Electronic Supplementary Information

Enzyme-mediated dynamic combinatorial chemistry allows out-of-equilibrium template-directed synthesis of macrocyclic oligosaccharides

Dennis Larsen and Sophie R. Beeren*

*Correspondence to: sopbee@kemi.dtu.dk

Contents:

S1. Materials, instrumentation and methods

S1.1 Materials

Except for $Cs_2[B_{12}I_{12}]$ (**4**), all chemicals were obtained from commercial suppliers and used as received. **G4** and **G8** were obtained from Carbosynth, Compton, United Kingdom. $Cs_2[B_{12}H_{12}]$ was purchased from Strem Chemicals, Inc., Newburyport (MA), United States. All other chemicals including HPLC grade acetonitrile and formic acid as well as LC-MS grade methanol and ammonium hydroxide were purchased from Sigma-Aldrich Chemie Gmbh, Munich, Germany. CGTase derived from *Bacillus macerans* was obtained as a kind gift from Amano Enzyme, Inc., Nagoya, Japan, in a stock solution that was used as supplied and stored at 5 °C. Water was purified on a Merck Millipore Synergy UV water purification system prior to use. Colorless Corning CoStar 0.65 mL centrifuge tubes were used for sample preparation (heating/acidification and centrifugation), while colorless 2 mL glass vials with PTFE-lined screw-cap septa (often with a 0.2 mL glass insert) were used for short term sample storage and injection on chromatographic equipment.

S1.2 Instrumentation

Chromatography was performed on either a Waters Alliance 2695 (high pressure) separation module with a Waters XBridge Amide 3.5 μm 4.6 × 150 mm column maintained at 50 °C or on a Thermo Scientific Dionex UltiMate 3000 HPLC (ultra-high pressure) system with a Waters Acquity UPLC BEH Amide 1.7 μm 2.1 ×150 mm column maintained at 30 °C. Both systems were equipped with an autosampler maintained at 20 °C. Detection was done with an Agilent Technologies 1260 Infinity ELSD, typically operating at evaporator and nebulizer temperatures of 90 and 70 °C, respectively, and a N_2 gas flow of 1.0 L/min. NMR spectroscopy was performed on a Bruker 400 MHz instrument. MALDI-TOF-MS experiments were carried out on a Bruker autoflex speed instrument. ESI-MS was performed on a Bruker MicrOTOF-QII system.

S1.3 Enzymatic reactions

CGTase-catalyzed reactions were performed in buffered water (50 mM sodium phosphate at pH 7.5). Solutions of the starting glucan (10 mg/mL) with the template (if any) at the desired concentrations were prepared in glass reaction chambers (usually 2 mL vials or NMR tubes) and left for ca. 5 minutes in a specially fitted aluminium block under thermostatic control at

25 °C. Reactions were initiated by adding CGTase stock solution (50 μL per mL of reaction mixture) directly to the reaction and mixing thoroughly. The reactions were monitored at various time points. Aliquots for analysis (typically 10 – 60 μL) were removed and the enzymatic reaction stopped either by rapidly heating to 100 °C (water bath) for ten minutes in a small capped centrifuge tube or by immediate 4-fold dilution with a 1% solution of trifluoroacetic acid in acetonitrile.

S1.4 Analysis of enzymatic reactions using HPLC

The reaction progress and glucan mixture compositions were analyzed using high performance liquid chromatography with evaporative light scattering detection (HPLC-ELS). Separation was performed using gradient elution on a HILIC-type column. For quantitative analysis, injection volumes were adjusted to stay within the approximately linear range of the detector to ensure that detector response was approximately proportional to the quantity by weight of each analyte. To prevent column blockage by insoluble species, such as enzyme and salts, all samples were centrifuged (10,000 RPM for 10 minutes) and the top fraction used for analysis. Comparison with non-centrifuged samples showed that no analytes (i.e. the glucans of interest in this study) were depleted by this centrifugation step. Eluents were acetonitrile (weak eluent) and water (strong eluent), both with either 0.1% by volume of formic acid for an acidic eluent system or 40 mM ammonium hydroxide for an alkaline eluent system. The alkaline eluent system allowed determination of concentration of **G1** (which overlapped with the buffer peak in eluents containing formic acid), but due to rapid column degradation under alkaline eluent conditions, the alkaline eluent was only used to perform the kinetics analysis described in the below. No significant changes in retention times or analyte response were observed upon switching from acidic to alkaline eluents. A typical gradient profile went from ca. 20% water to ca. 55% water without any curvature and at a constant flow rate over ca. 20 minutes. Peaks were identified by comparison to authentic samples of **G1** – **G8**, and α-, β-, and γ-CD obtained from commercial suppliers. Large-ring cyclodextrins CD9 and CD10 were identified by fractional collection of the eluting peaks and subsequent analysis by matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry.

S1.5 MALDI-TOF mass spectrometry

The isolated HPLC fractions, which consisted of a mixture of acetonitrile and water with 0.1 % formic acid and an unknown concentration of the species to be identified, were mixed 1:1 with a DHB (20 mg/mL in methanol) matrix in a centrifuge tube. Aliquots (usually 1-2 μL) of

the mixtures were transferred to a ground steel target plate and allowed to evaporate. Measurements were performed in the reflector mode in the positive direction. A mixture of large-ring cyclodextrins was used to calibrate the instrument using the $[M+Na]^+$-ions as reference points. Results (m/z): for CD9 1481.47 found, 1481.46 expected for $[M+Na]^+$; for CD10 1643.50 found, 1643.52 expected for $[M+Na]^+$.

S1.6 Kinetics analysis of the hydrolysis reaction.

To investigate the rate of hydrolysis during the enzymatic reactions, the glucose-units were first classed as belonging in one of two main groups: 1) the reducing-end (hemiacetal) glucose unit of linear species (G_r); and 2) all other glucose units (G_h), i.e. the glucose units that have a hydrolysable α(1→4) glycosidic bond from C1 to a neighbouring glucose unit. Note that the total concentration of glucose units ($[G_{total}]$), which is a known quantity based on the starting glucan concentration, can be expressed as $[G_{total}] = [G_r] + [G_h]$. Therefore, the concentration of G_h units ($[G_h]$) can be easily calculated after measuring the relative concentrations of all glucans with the HPLC-ELS method described above, using the knowledge that **G1** – **G8** all have exactly one G_r unit and that all of the CDs have zero G_r units. With each hydrolysis reaction, one G_h unit is converted to one G_r unit. Plotting the decreasing concentration of G_h units in the integrated rate equation assuming a second order reaction rate ($1/[G_h] - 1/[G_h]_0 = kt$, with t being the reaction time since addition of enzyme in seconds) gave linear fits with good determination coefficients (R^2 of 0.981 – 0.995 in the untemplated reactions, Supporting fig. 2a). Similar plots using the integrated equation for first order reaction rates ($\ln[G_h] - \ln[G_h]_0 = -kt$) gave poor fits with determination coefficients of 0.873 – 0.898. This led us to conclude that the overall hydrolysis reaction approximates second-order reaction kinetics, allowing us to determine the apparent second-order rate constants for overall hydrolysis when CGTase acts on different α-1,4-glucans (α-, β-, γ-CD or **G6**) in the presence or absence of 1-adamantane carboxylic acid (**2**) (Supporting fig. 2).

