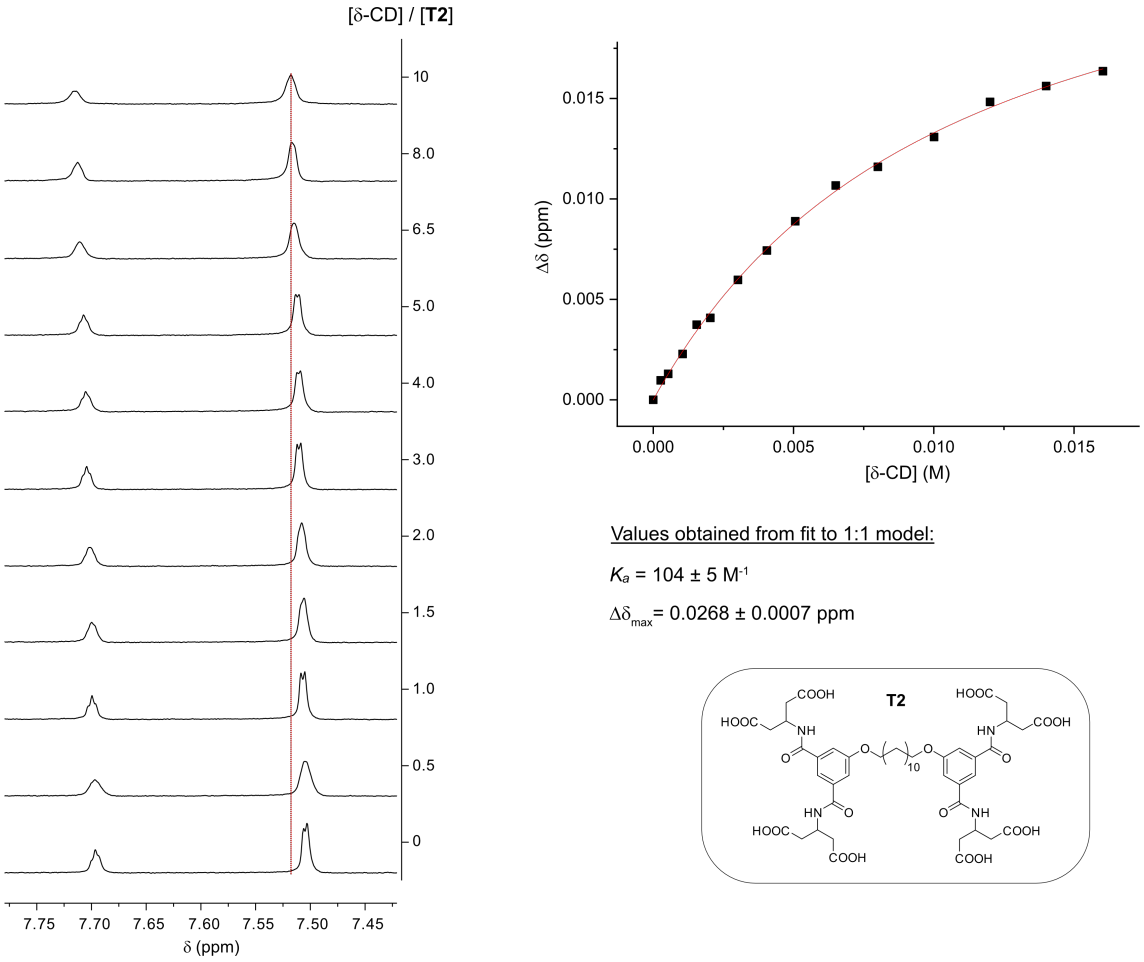


Figure S3.16. ^{1}H NMR spectroscopy titration of **T2** with δ-CD shows 1:1 binding in fast exchange with a fitted binding constant of K_a = 104 ± 5 M^{-1}. The titration was performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T2** at 1.0 mM. *Left*: Partial ^{1}H NMR spectra (400 MHz) of **T2** with increasing concentrations of δ-CD. *Right:* The change in chemical shift of the proton at 7.5 ppm and the resulting fit to a 1:1 binding model.

δ-CD and **T3**

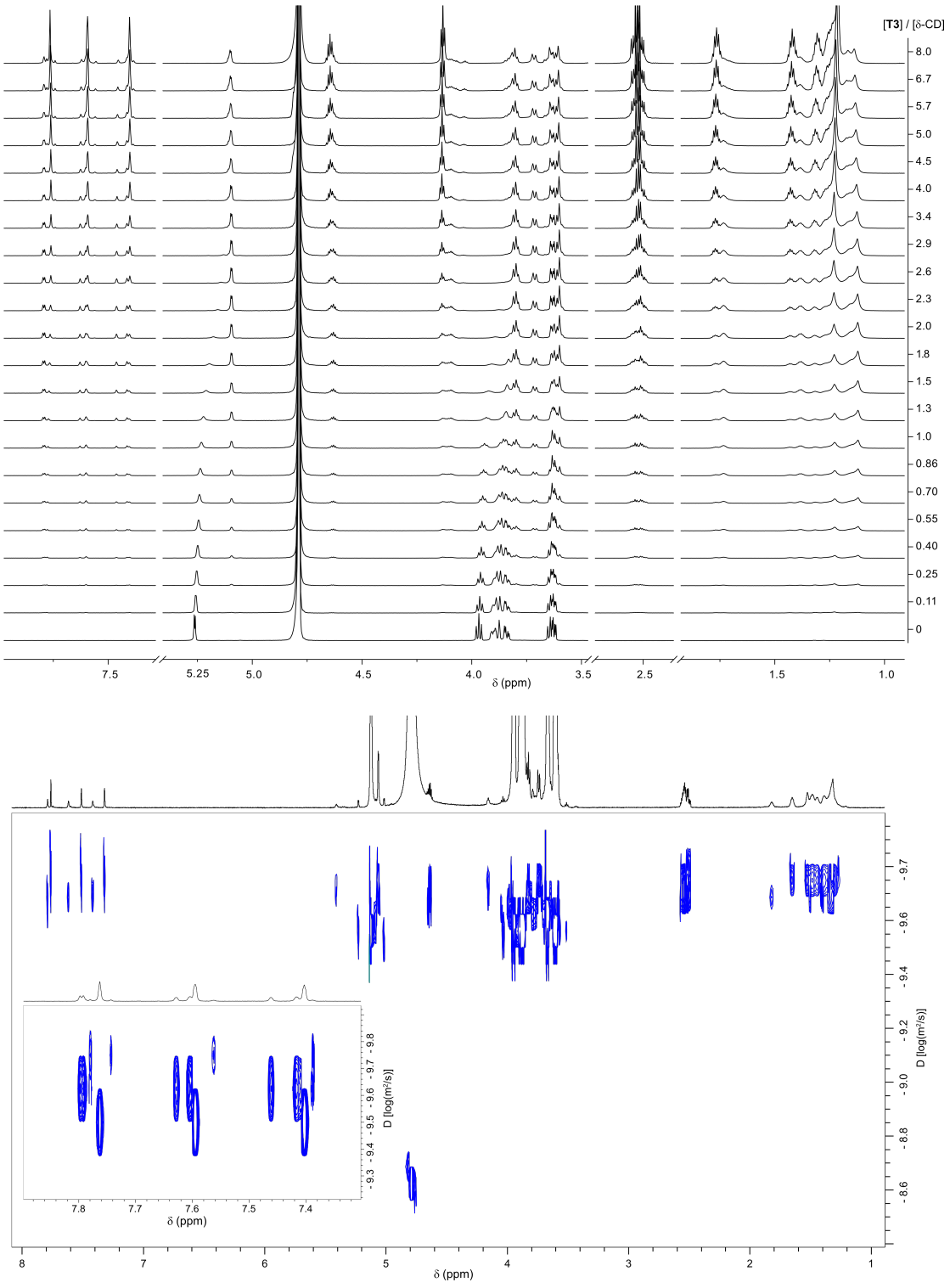


Figure S3.17. ^{1}H NMR spectroscopy titration of δ-CD with **T3** performed in sodium phosphate buffer (100 mM) at pH 7.5 in D_2O with a constant concentration of T3 at 1.0 mM. *Top:* ^{1}H NMR spectra (800 MHz) of the titration of **T3** with γ-CD. *Bottom:* DOSY spectrum (800 MHz) of the data point with 3.4 mM δ-CD.

δ-CD and **T4**

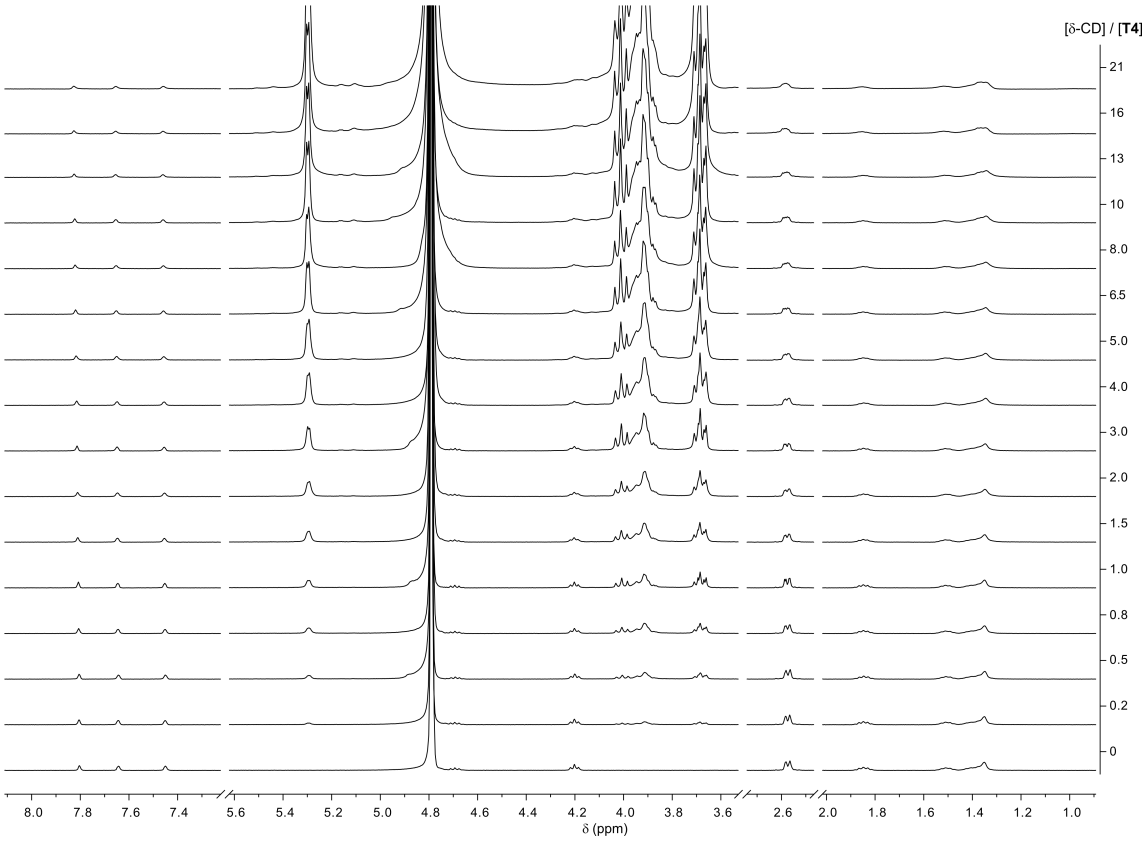


Figure S3.18. ^{1}H NMR spectra (400 MHz) of the titration of **T4** with δ-CD performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T4** at 1.0 mM.

δ-CD and **T4**

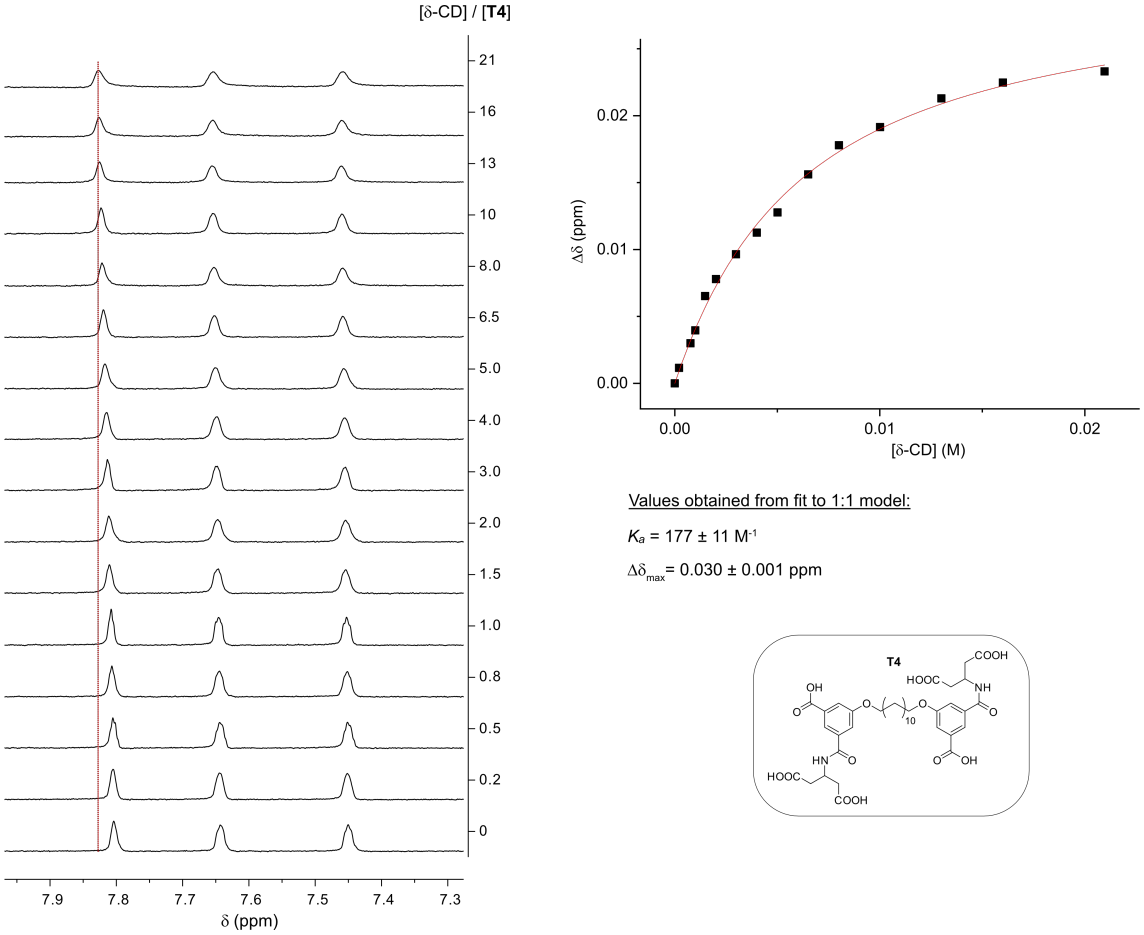


Figure S3.19. ^{1}H NMR spectroscopy titration of **T4** with δ-CD shows 1:1 binding in fast exchange with a fitted binding constant of K_a = 177 ± 11 M^{-1}. *Left*: Partial ^{1}H NMR (400 MHz) spectra of **T4** with increasing concentrations of δ-CD. The titration was performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O with a constant concentration of **T4** at 1.0 mM. *Right:* The change in chemical shift of the proton at 7.8 ppm and the resulting fit to a 1:1 binding model.

S3.2. NMR spectra of synthesized compounds

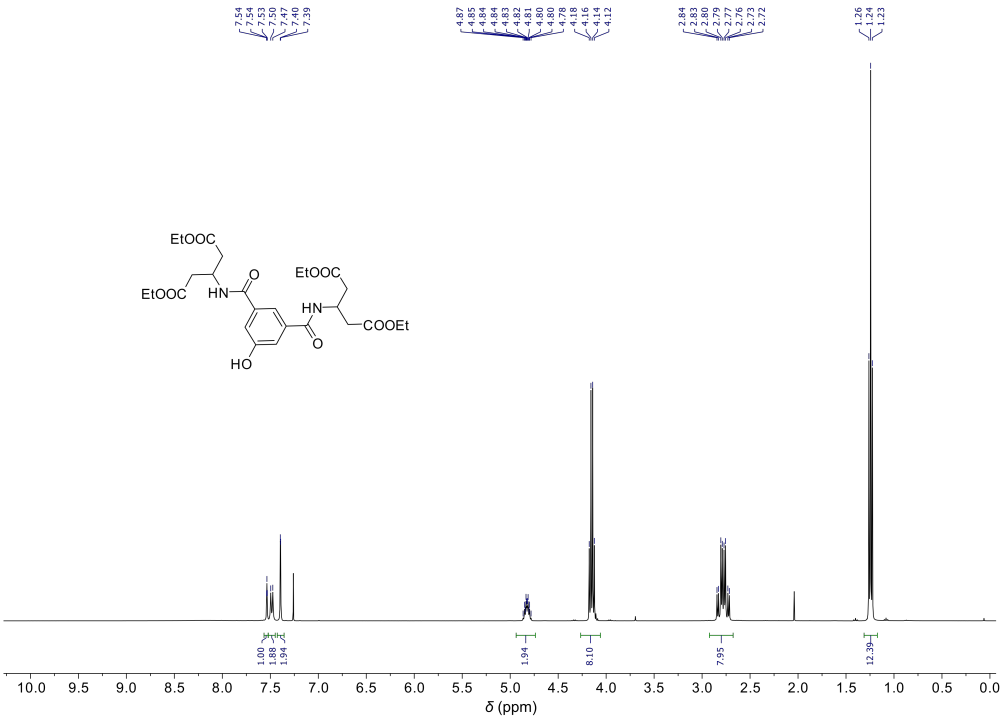


Figure S3.20. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of **3.4**.

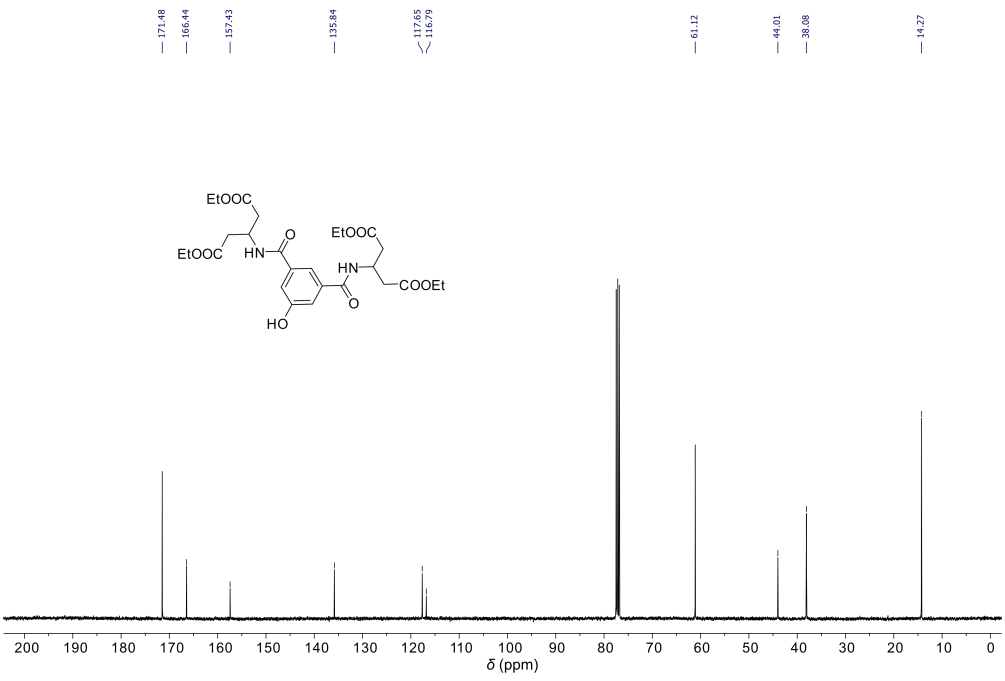


Figure S3.21. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of **3.4**.

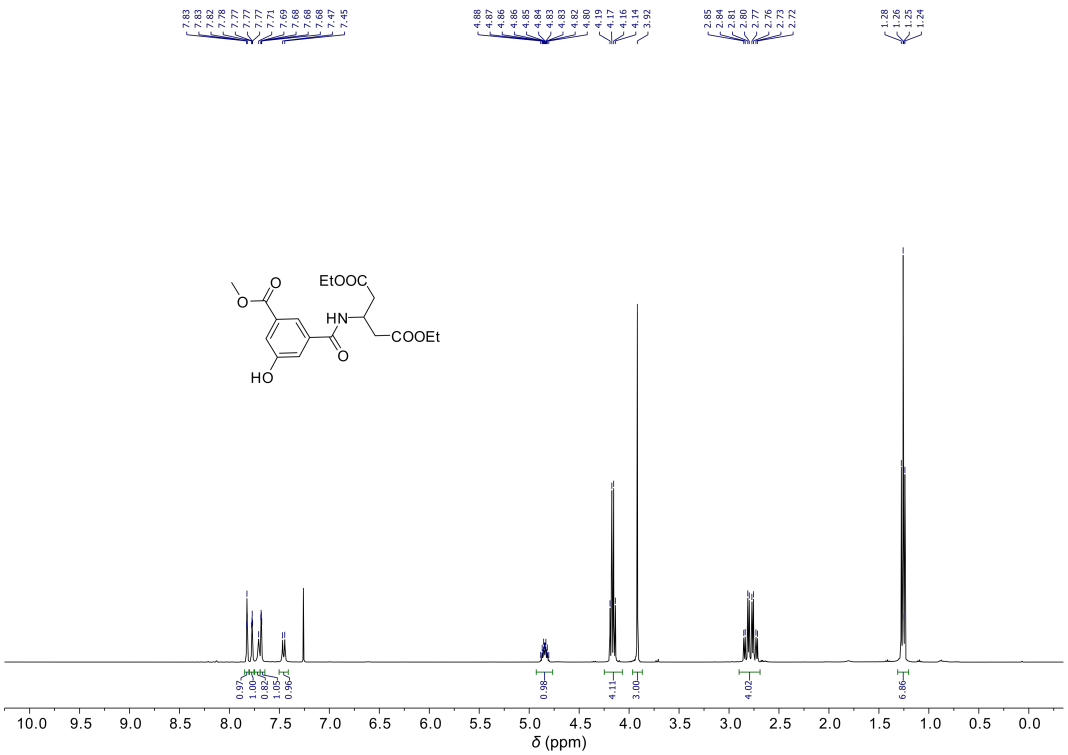


Figure S3.22. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of **3.7**.

