Electronic Supplementary Material (ESI) for Nanoscale.

Supporting Information

Computationally-Inspired Discovery of an Unsymmetrical Porous Organic Cage

Enrico Berardo,[†,§] Rebecca L. Greenaway,[‡,§] Lukas Turcani,[†] Ben M. Alston,[‡] Michael J. Bennison,[‡] Marcin Miklitz,[†] Rob Clowes,[‡] Michael E. Briggs,[‡] Andrew I. Cooper,[‡] Kim E. Jelfs[†*]

[†]Department of Chemistry, Imperial College London, South Kensington, London, SW7 2AZ, United Kingdom

[‡]Department of Chemistry and Materials Innovation Factory, University of Liverpool, 51 Oxford Street, Liverpool, L7 3NY, United Kingdom

[§]E.B. and R.L.G. contributed equally.

Contents

1. General synthetic and analytical methods

Materials: Chemicals were purchased from TCI UK, Fluorochem, Sigma-Aldrich, or Manchester Organics. Solvents were reagent or HPLC grade purchased from Fisher Scientific, with the exception of chloroform-D which was purchased from Apollo Scientific and used in the high-throughput screen. All chemicals and solvents were used as received unless specified.

Synthesis: All reactions requiring anhydrous or inert conditions were performed in oven-dried apparatus under an inert atmosphere of dry nitrogen, using anhydrous solvents introduced into the flask using disposable needles and syringes. All reactions were stirred magnetically using Teflon-coated stirring bars. Where heating was required, the reactions were warmed using a stirrer hotplate with heating blocks with the stated temperature being measured externally to the reaction flask with an attached probe. Removal of solvents was done using a rotary evaporator.

High-Throughput Cage Discovery: High-throughput automated synthesis was conducted using a Chemspeed Accelerator SLT-100 automated synthesis platform.

TLC and Column Chromatography: Reactions were monitored by thin layer chromatography (TLC), conducted on pre-coated aluminium-backed plates (Merck Kieselgel 60 with fluorescent indicator UV254). Spots were visualized either by quenching of UV fluorescence or by staining with potassium permanganate. Flash column chromatography was performed on a Biotage Isolera with KP-Sil Normal Phase disposable columns.

Melting points: Obtained using Griffin melting point apparatus and are uncorrected.

IR spectra: Infra-red (IR) spectra were recorded on a Bruker Tensor 27 FT-IR using ATR measurements for oils and solids as neat samples.

NMR Spectra: ^{1}H Nuclear magnetic resonance (NMR) spectra were recorded using an internal deuterium lock for the residual protons in $CDCl_3$ ($\delta = 7.26$ ppm) at ambient probe temperature using either a Bruker Avance 400 (400 MHz) or Bruker DRX500 (500 MHz) instrument. NMR data are presented as follows: chemical shift, integration, peak multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad, app = apparent) and coupling constants (*J* / Hz). Chemical shifts are expressed in ppm on a δ scale relative to δ_{CDCl_3} (7.26 ppm) and coupling constants, *J*, are given in Hz.

^{13}C NMR spectra were recorded using an internal deuterium lock using $CDCl_3$ ($\delta = 77.16$ ppm) at ambient probe temperatures using either a Bruker Avance 400 (101 MHz) or Bruker DRX500 (126 MHz) instrument.

2D NMR spectra (COSY, HSQC, HMBC) were recorded using an internal deuterium lock for the residual protons in $CDCl_3$ (δ_H = 7.26 ppm, δ_C = 77.16 ppm) at ambient probe temperature using a Bruker DRX500 instrument.

1D- and 2D-NOESY were recorded using an internal deuterium lock for the residual protons in $CDCl_3$ (δ = 7.26 ppm) at ambient probe temperature using a Bruker DRX500 instrument, with an optimised mixing time (D8) of 0.7 seconds.