S1.7 Synthesis of $Cs_2[B_{12}I_{12}]$

Literature procedure by Tiritiris and Schleid.[S1] $Cs_2[B_{12}H_{12}]$ (1.55 g) was dissolved in CH_2I_2 (50 mL) and I_2 (5.9 g) was added at room temperature. After stirring for ca. 10 minutes, ICl (11.2 g) dissolved in CH_2I_2 (10 mL) was added dropwise. The resulting mixture was equipped with a reflux condenser, heated till reflux, and then kept under refluxing conditions for 8 hours. Solvent and volatiles were removed under reduced pressure at 90 °C, and the residue was dried under a stream of N_2 overnight. The residue was heated in 10:1 water/acetonitrile (50 mL), a

blackish brown residue was filtered off, and the brown powder that resulted upon cooling the filtrate slowly to 5 °C was isolated. The powder was recrystallized thrice from hot water yielding 143 mg of an off-white powder. ESI-MS confirmed the presence of $[B_{12}I_{12}]^{2-}$ (m/z 826.49 found, 826.49 expected for $[M]^{2-}$) but also indicated presence of some $[B_{12}H_nI_{12-n}]$ with n = 1-5. The product was used as isolated.

S1.8 NMR spectroscopy titrations

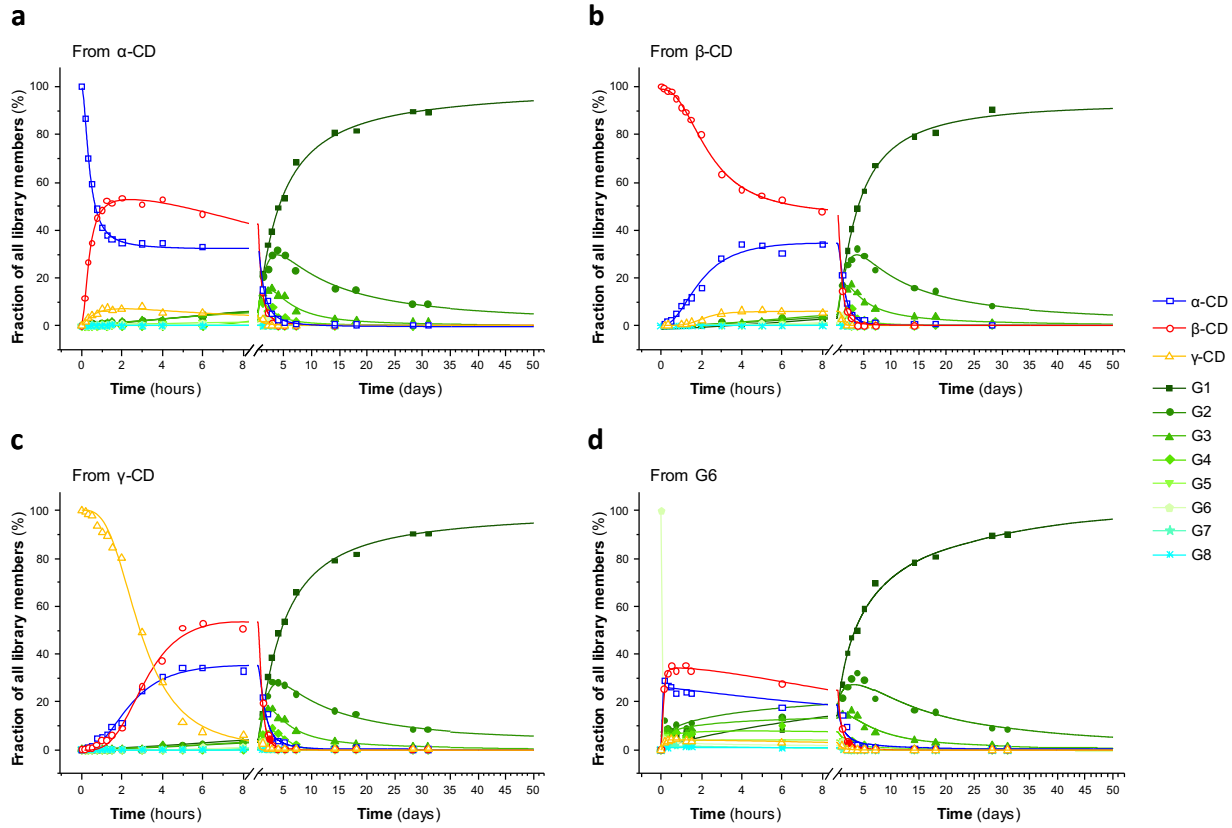
For each titration, two stock solutions were prepared. First, a stock solution of SDS (**1**) at 1.00 mM (for the α-CD titration) or 0.50 mM (for the β-CD and γ-CD titrations) in 50 mM sodium phosphate buffer at pH 7.5 in D_2O was prepared. From this, a second stock solution containing both SDS (**1**) and the desired cyclodextrin at 10.0 mM (α-CD), 20.0 mM (β-CD), or 25.0 mM (γ-CD) was prepared. These two stock solutions were then mixed in NMR tubes in the appropriate proportions to achieve the cyclodextrin concentrations listed in Supporting figs. 8 – 10. The ^{1}H NMR spectra of all the tubes were recorded at 25 °C using an autosampler setup that left the NMR tube in the heat-controlled probe for 10 minutes before measurements began to ensure uniform temperature. Water suppression was applied during the measurements. The resulting spectra were analyzed in MestReNova (version 11.0, Mestrelab Research S.L., Santiago de Compostela, Spain) and the resulting ppm values of the α-CH_2 group (for the α-CD titration) or the CH_3 group (for the β-CD and γ-CD titrations) of SDS (**1**) were tabulated along with the corresponding cyclodextrin concentrations. Using the data analysis software OriginPro 2018b (version 9.55, OriginLab Corp., Northhampton (MA), United States), the best fit to 1:1 and 2:1 binding models were fitted according to the iterative optimisation algorithms suggested by Anslyn and co-workers[S2] modified for the use with NMR data. α-CD and β-CD gave good fits with the 2:1 binding model, and γ-CD gave the best fit with the 1:1 model (errors on fit < 6 %) compared to a relatively poor fit with the 2:1 model (errors on fit > 50 %), leading us to conclude that the actual binding affinity of γ-CD towards SDS (**1**) could be best estimated using the 1:1 model. The partial ^{1}H NMR spectra, the resulting best fits, and the parameters from those fits are presented in Supporting figs. 8 – 10.

