Supplementary Information

Shape-persistent pyrrole-based covalent organic cages: Synthesis, structure and selective gas adsorption properties

Fei Wang,[a,b] Eric Sikma,[c] Zhiming Duan,[b] Tridib Sarma,[b] Chuanhu Lei,[b] Zhan Zhang,[*b,d] Simon M. Humphrey,[*c] and Jonathan L. Sessler[*b,c]

[a] School of Materials Science and Engineering

[b] Center for Supramolecular Chemistry & Catalysis, Shanghai University, Shanghai 200444, China.

[c] Department of Chemistry, The University of Texas at Austin, Texas 78712-1224, United States.

[d] College of Chemistry and Material Sciences, South-Central University for Nationalities, , Wuhan, Hubei 430074, China

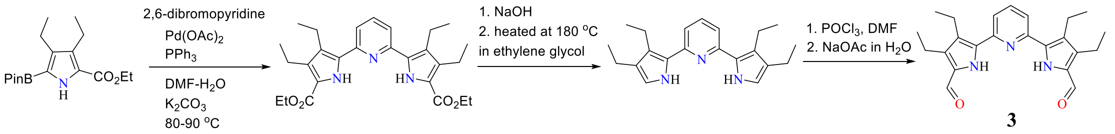
.

Table of Contents

I. Synthesis and Characterization

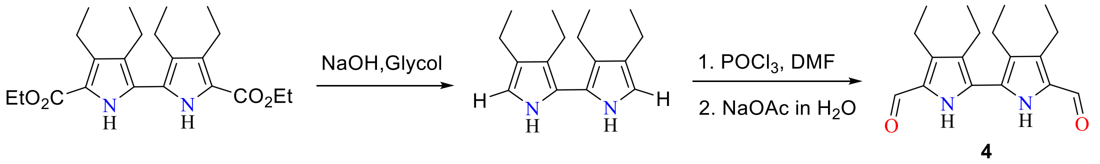
All chemicals used were purchased from Aldrich, TCI, Energy Chemicals, Adamas-beta® or Acros and used without further purification unless otherwise mentioned. All solvents used in the reactions were purchased from local suppliers and dried prior to use. Thin-layer chromatographic (TLC) analyses were carried out using silica gel pre-coated on glass sheets. Column chromatography was performed on silica gel (mesh size 200-300). ^{1}H and ^{13}C NMR spectra were recorded using Jeol 400, Bruker 500, or Bruker 600 MHz instruments at room temperature. Chemical shifts are reported in ppm using TMS or residual solvent signals as the internal reference standards. MALDI-TOF mass spectrometric measurements were made using a Bruker (Autoflex speed) matrix-assisted laser desorption ionization time-of-flight mass spectrometer.

Dialdehyde **3** was synthesized according to the literature procedure:[S1-2]

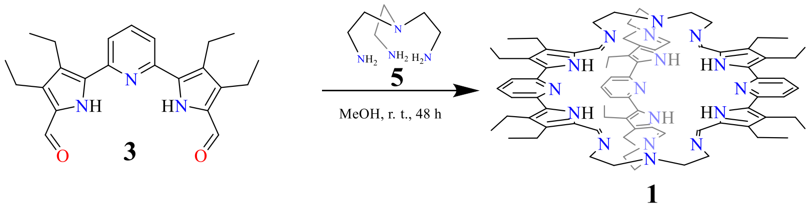


Compound **3**: White solid, yield: 88%. ^{1}H NMR (400 MHz, $CDCl_3$) δ 11.29 (s, 2H, NH), 9.62 (s, 2H, -CHO), 7.88 - 7.69 (m, 1H, pyridine), 7.54 (d, *J* = 8.0 Hz, 2H, pyridine), 2.77 (q, *J* = 7.6, 2.5 Hz, 8H, CH_2 alkyl), 1.25 (t, *J* = 7.6, 5.8 Hz, 12H, CH_3 alkyl). ^{13}C NMR (100 MHz, $CDCl_3$) δ 178.36 (s), 148.99 (s), 139.81 (s), 137.94 (s), 134.40 (s), 128.26 (s), 125.96 (s), 118.94 (s), 17.97 (s), 17.95 (s), 16.92 (s), 15.31 (s).

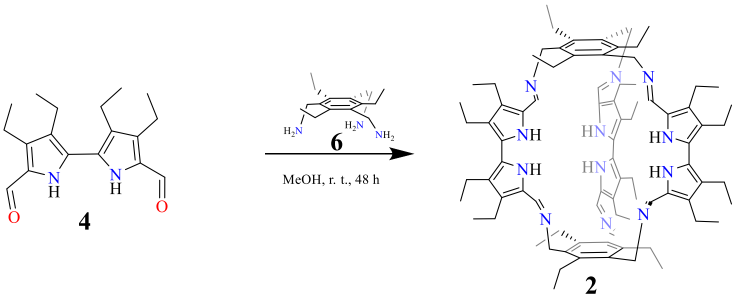
Precursor **4** was obtained according to the literature procedure:[S3]



Compound **4**: Off-white crystal, yield: 85%. ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.64 (s, 2H, -CHO), 9.39 (s, 2H, NH), 2.77 (q, *J* = 7.6 Hz, 4H, CH_2 alkyl), 2.51 (q, *J* = 7.6 Hz, 4H, CH_2 alkyl), 1.27 (t, J = 7.6 Hz, 6H, CH_3 alkyl), 1.06 (t, J = 7.6 Hz, 6H, CH_3 alkyl). ^{13}C NMR (100 MHz, $CDCl_3$) δ 177.53 (s), 137.94 (s), 129.26 (s), 127.63 (s), 126.64 (s), 17.65 (s), 17.59 (s), 17.20 (s), 15.80 (s).



*Synthesis of **1***. To a 500 mL round-bottom flask was added compound **3** (108.8 mg, 0.3 mmol) and dry methanol (300 mL). The solution was purged with N_2 for 20 min. Then, tris(2-aminoethyl)amine **5** (29.2 mg, 0.2 mmol) dissolved in 50 mL of dry methanol was added dropwise through a constant pressure dropping funnel over the course of 4 hours. The resulting mixture was stirred at room temperature for another 44 h. The solid was filtered off, washed with methanol, and dried under vacuum to give cage **1** as a saffron-yellow powder. Yield: 90%. ^{1}H NMR (600 MHz, $CDCl_3$) δ 11.59 (s, 6H, NH), 8.19 (s, 6H, CH imine), 7.65 (t, J = 7.8 Hz, 3H, pyridine), 7.37 (d, J = 7.9 Hz, 6H, pyridine), 3.85 - 3.70 (m, 12H, CH_2 alkyl amine), 3.32 - 3.03 (m, 12H, CH_2 alkyl amine), 2.75 (q, J = 7.4 Hz, 12H, CH_2 alkyl), 2.61 (q, J = 7.5 Hz, 12H, CH_2 alkyl), 1.22 (t, J = 7.5 Hz, 18H, CH_3 alkyl), 1.18 (t, J = 7.5 Hz, 18H, CH_3 alkyl). ^{13}C NMR (150 MHz, $CDCl_3$) δ 151.56 (s), 149.37 (s), 137.08 (s), 133.04 (s), 131.11 (s), 125.15 (s), 124.56 (s), 116.44 (s), 61.43 (s), 57.68 (s), 18.14 (s), 17.70 (s), 16.84 (s), 15.34 (s). MALDI-TOF MS Calcd for $C_{81}H_{105}N_{17}$ $[M + H]^+$ 1316.881. Found: 1316.728.



