Supporting Information for:

Crystalline Cyclophane-Protein Cage Frameworks

*Ngong Kodiah Beyeh[1,2,3], Nonappa[1], Ville Liljeström[1], Joona Mikkilä[4], Antti Korpi[4], Davide Bochicchio[5], Giovanni M. Pavan[5], Olli Ikkala[1], Robin H. A. Ras[1,4], Mauri A. Kostiainen[1,4]**

[1] HYBER Centre of Excellence, Department of Applied Physics, Aalto University, FI-00076 Aalto, Finland.

[2] Department of Chemistry and Biochemistry, University of Windsor, N9B 3P4 Windsor, Canada.

[3] Department of Chemistry, Oakland University, 146 Library Drive, Rochester, Michigan 48309-4479, USA

[4] Department of Bioproducts and Biosystems, Aalto University, FI-00076 Aalto, Finland.

[5] Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, CH-6928, Manno, Switzerland.

* Email: mauri.kostiainen@aalto.fi

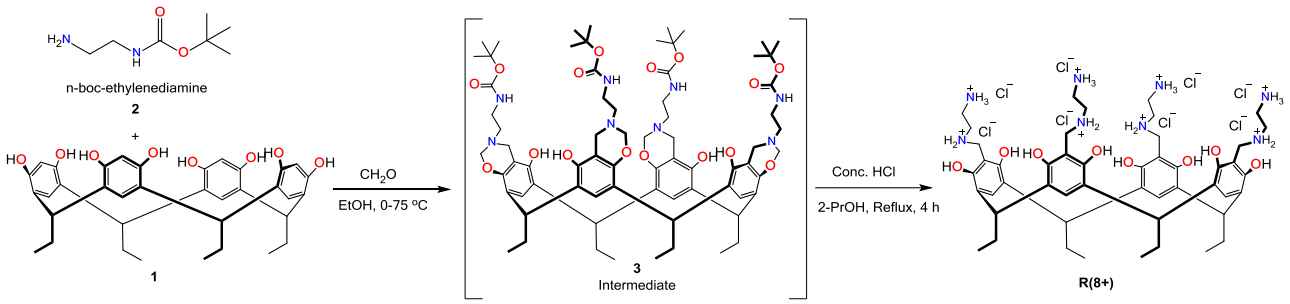
Table of Contents

1. Synthesis

The receptors **R(4-)**[1], **R(4+)**[2,3] and **P(10+)**[4] were synthesized according to reported procedures. Experimental details for the synthesis and characterization data for **R(8+)**, **R(16+)** and **P(5+)** are given below. All reagents were purchased from commercial sources and used as received unless otherwise stated. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker Avance DRX 400 (400 MHz for ^{1}H and 100 MHz for ^{13}C) spectrometer. All signals are given as δ values in ppm using residual solvent signals as the internal standard. Coupling constants are given in Hz. Melting points were determined with a Stuart SMP30 melting point apparatus. The mass spectrometric experiments were performed with a Bruker UltrafleXtreme MALDI-TOF/TOF-mass spectrometer and a Micromass LCT Electrospray ionization–Time-of-flight instrument. First, profile mass spectra were obtained (*e.g.* Figure S2 and S6) followed by high-resolution spectra using lock mass from an internal standard.

Calculated molecular weights for the studied hosts: **R(16+)** 2055.5 g mol^{-1}, **P(10+)** 2271.2 g mol^{-1}, **R(8+)** 1180.8 g mol^{-1}, **P(5+)** 2311.9 g mol^{-1}, **R(4+)** 1311.4 g mol^{-1}, **R(4-)** 1008.9 g mol^{-1}.

1.1 Synthesis of receptor R(8+)



Scheme S1. Synthesis of *N*-alkyl ammonium resorcinarene chloride **R(8+)**.

Into a solution of C_{ethyl}-resorcinarene **1** (3 g, 4.99 mmol) in ethanol (60 ml) and excess formaldehyde (4 ml), *N*-boc-ethylenediamine (3.23 ml, 20.47 mmol) in ethanol (10 ml) was added drop wise. The mixture was stirred at room temperature for 24 h. The precipitated crude product was not purified and used as such in the next step. Into a solution of the crude tetrabenzoxazine intermediate **3** (2.5 g, 1.86 mmol) in 2-propanol (50 ml), conc. HCl (10 ml) and H_2O (10 ml) were added. The mixture was refluxed for 4 h. After water and formaldehyde were removed by azeotropic distillation with chloroform (100 ml), the remaining 2-propanol was evaporated. The product was triturated with diethyl ether (100 ml). The

precipitated product was collected and dried *in vacuo*. This product was then recrystallized first in acetonitrile and then in methanol to give the final product **R(8+)** (1.14 g, 52 %). m.p. > 300 ºC; MALDI-TOF: *m/z* $C_{42}H_{49}N_2O_8^+$ 709.492 [**R(8+)**-7H-8Cl-3A]$^+$. ESI-TOF-HRMS: *m/z* Found 445.2827 $C_{48}H_{74}N_8O_8^{+2}$ [**R(8+)**-8Cl-6H]$^{+2}$, (-1.8 mDa, -4 ppm); calc. 445.2809. ^{1}H NMR (400 MHz, 300 K in [D_6]DMSO) δ (ppm): 0.86 (t, J=7.50 Hz, 12H, CH_3), 2.30 (m, 8H, CH_2), 3.20 (br, 16H, NCH_2CH_2N), 4.15 (br, 8H, $ArCH_2N$), 4.26 (t, J=7.50 Hz, 4H, CH), 7.67 (s, 4H, ArH), 8.40 (br, 12H, NH_3), 9.11 (s, 8H, NH_2), 9.54 (s, 8H, OH); ^{13}C NMR: (100 MHz, 300 K in [D_6]DMSO) δ (ppm) = 12.4, 25.5, 35.0, 36.4, 41.2, 43.8, 73.1, 108.5, 126.3, 150.2.