Analytical HPLC: HPLC was conducted on a Dionex UltiMate 3000 equipped with a diode array UV detector using a Thermo-Scientific Syncronis C8 column, 150x4.6 mm, 3 µm (SN 10136940, Lot 12459). The mobile phase was isocratic MeOH at a flow rate of 1 mL/min for a 20 min run time, and the column oven temperature was set to 30 °C. The injection volume was 10 µL and the sample concentration was approximately 1 mg/mL. Detection for UV analysis was conducted at 254 nm.

Preparative HPLC: The column used to purify the unsymmetrical cage was a Thermo Scientific Syncronis C8, 150x50mm, 5 µm (SN 97205-159370, Lot 12105). The mobile phase was isocratic MeOH at a flow rate of 40 mL/min for a 15 min run time, and the column oven temperature was set to 30 °C. The injection volume was 1500 µL and the sample concentration was approximately 20 mg/mL in a DCM/MeOH (1:1) solvent mixture. Detection for UV analysis was conducted at 254 nm.

HRMS: High resolution mass spectrometry (HRMS) was carried out using an Agilent Technologies 6530B accurate-mass QTOF Dual ESI mass spectrometer (MeOH + 0.1% formic acid, 0.25 mL/min, capillary voltage 4000 V, fragmentor 225 V) in positive-ion detection mode.

PXRD: Laboratory powder X-ray diffraction data were collected in transmission mode on samples held on thin Mylar film in aluminium well plates on a Panalytical X'Pert PRO MPD equipped with a high throughput screening (HTS) XYZ stage, X-ray focusing mirror and PIXcel detector, using Ni-filtered Cu Kα radiation. Data were measured over the range 4–50° in ~0.013° steps over 60 minutes.

TGA: Thermogravimetric analysis was carried out using a Q5000IR analyser (TA instruments) with an automated vertical overhead thermobalance. The sample was heated in aluminum pans under nitrogen at a rate of 10 °C/min.

Gas Sorption Analysis: Surface areas were measured by nitrogen sorption at 77.3 K. Powder samples were degassed offline at 90 °C for 15 hours under dynamic vacuum (10^{-5} bar) before analysis, followed by degassing on the analysis port under vacuum, also at 90 °C. Isotherms were measured using Micromeritics 2020 or 2420 volumetric adsorption analyzer. Gas uptake measurements (for N_2 and H_2) were taken at a temperature of 77 K, stabilized using a circulating water chiller/heater.

Diffusion NMR: All measurements were carried out non-spinning on a 400 MHz Bruker Avance 400 spectrometer, using a 5 mm indirect detection probe, equipped with a z-gradient coil producing a nominal maximum gradient of 34 G/cm. Diffusion data was collected using the Bruker pulse sequence 'gdstegp3s'. A diffusion encoding pulse δ of length 1–7 ms, and diffusion delay D of 0.1–0.25 s were used. Gradient amplitudes were equally spaced between 1.70 and 32.4 G/cm. Each FID was acquired using 16 k data points. All experiments were carried out at a nominal probe temperature of 298 K, with an air flow of 800 L/h to minimise convection. All diffusion co-efficients were calculated using measurements from multiple peak areas in the ^{1}H NMR spectra and the numbers quoted represent the mean.

Diffusion coefficients were calculated from signal intensities using the Skejskal-Tanner equation[1]:

$$I = I_0 e^{\gamma^2 g^2 \delta^2 \left(\Delta - \delta/_3\right) D}$$

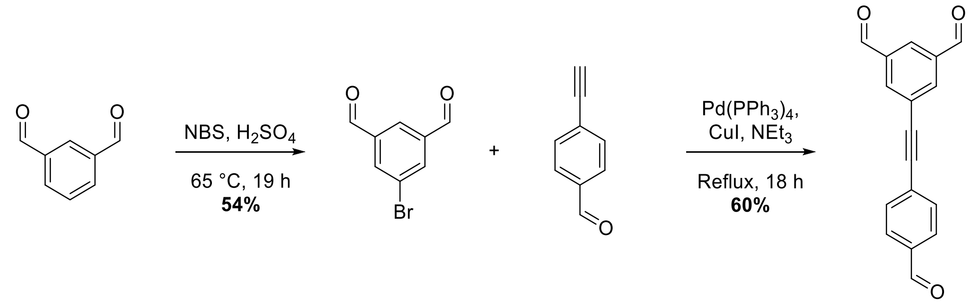
Where I is the signal intensity, I_0 is the signal intensity at a gradient strength of zero, g is the gradient strength, and D is the diffusion coefficient (D = m^2/s). Solvodynamic radii, R_S (nm), of solution-phase species were calculated from the Stokes-Einstein equation assuming molecules have a spherical geometry:

$$D = \frac{kT}{6\pi\eta R_S}$$

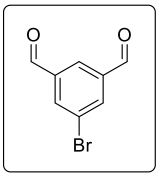
2. Synthesis and characterisation of cage precursors

2.1. Synthesis of trialdehyde precursor

Trialdehyde **1** was synthesised according to the following scheme, with the individual reaction steps described below:



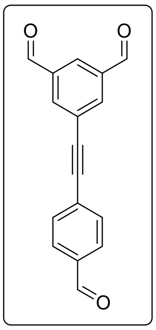
5-Bromoisophthalaldehyde, S1



To a round-bottomed flask, equipped with stirrer bar, was added isophthalaldehyde (50.00 g, 372.7 mmol, 1.0 eq.) followed by concentrated sulphuric acid (200 mL). The resulting mixture was heated to 65 °C, before direct heating was removed for the portionwise addition of *N*-bromosuccinimide (72.98 g, 410.1 mmol, 1.1 eq.) over 20 min. After complete addition, heating was resumed and the reaction was stirred at 65 °C for 19 hours. The reaction was allowed to cool to RT, and was poured into ice (~1 L) and stirred. The mixture was left for 1 hour before the resulting precipitate was collected by filtration. The collected solid was dissolved in DCM (1 L) and washed with water (2 x 200 mL). The organic layer was dried ($MgSO_4$) and hexane (500 mL) added, before the DCM was carefully removed *in-vauo* to afford a beige precipitate which was collected by filtration. The resulting solid was washed with a 1:2 methanol/hexane mixture (300 mL) and dried in vacuo to afford 5-bromoisophthalaldehyde **S1** which was used without further purification (43.14 g, 202.5 mmol, 54%).

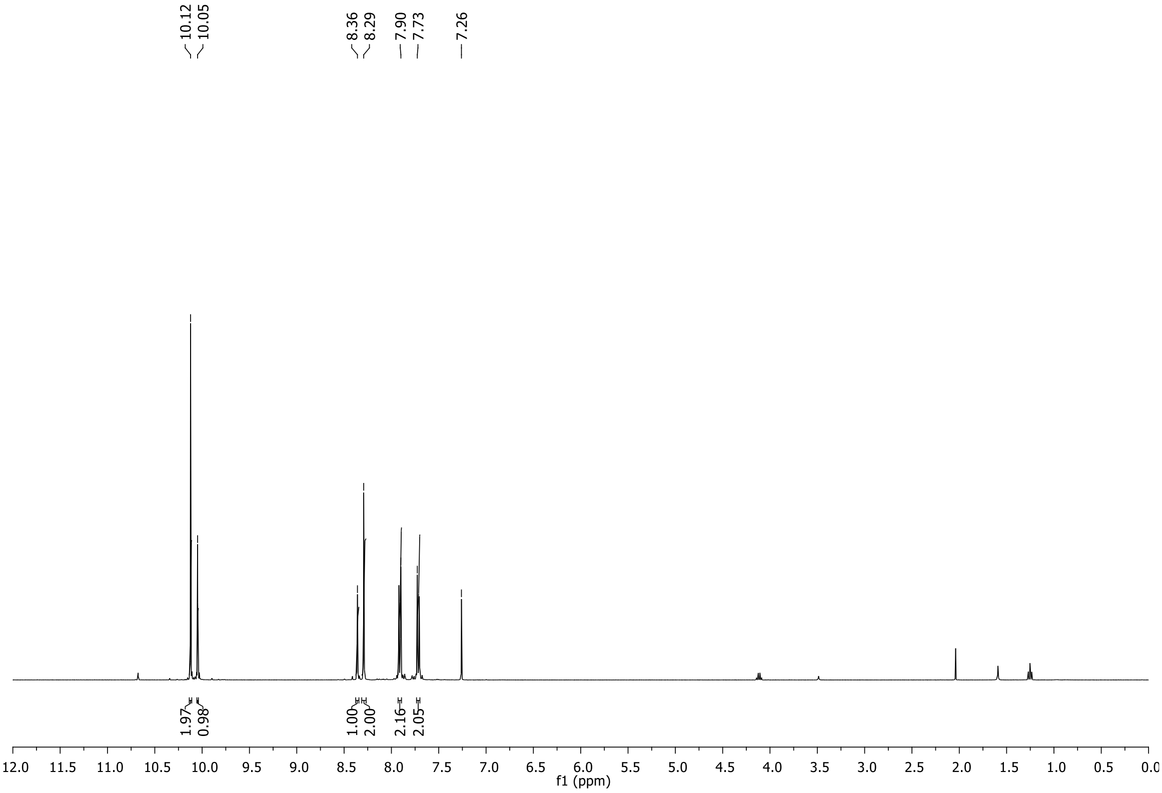
^{1}**H NMR** (400 MHz, $CDCl_3$) δ_H 10.06 (2H, s), 8.30 (1H, t, J = 1.4 Hz), 8.26 (2H, d, J = 1.4 Hz). Data in accordance with literature values.[2]