*Synthesis of **2***. To a 500 mL round-bottom flask was added compound **4** (90.1 mg, 0.3

mmol) and dry methanol (300 mL). The solution was purged with N_2 for 20 min. Then, 1,3,5-tris(aminomethyl)-2,4,6-triethylbenzene **6** (49.9 mg, 0.2 mmol) dissolved in 50 mL of dry methanol was added dropwise through a constant pressure dropping funnel over the course of 4 hours. The resulting mixture was stirred at room temperature for another 44 h. The solid was filtered off, washed with methanol, and dried under vacuum to give cage **2** as a saffron-yellow powder. Yield: 85%. ^{1}H NMR (600 MHz, $CDCl_3$) δ 8.24 (s, 6H, imine), 4.69 (s, 12H, CH_2 alkyl benzene), 3.02 (s, 12H, CH_2 alkyl benzene), 2.53 (q, J = 7.3 Hz, 12H, CH_2 alkyl pyrrole), 2.31 (q, J = 7.1 Hz, 12H, CH_2 alkyl pyrrole), 1.32 (t, J = 7.4 Hz, 18H, CH_3 alkyl benzene), 1.11 (t, J = 7.5 Hz, 18H, CH_3 alkyl pyrrole), 0.92 (t, J = 7.4 Hz, 18H, CH_3 alkyl pyrrole). ^{13}C NMR (150 MHz, $CDCl_3$) δ 148.41 (s), 142.25 (s), 133.67 (s), 129.18 (s), 126.14 (s), 123.91 (s), 123.78 (s), 58.82 (s), 21.63 (s), 17.71 (s), 17.30 (s), 17.00 (s), 16.76 (s), 16.00 (s). MALDI-TOF MS Calcd for $C_{84}H_{114}N_{12}$ $[M + H]^+$ 1291.936. Found: 1292.242.

II. NMR & MALDI-TOF mass spectra

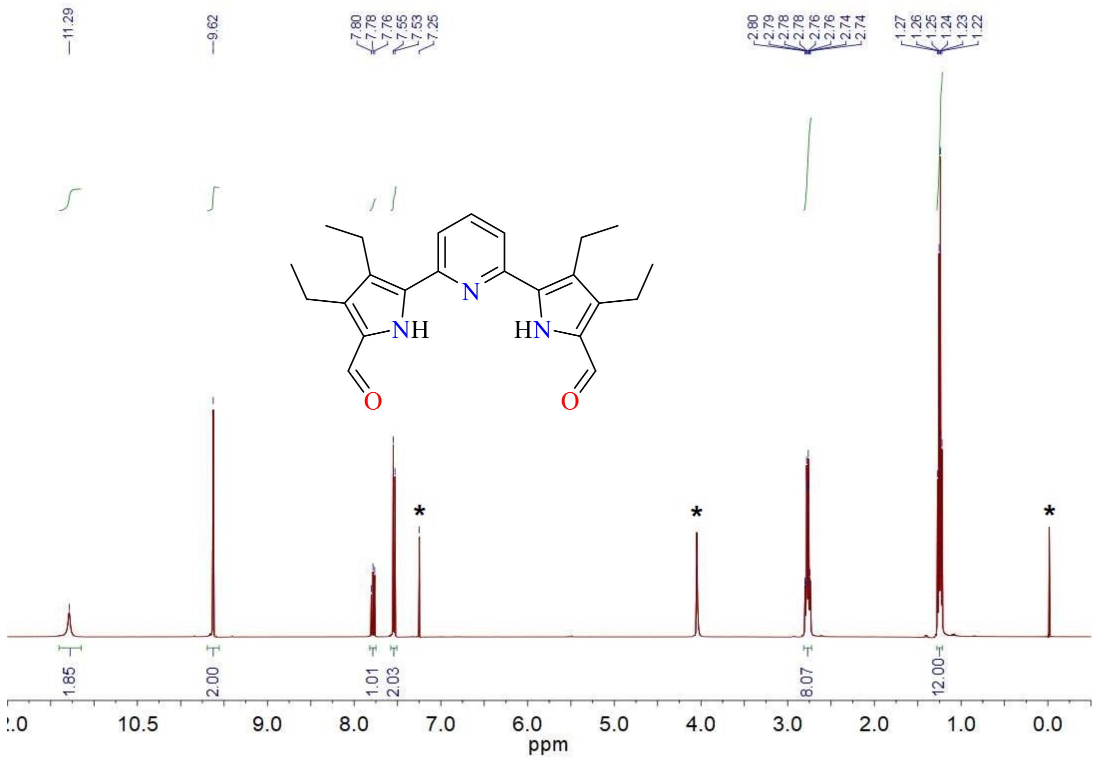


Figure S1. ^{1}H NMR spectrum of **3** recorded in $CDCl_3$. *Asterisk indicates solvent residual impurities.