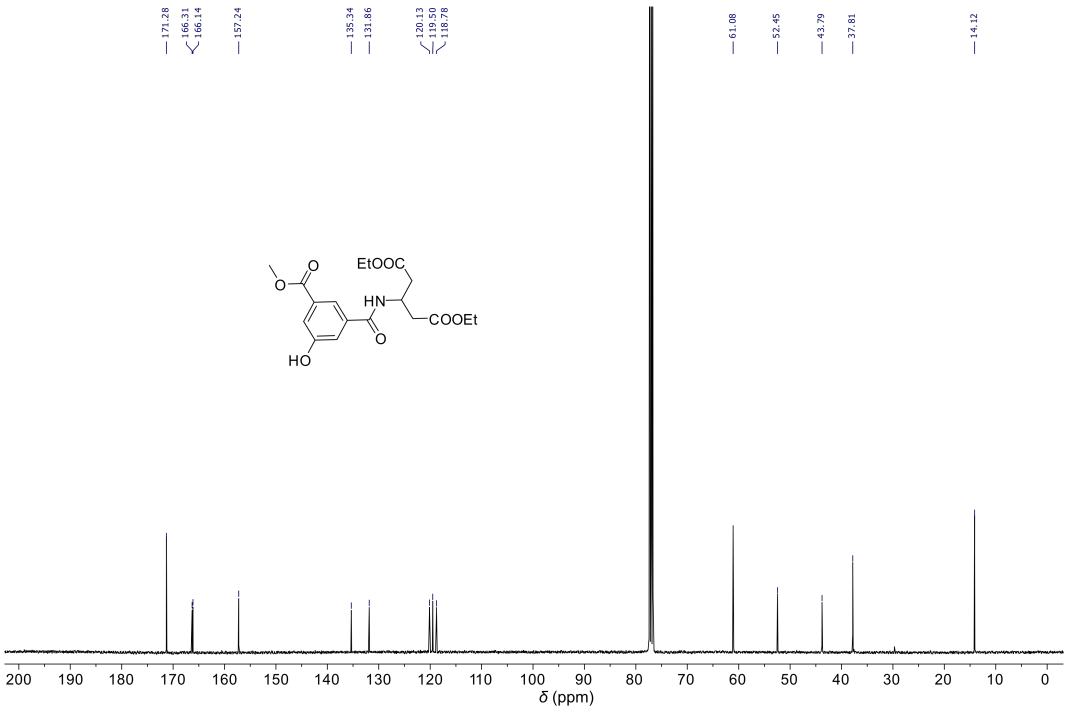


Figure S3.23. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of **3.7**.

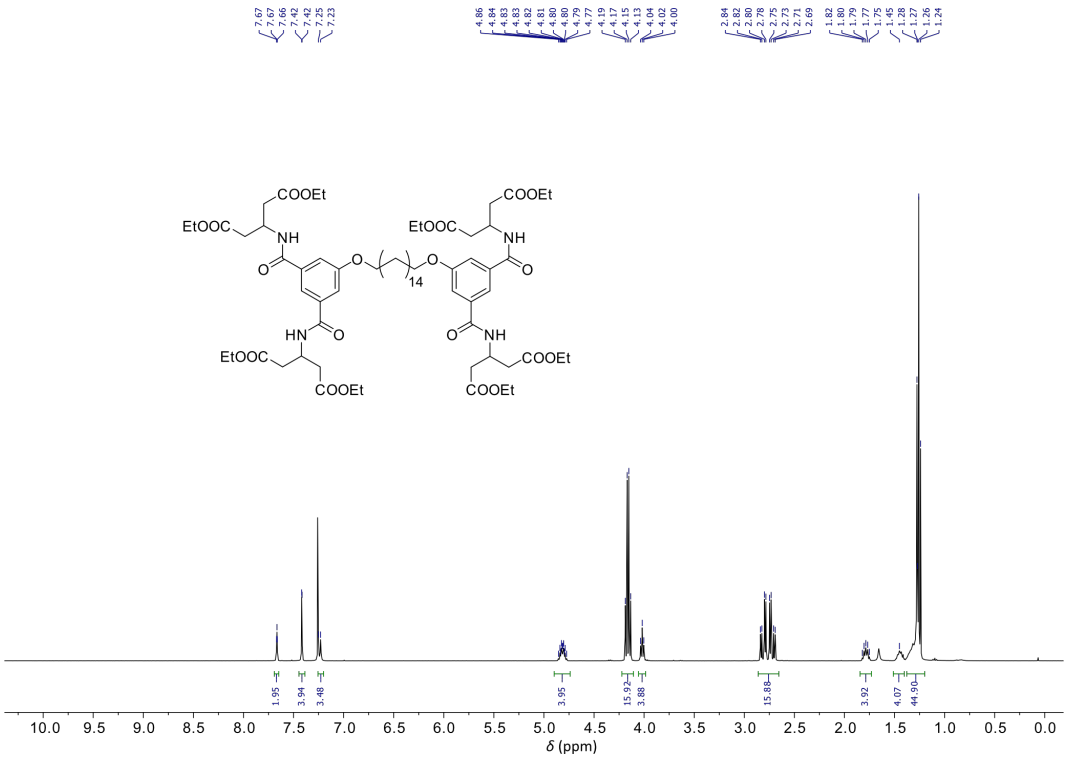


Figure S3.24. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of **3.9**.

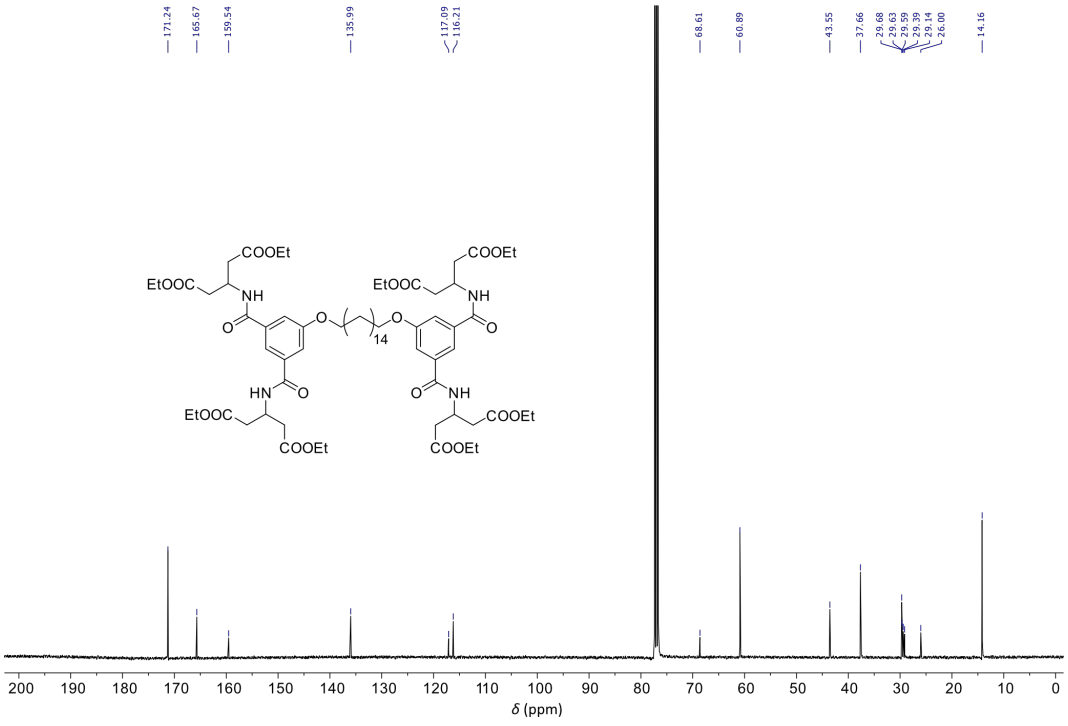


Figure S3.25. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of **3.9**.

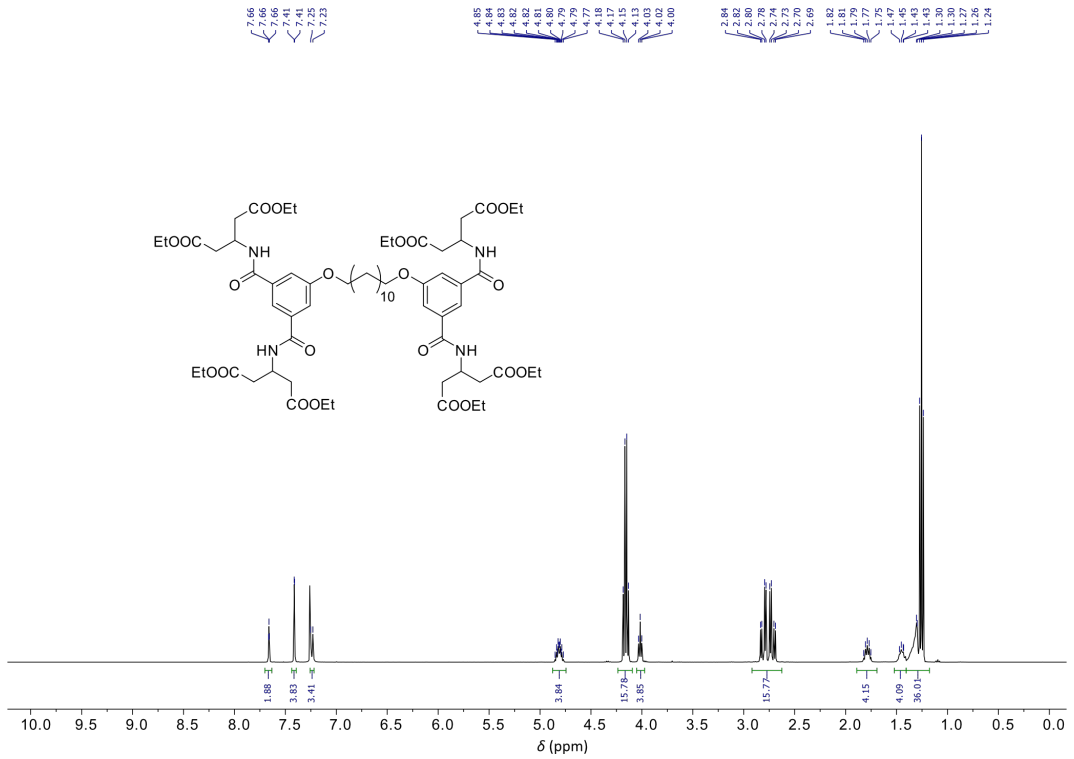


Figure S3.26. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of **3.10**.

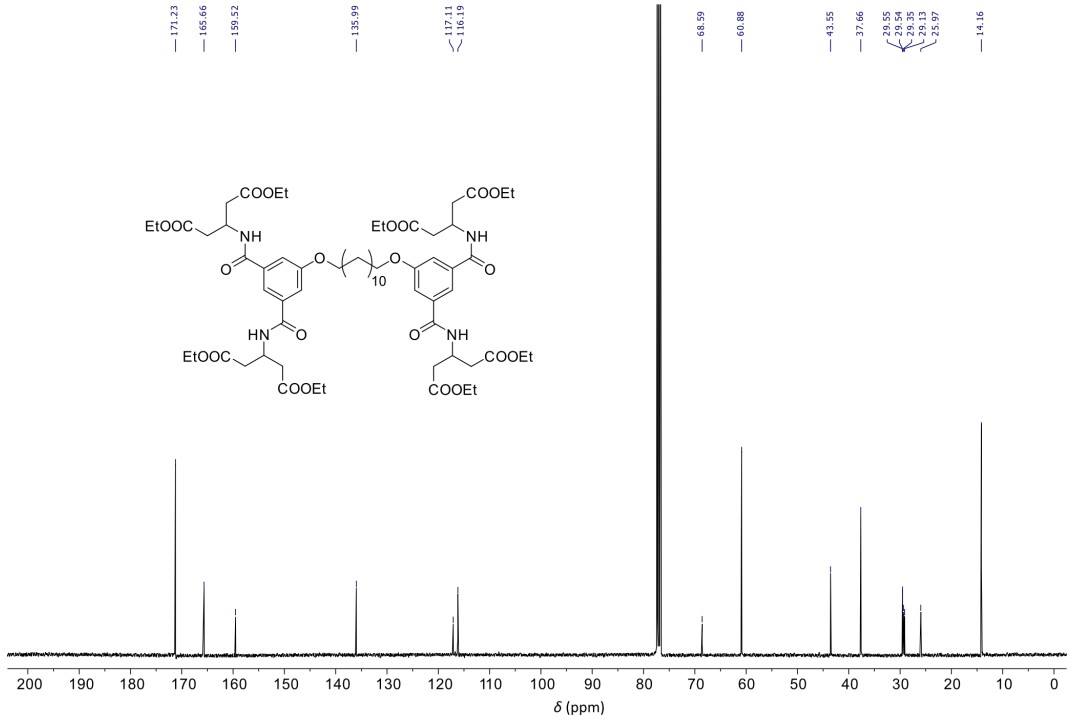


Figure S3.27. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of **3.10**.

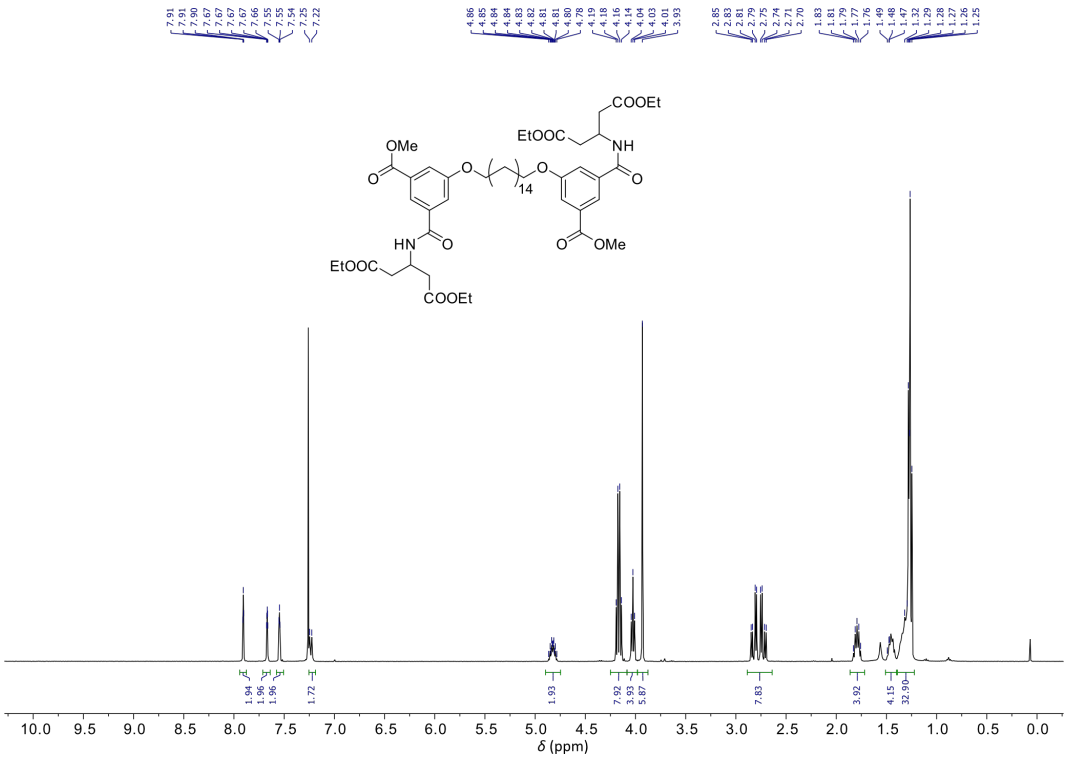


Figure S3.28. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of **3.11**.

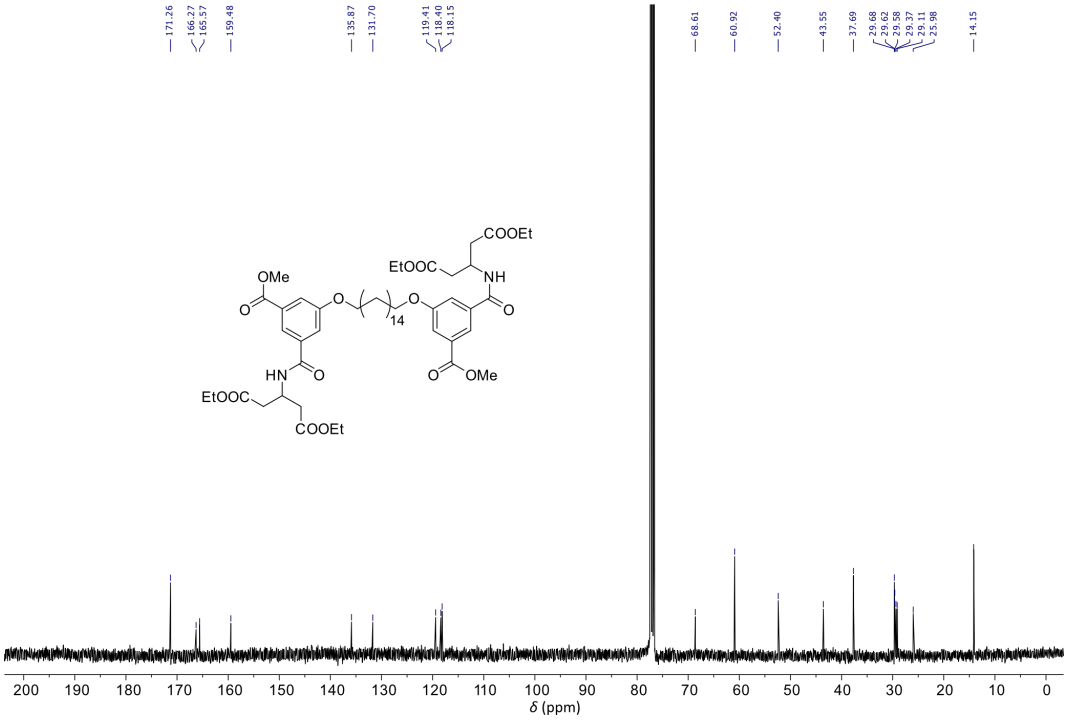


Figure S3.29. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of **3.11**.

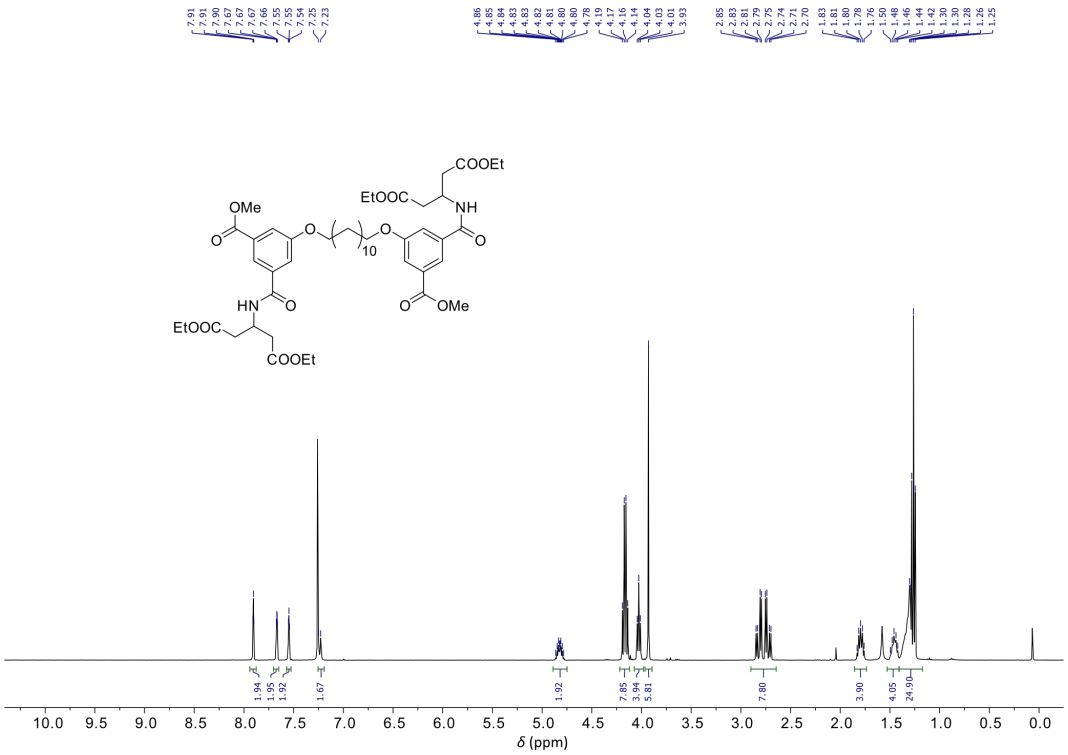


Figure S3.30. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of **3.12**.