5-((4-Formylphenyl)ethynyl)isophthalaldehyde, 1



To an oven dried round-bottomed flask equipped with stirrer bar, was added 5-bromoisophthalaldehyde **S1** (1.17 g, 5.49 mmol, 1.0 eq.), 4-ethynylbenzaldehyde (1.00 g, 7.68 mmol, 1.4 eq.) and copper iodide (104 mg, 0.55 mmol, 0.1 eq.) before the flask was evacuated for 10 min and refilled with N_2. Triethylamine (40 mL) was added, and the solution degassed (N_2 bubbling, 15 min) prior to the addition of $Pd(PPh_3)_4$ (317 mg, 0.27 mmol, 0.05 eq.). The resulting suspension was stirred at reflux for 18 h before being allowed to cool to room temperature. The reaction was diluted with water (200 mL) and the aqueous layer extracted with DCM (3 × 200 mL). The combined organic layers were dried ($MgSO_4$), passed through a small Celite plug and concentrated *in vacuo*. Purification *via* column chromatography (0–50% EtOAc in hexane) afforded the desired product **1** as a yellow solid (865 mg, 3.29 mmol, 60%).

R$_f$ 0.65 (hexane/EtOAc (1:1)); **mpt** 165–168 °C; **IR** (ν_{max}/cm^{-1}) 3058, 2854, 1694, 1603, 1587, 1561, 1384, 1331, 1210, 1167, 1126, 1103; **^{1}H NMR** (400 MHz, $CDCl_3$) δ_H 10.12 (2H, s), 10.05 (1H, s), 8.36 (1H, t, *J* = 1.5 Hz), 8.29 (2H, d, *J* = 1.5 Hz), 7.91 (2H, d, *J* = 8.4 Hz), 7.72 (2H, d, *J* = 8.2 Hz); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C 191.40, 190.26, 137.42, 137.39, 136.22, 132.50, 130.34, 129.82, 128.32, 125.27, 91.49, 90.24; **HRMS** (ES+) calc. for $C_{17}H_{10}O_3$ 262.0630, found $[M+H]^+$ 263.0710.