S1.9 DCL simulations

The software *DCLsim* developed in the group of Otto was used (see reference 28 of the main text). The software requires input of the concentrations of different building blocks (in this case only one building block is present in the system, namely glucose), the identity of oligomers (library members) formed in the DCL (α-, β- and γ-CD in this case), the number of equivalents

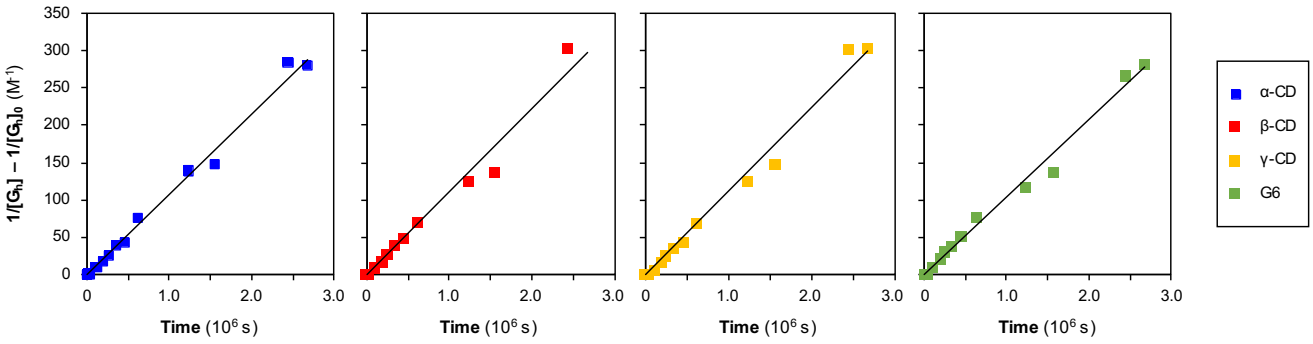
of building blocks used for each of these library members (6, 7 and 8 in this case), their relative formation constants, $K_{f(rel)}$ (determined from the *pseudo*-equilibrium composition in an untemplated library), and their individual association constants, K_a, to the template (determined using ^{1}H NMR spectroscopy titrations in this case). To simulate the 2:1 binding events, two mock library members defined as $(\alpha\text{-CD})_2$ and $(\beta\text{-CD})_2$ were implemented in the model, and their relative formation constants were defined as $(K_{f(rel)}{}^{\alpha\text{-CD}})^2$ and $(K_{f(rel)}{}^{\beta\text{-CD}})^2$, respectively. The association constant for these mock library members towards SDS (**1**) were defined by the product of both the first and second binding constant (i.e. $K_{a1} \cdot K_{a2}$).

S2. Supporting Figures



Supporting Figure 1. Distribution of both cyclic (α-, β-, γ-CD) and linear (G1-G8) α-1,4-glucans generated when CGTase acts upon different starting materials.
Relative concentrations by weight of α-1,4-glucans as a function of time, determined by HPLC-ELS analysis, in enzymatic reactions starting from **a**, α-CD; **b**, β-CD; **c**, γ-CD; **d**, **G6**. Lines are merely to guide the eye. Reaction conditions as in Fig. 2 of the main article.

a

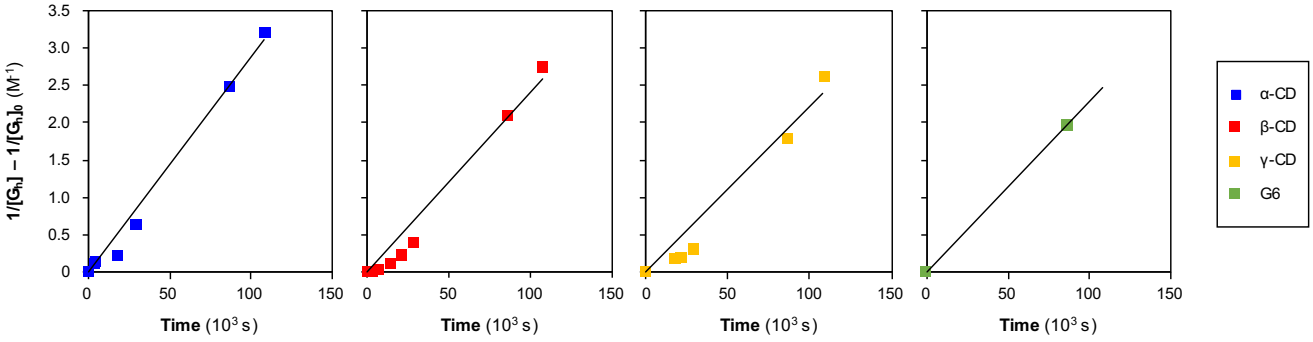


Apparent second-order rate constants ($k_{2(obs)}$) for the hydrolysis reaction by CGTase in absence of template*

Starting material	$k_{2(obs)}$ ($M^{-1}s^{-1}$)†	R^2
α-CD	$(1.081 \pm 0.017) \cdot 10^{-3}$	0.995
β-CD	$(1.11 \pm 0.03) \cdot 10^{-3}$	0.981
γ-CD	$(1.12 \pm 0.02) \cdot 10^{-3}$	0.992
G6	$(1.04 \pm 0.02) \cdot 10^{-3}$	0.994

* As determined from linear regression to $(1/[G_h] - 1/[G_h]_0)$ vs. time. † Listed as slope ± error on slope from the linear regression fit.