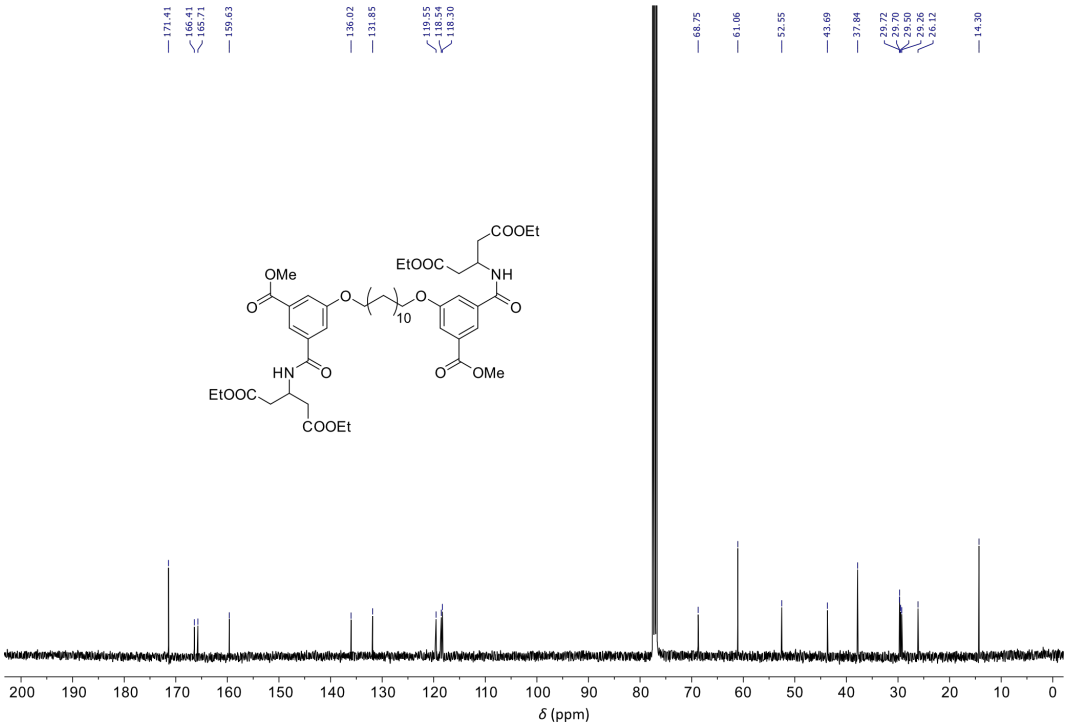


Figure S3.31. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of **3.12**.

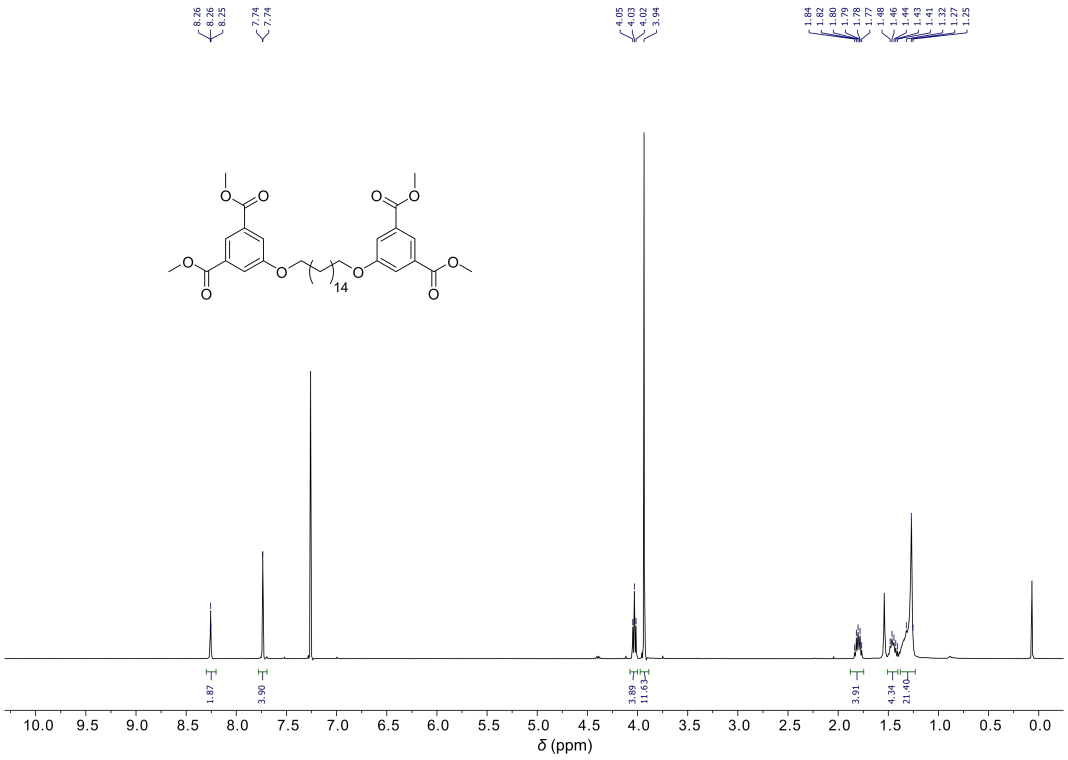


Figure S3.32. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of **3.13**.

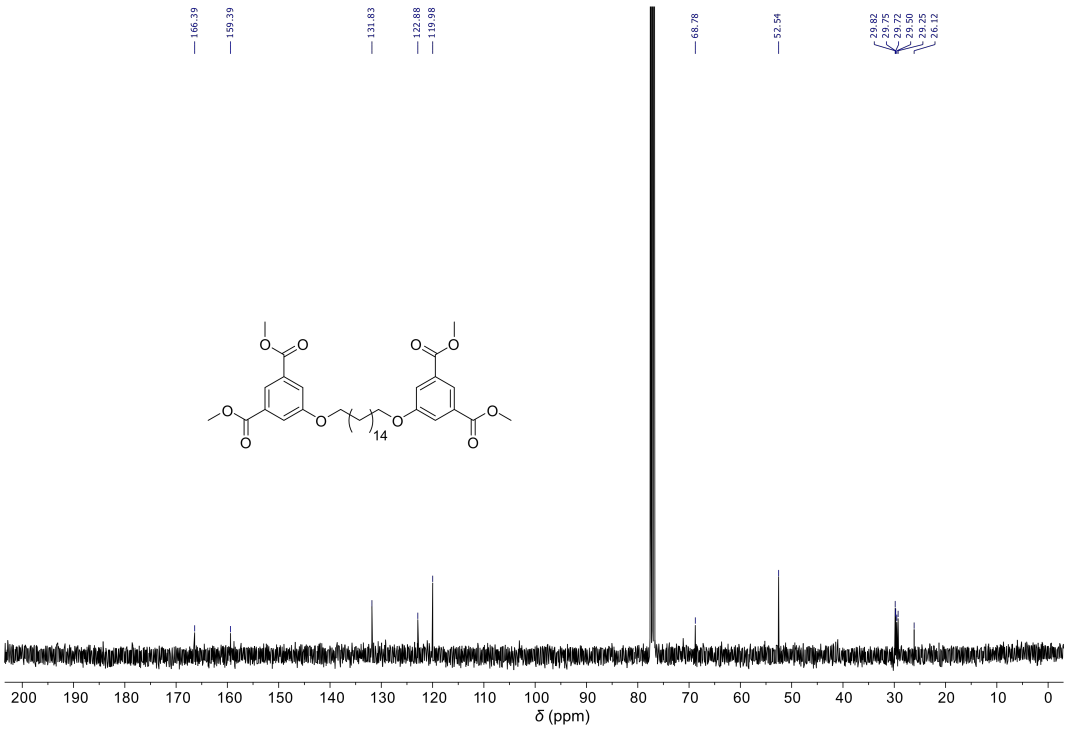


Figure S3.33. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of **3.13**.

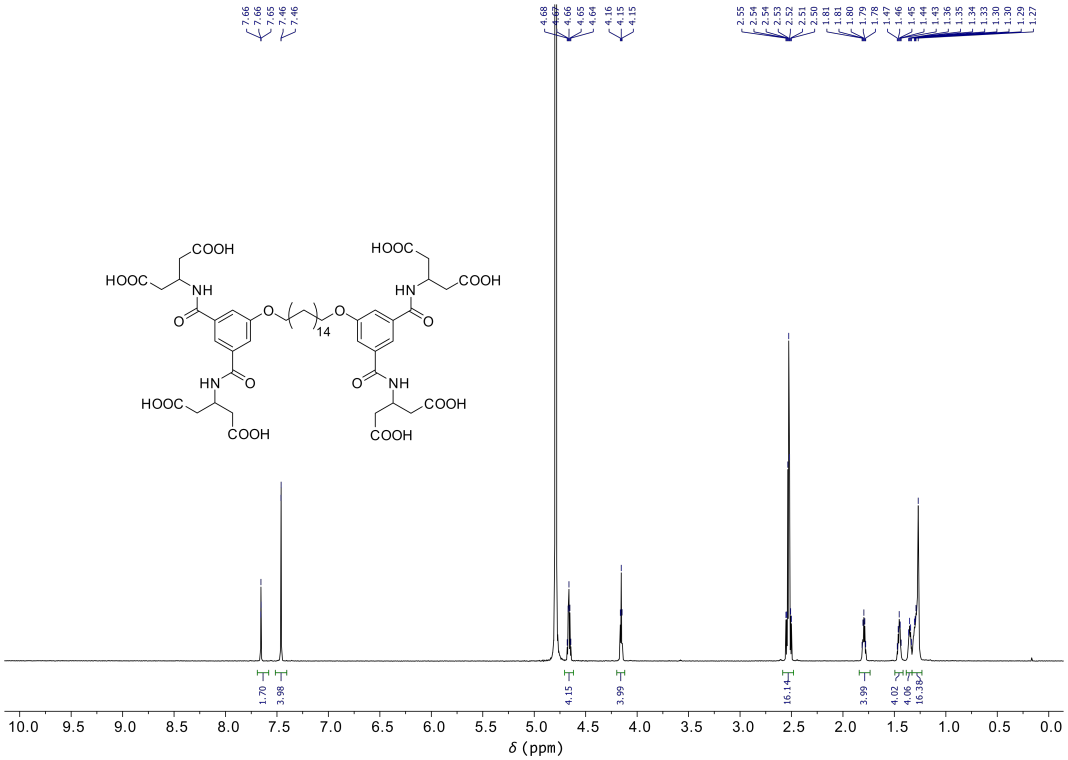


Figure S3.34. ^{1}H NMR (800 MHz) spectrum of **T1** in D_2O phosphate buffer, 50 mM, pH 7.5.

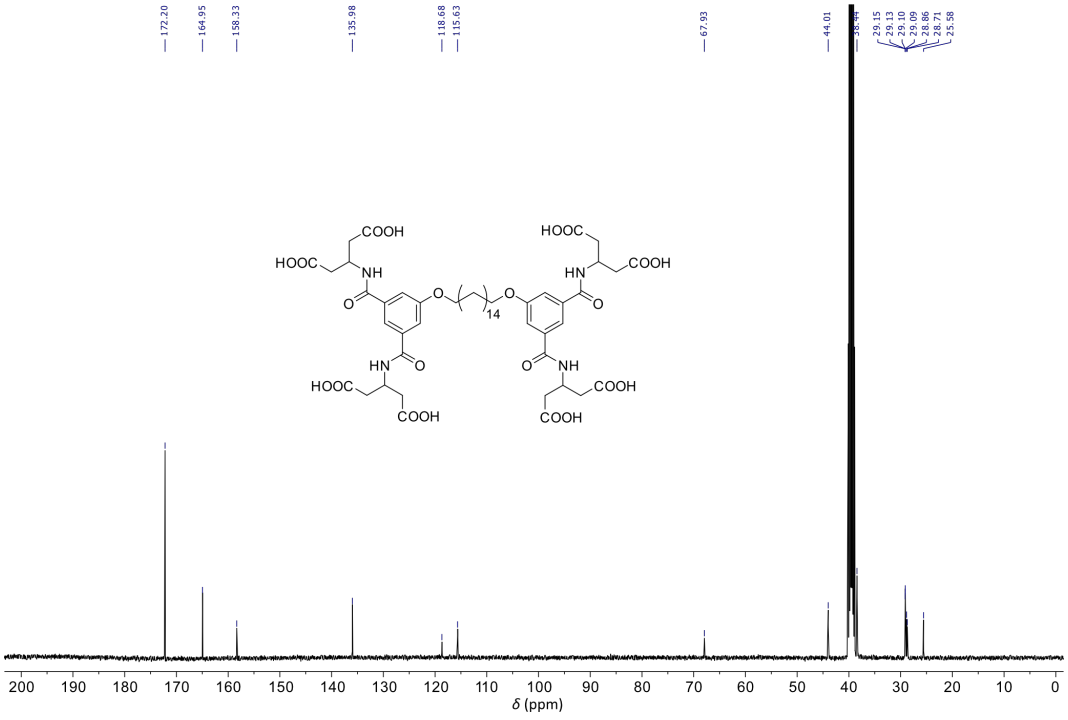


Figure S3.35. ^{13}C NMR (101 MHz, DMSO-*d6*) spectrum of **T1**.

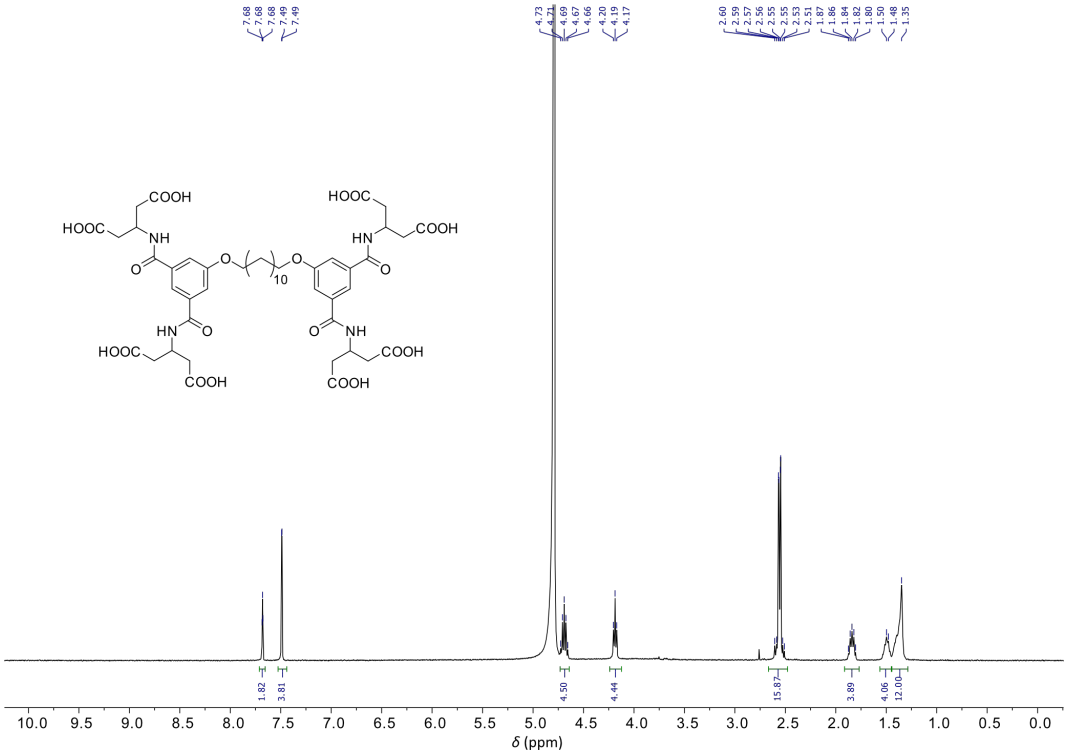


Figure S3.36. ^{1}H NMR (400 MHz) spectrum of **T2** in D_2O phosphate buffer, 50 mM, pH 7.5.

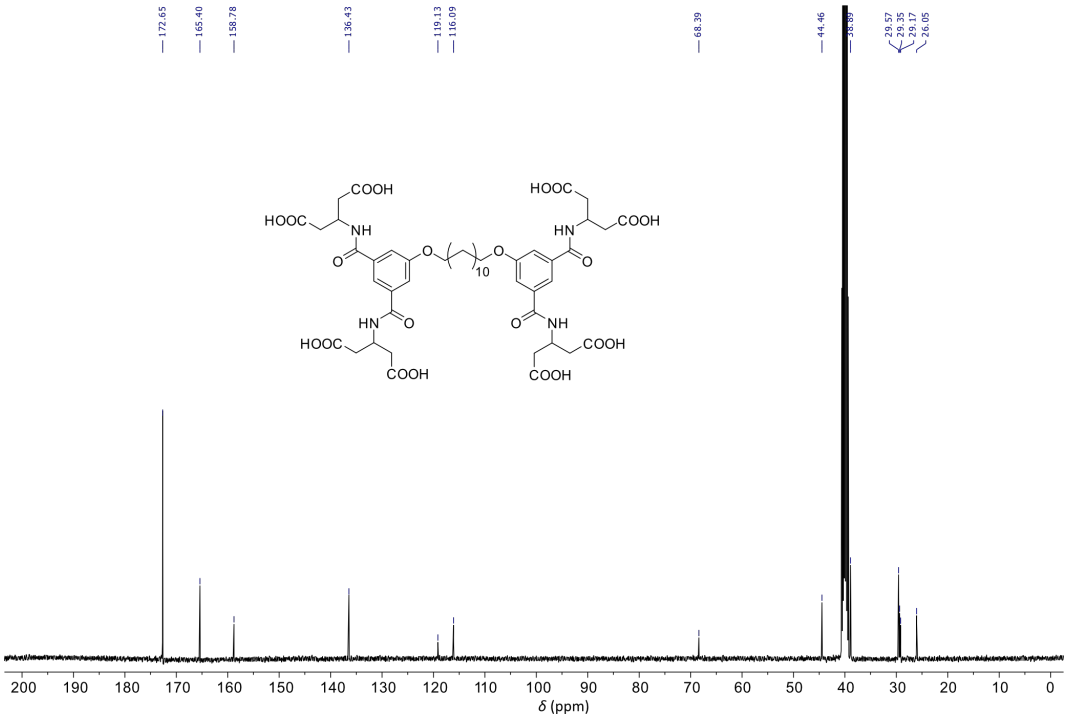


Figure S3.37. ^{13}C NMR (101 MHz, DMSO-*d6*) spectrum of **T2**.

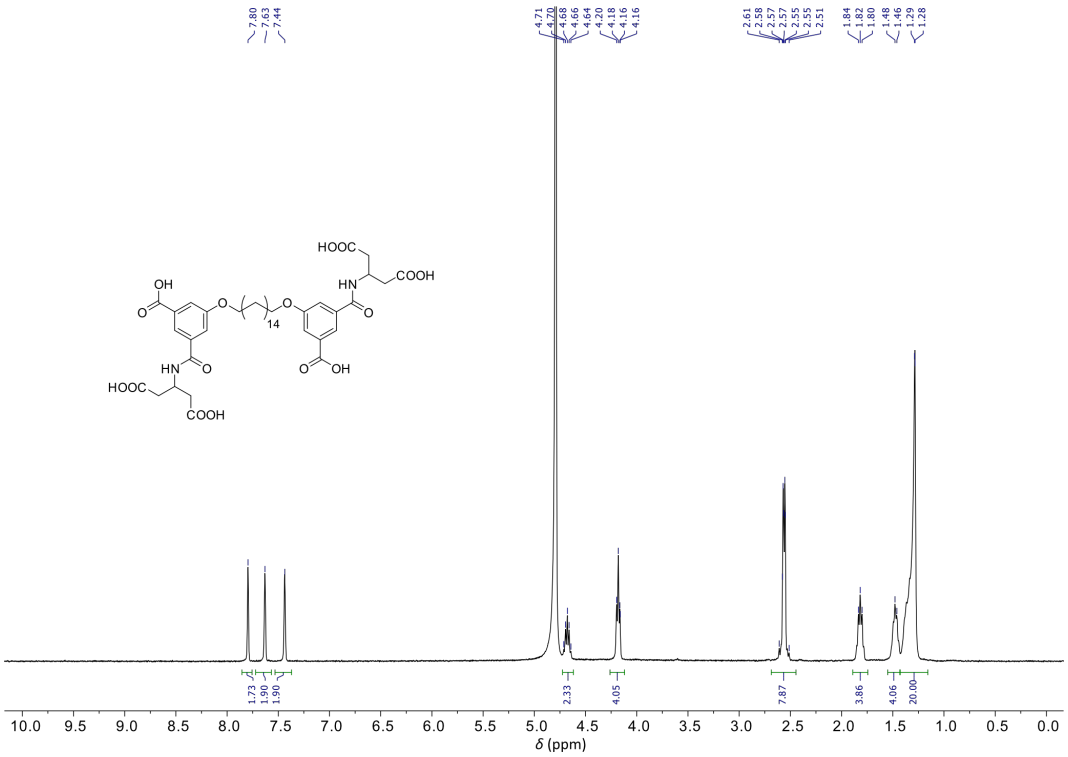


Figure S3.38. ^{1}H NMR (400 MHz) spectrum of **T3** in D_2O phosphate buffer, 50 mM, pH 7.5.