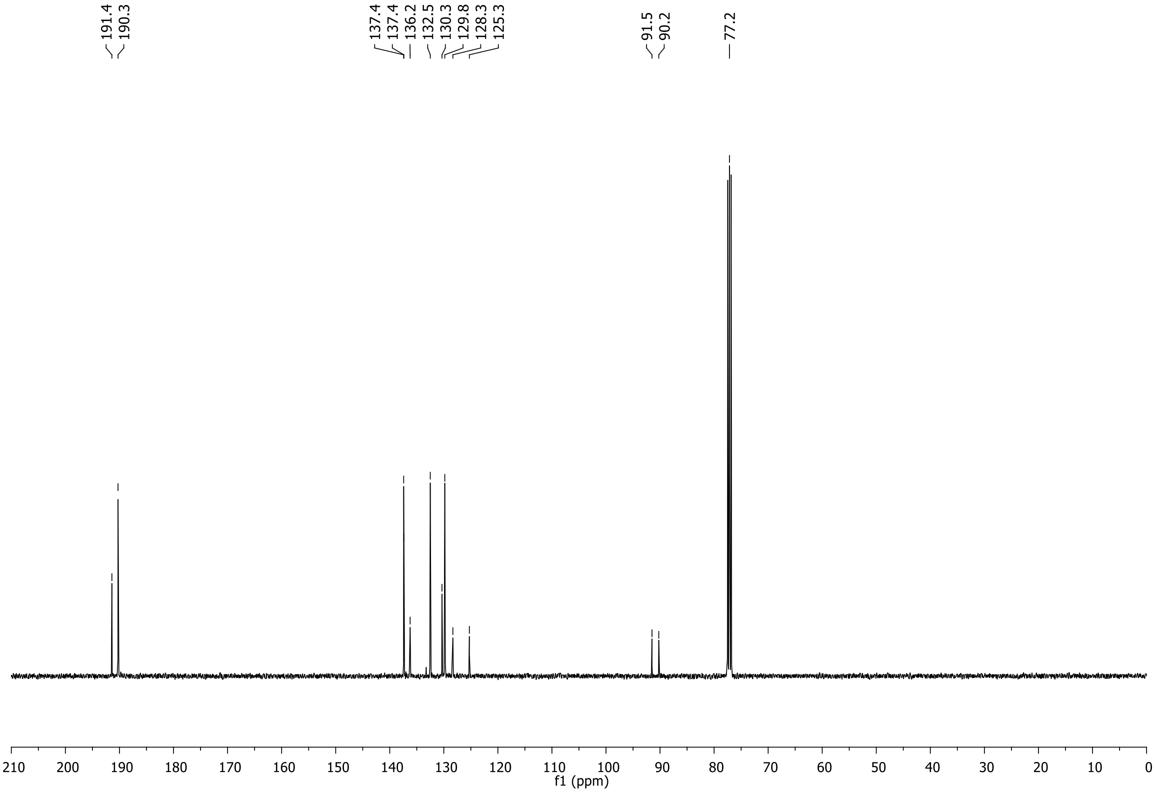
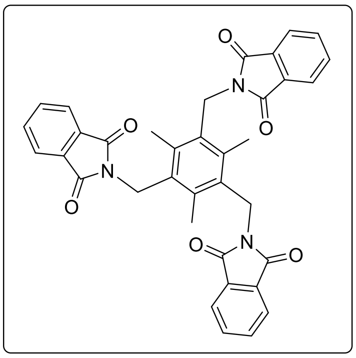


Figure S1: ^1H NMR ($CDCl_3$; upper) and ^{13}C NMR ($CDCl_3$; lower) for 5-((4-formylphenyl)ethynyl)isophthalaldehyde, **1**

2.2. Synthesis of triamine precursors

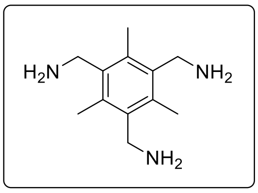
2,2',2''-((2,4,6-Trimethylbenzene-1,3,5-triyl)tris(methylene))tris(isoindoline-1,3-dione), S2



A modification of the procedure of Grawe *et al*. was used for this reaction.[3] To a solution of 1,3,5-tris(bromomethyl)-2,4,6-trimethylbenzene (20 g, 50.13 mmol, 1.0 eq.) and 18-crown-6 (3.97 g, 15.03 mmol, 0.3 eq.) in toluene (640 mL) was added potassium phthalamide (33.42 g, 180.47 mmol, 3.6 eq.). The mixture was heated at 100 °C under N_2 for 24 h before being allowed to cool to room temperature. The mixture was concentrated *in vacuo* and the resulting solid suspended in water (400 mL) and collected by filtration. The resulting solid was further washed with water (2 × 400 mL) and MeOH (400 mL) before being dried *in vacuo* to afford the desired product **S2** as a colourless solid which was used without further purification (30.45 g, quant).