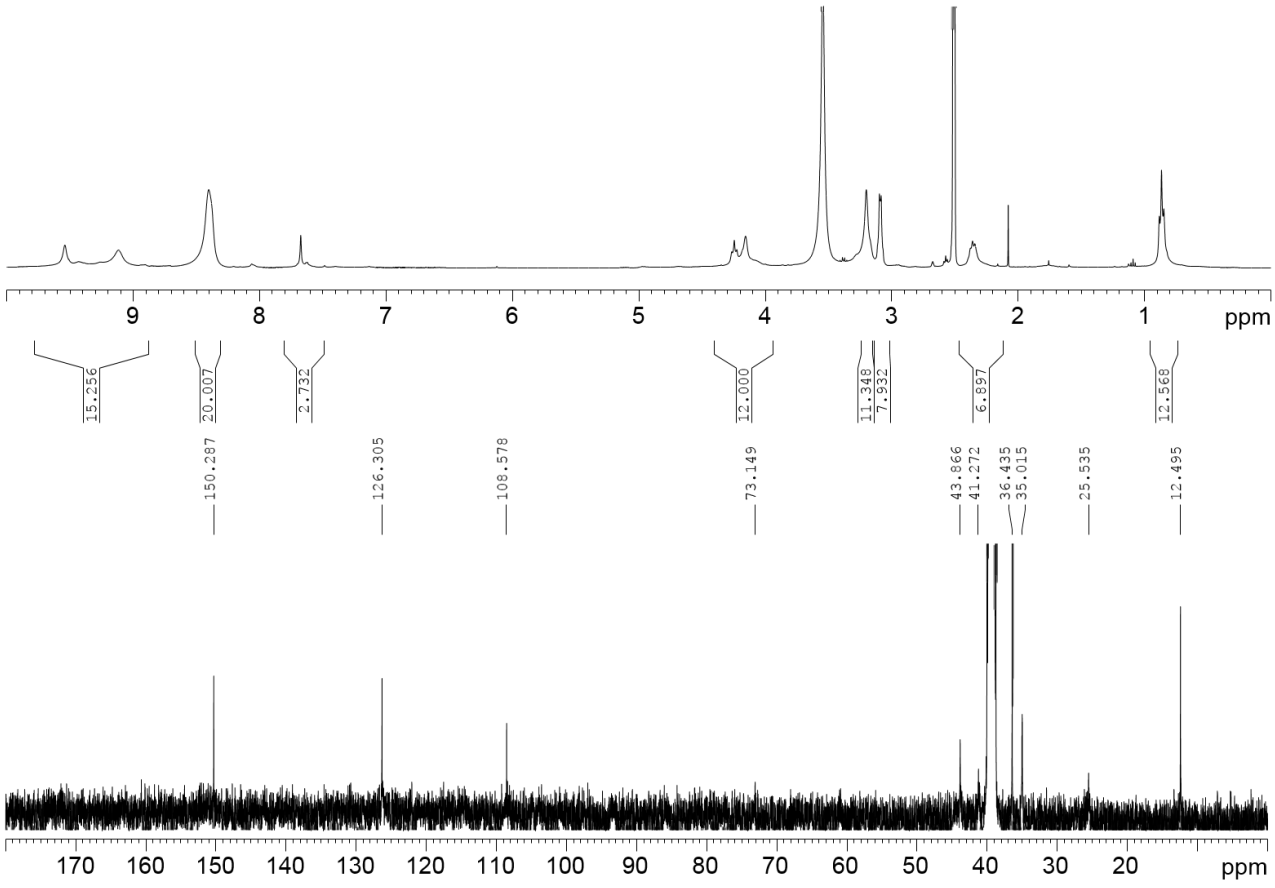


Figure S1. ^{1}H and ^{13}C NMR spectra of *N*-alkyl ammonium resorcinarene chloride **R(8+)** in [D_6]DMSO at 300 K.

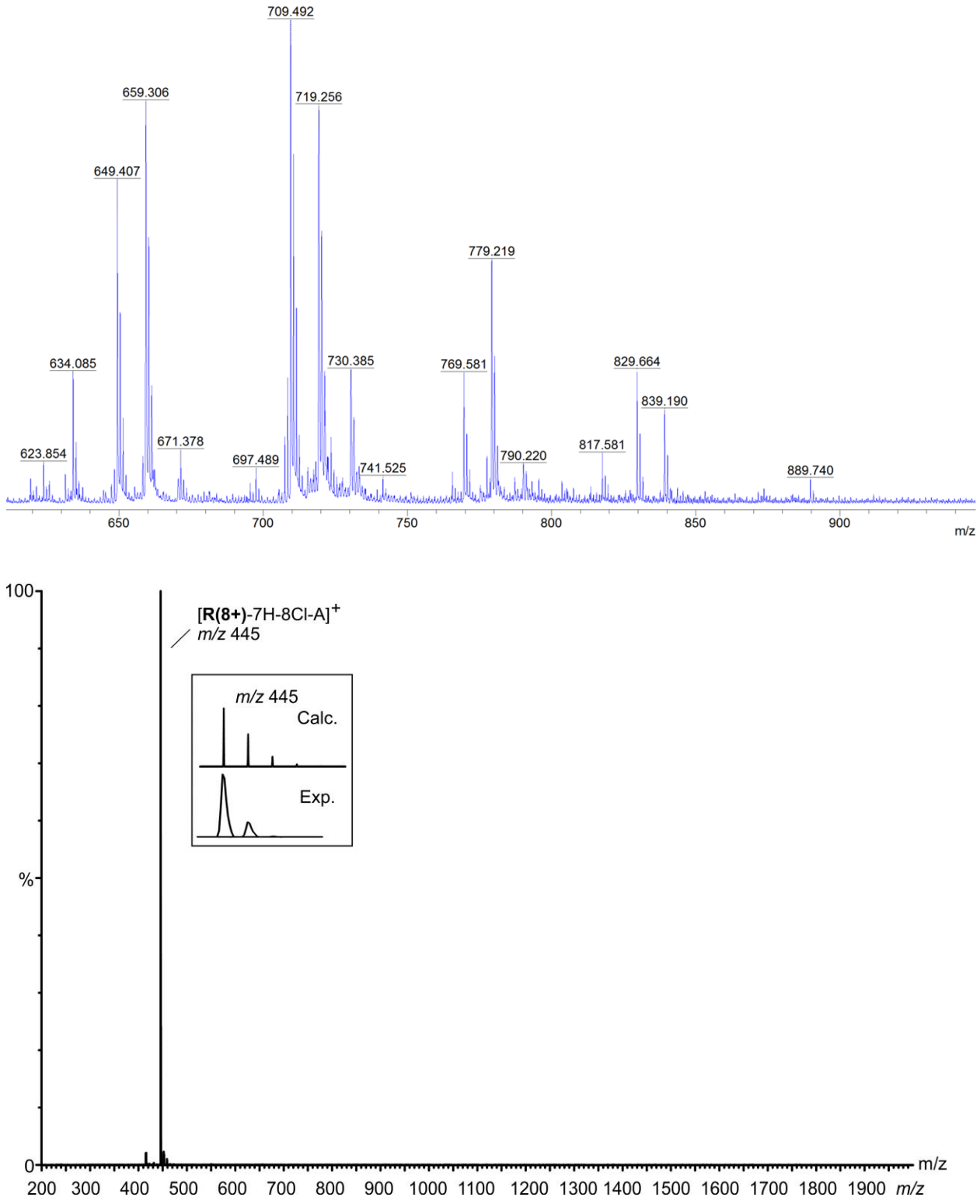
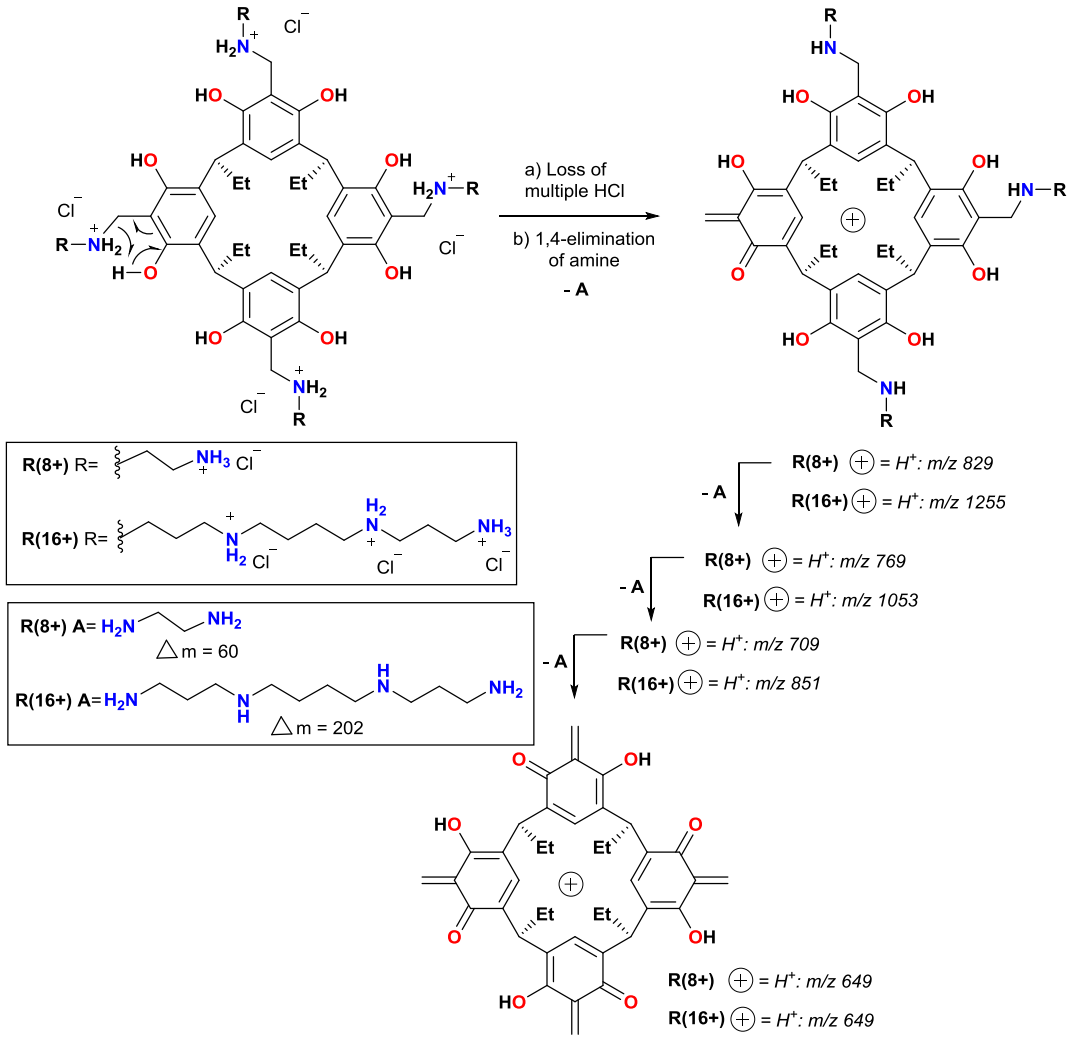
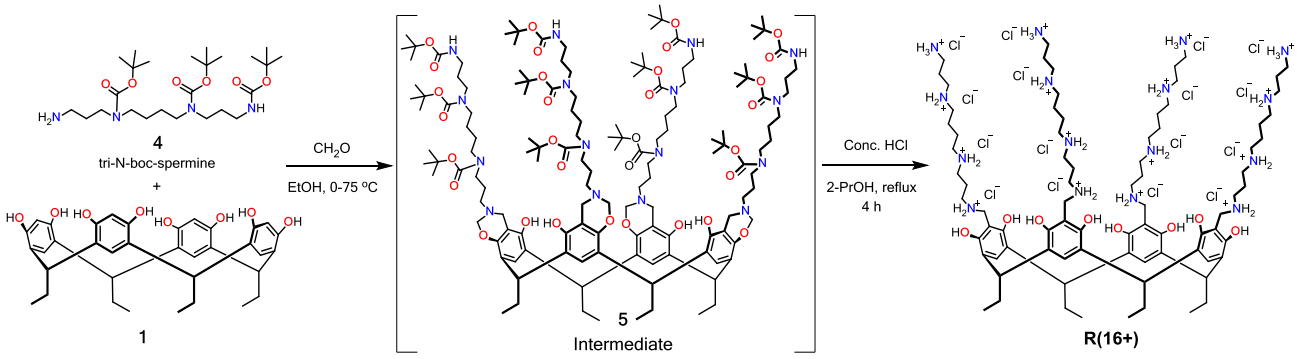