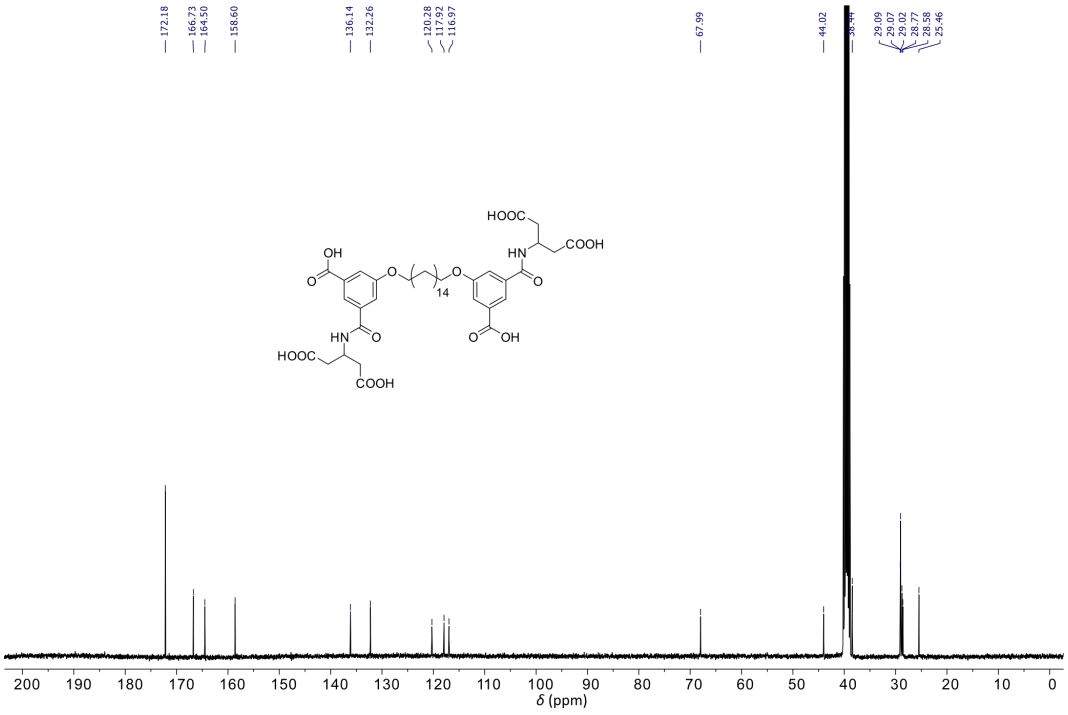


Figure S3.39. ^{13}C NMR (101 MHz, DMSO-*d6*) spectrum of **T3**.

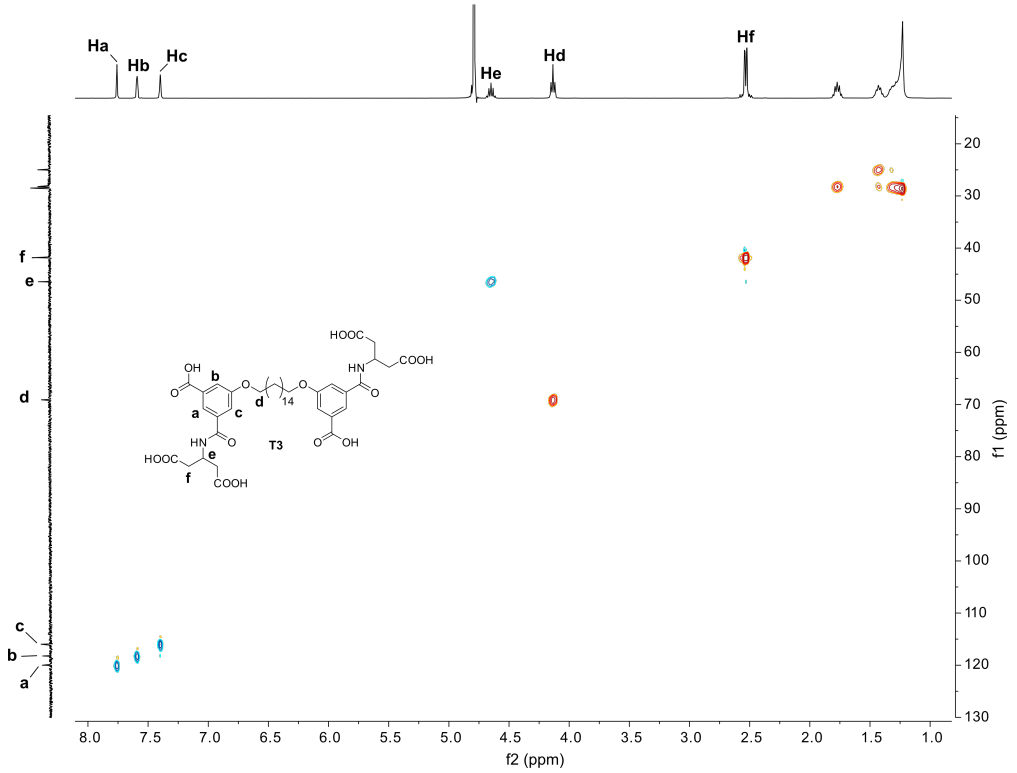


Figure 3.40. Multiplicity-edited ^{1}H-^{13}C HSQC (400 / 101 MHz) spectrum of **T3** in D_2O sodium phosphate buffer (50 mM, pH 7.5).

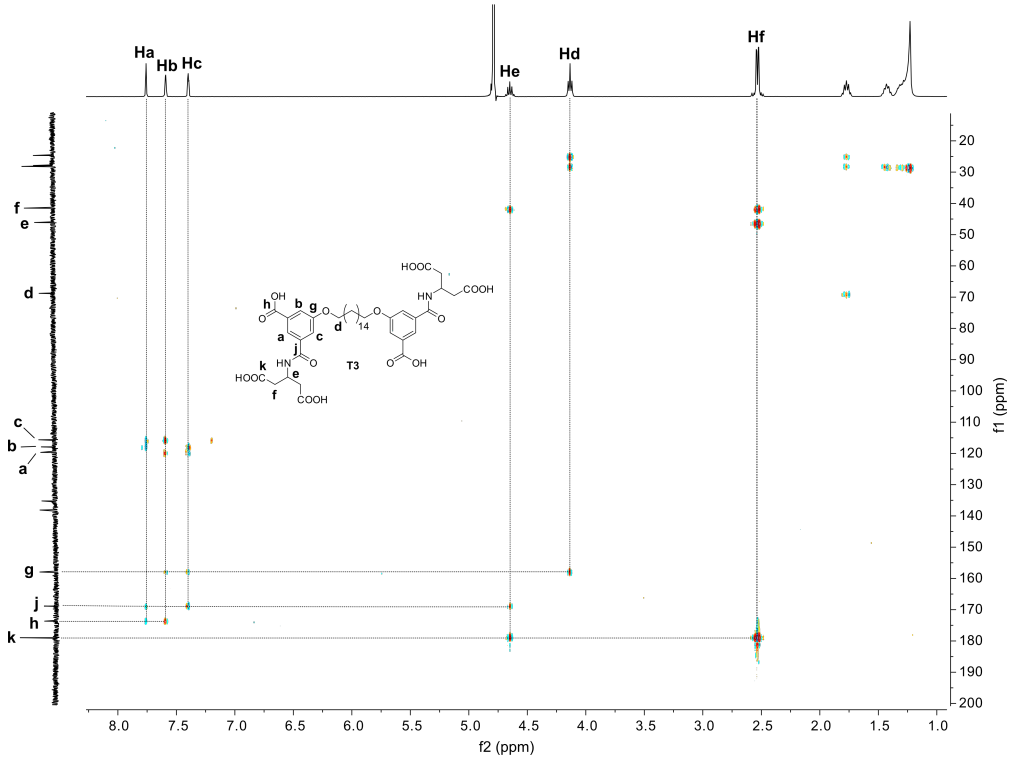


Figure 3.41. ^{1}H-^{13}C HMBC (400 / 101 MHz) spectrum of **T3** in D_2O sodium phosphate buffer (50 mM, pH 7.5).

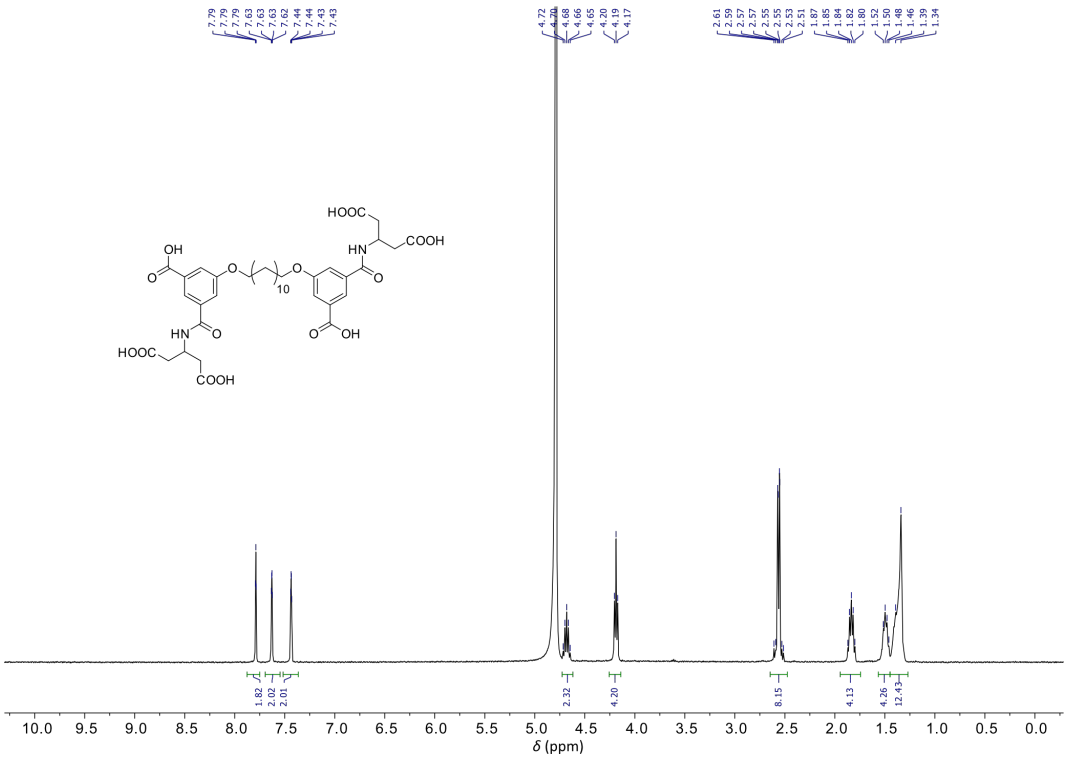


Figure S3.42. ^{1}H NMR (400 MHz) spectrum of **T4** in D_2O phosphate buffer, 50 mM, pH 7.5.

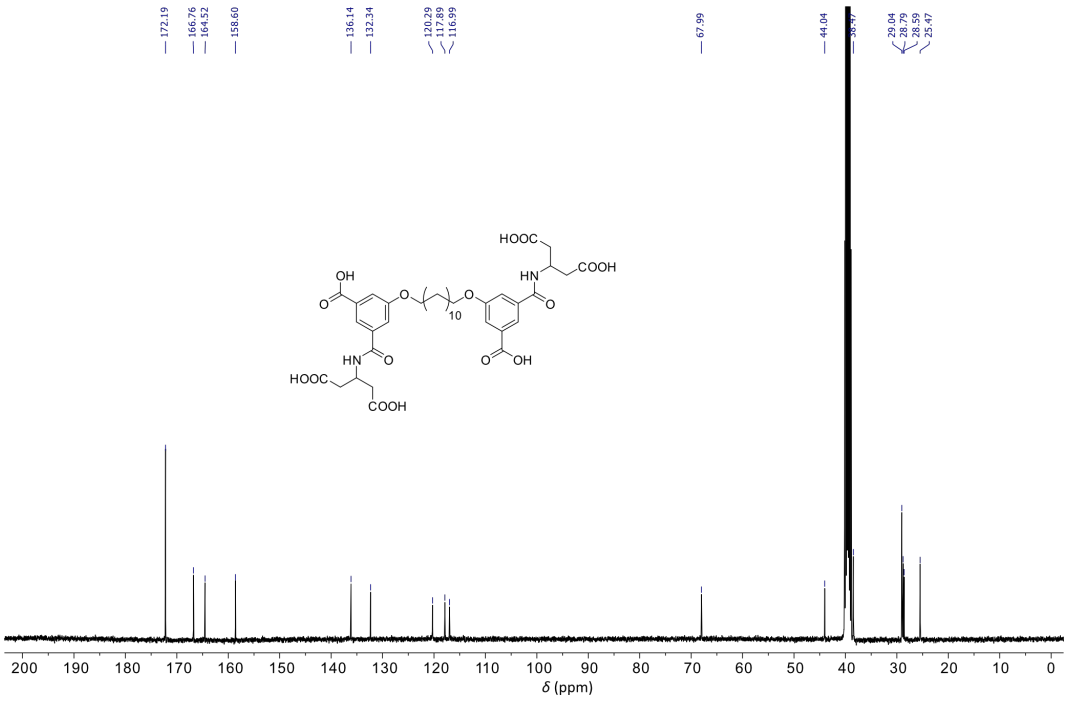


Figure S3.43. ^{13}C NMR (101 MHz, DMSO-*d6*) spectrum of **T4**.

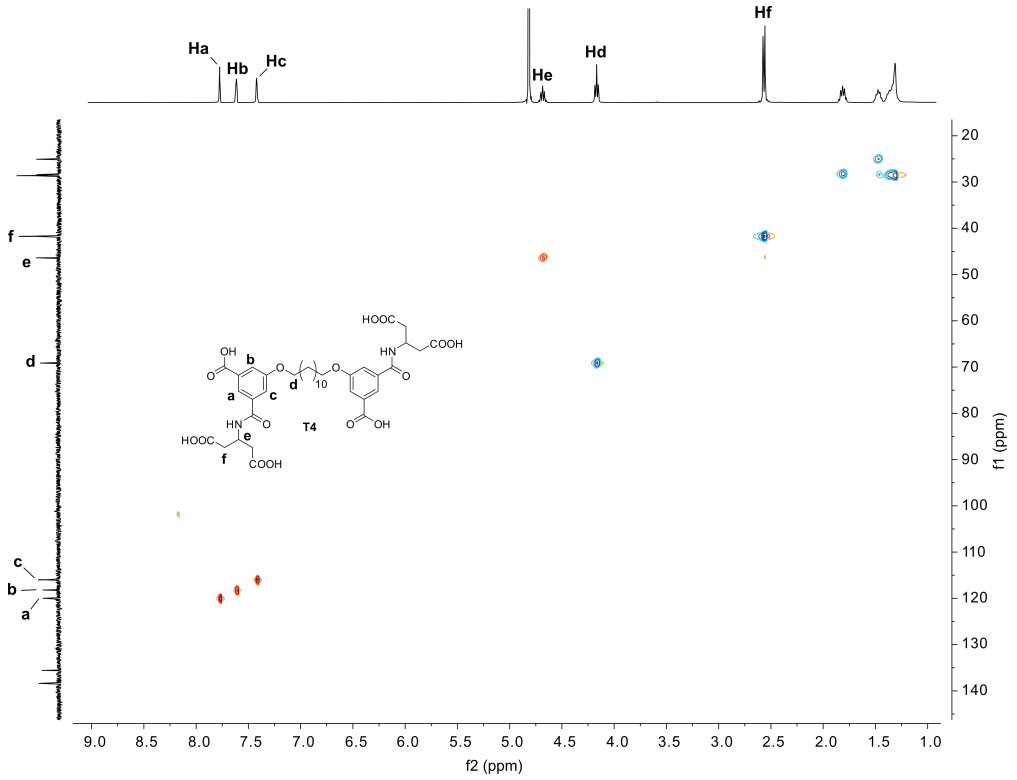


Figure 3.44. Multiplicity-edited ^{1}H-^{13}C HSQC (400 / 101 MHz) spectrum of **T4** in D_2O sodium phosphate buffer (50 mM, pH 7.5).

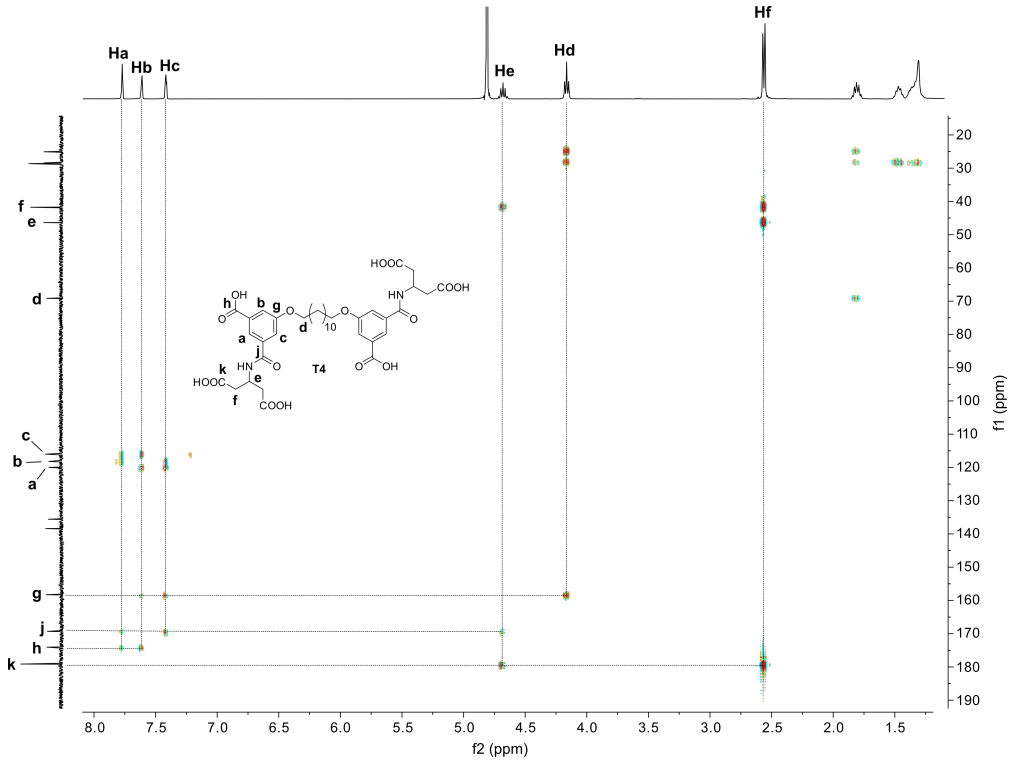


Figure 3.45. ^{1}H-^{13}C HMBC (400 / 101 MHz) spectrum of **T4** in D_2O sodium phosphate buffer (50 mM, pH 7.5).

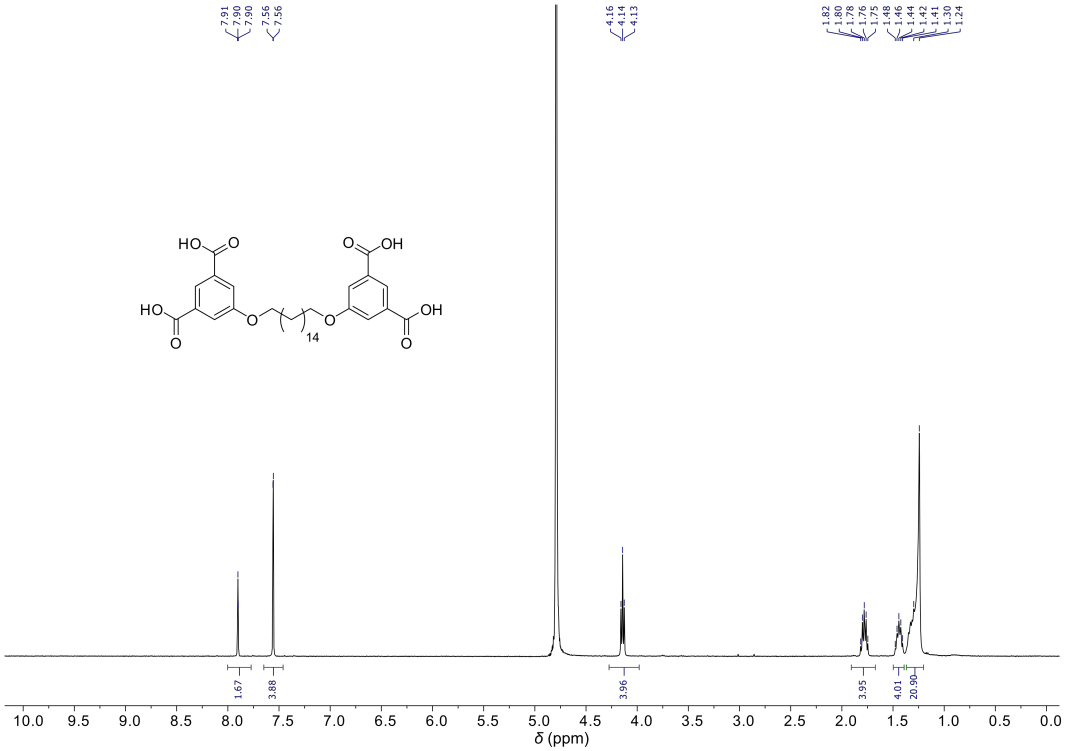


Figure S3.46. ^{1}H NMR (400 MHz) spectrum of **T5** in D_2O phosphate buffer, 50 mM, pH 7.5.