b



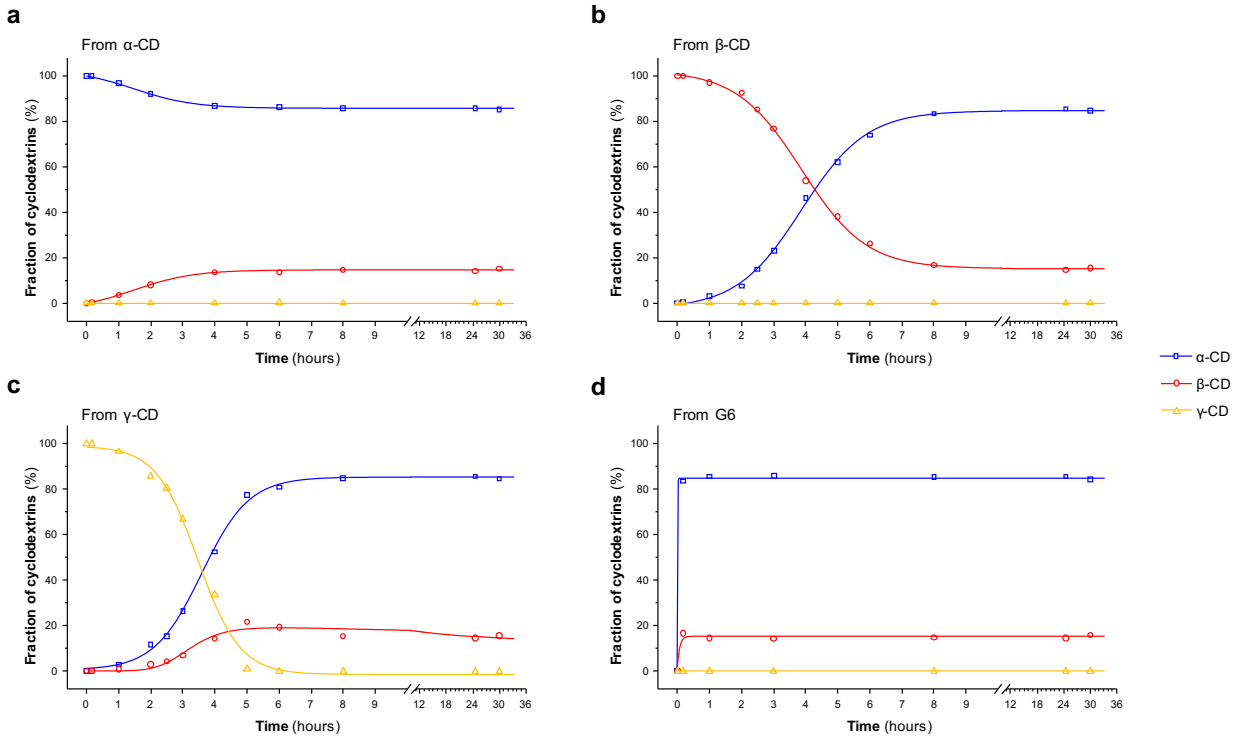
Apparent second-order rate constants ($k_{2(obs)}$) for the hydrolysis reaction by CGTase in presence of 2 (10 mM)*

Starting material	$k_{2(obs)}$ ($M^{-1}s^{-1}$)†	R^2
α-CD	$(2.89 \pm 0.10) \cdot 10^{-5}$	0.992
β-CD	$(2.40 \pm 0.12) \cdot 10^{-5}$	0.978
γ-CD	$(2.20 \pm 0.17) \cdot 10^{-5}$	0.971
G6	$2.28 \cdot 10^{-5}$ ‡	1 ‡

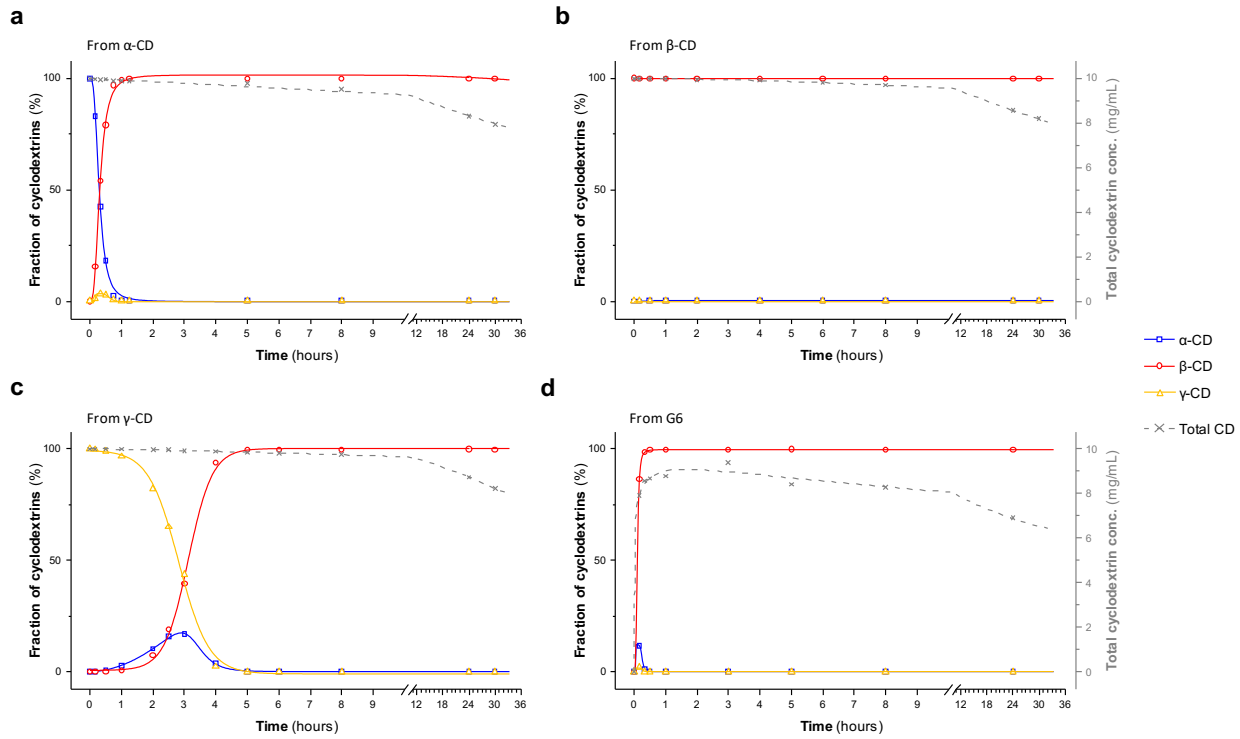
* As determined from linear regression to $(1/[G_h] - 1/[G_h]_0)$ vs. time (only data points after reaching equilibrium taken into account). † Listed as slope ± error on slope from the linear regression fit. ‡ Assessed from a single data point due to lack of sample material.