Figure S2. Mass spectra of *N*-alkyl ammonium resorcinarene chloride **R(8+).** Top: MALDI-TOF spectrum showing loss of HCl and multiple fragmentation through 1,4-elimination of amine. Bottom: ESI-TOF mass spectrum.



Scheme S2. Loss of multiple HCl and fragmentation mechanism and pathways for the consecutive 1,4-eliminations of the amine from **R(8+)** and **R(16+)**.

1.2. Synthesis of receptor R(16+)



Scheme S3. Synthesis of *N*-alkyl ammonium resorcinarene chloride **R(16+)**.

Into a solution of C_{ethyl}-resorcinarene (0.46 g, 0.776 mmol) in ethanol (25 ml) and excess formaldehyde (2 ml), the tri-N-boc-spermine (1.6 g, 3.18 mmol) in ethanol (20 ml) was added drop wise. The mixture was stirred at room temperature for 24 h. The mixture was partially purified *via* column chromatography (Eluent: DCM/MeOH/EtOAc 85/10/5 v/v). The semi-purified intermediate was used as such in the next step. Into a solution of the tetrabenzoxazine intermediate **5** (1.27g, 0.47 mmol) in 2-propanol (15 ml), conc. HCl (6 ml) and H_2O (6 ml) were added. The mixture was refluxed for 4 h. After the water and formaldehyde were removed by azeotropic distillation with chloroform (100 ml), the remaining 2-propanol was evaporated. The product was a molten smear on the flask. The crude product was dried *in vacuo* and then recrystallized in MeCN to give the final product **R(16+)** (0.77 g, 82 %). m.p. > 300 °C; MALDI-TOF: *m/z* $C_{50}H_{67}N_4O_8^+$ 851.505 [**R(16+)**-15H-16Cl-3A]$^+$. ^{1}H NMR (400 MHz, 300 K in D_2O) δ (ppm): 0.78 (br, 12H, CH_3), 1.67 (m, 24H, CH_2), 1.97 (m, 24H, NCH_2), 3.00 (m, 68H, CH_2N), 4.14 (m, 12H, CH,$ArCH_2$), 7.09 (br, 4H, ArH); ^{13}C NMR: (100 MHz, 300 K in D_2O) δ (ppm) = 11.6, 22.5, 22.7, 23.7, 36.5, 43.9, 44.5, 44.6, 46.9, 81.7, 125.9, 126.7, 150.4.