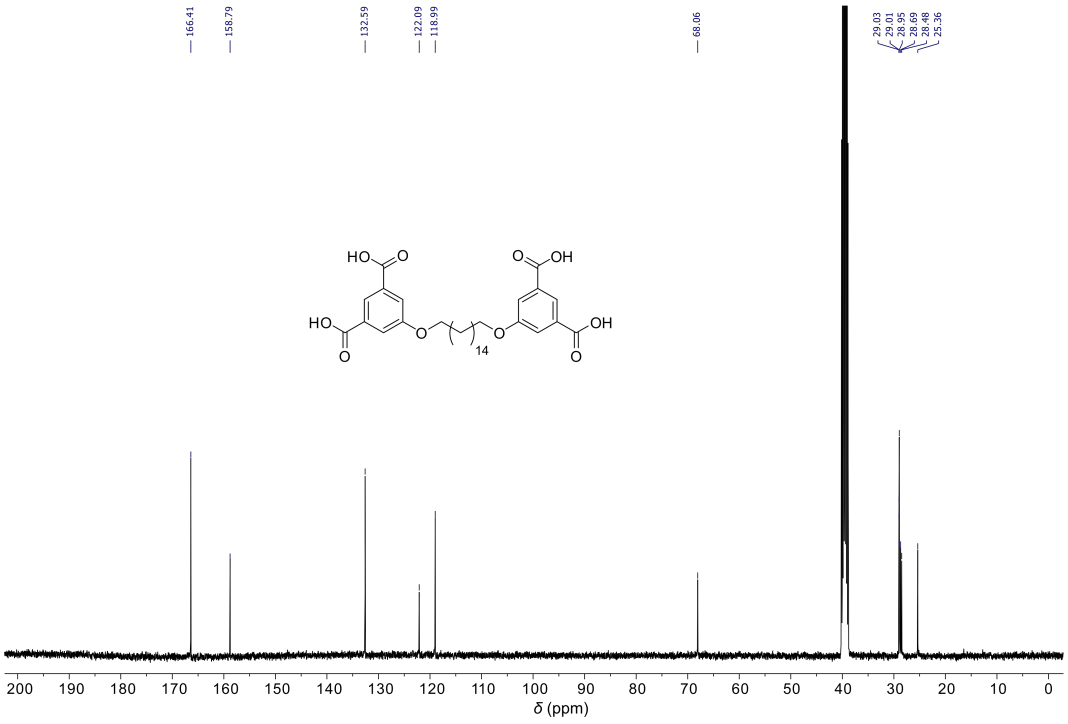


Figure S3.47. ^{13}C NMR (101 MHz, DMSO-*d6*) spectrum of **T5**.

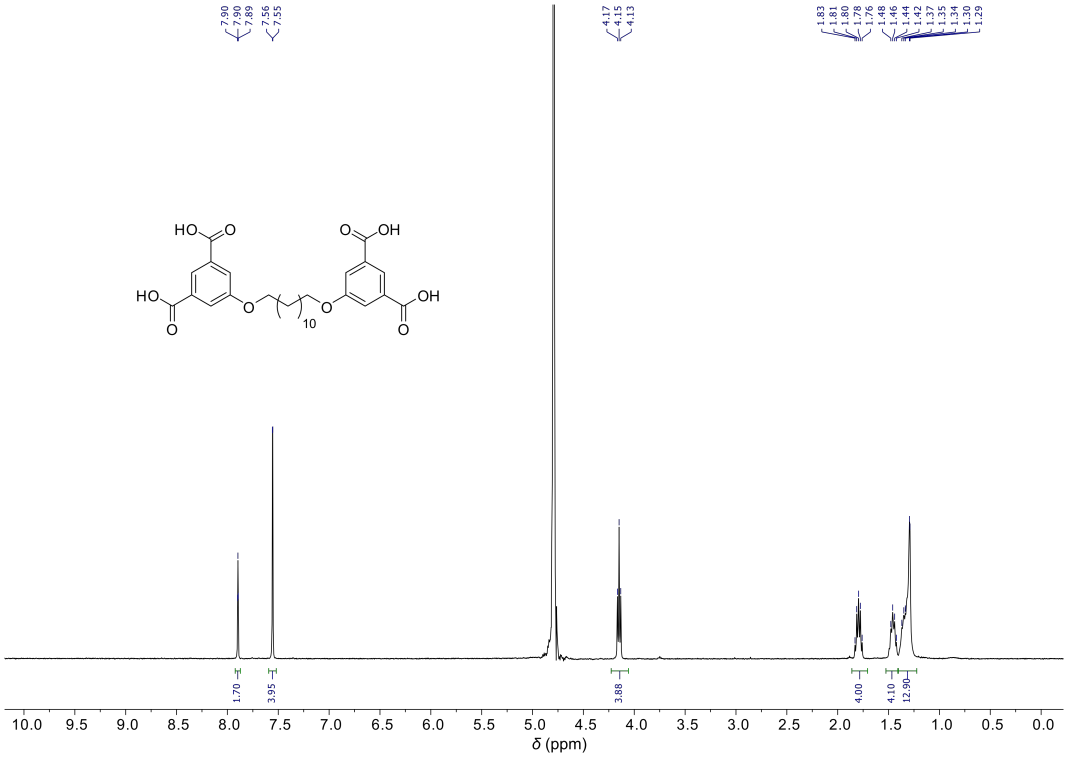


Figure S3.48. ^{1}H NMR (400 MHz) spectrum of **T6** in D_2O phosphate buffer, 50 mM, pH 7.5.

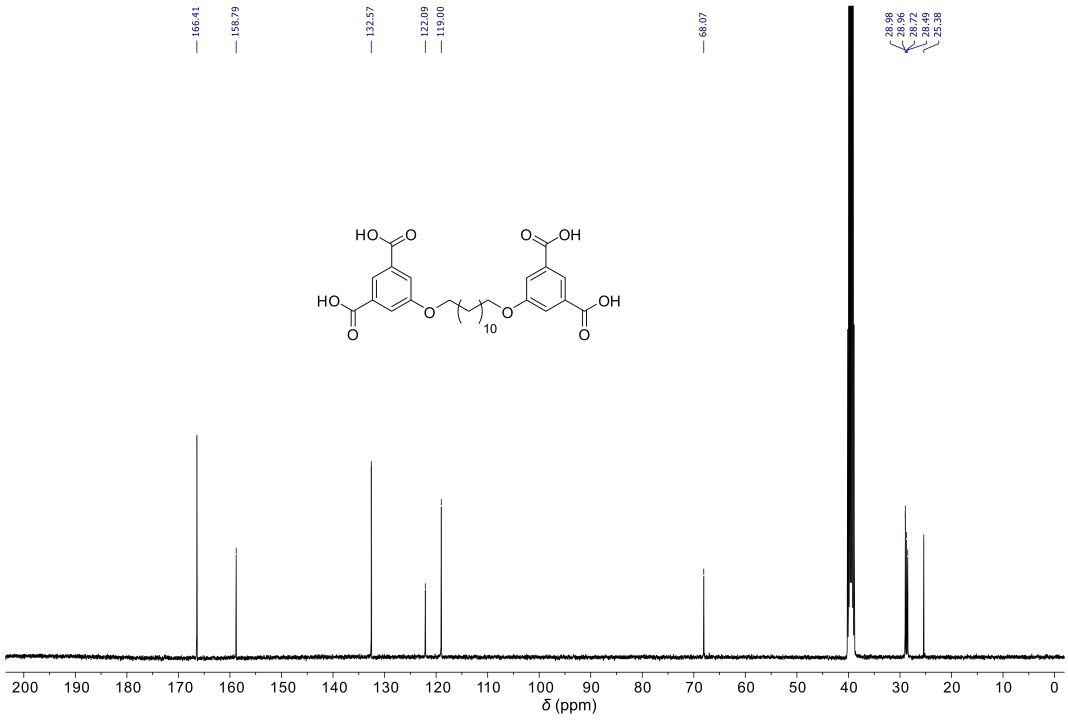


Figure S3.49. ^{13}C NMR (101 MHz, DMSO-*d6*) spectrum of **T6**.

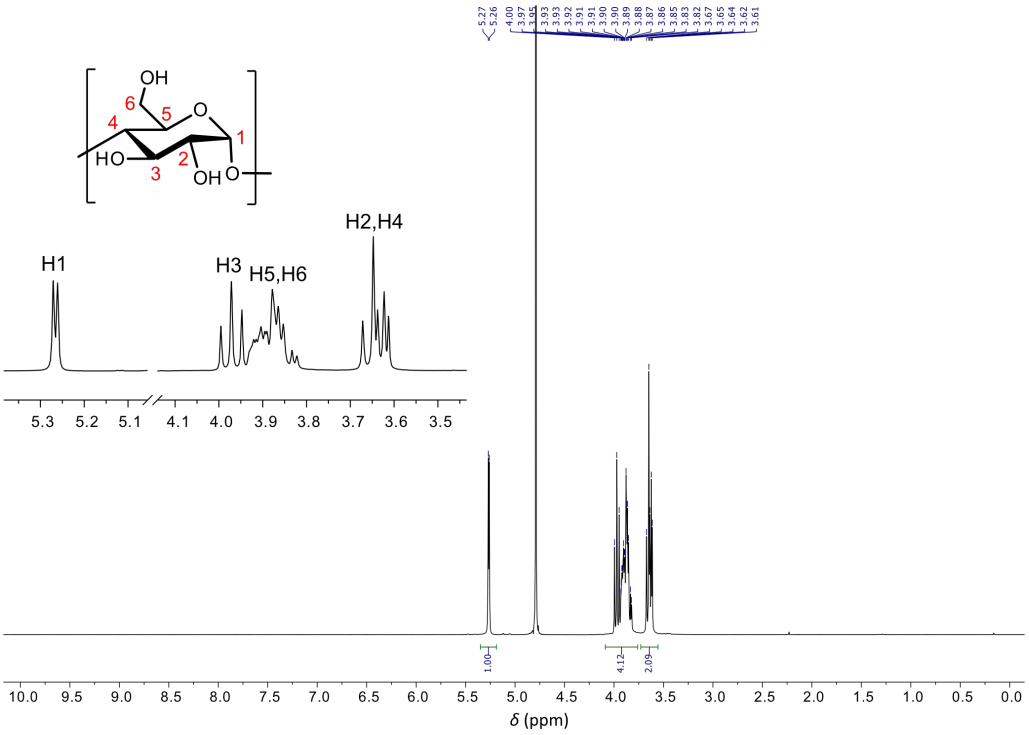


Figure S3.50. ^{1}H NMR (400 MHz) spectrum of δ-CD in D_2O.

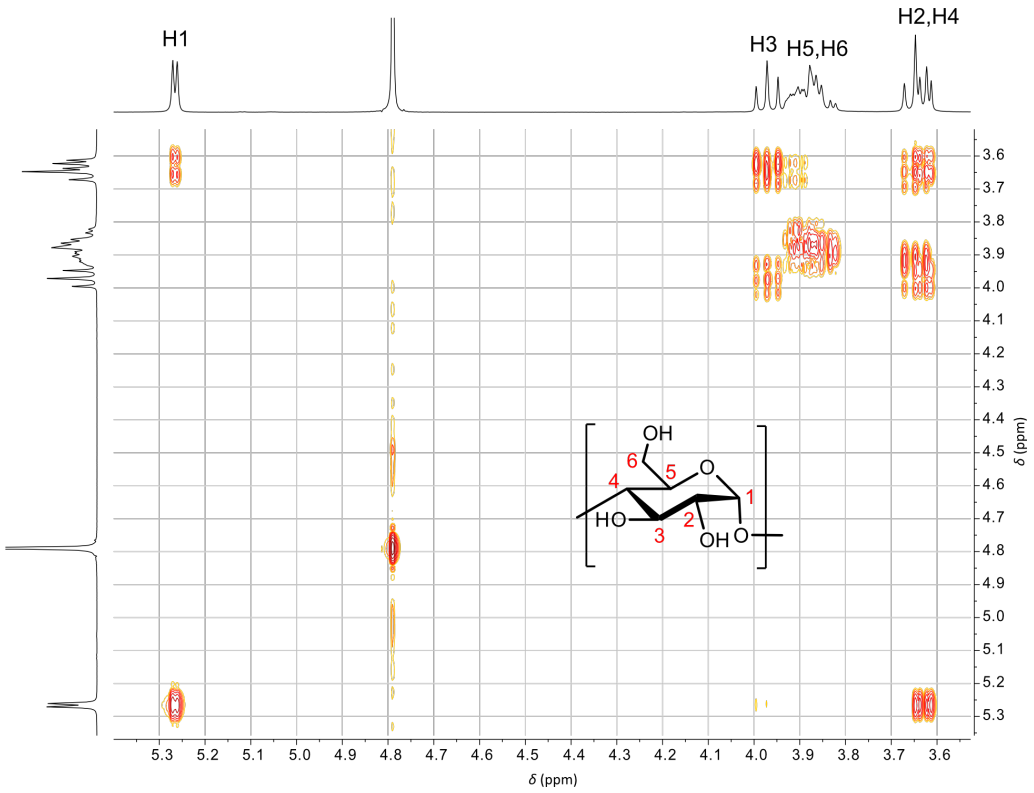


Figure S3.51. ^{1}H-^{1}H COSY (400 MHz, D_2O) spectrum of δ-CD.

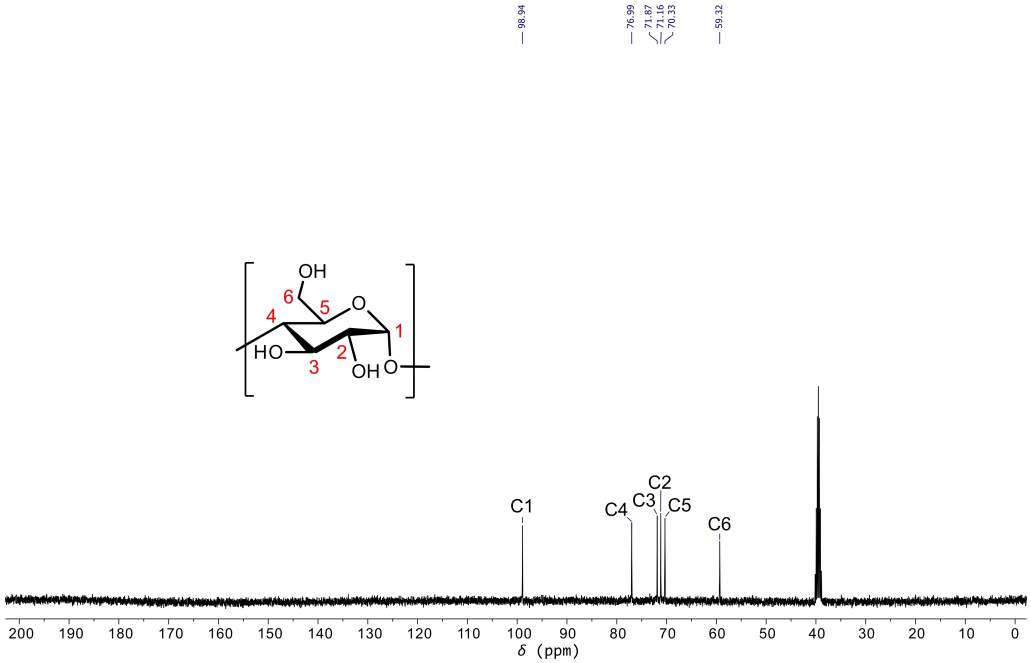


Figure S3.52. ^{13}C NMR (101 MHz, D_2O) spectrum of δ-CD, spectrum referenced to residual solvent in DMSO-*d6* lock tube.

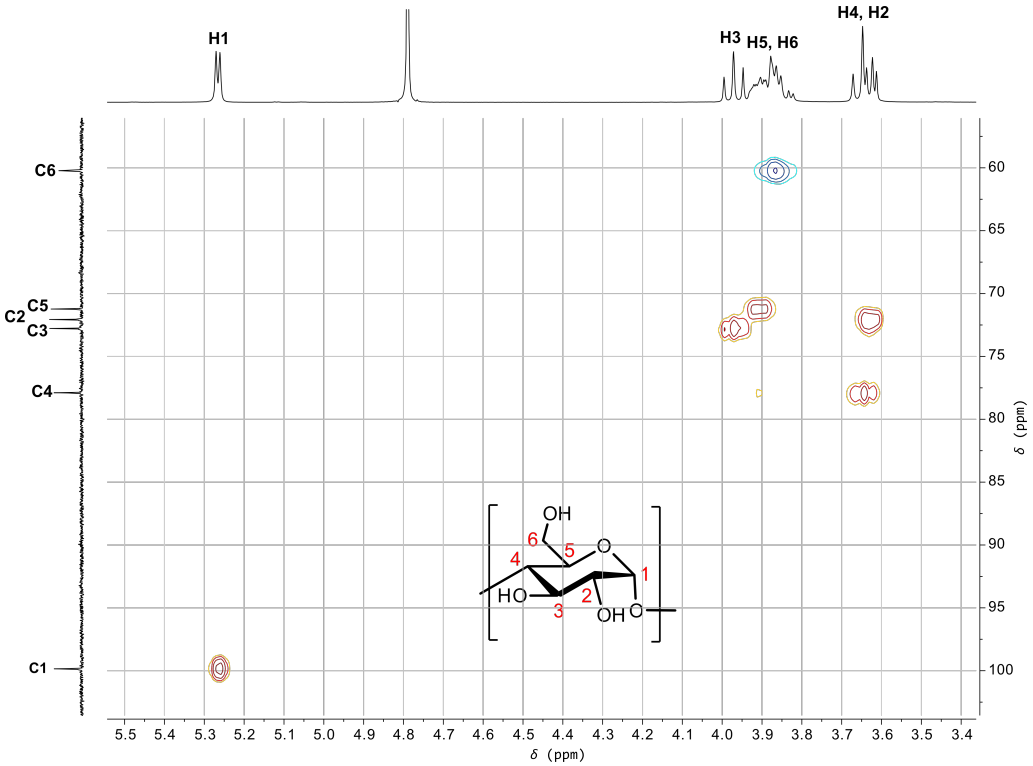


Figure S3.53. ^{1}H-^{13}C multiplicity-edited HSQC (400 / 101 MHz, D_2O) spectrum of δ-CD.

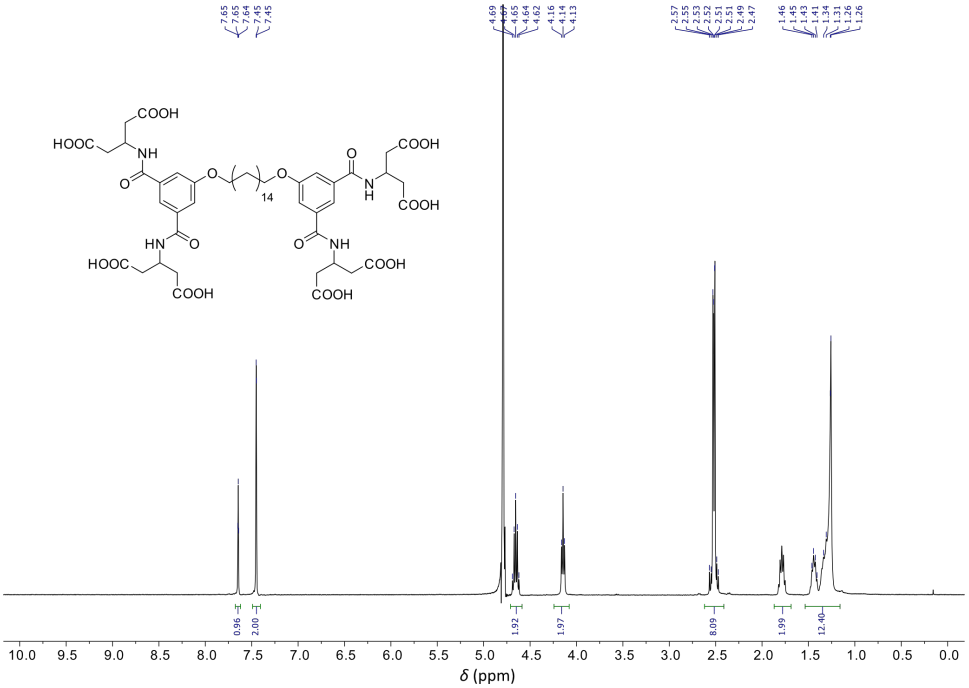


Figure S3.54. ^{1}H NMR (400 MHz) spectrum in phosphate buffered D_2O (50 mM, pH 7.5) of template **T1** recovered after 5 cycles of the templated enzymatic synthesis of δ-CD.