Supporting Figure 2. Kinetic analysis of the overall rate of the hydrolysis reactions afforded by CGTase on a range of starting materials.
a, Linear regression plots of $(1/[G_h] - 1/[G_h]_0)$ vs. time, t, for the untemplated libraries. According to the integrated second-order rate law $1/[G_h] - 1/[G_h]_0 = k_{2(obs)}t$, and thus the resulting slopes correspond to the apparent second-order rate constants for the sum of all hydrolysis reactions, which are tabulated along with the R^2-values from the best fit. Analysis performed on the data that is presented in Supporting Fig. 1. **b**, Linear regression plots of $(1/[G_h] - 1/[G_h]_0)$ vs. time for the libraries with 1-adamantane carboxylic acid **2** (10 mM) as template. Analysis performed on the data that is from the experiment presented in Supporting

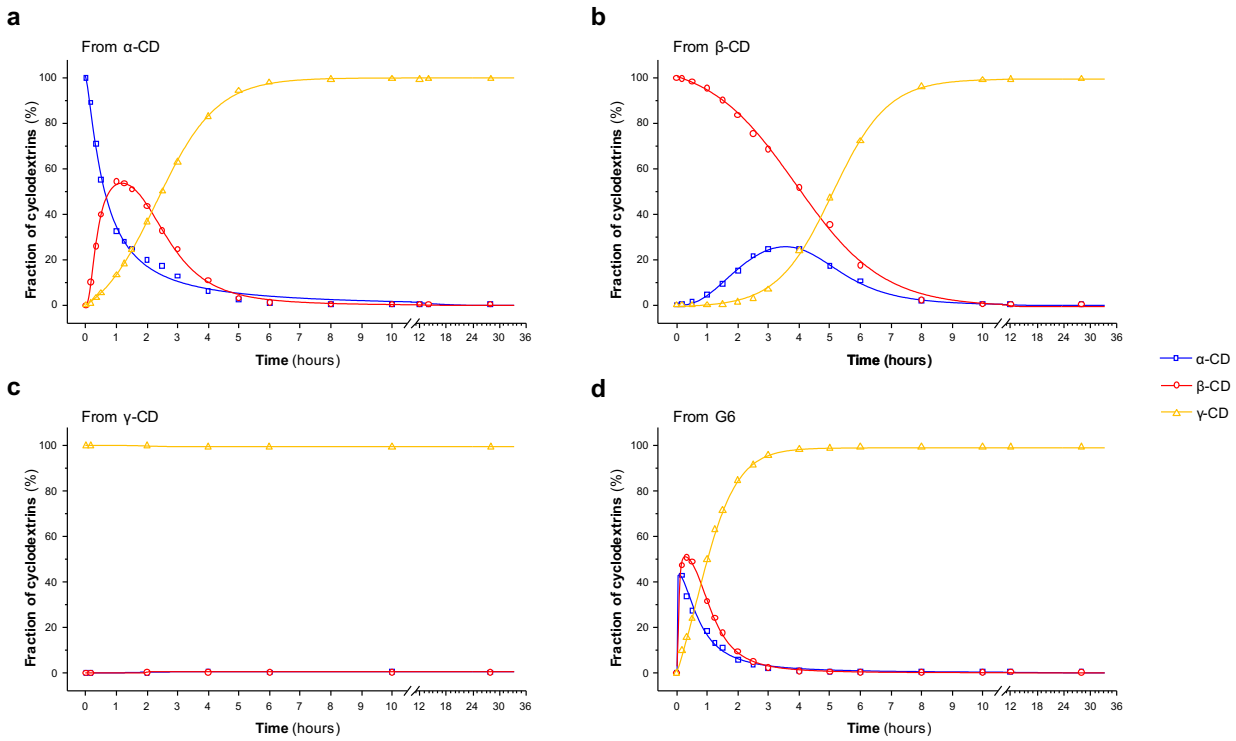
Fig. 4. Only data points after equilibrium was reached are taken into account (t ≥ 60 min for α-CD; all data points for β-CD; t ≥ 300 min for γ-CD; t ≥ 30 min for **G6**).



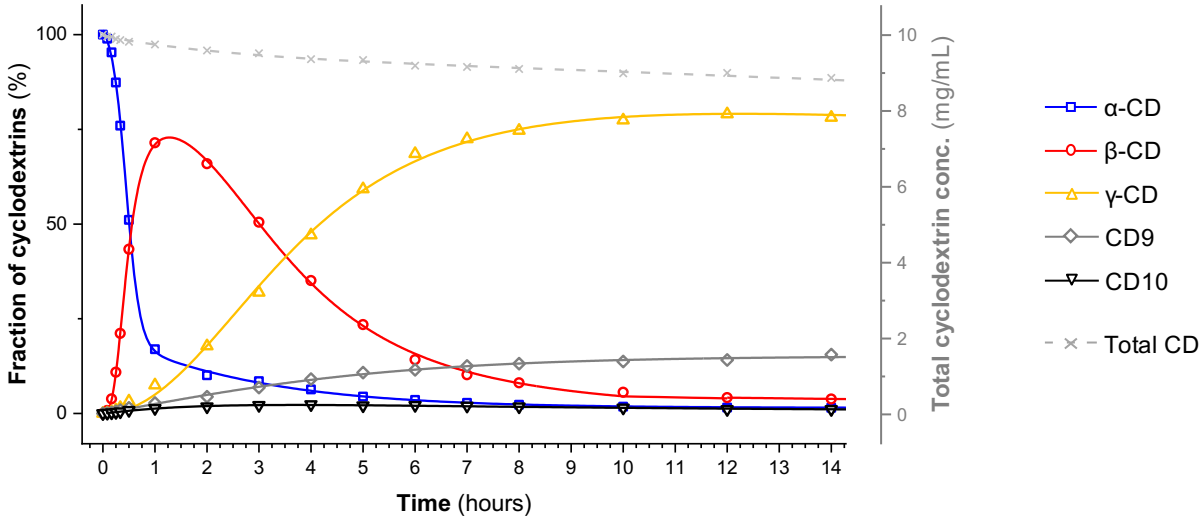
Supporting Figure 3. Sodium dodecyl sulfate (1) acts as a template for the formation of α-CD.
Distribution of α-, β-, and γ-CD formed in the presence of 10 mM sodium dodecyl sulfate (**1**) as a function of time when CGTase acts upon different α-1,4-glucan starting materials: **a**, α-CD; **b**, β-CD; **c**, γ-CD; **d**, G6. Lines are merely to guide the eye. Reaction conditions as in Fig. 2 of the main article.