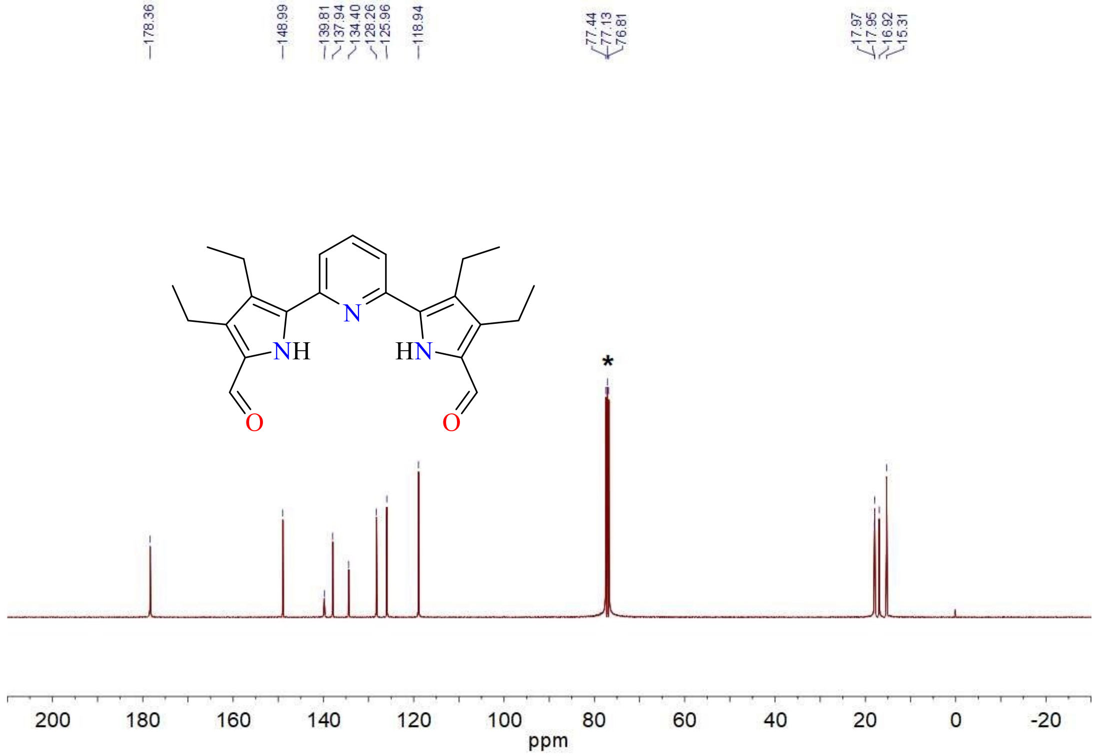


Figure S2. ^{13}C NMR spectrum of **3** recorded in $CDCl_3$. *Asterisk indicates solvent residual impurities.

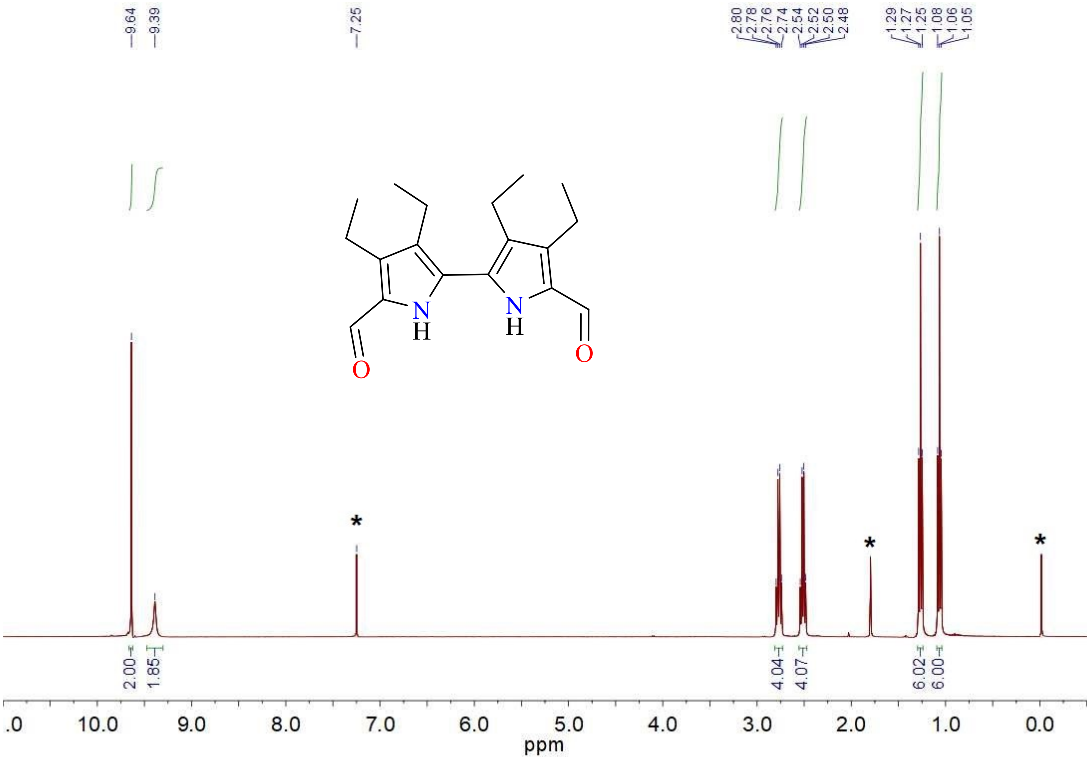


Figure S3. ^{1}H NMR spectrum of **4** recorded in $CDCl_3$. *Asterisk indicates solvent residual impurities.

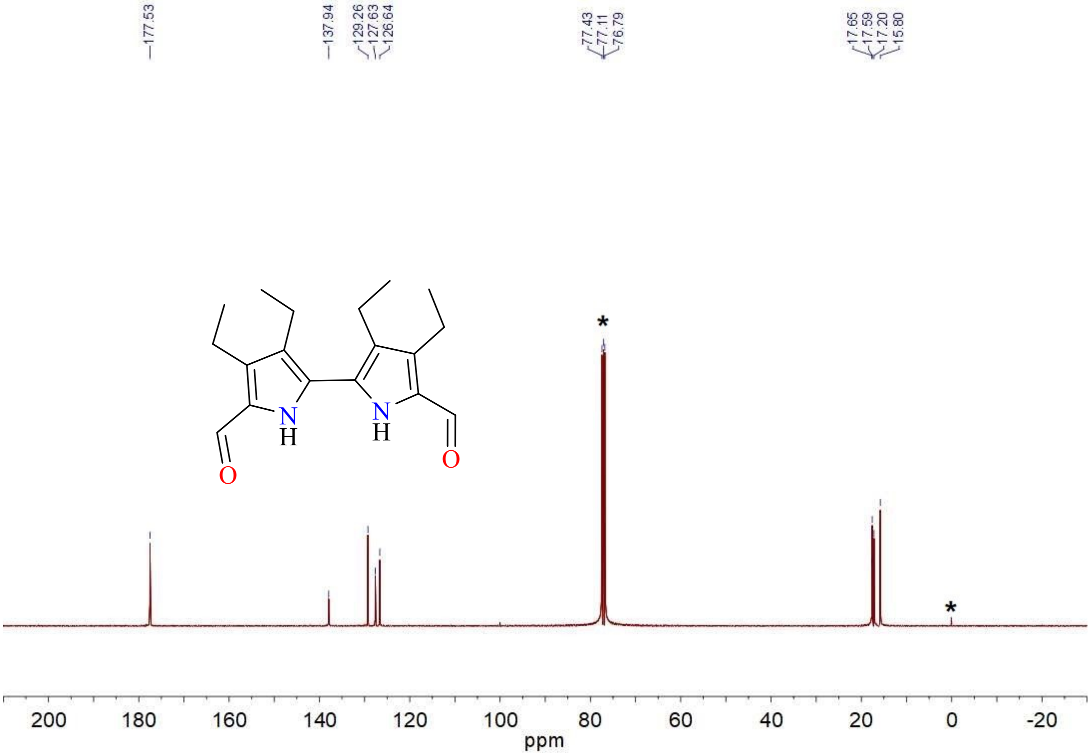


Figure S4. ^{13}C NMR spectrum of **4** recorded in CDCl$_3$. *Asterisk indicates solvent residual impurities.