S3.3. Optimization of glucan and template concentrations to maximize δ-CD yield

A series of reactions were set up with α-CD (at various concentrations), template **T1** (at various concentrations) and glycerol-free CGTase at room temperature in sodium phosphate buffer (0.2 M, pH 7.5). The reactions were monitored by HPLC with an ELS detector. Aliquots of reaction mixture (6 or 12 µl, depending on the glycan concentration – smaller volume for higher concentrations) were diluted in a solution of 2% TFA in H_2O (30 or 24 µl) for a total volume of 36 µl. These were centrifuged before injection on the HPLC as described previously. The distributions of CDs (α, β, γ, δ-CD) and linear α-1,4-glucans (**G2–G7** and **G8–G20**) as a function of time are plotted in supporting figures S3.55–S3.57. Glucose (**G1**) is not plotted due to overlap with buffer/salt peaks in the chromatograms.

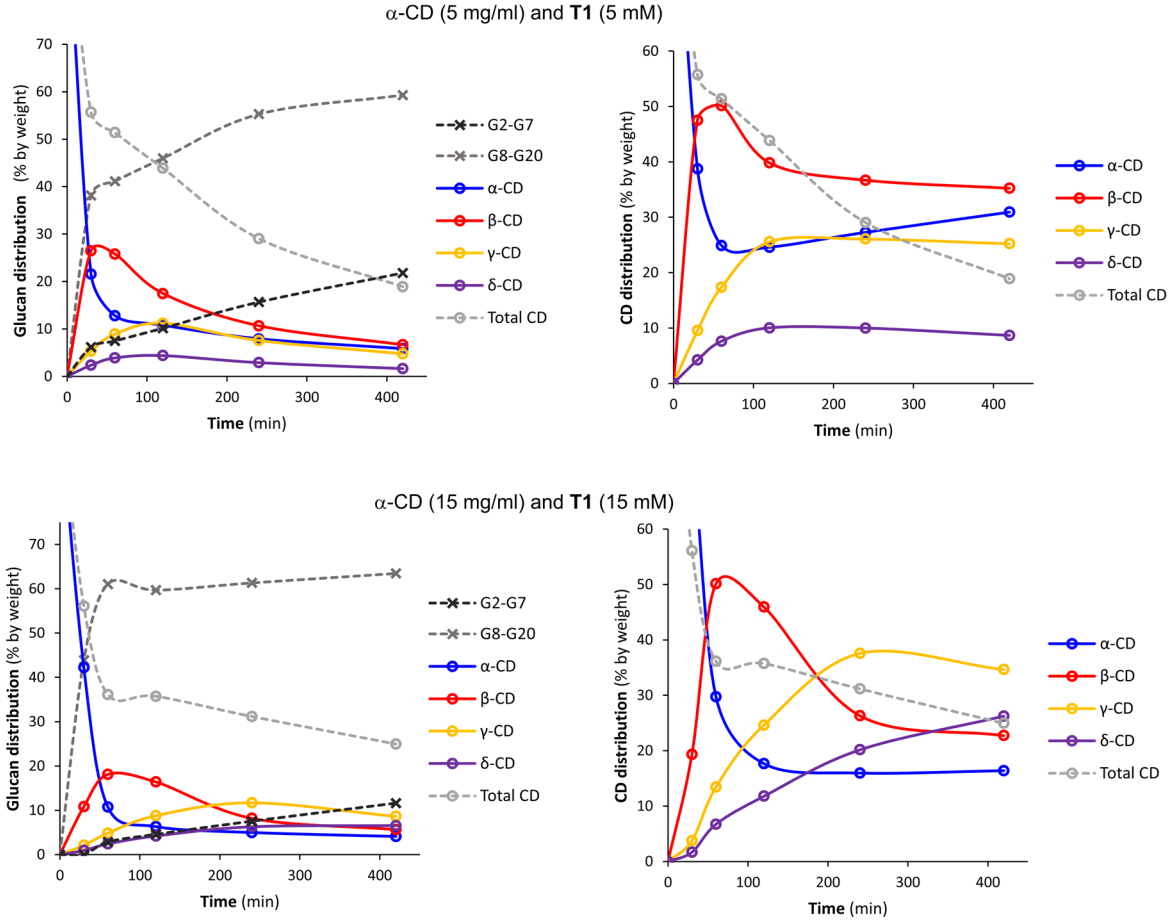


Figure S3.55. Glucan distributions (% by weight) as a function of time in reactions with α-CD and template **T1** (at concentrations as indicated on figure) with glycerol-free CGTase (65 µl per ml reaction mixture) in sodium phosphate buffer (200 mM, pH 7.5). *Left:* Distribution of all glucans, CDs (α, β, γ, δ-CD) and linear α-1,4-glucans (**G2**–**G7** and **G8**–**G20**) and % of CDs in total library. *Right:* Distribution within CD subsystem (α, β, γ, δ-CD) and % of CDs in total library.

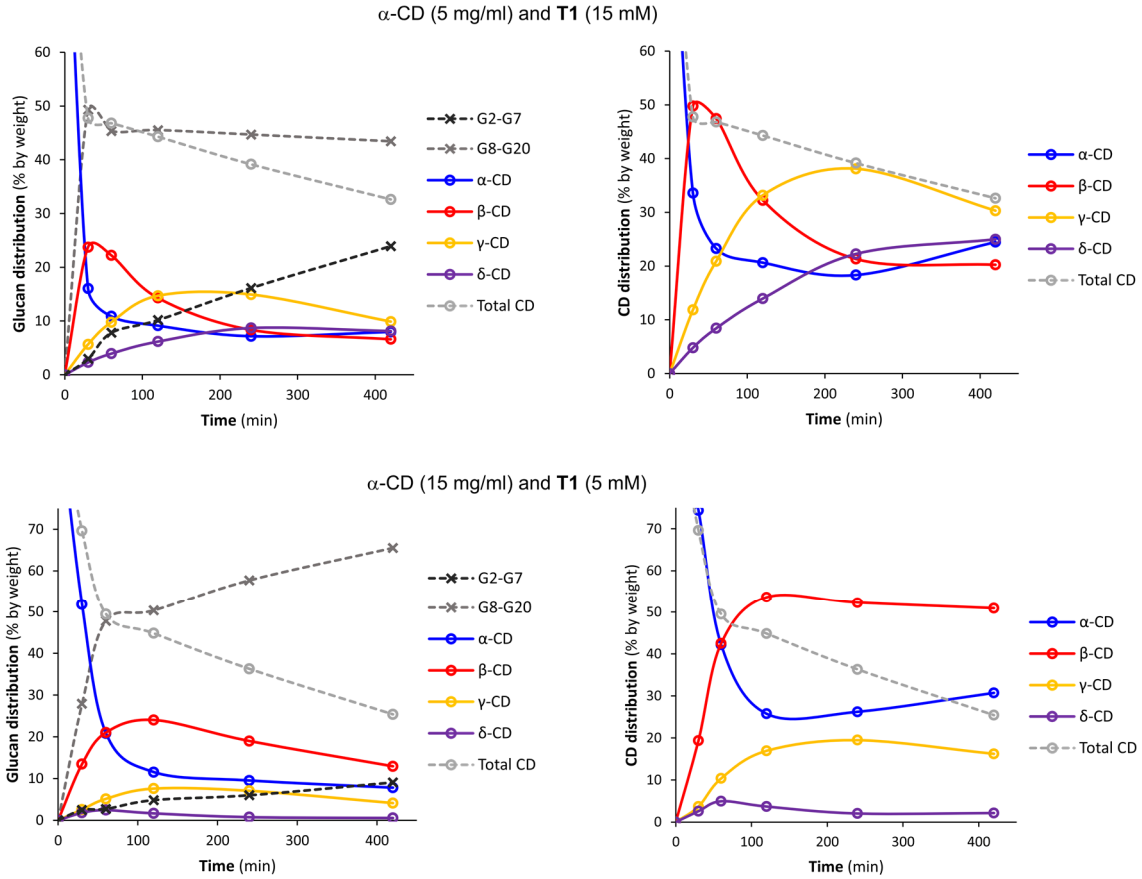


Figure S3.56. Glucan distributions (% by weight) as a function of time in reactions with α-CD and template **T1** (at concentrations as indicated on figure) with glycerol-free CGTase (65 µl per ml reaction mixture) in sodium phosphate buffer (200 mM, pH 7.5). *Left:* Distribution of all glucans, CDs (α, β, γ, δ-CD) and linear α-1,4-glucans (**G2**–**G7** and **G8**–**G20**) and % of CDs in total library. *Right:* Distribution within CD subsystem (α, β, γ, δ-CD) and % of CDs in total library.

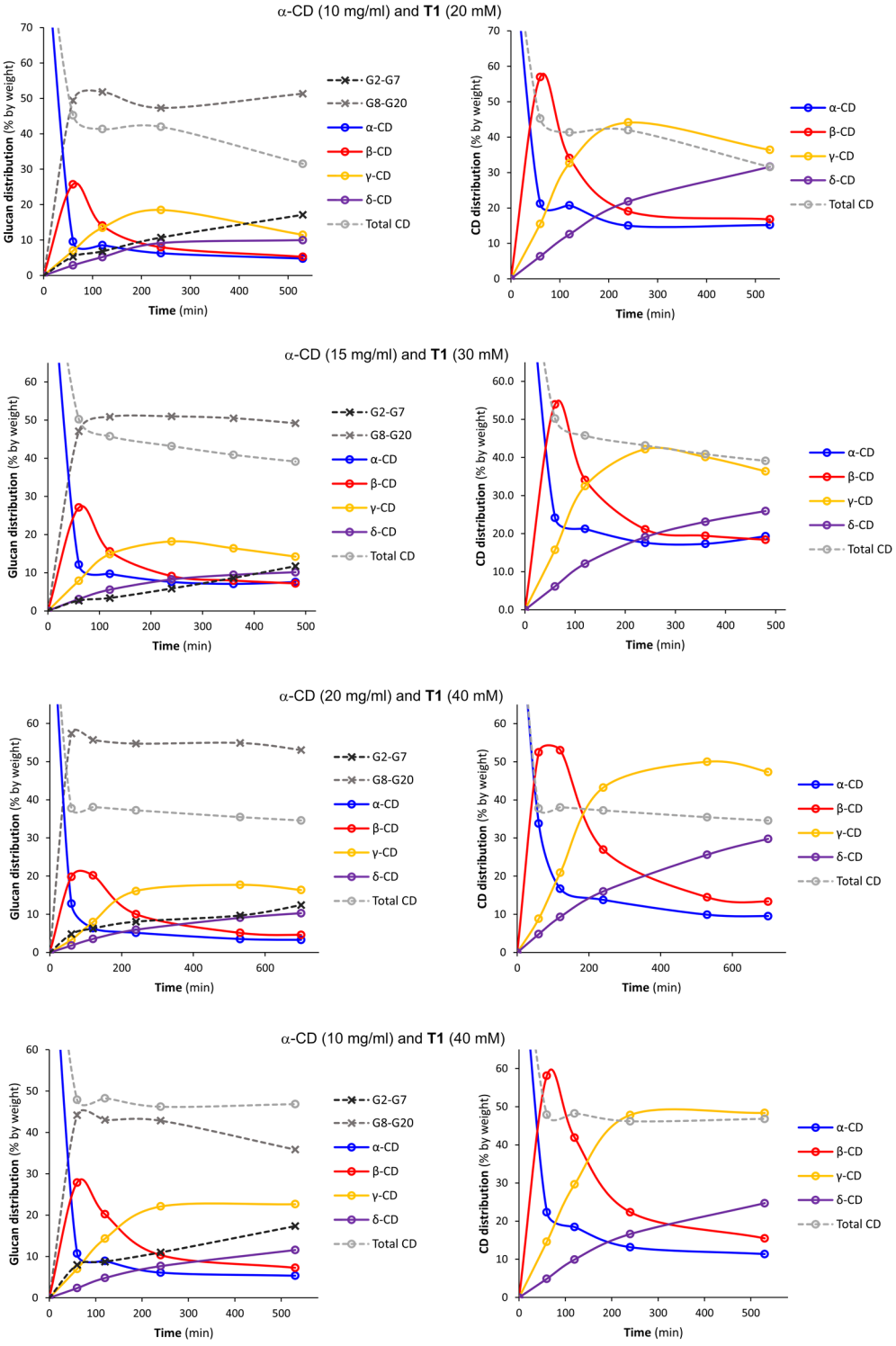


Figure S3.57. Glucan distributions (% by weight) as a function of time in reactions with α-CD and template **T1** (at concentrations as indicated on figure) with glycerol-free CGTase (98 µl per ml reaction mixture) in sodium phosphate buffer (200 mM, pH 7.5). *Left:* Distribution of all glucans, CDs (α, β, γ, δ-CD) and linear α-1,4-glucans (**G2**–**G7** and **G8**–**G20**) and % of CDs in total library. *Right:* Distribution within CD subsystem (α, β, γ, δ-CD) and % of CDs in total library.

SI.4: Supporting information for Chapter 4

S4.1. 1:2 binding model for NMR titrations in fast exchange

For the 1:2 binding of a guest (G) to a host (H) to form the species HG and HG_2 in an NMR titration under fast exchange, the following equation (S4.1) can be derived:[2]

$$\Delta\delta_{obs} = \frac{\Delta\delta_{max1}K_1[G]+\Delta\delta_{max2}K_1K_2[G]^2}{1+K_1[G]+K_1K_2[G]^2} \quad \text{(S4.1)}$$

where $\Delta\delta_{obs}$ is the observed change in chemical shift for the host upon addition of guest, [G] is the concentration of unbound guest, K_1 and K_2 are the first and second binding constants and $\Delta\delta_{max1} = \delta_{HG} - \delta_H$ and $\Delta\delta_{max2} = \delta_{HG_2} - \delta_H$ are the maximum possible changes in chemical shift upon addition of guest for each of the two bound species HG and HG_2.

The concentration of unbound guest, [G], is unknown. A cubic equation (S4.2) for [G] can be derived:[2]

$$A[G]^3 + B[G]^2 + C[G] + D = O \quad \text{(S4.2)}$$

wherein

$A = K_1K_2$

$B = K_1 + 2\,K_1K_2[H]_t - K_1K_2[G]_t$

$C = 1 + K_1[H]_t - K_1[G]_t$

$D = -[G]_t$

and $[H]_t$ and $[G]_t$ are the total concentrations of H and G. The following script written in the nonlinear curve fitter of Origin using Labscript can then be used to solve equation (S4.2) numerically while fitting the data to the unknown parameters K_1, K_2, $\Delta\delta_{max1}$, $\Delta\delta_{max2}$ using equation (S4.1):[3]

```
A=K1*K2;
B=K1+2*K2*K1*Ht-K1*K2*x;
C=1+K1*Ht-K1*x;
D=-x;
for (G=x, step=1; abs(step)>1e-15; G=G-step){
step=(A*G*G*G+B*G*G+C*G+D)/(3*A*G*G+2*B*G+C);};
y=(D1*K1*G+D2*K1*K2*G*G)/(1+K1*G+K1*K2*G*G)
```

where K1, K2, D1, D2, Ht, x and y represent K_1, K_2, $\Delta\delta_{max1}$, $\Delta\delta_{max2}$, $[H]_t$, $[G]_t$, and $\Delta\delta_{obs}$, respectively.

S4.2. NMR spectra

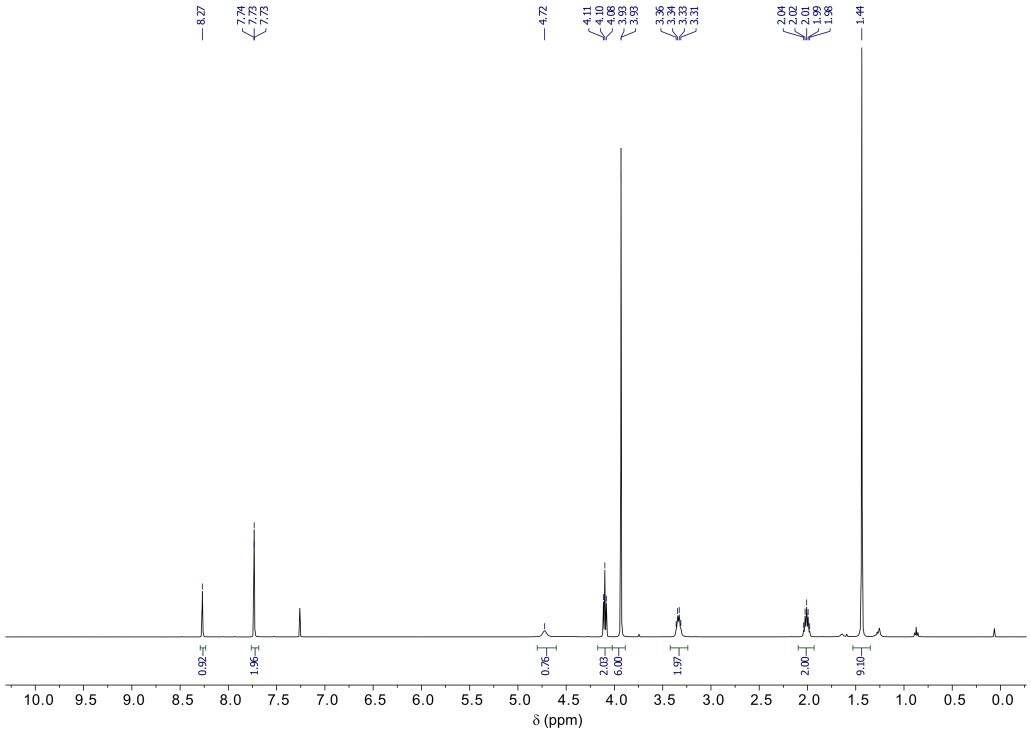


Figure S4.1. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of compound **4.2**.

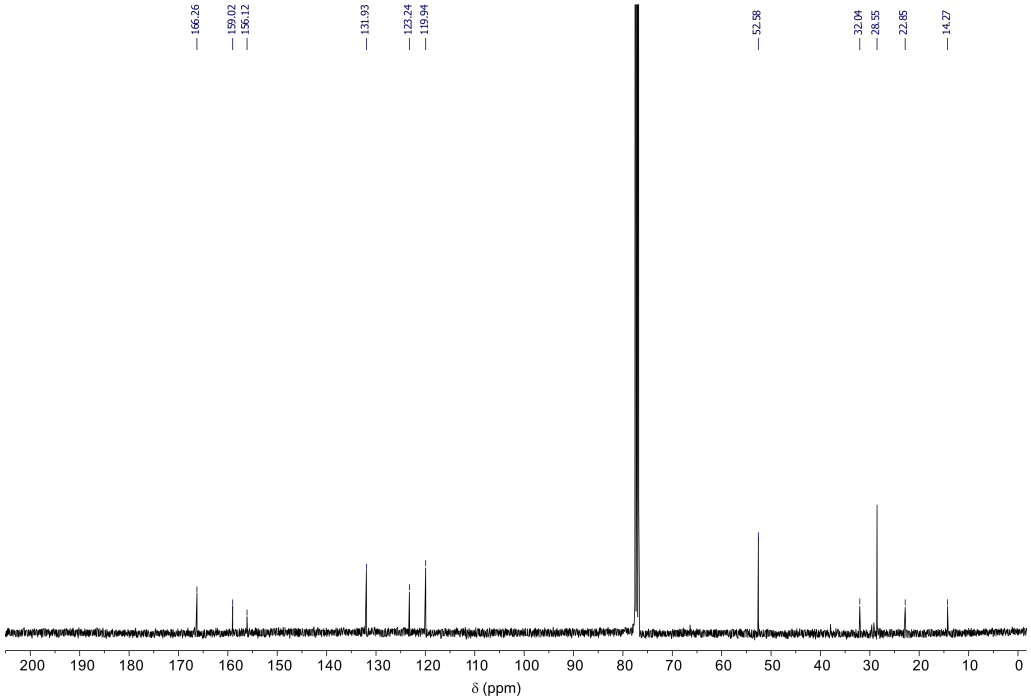


Figure S4.2. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of compound **4.2**.