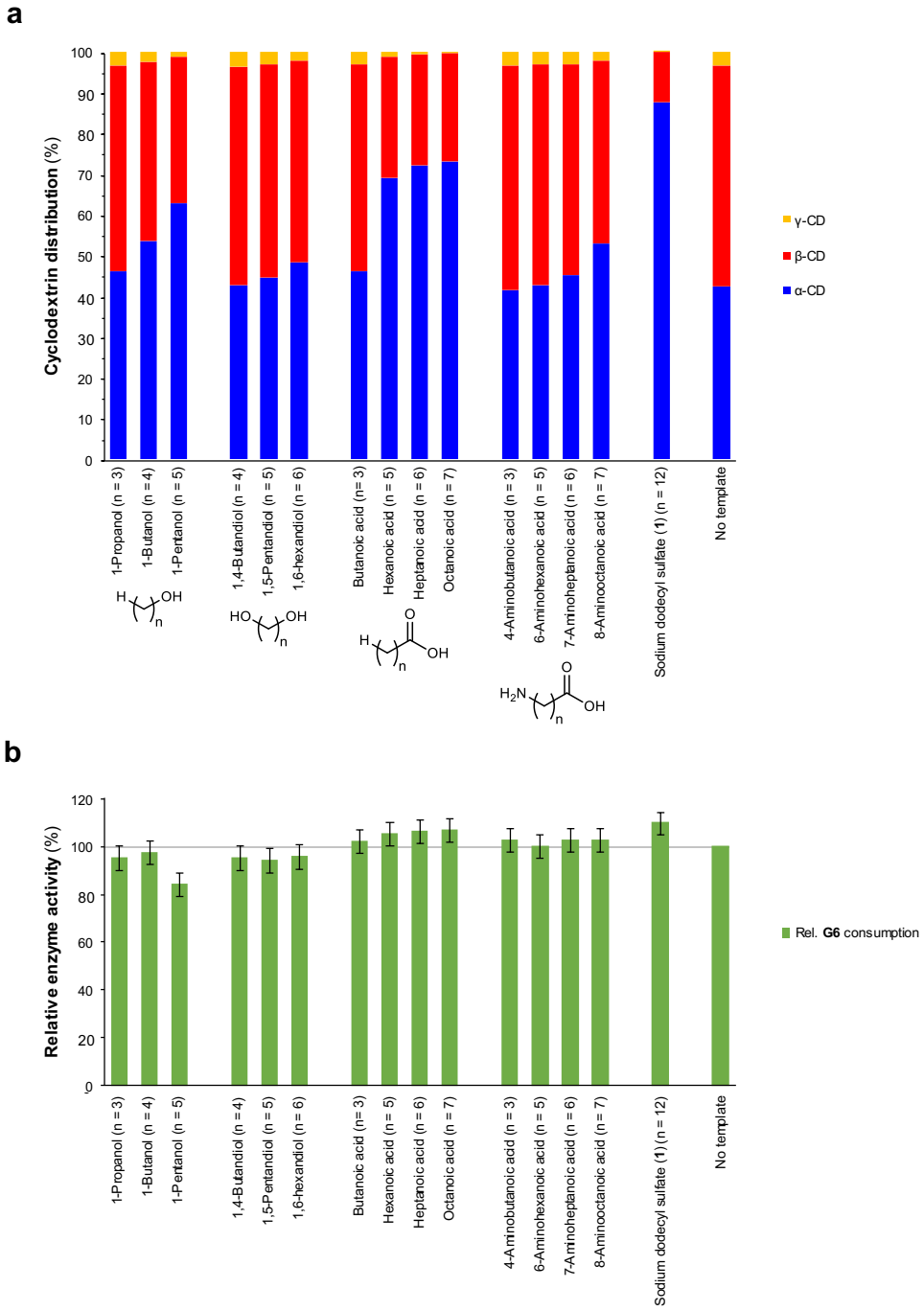
Supporting Figure 4. 1-Adamantane carboxylic acid (2) acts as a template for the formation of β-CD.
Distribution of α-, β-, and γ-CD (left axis) and total concentration of cyclodextrins (right axis) formed in the presence of 10 mM adamantane carboxylate (2) as a function of time when CGTase acts upon different α-1,4-glucan starting materials: a, α-CD; b, β-CD; c, γ-CD; d, G6. Lines are merely to guide the eye. Reaction conditions as in Fig. 2 of the main article.