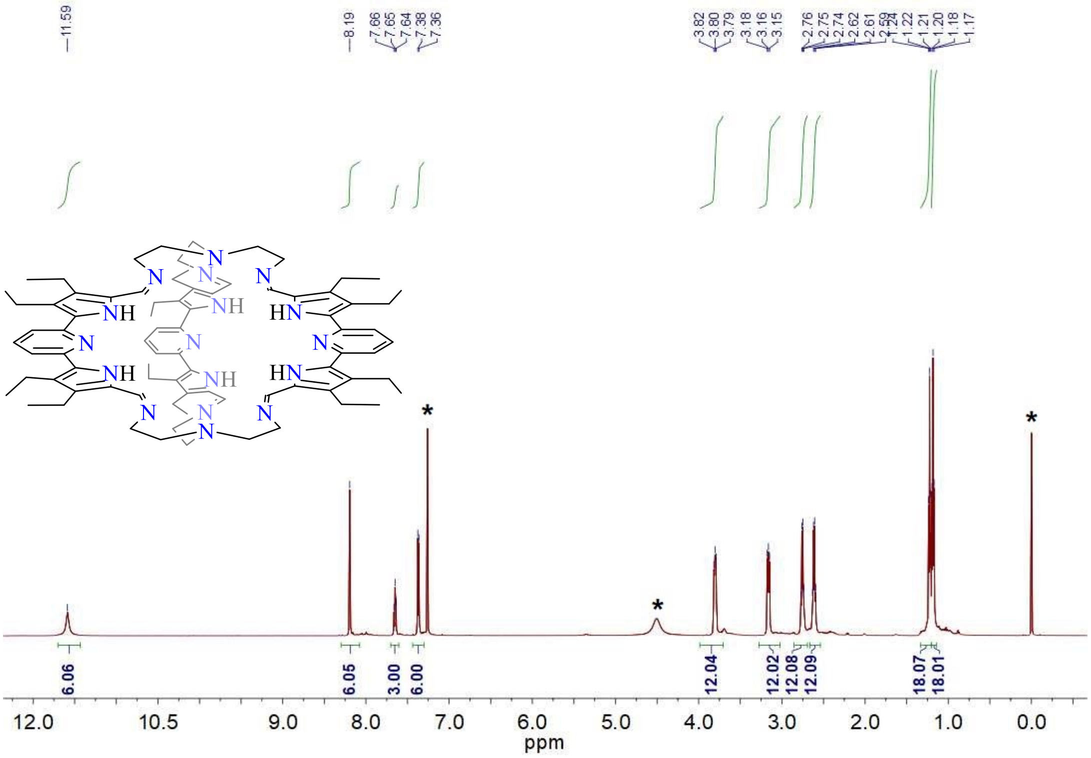


Figure S5. ^{1}H NMR spectrum of cage **1** recorded in CDCl$_3$. *Asterisk indicates solvent residual impurities.

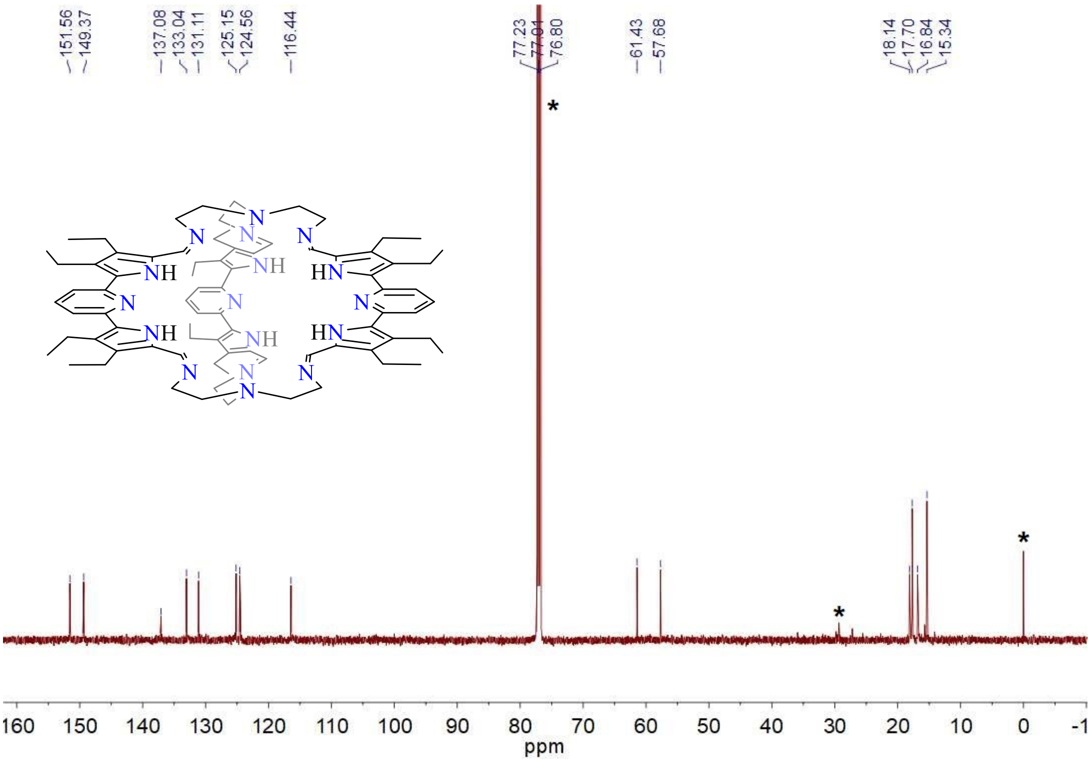


Figure S6. ^{13}C NMR spectrum of cage **1** recorded in $CDCl_3$. *Asterisk indicates solvent residual impurities.