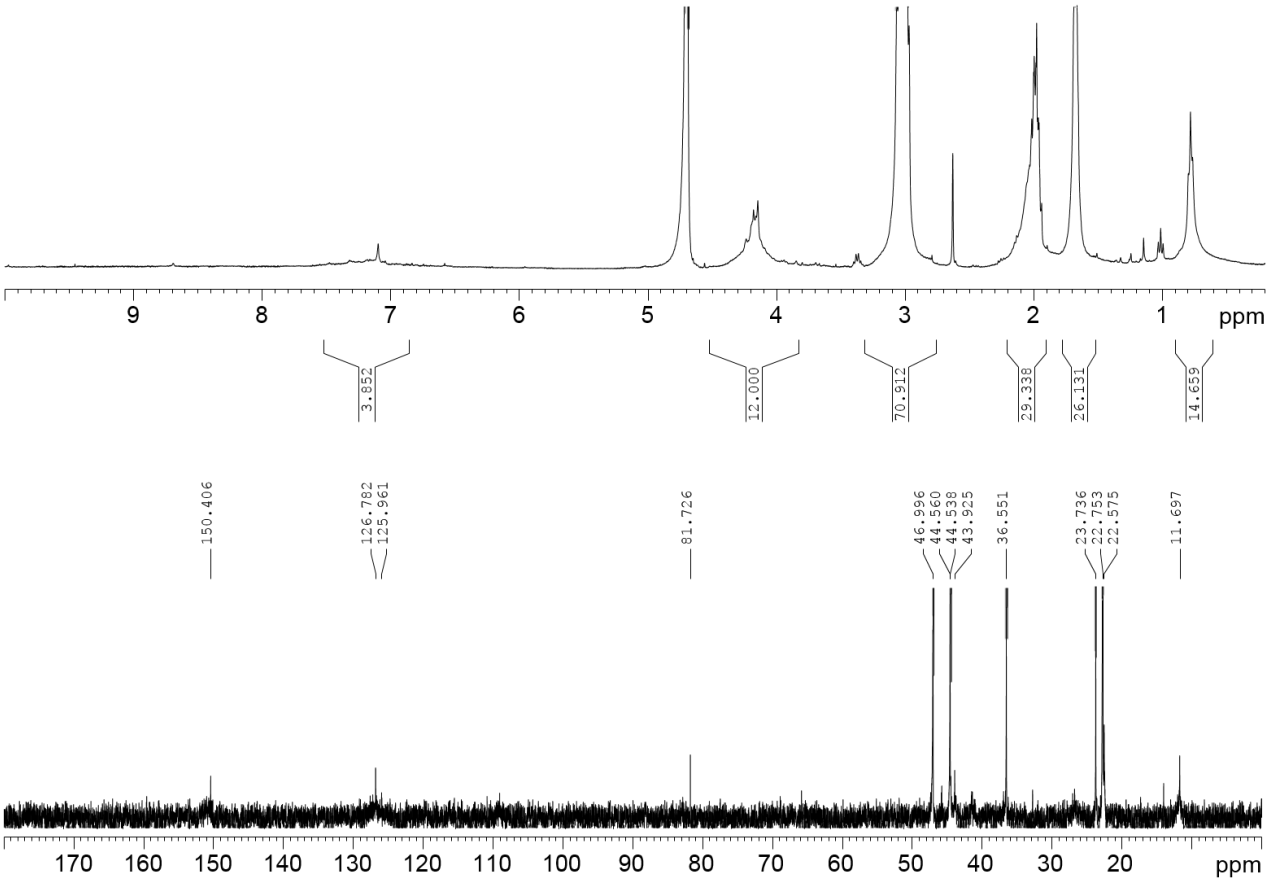


Figure S3. ^{1}H and ^{13}C NMR spectra of *N*-alkyl ammonium resorcinarene chloride **R(16+)** D_2O at 300 K.

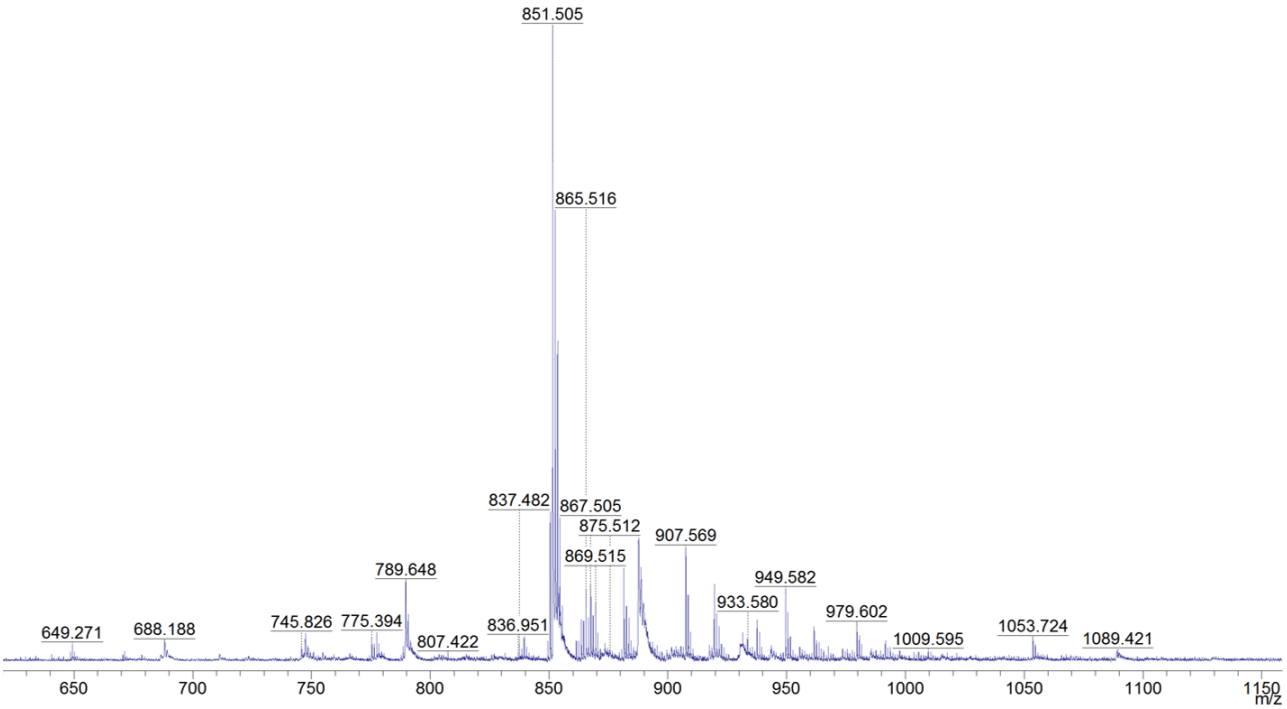
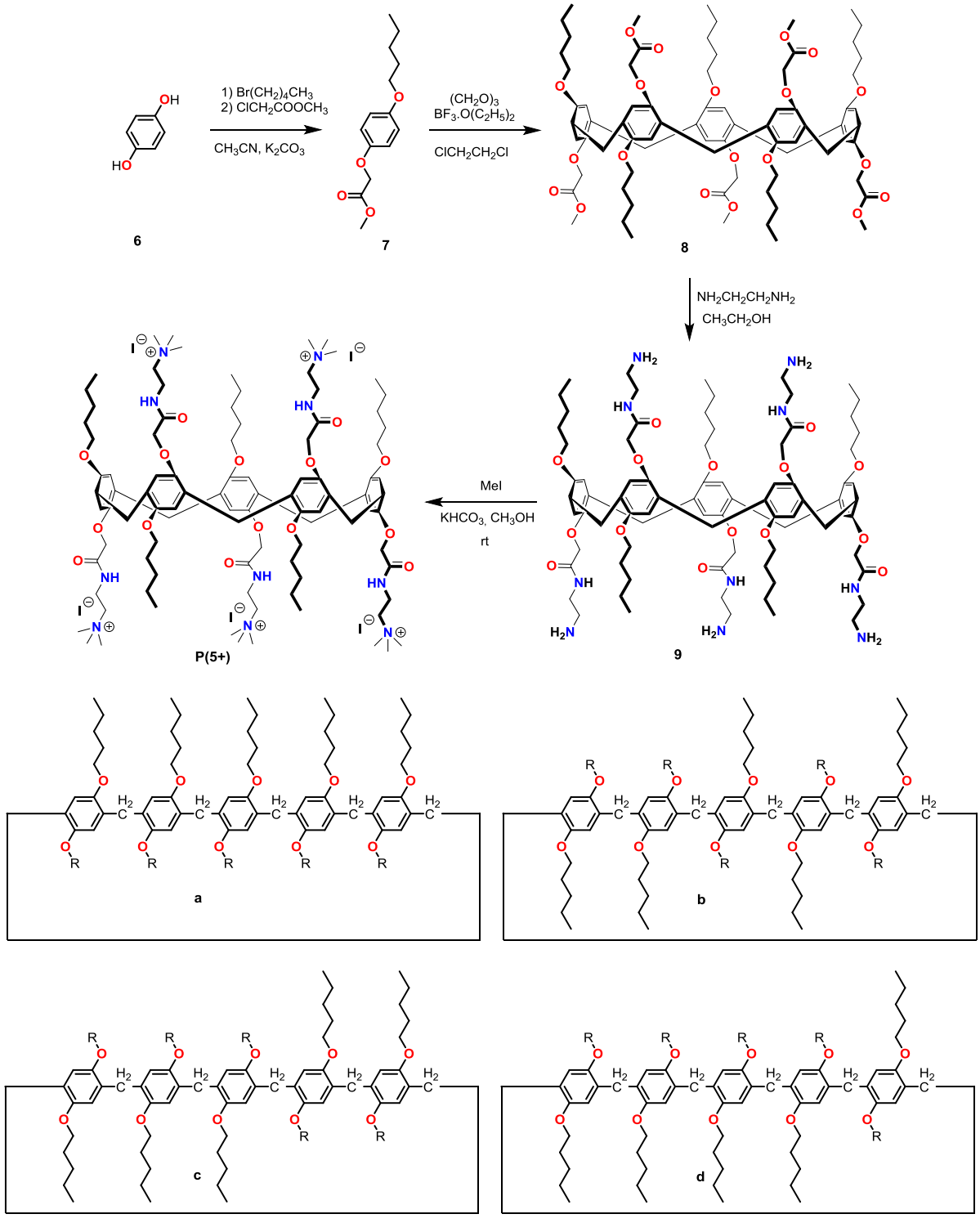


Figure S4. MALDI-TOF mass spectrum of *N*-alkyl ammonium resorcinarene chloride **R(16+)** showing loss of HCl and multiple fragmentation through 1,4-elimination of amine.