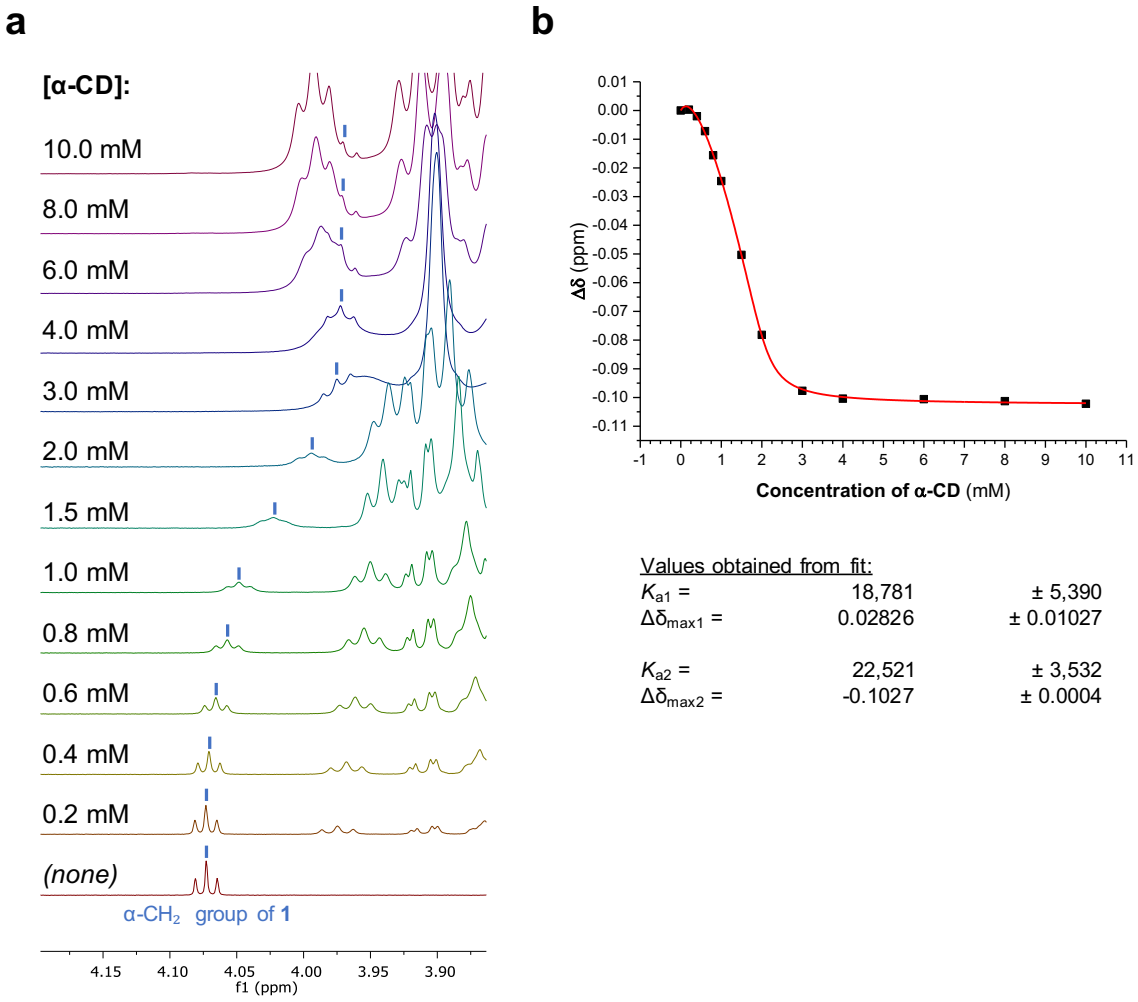
Supporting Figure 5. Sodium tetraphenylborate (3) acts as a template for the formation of γ-CD.
Distribution of α-, β-, and γ-CD formed in the presence of 10 mM sodium tetraphenylborate (**3**) as a function of time when CGTase acts upon different α-1,4-glucan starting materials: **a**, α-CD; **b**, β-CD; **c**, γ-CD; **d**, **G6**. Lines are merely to guide the eye. Reaction conditions as in Fig. 2 of the main article.



Supporting Figure 6. Cesium dodecaiodo dodecaborate (4) acts as a template for the formation of "non-native" large-ring cyclodextrins CD9 and CD10.
Distribution of cyclodextrins (α-, β-, and γ-CD as well as CD9 and CD10) (left axis) and total concentration of cyclodextrins (right axis) formed in the presence of 10 mM $Cs_2[B_{12}I_{12}]$ (**4**) when CGTase acts upon α-CD. Lines are merely to guide the eye. Reaction conditions as in Fig. 2 of the main article.

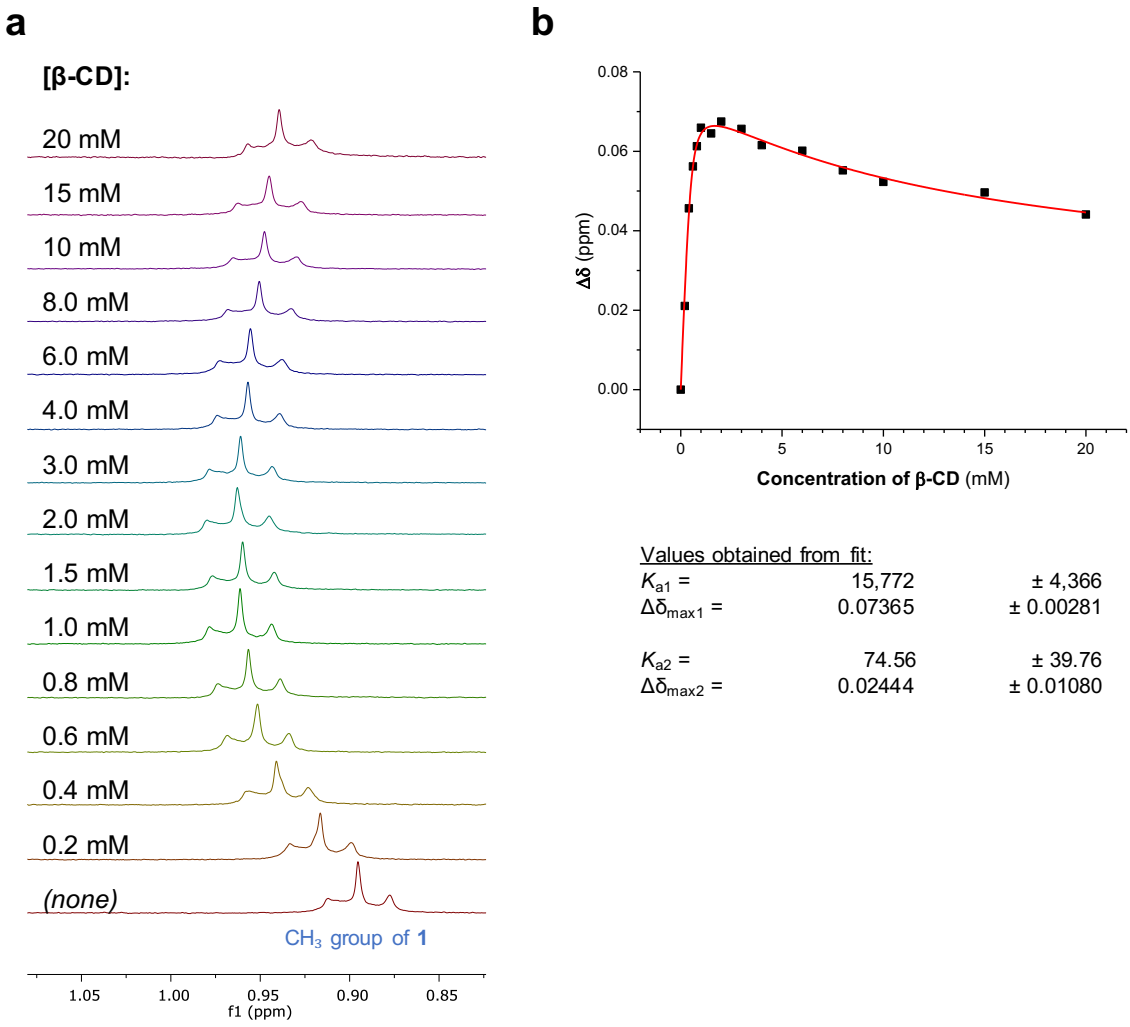


Supporting Figure 7. Template screening for the purpose of amplifying α-CD production. **a**, Distribution of cyclodextrins formed when CGTase acts on **G6** in presence of the indicated template (10 mM) monitored after 240 mins. Conditions and analysis otherwise as in Fig. 2 of the main article. **b**, Enzyme activity in % measured as consumption of **G6** relative to a reference (no template) during five minutes after incubation of CGTase with the indicated template (10 mM) for at least three hours at 25 °C (estimated 5 percentage point error). Some of the stronger amplifiers led to a perceived increase in activity, likely due to the kinetic trapping of the α-CD complex formed, leading to an apparent higher **G6** consumption.