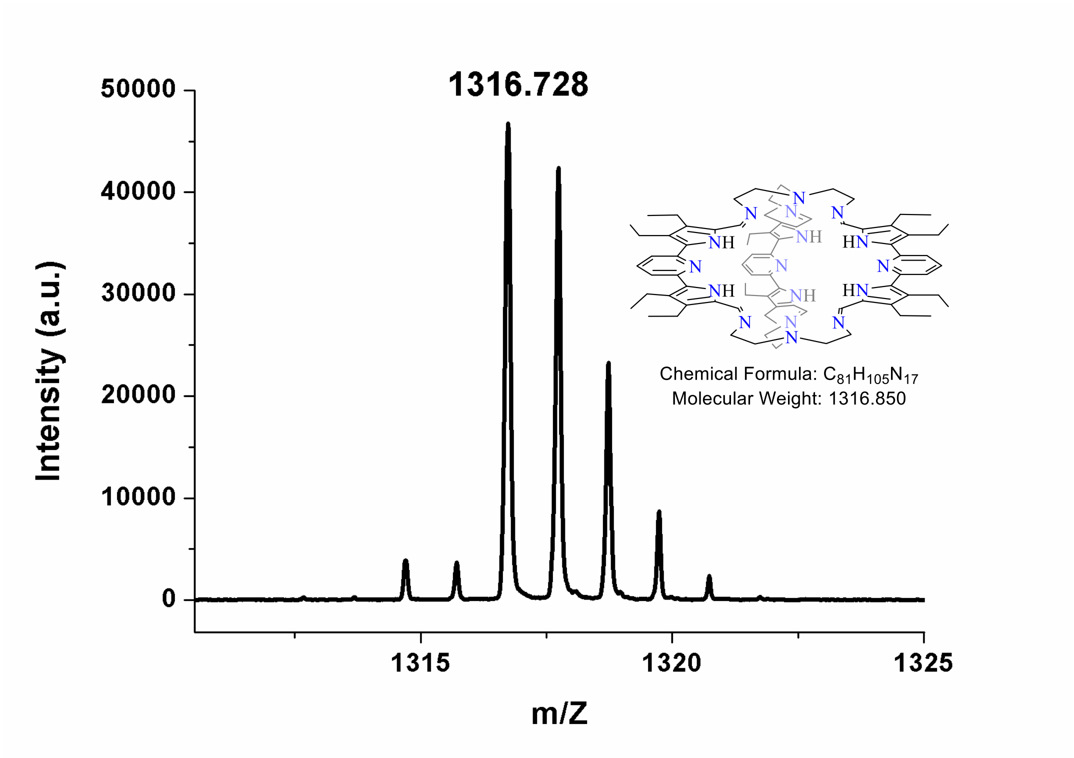


Figure S7. MALDI-TOF mass spectrum of cage **1**.

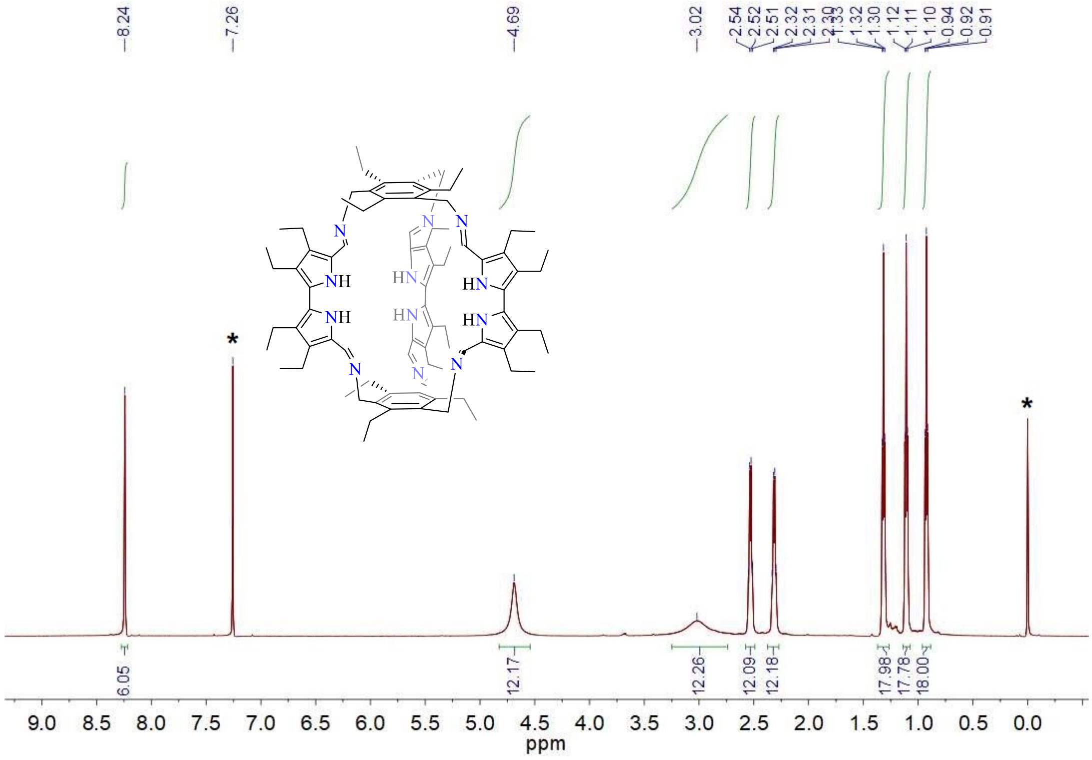


Figure S8. ^{1}H NMR spectrum of cage **2** recorded in $CDCl_3$. *Asterisk indicates solvent residual impurities.