1.3 Synthesis of receptor P(5+)



Scheme S4. Synthesis of penta-cationic pillar[5]arene **P(5+)**.

The compounds **7**, **8** and **9** were synthesized according to reported procedures.[5] Pillar[5]arene **8**, **9** and **P(5+)** has 4 constitutional isomers **a**, **b**, **c**, and **d** (Scheme S4). The isomers were not separated. Methyl iodide (0.033 g, 0.236 mmol) and $KHCO_3$ (0.023 g, 0.236 mmol) were added to the Pillar[5]arene **9** (0.069 g, 0.0472 mmol) in methanol and stirred for 24 hours at room temperature under N_2 atmosphere. The excess methyl iodide was distilled off and neutralized in 10 % NH_4OH solution. The solvent was evaporated and the mixture triturated with diethyl ether (20 ml) to give the penta-quaternized pillar[5]arene iodide **P(5+)** (0.091 g, 84 %). m.p. > 300 ^oC; ESI-TOF-HRMS: *m/z* Found 509.9956 $C_{87}H_{133}N_8O_{15}^{+3}$ [**P(5+)**-5I-A-B]$^{+3}$, (0.2 mDa, 0.4 ppm); calc. 509.9958. ^1H NMR (400 MHz, 298 K in CD_3OD) δ (ppm): 0.93-2.0 (m, 85H, CH_3, CH_2, OCH_2), 3.30-4.80 (br, 55H, Ar-CH_2, NCH_3), 6.89 (br, 10H, ArH).

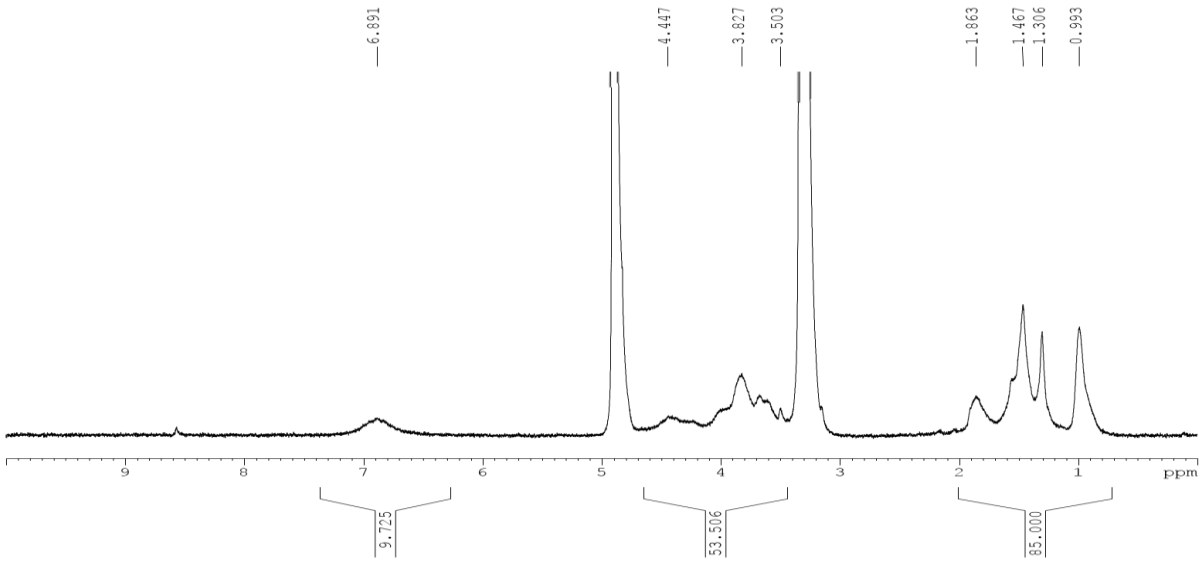


Figure S5. ^1H NMR spectra of penta-quaternized pillar[5]arene iodide **P(5+)** in CD_3OD at 298 K.

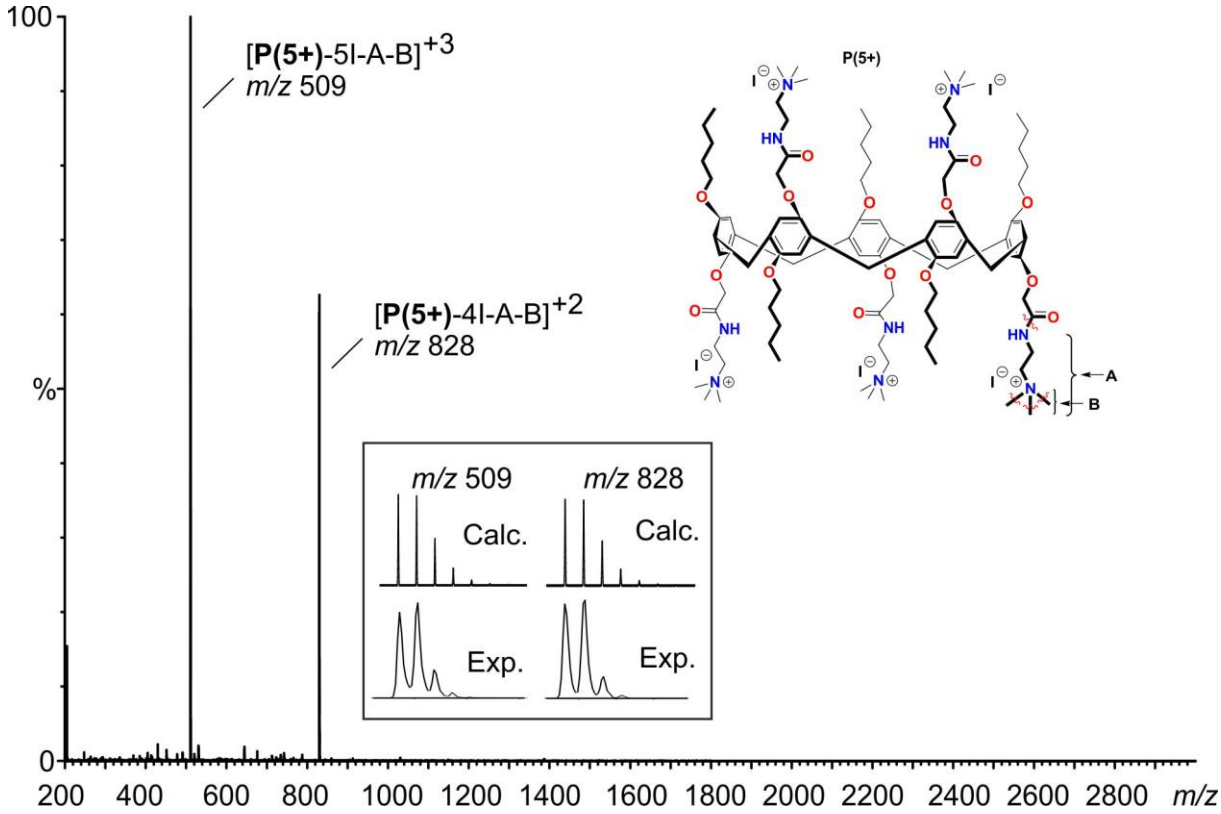


Figure S6. ESI-TOF mass spectrum of penta-quaternized pillar[5]arene iodide **P(5+)**.