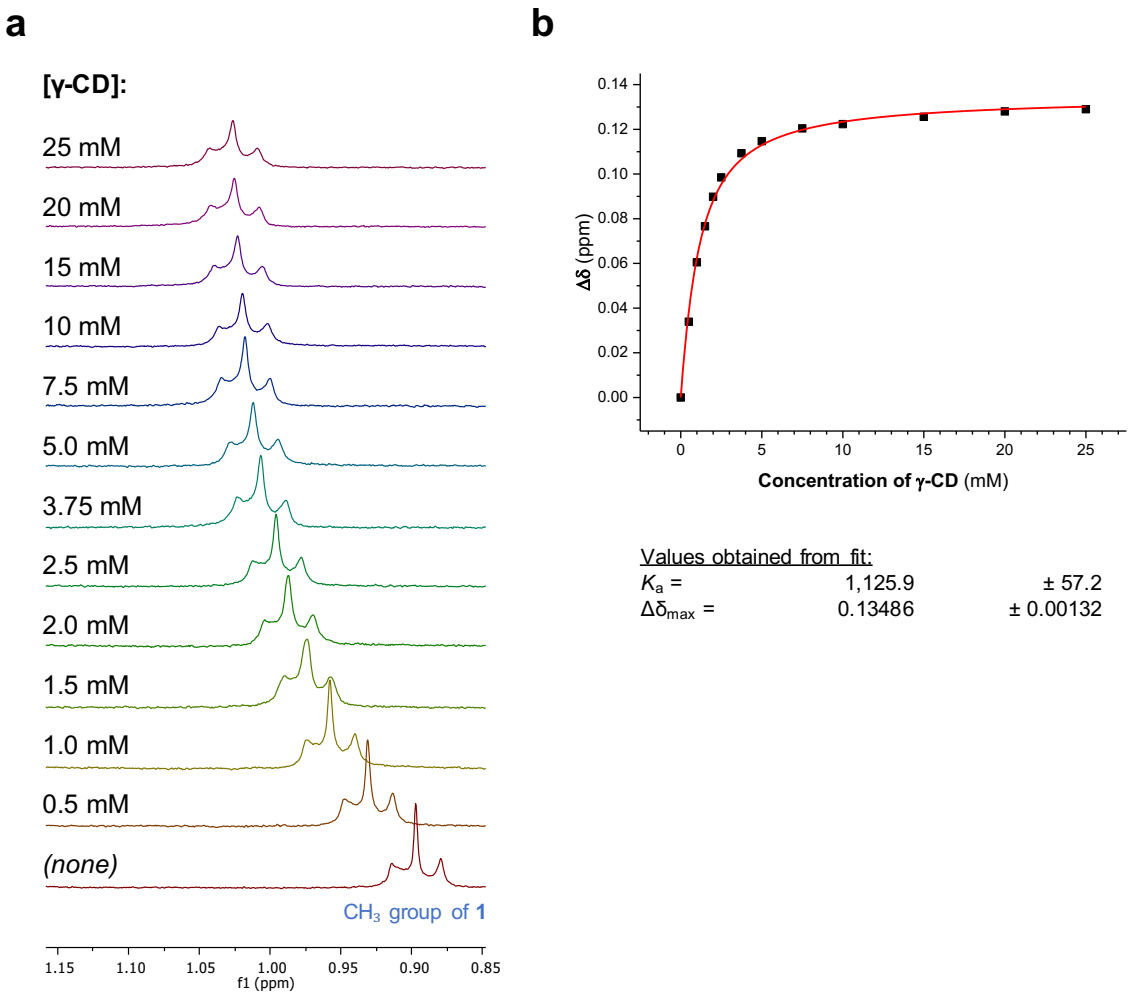
Supporting Figure 8. ^{1}H NMR spectroscopy titration of sodium dodecyl sulfate (1) with α-CD shows 2:1 binding.

a, Partial ^{1}H NMR spectra showing a change in the chemical shift of the α-CH_2 group of **1** with increasing concentrations of α-CD. The α-CH_2 group of **1** was followed, and the assigned peaks are shown with a blue marker. **b**, The resulting best fit to a 2:1 α-CD/**1** binding model. The titration was performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O. The concentration of **1** was kept constant at 1.00 mM.



Supporting Figure 9. ^{1}H NMR spectroscopy titration of sodium dodecyl sulfate (1) with β-CD shows 2:1 binding.

a, Partial ^{1}H NMR spectra showing a change in the chemical shift of the CH_3 group of **1** with increasing concentrations of β-CD. **b**, The resulting best fit to a 2:1 β-CD/**1** binding model. The titration was performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O. The concentration of **1** was kept constant at 0.50 mM.



Supporting Figure 10. ^{1}H NMR spectroscopy titration of sodium dodecyl sulfate (1) with γ-CD shows 1:1 binding.
a, Partial ^{1}H NMR spectra showing a change in the chemical shift of the CH_3 group of **1** with increasing concentrations of γ-CD. **b**, The resulting best fit to a 1:1 γ-CD/**1** binding model. The titration was performed in sodium phosphate buffer (50 mM) at pH 7.5 in D_2O. The concentration of **1** was kept constant at 0.50 mM.

S3. Supporting References

[S1] Tiritiris, I. & Schleid, T. Die Kristallstrukturen der Dicaesium-Dodekahalogeno-closo-Dodekaborate $Cs_2[B_{12}X_{12}]$ (X = Cl, Br, I) und ihrer Hydrate. *Zeitschrift fur Anorg. und Allg. Chemie* **630**, 1555–1563 (2004).

[S2] Hargrove, A. E., Zhong, Z., Sessler, J. L. & Anslyn, E. V. Algorithms for the determination of binding constants and enantiomeric excess in complex host:guest equilibria using optical measurements. *New J. Chem.* **34**, 348 (2010).