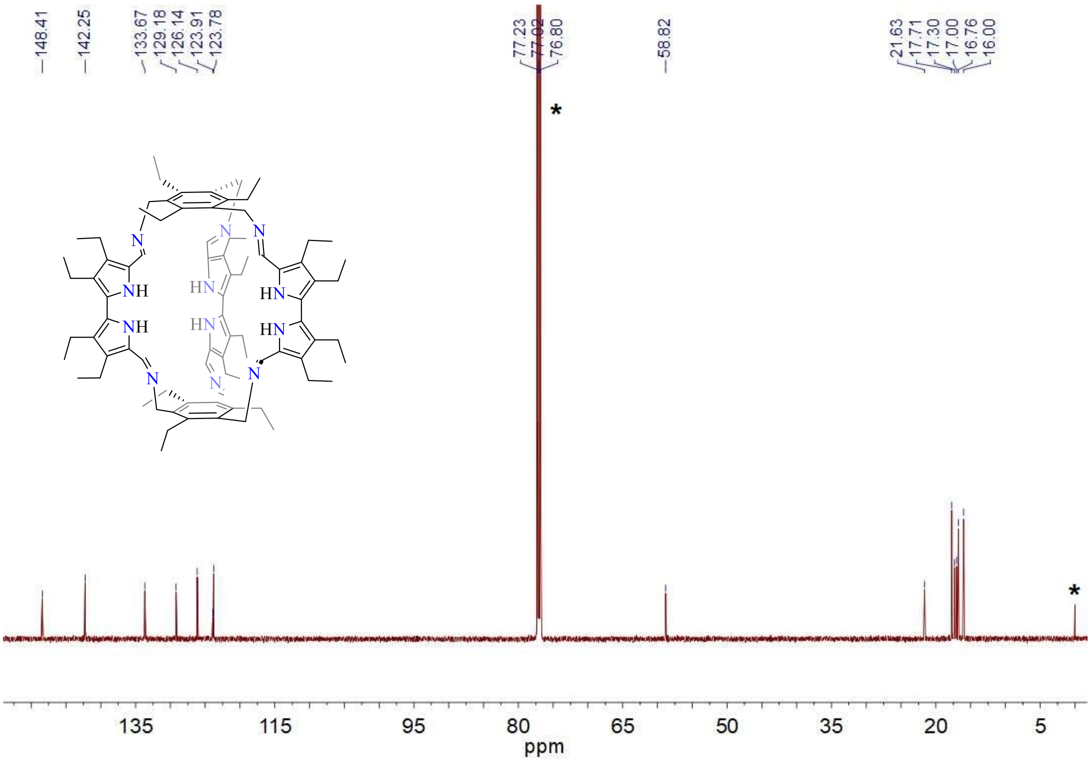


Figure S9. ^{13}C NMR spectrum of cage **2** recorded in $CDCl_3$. *Asterisk indicates solvent residual impurities.

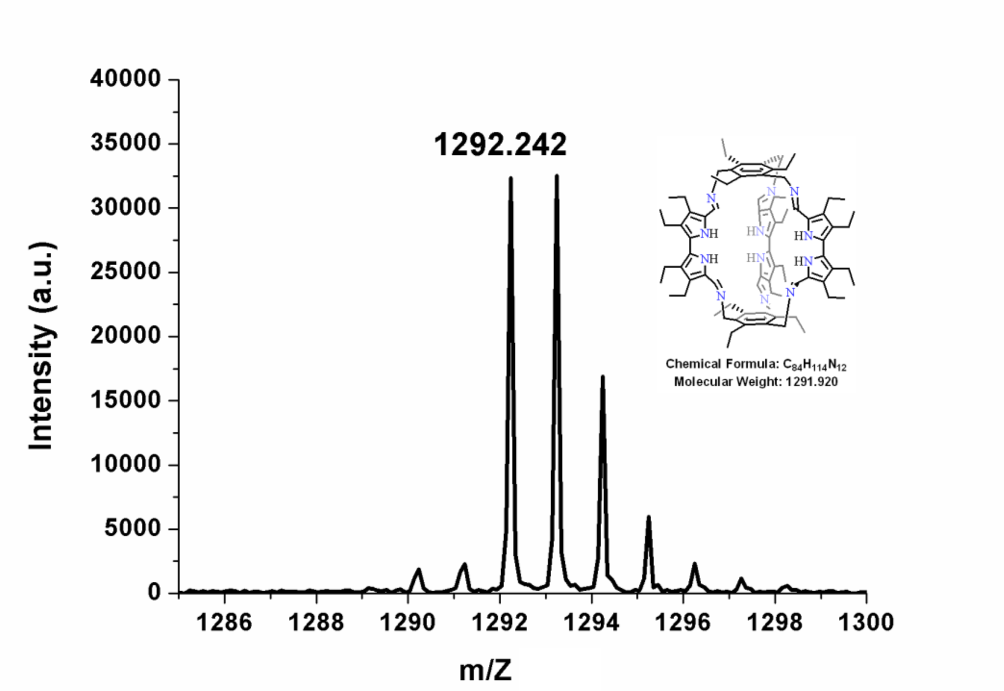


Figure S10. MALDI-TOF mass spectrum of **2**.

III. X-ray crystallography

X-ray diffraction analysis of single crystals of cage $1\cdot3H_2O$: The data were collected on an Agilent Technologies SuperNova Dual Source diffractometer using a μ-focus Cu Kα radiation source ($\lambda = 1.5418$ Å) with collimating mirror monochromators. A total of 581 frames of data were collected using ω-scans with a scan range of 1° and a counting time of 23 seconds per frame using a detector offset of +/- 41.3° and a counting time of 65 seconds per frame using a detector offset of +/- 108.3°. The data were collected at 173 K using an Oxford Cryostream low temperature device. Details of crystal data, data collection and structure refinement are listed in the Tables below. Data collection, unit cell refinement and data reduction were performed using the Agilent Technologies' CrysAlisPro V 1.171.38.43f software.[S4] The structure was solved by direct methods using SHELXT[S5] and refined by full-matrix least-squares on F^2 with anisotropic displacement parameters for the non-H atoms using SHELXL-2016/6.[S6] Structure analysis was aided by use of the programs PLATON[S7] and WinGX.[S8] The hydrogen atoms were calculated in ideal positions with isotropic displacement parameters set to 1.2 × Ueq of the attached atom (1.5 × Ueq for methyl hydrogen atoms). Non-hydrogen atoms were refined with anisotropic displacement parameters. Contributions of highly disordered solvents were removed by SQUEEZE routine using PLATON.[S9] DISP instruction was used for all datasets.

The function, $\Sigma w(|F_o|^2 - |F_c|^2)^2$, was minimized, where $w = 1/[(\sigma(F_o))^2 + (0.0819{*}P)^2]$ and $P = (|F_o|^2 + 2|F_c|^2)/3$. $R_W(F^2)$ refined to 0.172, with R(F) equal to 0.052 and a goodness of fit, S, = 1.07. Definitions used for calculating R(F), $R_W(F^2)$ and the goodness of fit, S, are given below.[S10]The data were checked for secondary extinction effects but no correction was necessary. Neutral atom scattering factors and values used to calculate the linear absorption coefficient are from the International Tables for X-ray Crystallography (1992).[S11] All figures were generated using SHELXTL/PC.[S12] Crystallographic data has been deposited in the Cambridge Crgystallographic Data Center with CCDC number: 1904276. This data can be obtained free of charge at http://www.ccdc.cam.ac.uk/data_request/cif.

Table S1. Crystal data and structure refinement for cage **1**.