2. Other experimental methods

Dynamic light scattering (DLS) and ζ-potential measurements

The hydrodynamic diameter (D_h) of the assemblies was measured using a Malvern Instruments DLS device (Zetasizer Nano ZS Series) with a 4 mW He-Ne ion laser at a wavelength of 633 nm and an avalanche photodiode detector at an angle of 173°. Experiments were carried out at 25 °C. Stability at different temperatures was studied between 10–90 °C. Plastibrand semi-micro PMMA cuvettes and Malvern disposable ζ-cells were used for the size and electrophoretic mobility measurements respectively. Zetasizer software (Malvern Instruments) was used to obtain the particle size distributions and ζ-potentials using the following sample preparation: 500 µl of aqueous aFT solution (100 mg ml^{-1}) was titrated with the host (10–0.1 mg ml^{-1}) to reach the desired ratio (addition did not exceed 5 % of sample volume and hence no correction for dilution was made), which was finally titrated with 0.1-5 M NaCl to disassemble the complex. 10 mM NaAc and Tris-HCl buffers were used in studying the stability at different pH values.

Cryogenic transmission electron microscopy (cryo-TEM)

The cryo-TEM images were collected using JEM 3200FSC field emission microscope (JEOL) operated at 300 kV in bright field mode with Omega-type Zero-loss energy filter. The images were acquired with Gatan Digital Micrograph software while the specimen temperature was maintained at -187 °C. The cryo-TEM samples were prepared by placing 3 µl aqueous dispersion of the sample on a 200 mesh copper grid with holey carbon support film (CF-Quantifoil) and plunge-frozen using Vitrobot with 3 s blotting time under 100 % humidity.

Small-angle X-ray scattering (SAXS)

The aqueous samples were sealed between two Kapton foils inside a metal washer. The sample environment was evacuated to reduce background scattering from air. The SAXS was measured using a Bruker Microstar microfocus rotating anode X-ray source (Cu Kα radiation, λ = 1.54 Å). The beam was monochromated and focused by a Montel multilayer focusing monochromator (Incoatec). The X-ray beam was further collimated by four collimation slits (JJ X-ray) resulting in a final spot size of less than

1 mm at the sample position. A Hi-Star 2D area detector (Bruker) was used to collect the scattered intensity. Sample-to-detector distance was 1.59 m, and a silver behenate standard sample was used for the calibration of the length of the scattering vector q. One-dimensional SAXS data was obtained by azimuthally averaging the 2D scattering data and the magnitude of the scattering vector q is given by $q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle. The structure factor in Fig. 2e was calculated with PowderCell[6], and the fitted model consists of a constant background, a power law ($const. \cdot q^{-\alpha}$) and Gaussian peaks fitted to the peak positions of the structure factor $S(q)$.

Sample preparation for SAXS, cryo-TEM, UV-Vis and optical microscopy

Cyclophane-aFT samples were prepared by combining 6 µl of aqueous aFT solution (15 mg ml^{-1}, *Pyrococcus furiosus*, Molecular Links Rome), 3.5 µl 20 mM Tris pH 8.5 buffer, 1.5 µl of 0–400 mM NaCl solution and 4 µl of 10 mg ml^{-1} cyclophane in water, in this order. The samples were gently mixed with a pipette and allowed to set.

All-atom simulations

All simulations were conducted using the GROMACS 5.1.2 software and the PLUMED 2 plugin.[7,8] The atomistic model for **P(10+)** and aFT were parametrized with the General Amber Force Field (GAFF) and the Amber Force Field *ff12SB*, respectively.[9] The atomistic model for the ferritin cage was obtained from the Protein Data Bank (PDB entry: 2JD6). The electrostatic potential for the aFT protein trimer was calculated with the PDB2PQR web portal.[10] For the **P(10+)**-aFT molecular recognition modelling studies we selected a portion of the full cage enclosing one pore in the threefold symmetry axis and its immediate neighbourhood (chains 4, 8, 9 M, Q, R from original 2JD6 structure). The 1:1 **P(10+)**-aFT system was then immersed into a periodic simulation box filled with explicit TIP3P[11] water molecules and Na+ counterions (Table S1).

Table S1: Details of the simulated systems.

System	N° of simul.	N° of atoms	N° of P(10+)	N° of water molecules	N° of ions	Simulation time (ns)
In water:						
1 aFT pore + P(10+)	1	150986	1	44755	50	30
1 aFT pore + P(10+)	4	150986	1	44755	50	8
2 aFT portions + P(10+)	1	232562	1	66455	110	30
In vacuum:						
1 aFT + 8 P(10+)	1	67864	8	-	160	150
2 aFT + 15 P(10+)	1	135483	15	-	330	30

The related 1:2 **P(10+)**-aFT model was obtained by addition of a mirror pore section on the opposite side of **P(10+)** starting from the pre-equilibrated 1:1 systems, and following re-solvation. All these AA-MD simulations lasted 30 ns and were conducted in NPT conditions (constant N: number of particles, P: pressure and T: temperature) at the temperature of 300 K and 1 atm of pressure. The **P(10+)**-aFT pore interaction non-bond energy data extracted directly from the AA-MD trajectories, were used to assess binding equilibration during the runs (Figure 4). The **P(10+)** binding energy (ΔG) was calculated from metadynamics[12] simulations biasing **P(10+)** disassembly from the cage run with GROMACS 5.1.2 and the PLUMED 2 plugin.[13] We used the distance of **P(10+)** from the pore (measured perpendicularly to the surface) as the collective variable (CV) in the metadynamics simulation (HILLS height of 0.1 kcal mol^{-1}, Gaussian SIGMA of 0.05 nm, deposition time of 4 ps).[11] The free-energy profile as a function of the CV was obtained from the average of four metadynamics runs (Figure 4). All explicit solvent NPT simulations used the v-rescale[14] thermostat and the Berendsen barostat[15] with isotropic pressure scaling. Long-range electrostatics were treated using the particle mesh Ewald (PME)[16] approach. All bonds involving hydrogens were constrained using the LINCS algorithm.[17]

Optical microscopy

Sample solutions (20 µl) were placed on a Marienfeld-superior microscope slide (76×26×1.35 mm with a round cavity of 15–18 mm diameter and 0.6–0.8 mm depth) and a cover glass was placed on the sample. Samples were imaged with Leica DM4500 P optical microscope equipped with Canon EOS 60D camera using transmission mode. A ruler at the sample position was imaged to calibrate the scale of the images.