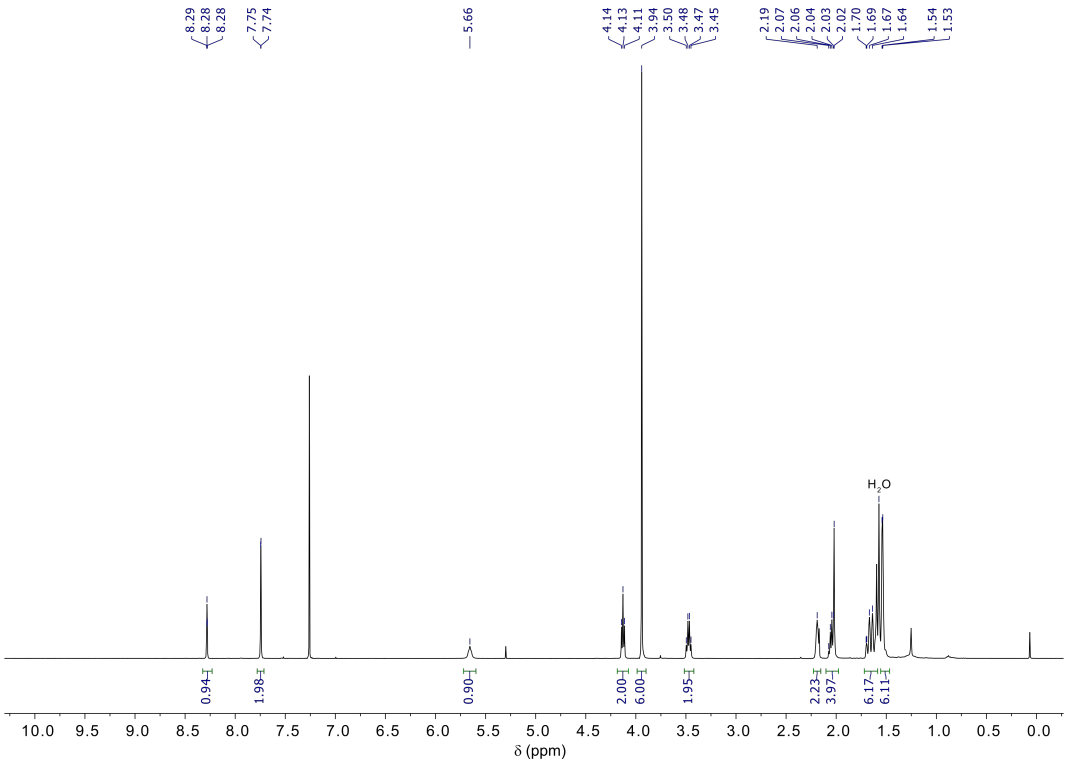


Figure S4.3. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of compound **4.4**.

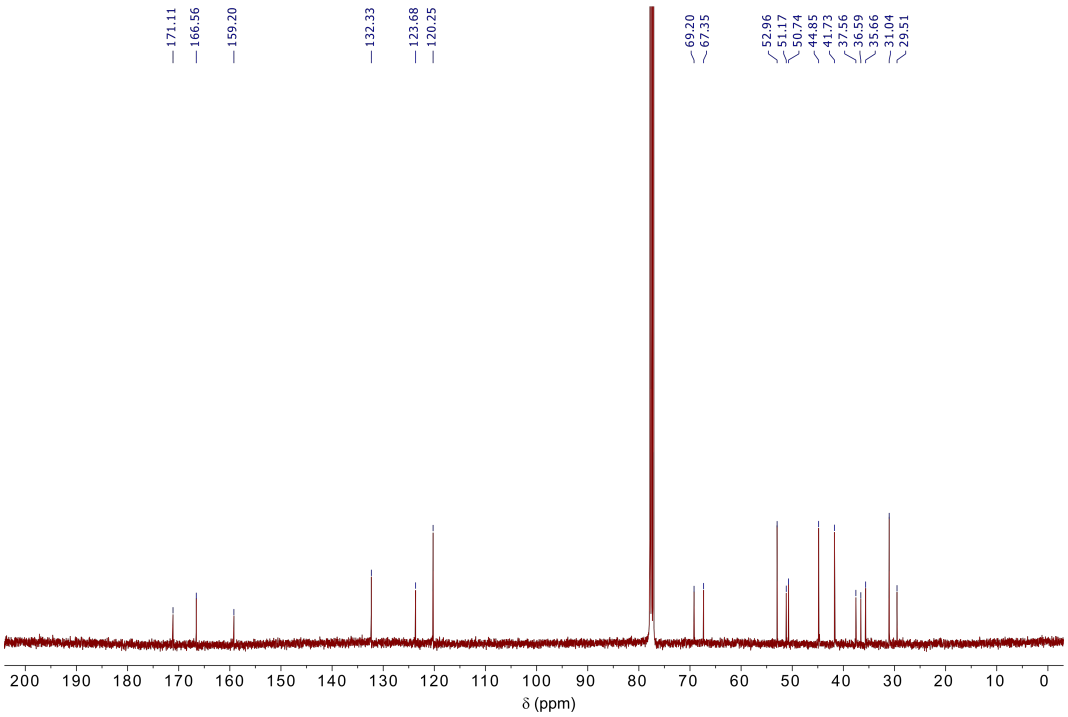


Figure S4.4. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of compound **4.4**.

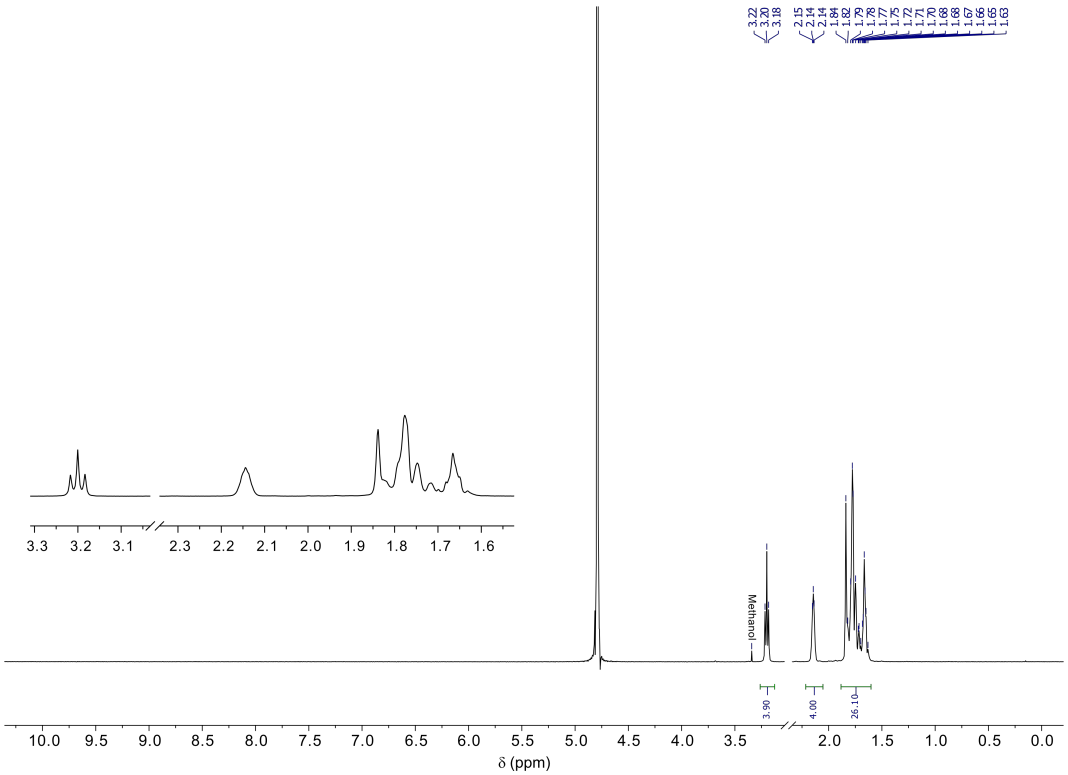


Figure S4.5. [1]H NMR (400 MHz, D2O phosphate buffer, 0.n1 M, pH 7.5) of compound **4.9**.

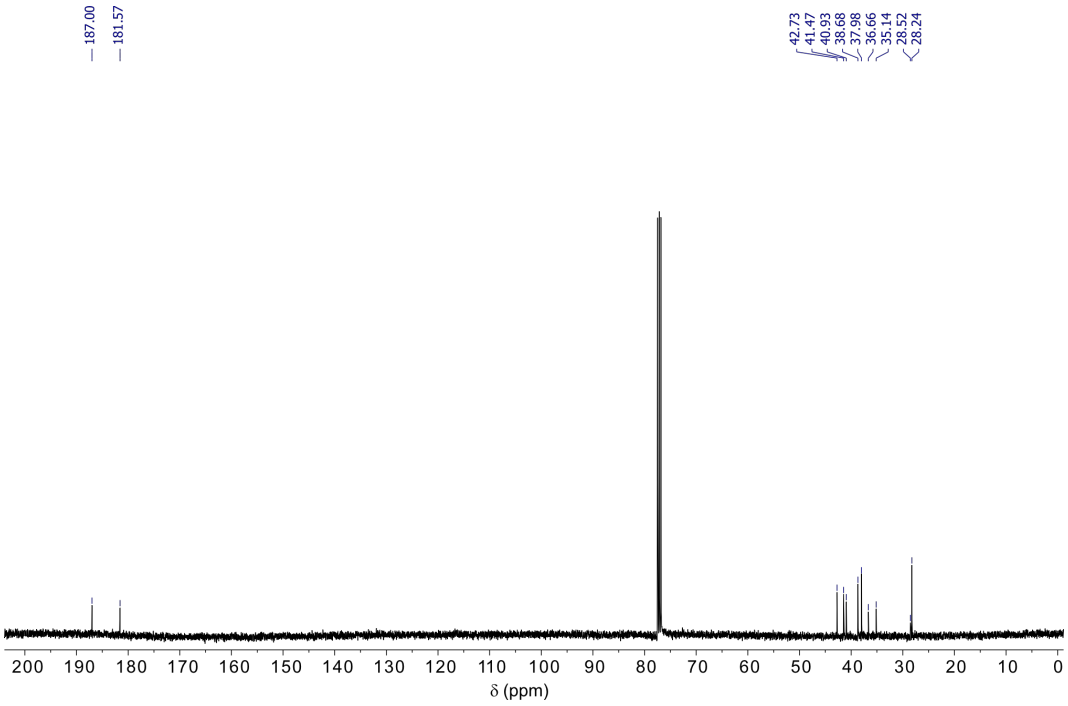


Figure S4.6. [13]C NMR (101 MHz, D2O phosphate buffer, 0.1 M, pH 7.5) of compound **4.9**. Referenced to residual solvent in $CDCl_3$ locktube.

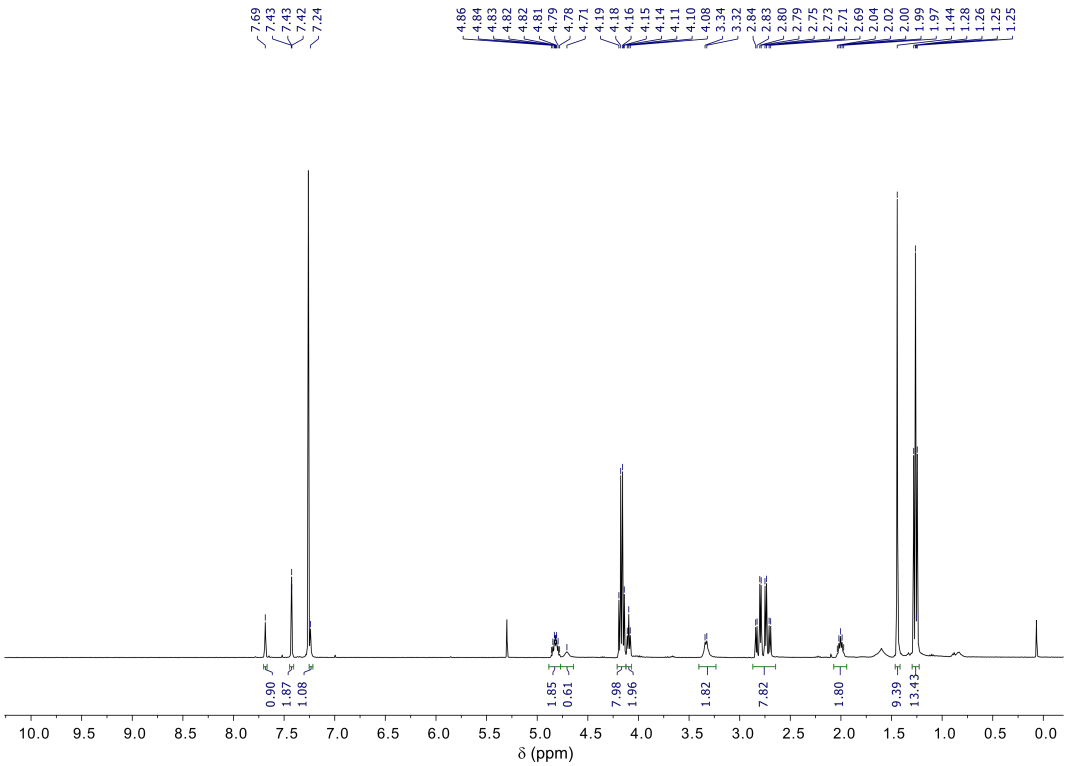


Figure S4.7. ^{1}H NMR (400 MHz, $CDCl_3$) of compound **4.11**.

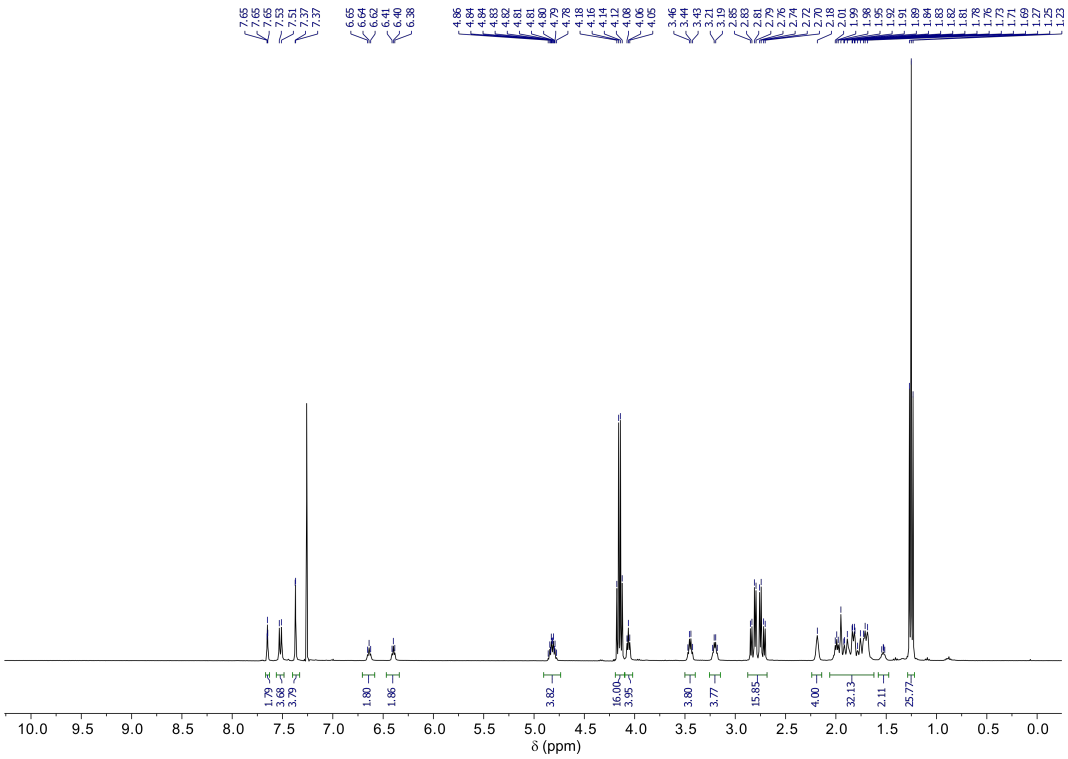


Figure S4.8. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of compound **4.13**.

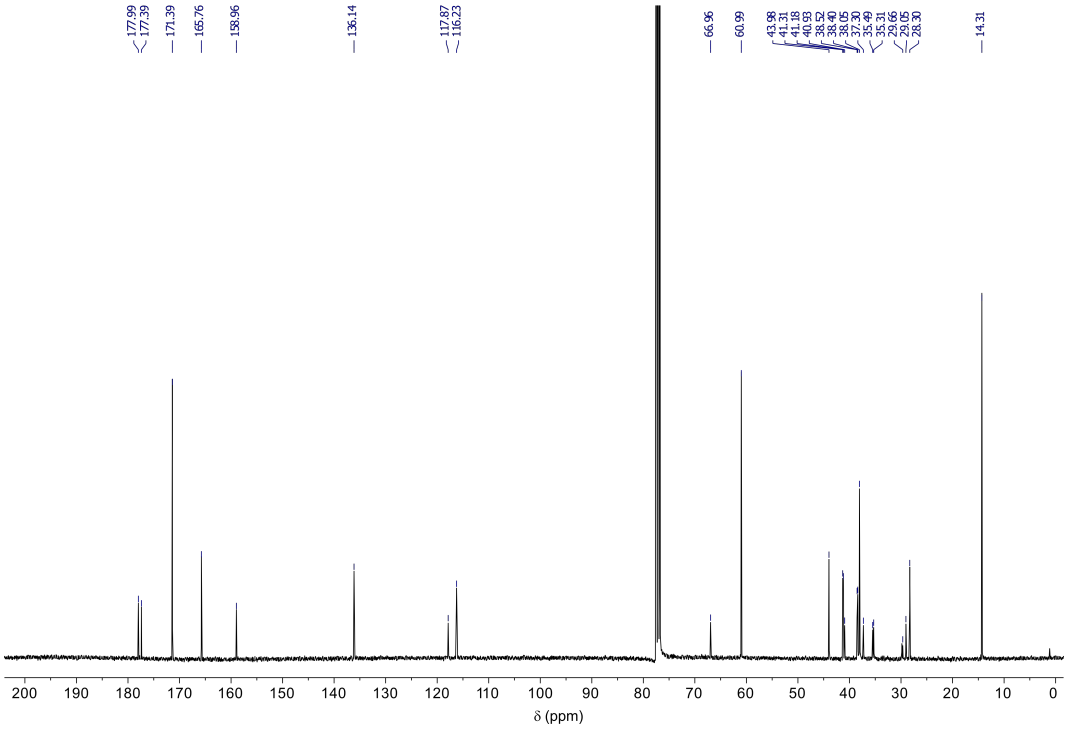


Figure S4.9. ^{13}C NMR (101 MHz, $CDCl_3$) spectrum of compound **4.13**.

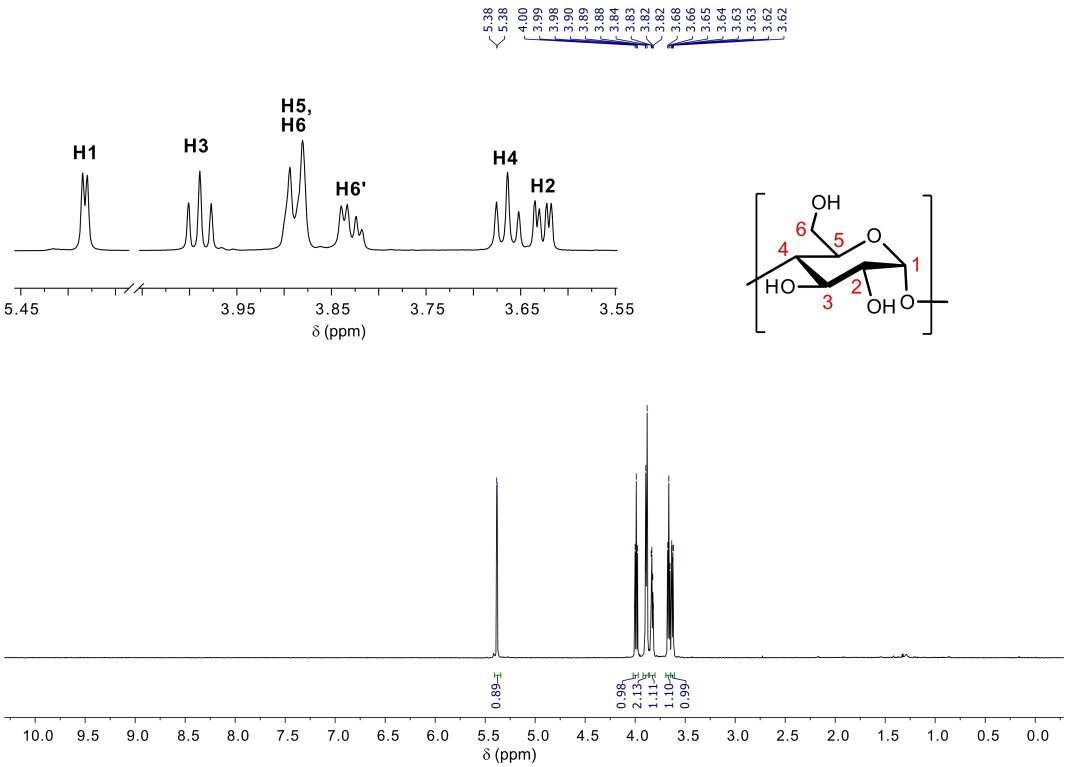


Figure S4.10. ^{1}H NMR (800 MHz) spectrum with water suppression of CD16 in D_2O.