Identification code	**1•3H$_2$O**	
Empirical formula	$C_{81}H_{111}N_{17}O_3$	
Formula weight	1370.86	
Temperature	173(2) K	
Wavelength	1.54184 Å	
Crystal system	Hexagonal	
Space group	P6(3)/m	
Unit cell dimensions	a = 17.3339(3) Å	α = 90°
	b = 17.3339(3) Å	β = 90°
	c = 23.8668(4) Å	γ = 120°
Volume	6210.4(2) Å^3	
Z	2	
Calculated density	0.733 Mg/m^3	
Absorption coefficient	0.361 mm^{-1}	
F(000)	1480	
Crystal size	0.340 × 0.240 × 0.220 mm	
Theta range for data collection	3.478 to 67.054°	
Limiting indices	-20<=h<=20, -19<=k<=20, -28<=l<=28	
Reflections collected/unique	68031 / 3807 [R(int) = 0.0703]	
Completeness to θ = 67.054°	99.9%	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.925 and 0.887	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	3807 / 1 / 162	
Goodness-of-fit on F^2	1.068	
Final R indices [$I>2\sigma(I)$]	R_1 = 0.0521, wR_2 = 0.1644	
R indices (all data)	R_1 = 0.0592, wR_2 = 0.1715	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.273 and -0.228 e.Å^{-3}	

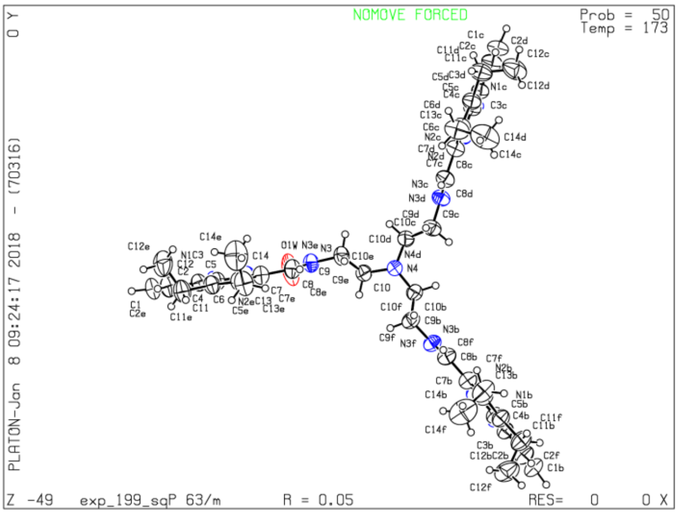


Figure S11. Crystal structure of cage **1•3H2O** viewed from the top.

X-ray diffraction analysis of single crystals of cage 2: The data collection instrument and analysis methods were the same as for the structure of cage **1**. In the case of cage **2**, $R_W(F^2)$ refined to 0.234, with R(F) equal to 0.081 and a goodness of fit, S, = 0.99. The crystallographic data may be obtained from the Cambridge Crystallographic Data Centre by making reference to CCDC number 1904275.

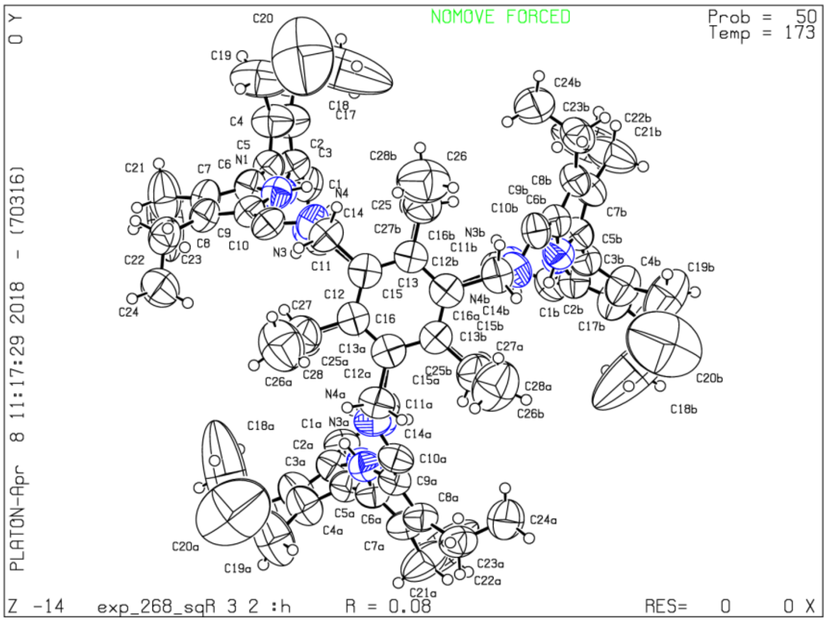
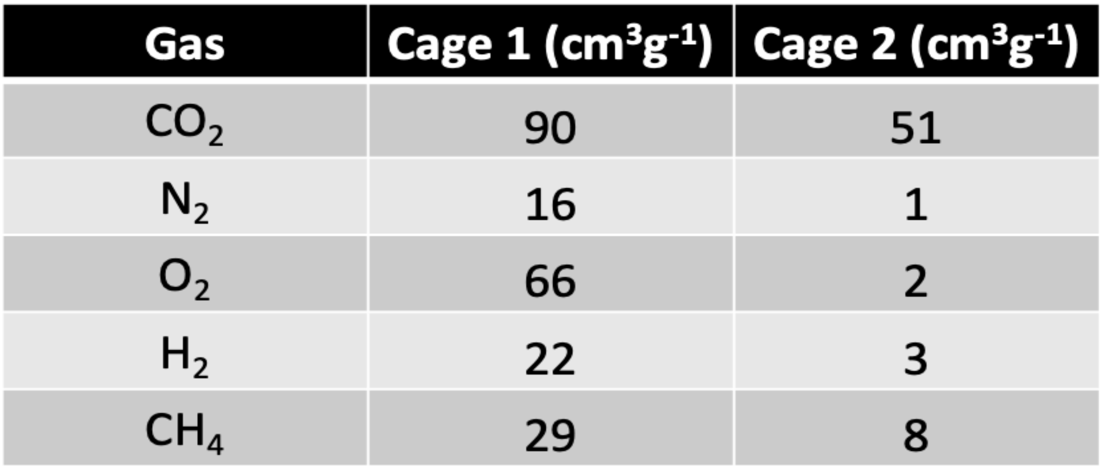


Figure S12. Crystal structure of cage **2** viewed from the top.

Table S2. Crystal data and structure refinement for cage **2**.