3. Additional DLS and ζ-potential measurements

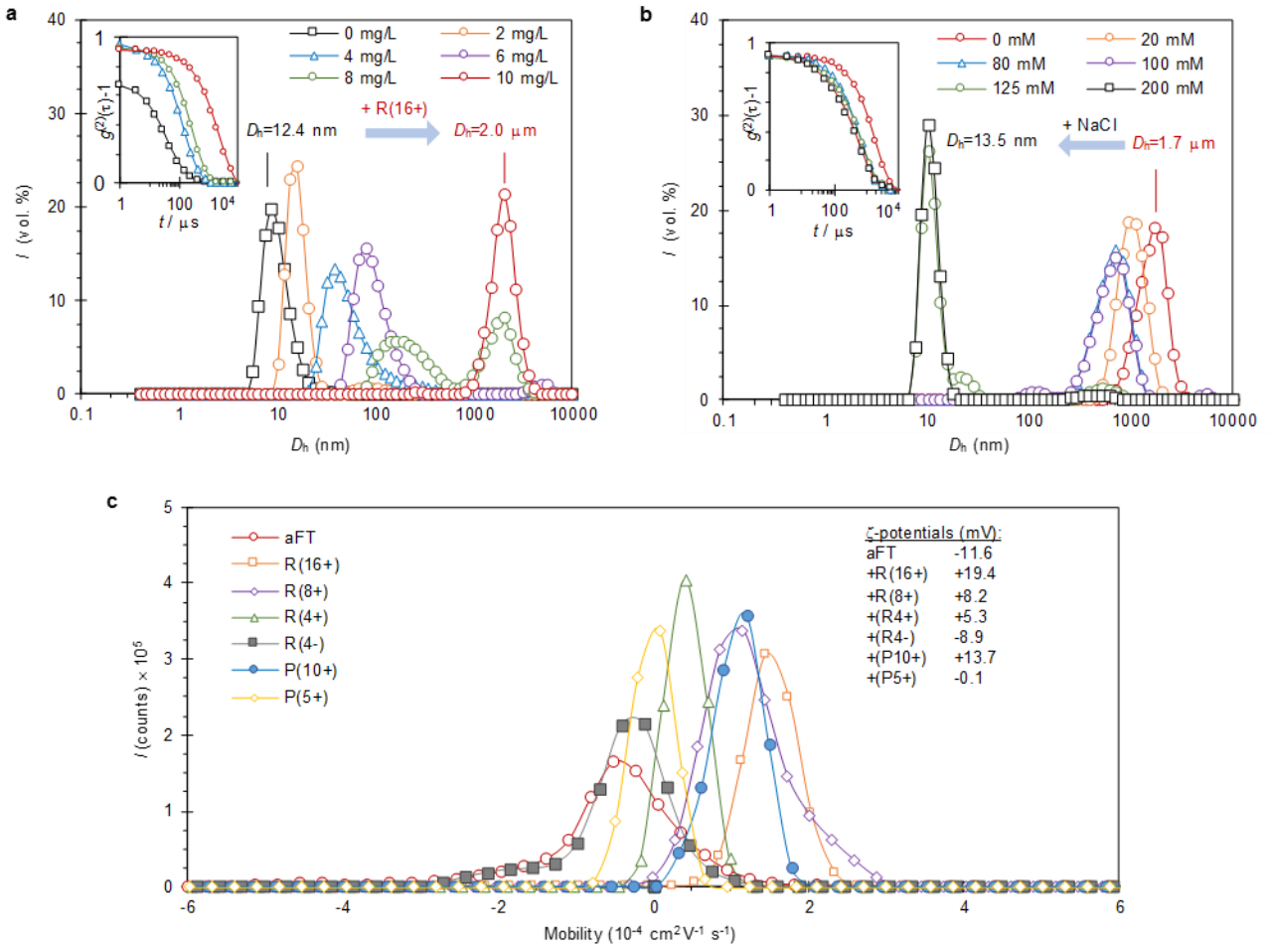


Figure S7. DLS size distribution profiles and ζ-potentials. a) Volume-averaged size distribution of free aFT titrated with **R(16+)**. b) The resulting complexes disassembled with NaCl. Insets: second order autocorrelation functions of the corresponding measurements. c) ζ-potential measured from all aFT-host complexes.

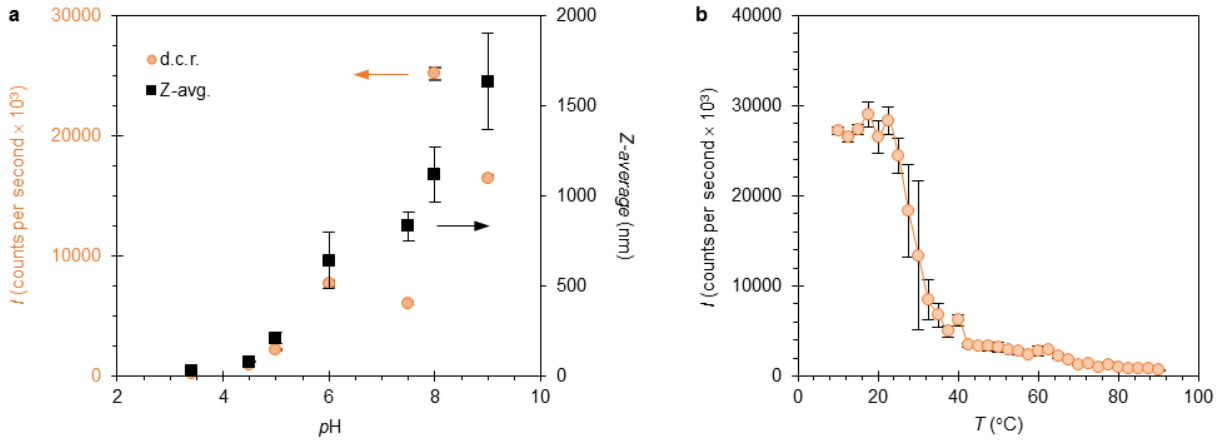


Figure S8. Complex stability studied by DLS. a) Size (Z-average) and scattering intensity (count rate) measured at different pH values. Complex stability is reduced significantly when pH drops below 5. Isoelectric point of aFT is approximately 4.5 and therefore electrostatic interactions are largely lost when pH approaches this value. b) Count rate monitored with increasing temperature. Heavily scattering complexes are disassembled when temperature increases above 25 °C and all assemblies are lost at T > 65 °C.

4. Low-magnification TEM images

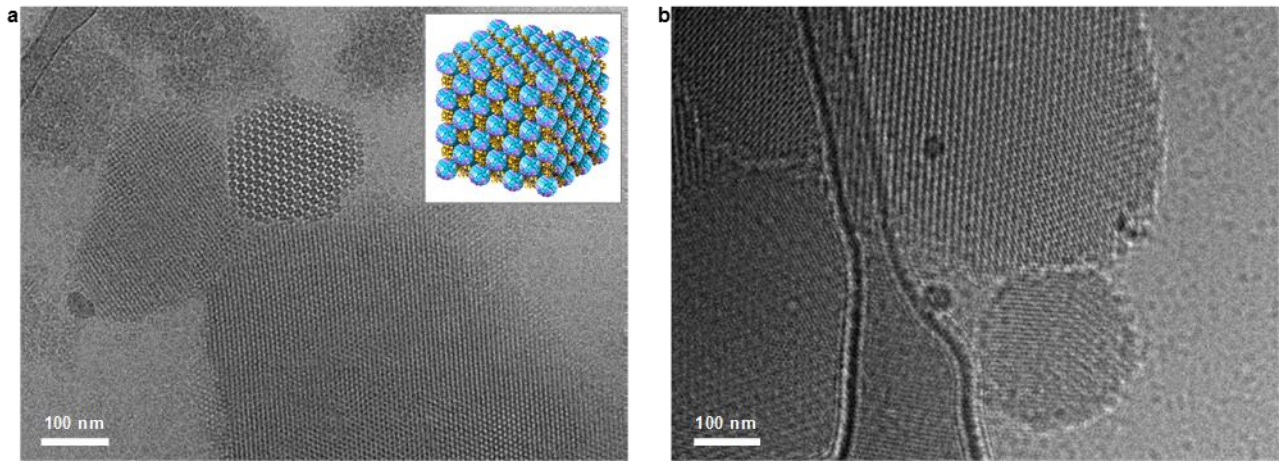


Figure S9. Low-magnification cryo-TEM images. a,b) aFT-**P(10+)** crystals imaged with cryo-TEM show clearly multiple highly-ordered complexes viewed from random orientations. Inset shows a schematic image of 3×3 unit cell crystal.