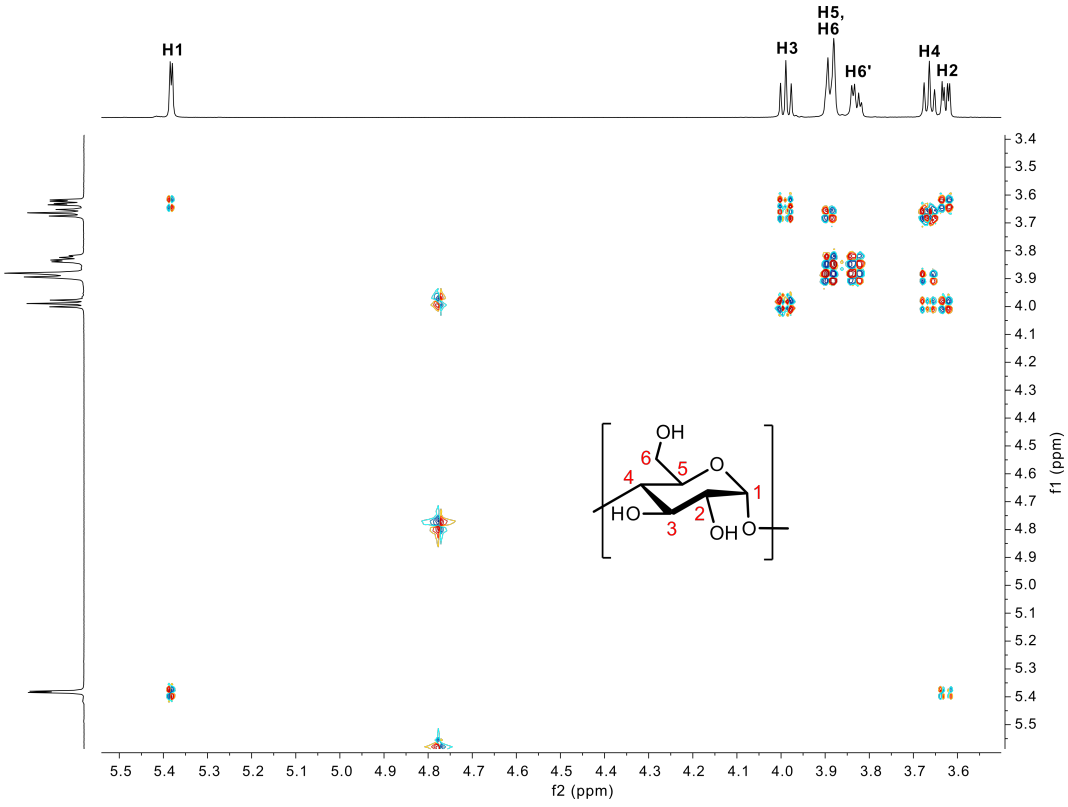


Figure S4.11. ^{1}H-^{1}H DQF-COSY (800 MHz) spectrum of CD16 in D_2O.

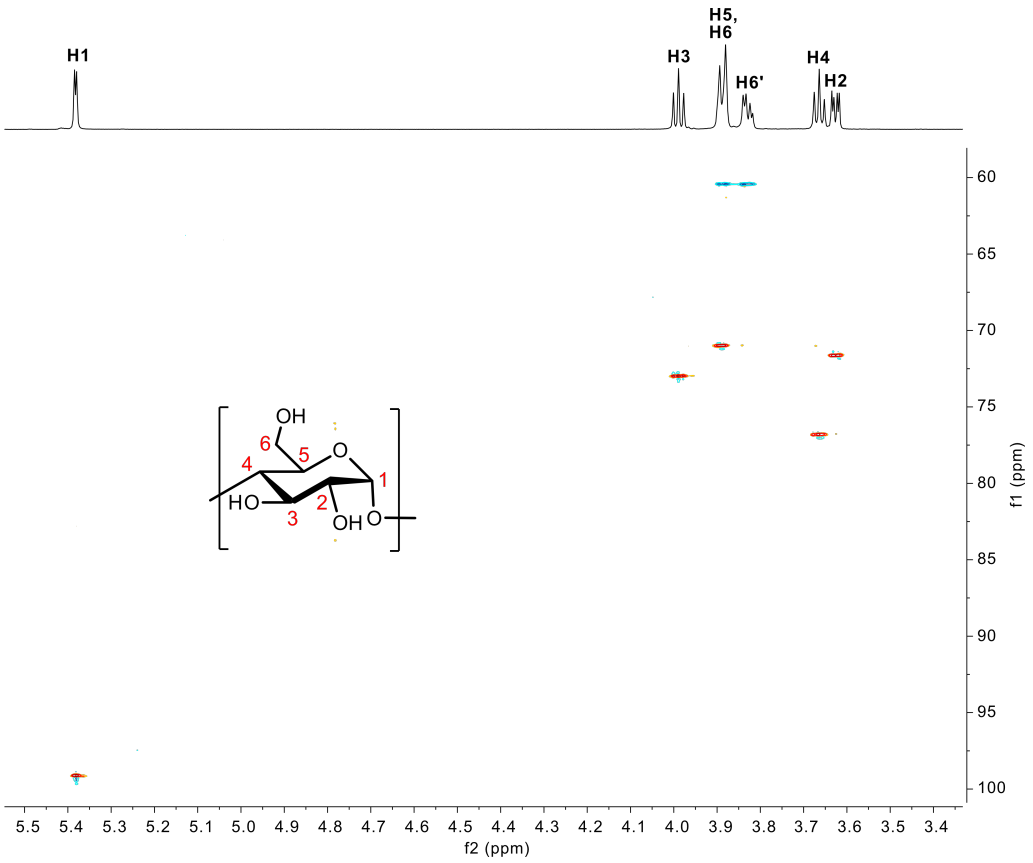


Figure S4.12. ^{1}H-^{13}C multiplicity-edited HSQC (800 / 201 MHz) spectrum of CD16 in D_2O.

SI.5: Supporting information for Chapter 5

S5.1. NMR spectra

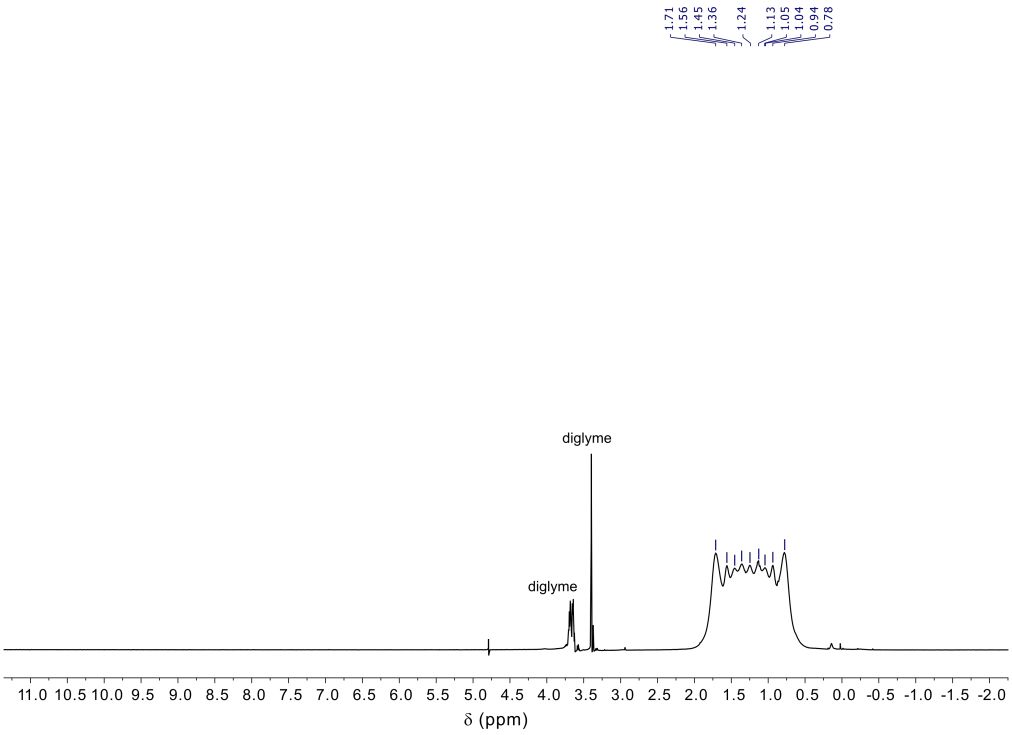


Figure S5.1. ^{1}H NMR (400 MHz) spectrum with water suppression of $Na_2[B_{12}H_{12}]$ in D_2O

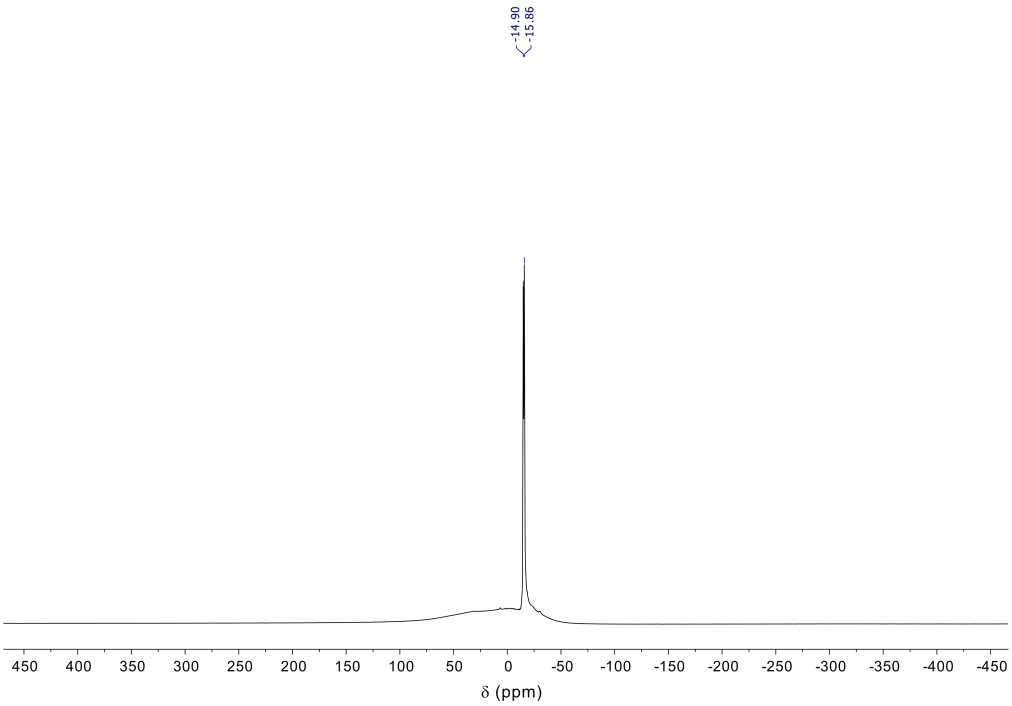


Figure S5.2. ^{11}B NMR (128 MHz) spectrum of $Na_2[B_{12}H_{12}]$ in D_2O. Broad signal due to borosilicate glass NMR tube.

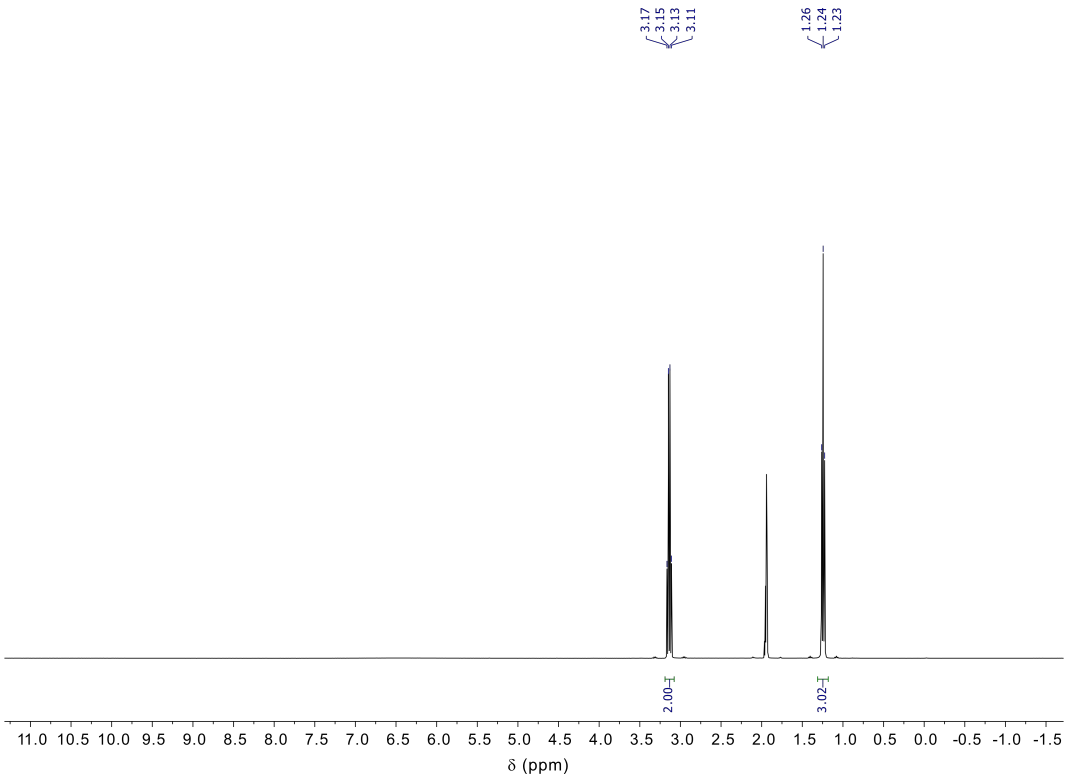


Figure S5.3. ^{1}H NMR (400 MHz) spectrum of $[Et_3NH]_2[B_{12}Cl_{12}]$ in CD_3CN.

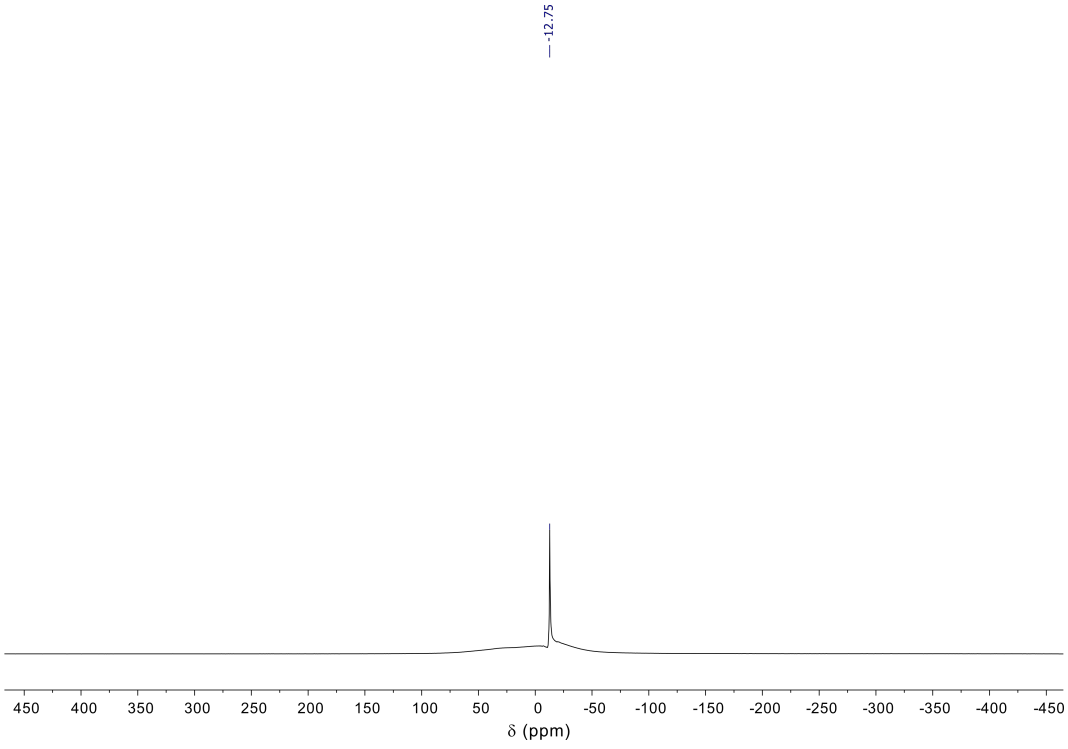


Figure S5.4. ^{11}B NMR (128 MHz) spectrum of $[Et_3NH]_2[B_{12}Cl_{12}]$ in CD_3CN. Broad signal due to borosilicate glass NMR tube.

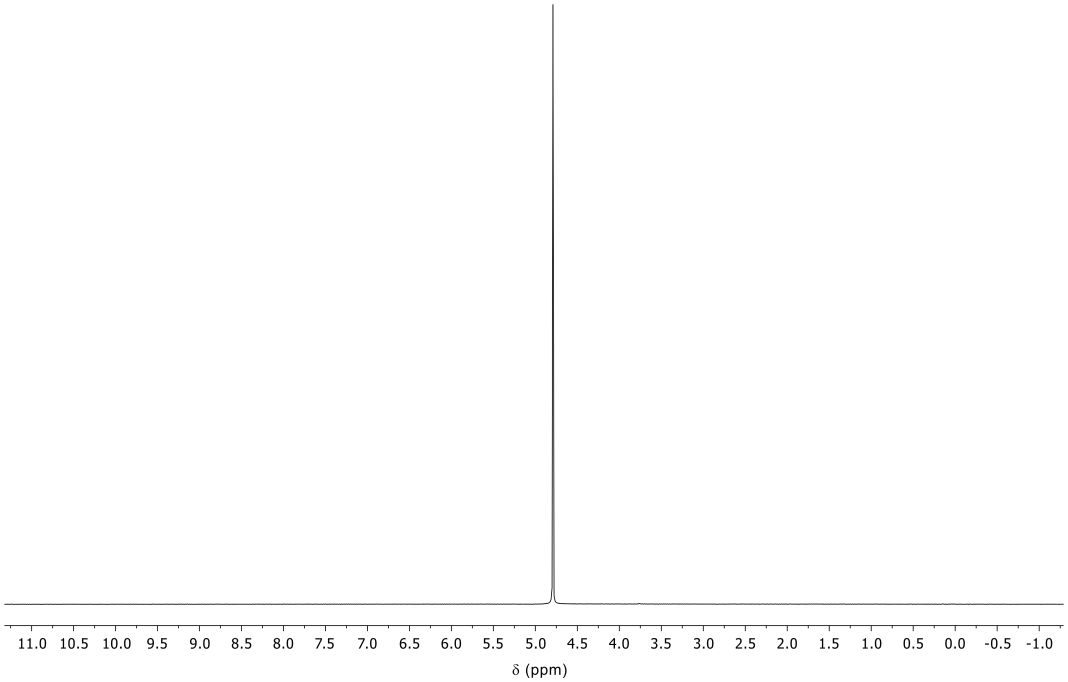


Figure S5.5. ^{1}H NMR (400 MHz) spectrum of $Na_2[B_{12}Cl_{12}]$ in D_2O

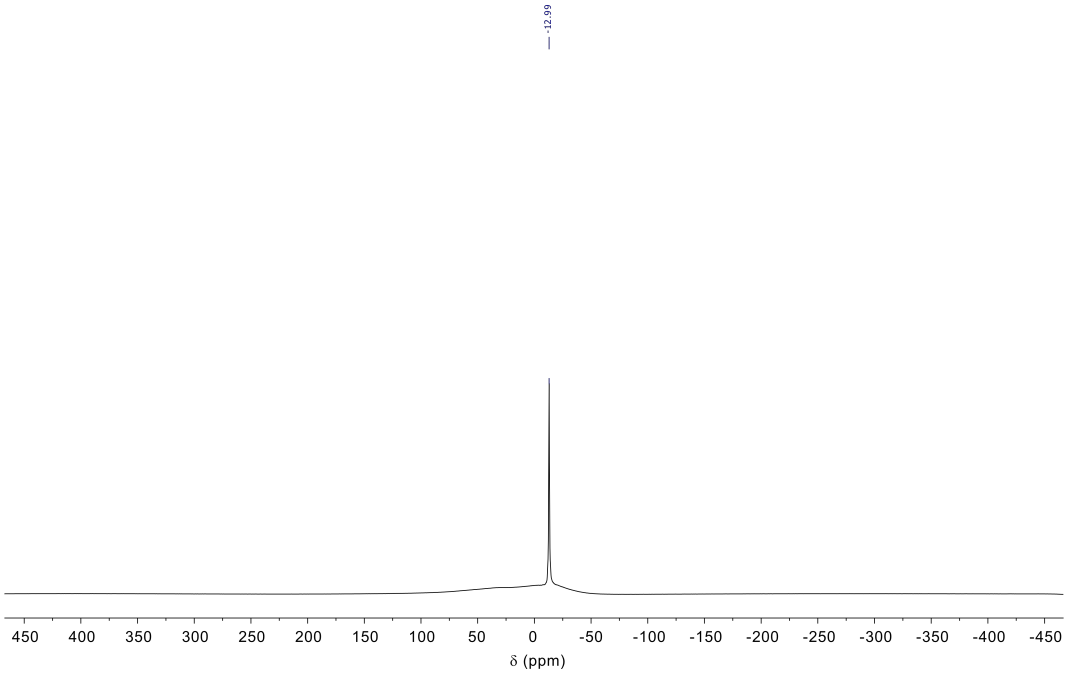


Figure S5.6. ^{11}B NMR (128 MHz) spectrum of $Na_2[B_{12}Cl_{12}]$ in D_2O. Broad signal due to borosilicate NMR tube.

Supporting Bibliography

[1] A. Erichsen, D. Larsen, S. R. Beeren, *Front. Chem.* **2021**, *9*, 1–8.

[2] P. Thordarson, *Chem. Soc. Rev.* **2011**, *40*, 1305–1323.

[3] A. E. Hargrove, Z. Zhong, J. L. Sessler, E. V. Anslyn, *New J. Chem.* **2010**, *34*, 348–354.