Identification code	**2**
Empirical formula	$C_{84}H_{114}N_{12}$
Formula weight	1291.92
Temperature	173(2) K
Wavelength	1.54184 Å
Crystal system	Trigonal
Space group	R32
Unit cell dimensions	a = 16.3754(4) Å α = 90°
	b = 16.3765(4) Å β = 90°
	c = 61.2246(15) Å γ = 120°
Volume	14220.0(8) Å^3
Z	6
Calculated density	0.884 Mg/m^3
Absorption coefficient	0.409 mm^{-1}
F(000)	4032
Crystal size	0.340 × 0.290 × 0.270 mm
Theta range for data collection	3.435 to 76.320°
Limiting indices	-20<=h<=20, -20<=k<=12, -71<=l<=76
Reflections collected/unique	47303 / 6410 [R(int) = 0.0603]
Completeness to θ = 67.684°	99.9%
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.898 and 0.874
Refinement method	Full-matrix least-squares on F^2
Data/restraints/parameters	6410 / 1 / 294
Goodness-of-fit on F^2	0.991
Final *R* indices [$I>2\sigma(I)$]	R_1 = 0.0806, wR_2 = 0.2201
R indices (all data)	R_1 = 0.1035, wR_2 = 0.2339
Absolute structure parameter	0.0(11)
Extinction coefficient	n/a
Largest diff. peak and hole	0.301 and -0.219 e.Å^{-3}

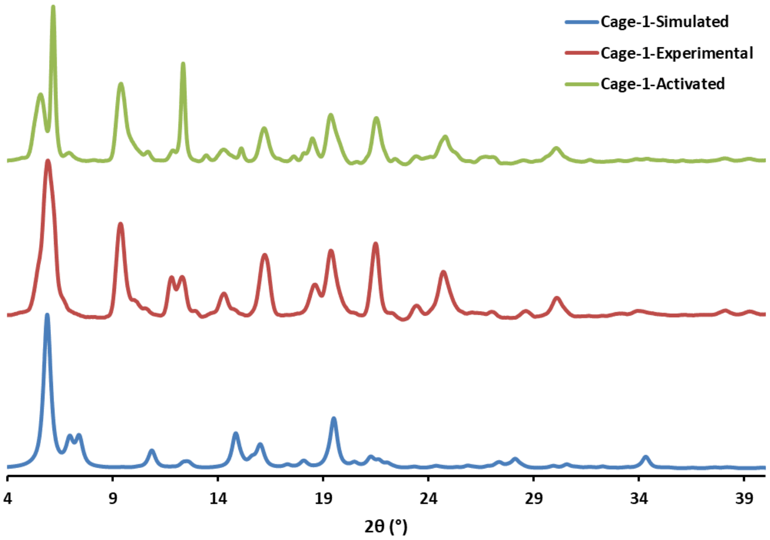
Gas Adsorption Studies



Gas	Cage 1 (cm^3g^{-1})	Cage 2 (cm^3g^{-1})
CO_2	90	51
N_2	16	1
O_2	66	2
H_2	22	3
CH_4	29	8

Table S3: Measured adsorption capacities of Cage **1** and Cage **2** for various probe gases.

PXRD Studies



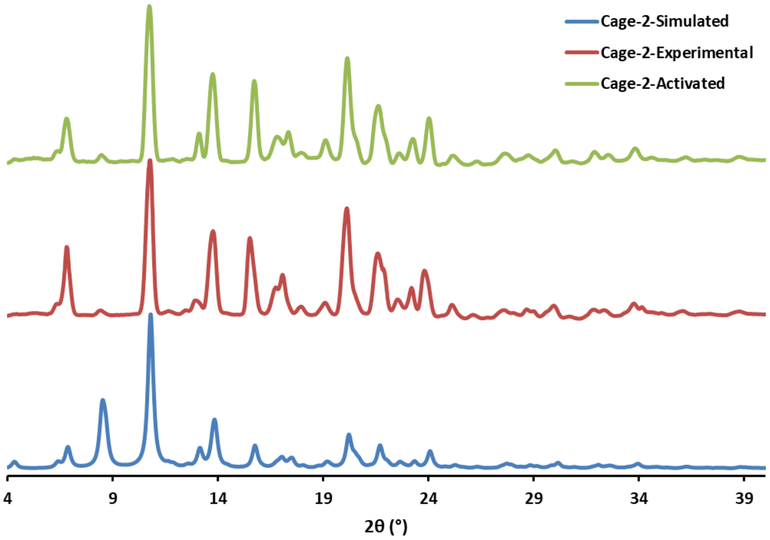


Figure S13. PXRD spectra for cages **1** and **2** recorded before and after activation (desolvation) per the conditions used to activate the samples for the BET analyses. Also shown are simulated patterns based on the single molecule X-ray diffraction analyses.

References & Notes

S1. J.-i. Setsune, M. Toda, K. Watanabe, P. K. Panda and T. Yoshida, Synthesis of bis (pyrrol-2-yl) arenes by Pd-catalyzed cross coupling, *Tetrahedron Lett.*, 2006, **47**, 7541-7544.

S2. J.-i Setsune and K. Watanabe, Cryptand-like Porphyrinoid Assembled with Three Dipyrrylpyridine Chains: Synthesis, Structure, and Homotropic Positive Allosteric Binding of Carboxylic Acids, *J. Am. Chem. Soc.*, 2008, **130**, 2404-2405.

S3. S. V. Shevchuk, J. M. Davis, J. L. Sessler, Synthesis of Sapphyrins via a '3+1+1' Procedure, *Tetrahedron Lett.*, 2001, **42**, 2447-2450.

S4. CrysAlisPro. Agilent Technologies (2013). Agilent Technologies UK Ltd., Oxford, UK, SuperNova CCD System, CrysAlicPro Software System, 1.171.38.43f.

S5. G. M. Sheldrick, SHELXT-Integrated Space-Group and Crystal-Structure Determination, *Acta Cryst. A*, 2015, **719**, 3-8.

S6. G. M.Sheldrick, Crystal Structure Refinement with SHELXL, *Acta Cryst. C*, 2015, **71**(1), 3-8.

S7. A. L. Spek, PLATON SQUEEZE: A Tool for the Calculation of the Disordered Solvent Contribution to the Calculated Structure Factors, *Acta Cryst. C*, 2015, **71**, 9-18.

S8. L. J. Farrugia, G. X. Win, Program Features, *J. Appl. Crystallogr.*, 1999, **32**, 837-838.

S9. P. Van der Sluis,; A. L. Spek, BYPASS: an Effective Method for the Refinement of Crystal Structures Containing Disordered Solvent Regions, *Acta Cryst. A*, 1990, **46**, 194-201.

S10. $R_W(F^2)=\{\Sigma w(|F_o|^2-|F_c|^2)^2/\Sigma w(|F_o|)^4\}^{1/2}$ where w is the weight given each reflection.

$R(F) = \Sigma(|F_o| - |F_c|)/\Sigma|F_o|\}$ for reflections with $F_o > 4(\sigma(F_o))$.

$S = [\Sigma w(|F_o|^2 - |F_c|^2)^2/(n - p)]^{1/2}$, where n is the number of reflections and p is the number of refined parameters.

S11. A. J. C. Wilson, International Tables for X-ray Crystallography, Vol. C, Tables 4.2.6.8 and 6.1.1.4, *Kluwer Academic Press, Boston*, 1992.

S12. G. M. Sheldrick, SHELXTL/PC (Version 5.03). Siemens Analytical X-ray Instruments, Inc., Madison, Wisconsin, USA, 1994.