5. Macroscopic crystal habit

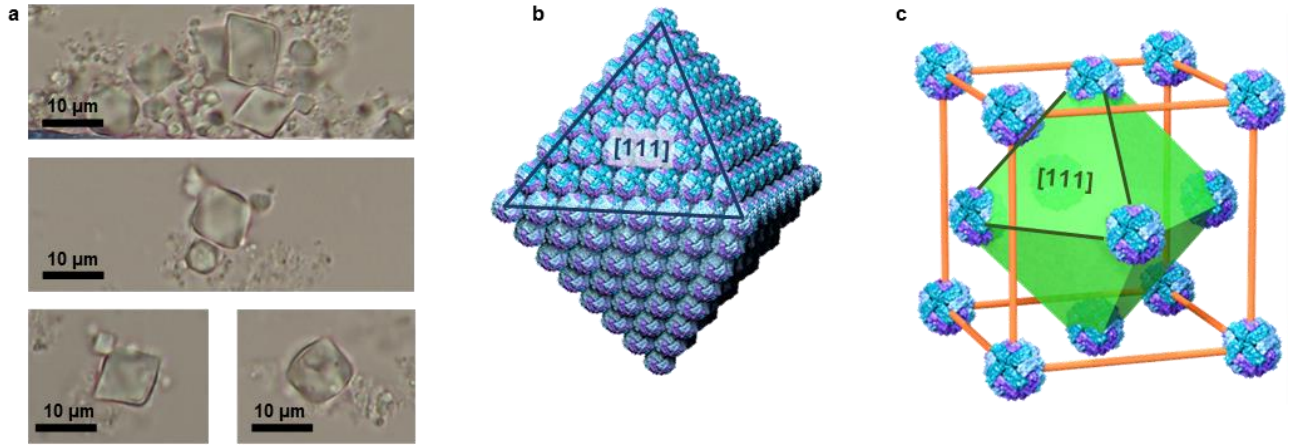


Figure S10. Optical imaging and habit of aFT-**P(10+)** crystals. a) Crystals with octahedral habit and sizes over 10 µm observed with an optical microscope. b) A schematic model of a face centred cubic packed aFT cages with octahedral shape. c) The unit cell (aFT diameter reduced and **P(10+)** omitted for clarity) showing the [111] planes in green. Growth in the [111] directions is thermodynamically favourable and results in the octahedral crystal habit.

6. Crystal voids

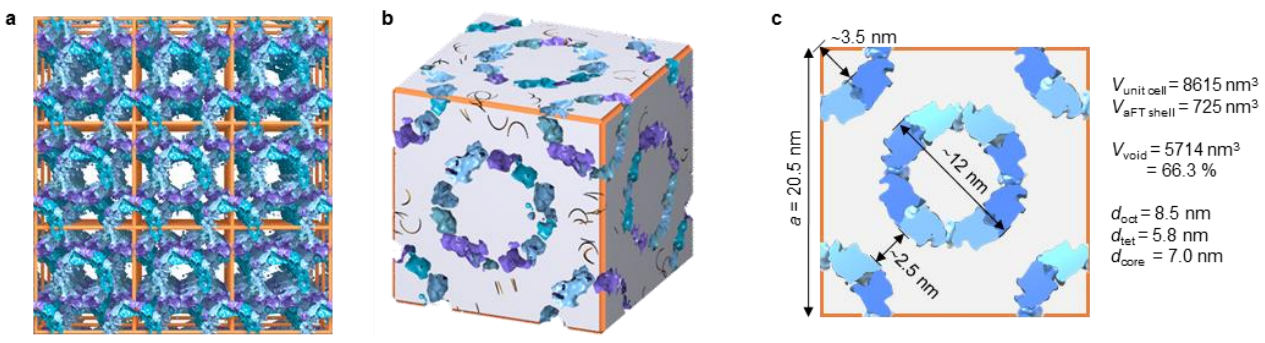


Figure S11. Porosity of the aFT-**P(10+)** crystals. a) An image of the 3×3 fcc packed unit cell of the aFT cages demonstrates the high void volume of the protein cage crystal. **P(10+)** are omitted for clarity and aFT is shown as a ring (centre slice). b) Free space in the unit cell (grey). c) [100] cross-section of the unit cell and the approximate dimensions used to calculate the free space (66.3 %) of the protein crystals.

7. Supporting information references

1. Kobayashi, K., Asakawa, Y., Kato, Y. & Aoyama, Y. Complexation of Hydrophobic Sugars and Nucleosides in Water with Tetrasulfonate Derivatives of Resorcinol Cyclic Tetramer Having a Polyhydroxy Aromatic Cavity: Importance of Guest-Host CH-π Interaction. *J. Am. Chem. Soc.* **114**, 10307–10313 (1992).

2. Beyeh, N. K., Pan, F. & Ras, R. H. A. *N*-Alkyl Ammonium Resorcinarene Chloride Receptors for Guest Binding in Aqueous Environment. *Asian J. Org. Chem.* **5**, 1027–1032 (2016).

3. Beyeh, N. K., Jo, H. H., Kolesnichenko, I., Pan, F., Kalenius, E., Anslyn, E. V., Ras, R. H. A. & Rissanen, K. Recognition of Viologen Derivatives in Water by *N*-Alkyl Ammonium Resorcinarene Chlorides. *J. Org. Chem.*, **82**, 5198–5203 (2017).

4. Montes-Garcia, V., Fernandez-Lopez, C., Gomez, B., Perez-Juste, I., Garcia-Rio, L., Liz-Marzan, L., Perez-Juste, J. & Pastoriza-Santos, I. Pillar[5]arene-Mediated Synthesis of Gold Nanoparticles: Size Control and Sensing Capabilities. *Chem. Eur. J.*, **20**, 8404–8409 (2014).

5. Yao, Y., Xue, M., Chen, J., Zhang, M. & Huang, F. An Amphiphilic Pillar[5]arene: Synthesis, Controllable Self-Assembly in Water, and Application in Calcein Release and TNT Adsorption. *J. Am. Chem. Soc.*, **134**, 15712−15715 (2012).

6. Kraus, W. & Nolzeb, G. POWDER CELL - a Program for the Representation and Manipulation of Crystal Structures and Calculation of the Resulting X-Ray Powder Patterns. *J. Appl. Cryst.*, **29**, 301–303 (1996).

7. Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B. & Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX*, **1–2**, 19–25 (2015).

8. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.*, **25**, 1157–1174 (2004).

9. Case, D. A. *et al*. AMBER 12, University of California: San Francisco (2012).

10. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an Automated Pipeline for the Setup of Poisson–Boltzmann Electrostatics Calculations. *Nucl. Acids Res.*, **32**, W665-W667 (2004).

11. Bochicchio, D., Salvalaglio, M. & Pavan, G. M. Into the Dynamics of a Supramolecular Polymer at Submolecular Resolution. *Nat. Commun.*, **8**, 147 (2017).

12. Laio, A. & Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl Acad. Sci. USA*, **99**, 12562–12566 (2002).

13. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.*, **185**, 604–613 (2014).

14. Bussi, G., Donadio, D. & Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.*, **126**, 14101 (2007).

15. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, **81**, 3684–3690 (1984).

16. Darden, T., York, D. & Pedersen, L. Particle Mesh Ewald: An $N\cdot\log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.*, **98**, 10089–10092 (1993).

17. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.*, **18**, 1463–1472 (1997).