^{1}H NMR (500 MHz, $CDCl_3$) δ_H 7.78 (6H, dd, *J* = 5.5, 3.0 Hz), 7.67 (6H, dd, *J* = 5.4, 3.1 Hz), 4.95 (6H, s), 2.50 (9H, s); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C 168.3, 138.7, 134.0, 132.1, 130.3, 123.4, 38.7, 17.5. Data in accordance with literature values.[4]

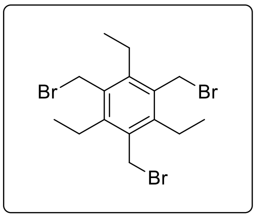
(2,4,6-Trimethylbenzene-1,3,5-triyl)trimethanamine, 15



A modification of the procedure of Grawe *et al*. was used for this reaction.[3] To a suspension of 2,2',2''-((2,4,6-trimethylbenzene-1,3,5-triyl)tris(methylene))tris(isoindoline-1,3-dione) **S2** (37.43 g, 62.63 mmol, 1.0 eq.) in a mixture of toluene (550 mL) and EtOH (1100 mL) was added hydrazine hydrate in a single portion (23.4 mL, 50 wt% solution in water, 375.79 mmol, 6.0 eq.). The resulting mixture was heated at 90 °C for 5 days, at which point a large amount of solid had precipitated, before being allowed to cool to room temperature. The reaction mixture was concentrated *in vacuo* (not to dryness) and partitioned between an aqueous KOH solution (200 mL, 40 wt%) and $CHCl_3$ (500 mL). The aqueous layer was further extracted with $CHCl_3$ (2 × 300 mL) before the combined organic layers were dried (Na_2SO_4) and concentrated *in vacuo* to afford the desired triamine **15** as a pale yellow solid which was used without further purification (12.20 g, 58.87 mmol, 94%).

mpt 126-128 °C; **IR** (ν_{max}/cm^{-1}) 3355 (br), 2907 (br), 2662 (br), 1565, 1455, 1376, 1299; **^{1}H NMR** (500 MHz, $CDCl_3$) δ_H 3.93 (6H, s), 2.45 (9H, s), 1.32 (6H, br s); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C 138.3, 133.6, 41.0, 15.6.

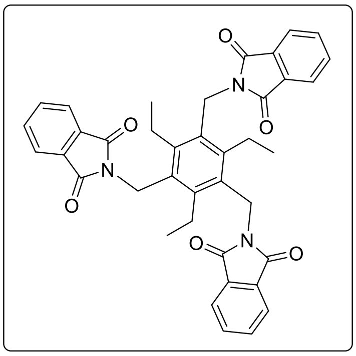
1,3,5-Tris(bromomethyl)-2,4,6-triethylbenzene, S3



The procedure of Vacca *et al*. was used for this reaction.[5] To a mixture of paraformaldehyde (16.7 g, 556.11 mmol, 10.5 eq.) and triethylbenzene (10 mL, 53.12 mmol, 1.0 eq.) in a solution of HBr in AcOH (100 mL, 30 wt%) was added zinc bromide (19.7 g, 87.47 mmol, 1.65 eq.) portionwise at room temperature. After complete addition, the reaction was heated to 90 °C under N_2 for 22 h before being allowed to cool to room temperature. The precipitated solid was collected by filtration, washed with water (3 × 400 mL) and dried *in vacuo* to afford the desired product **S3** as an off-white solid (19.23 g, 43.60 mmol, 82%).

^{1}H NMR (500 MHz, $CDCl_3$) δ_H 4.58 (6H, s), 2.94 (6H, q, *J* = 7.6 Hz), 1.34 (9H, t, *J* = 7.6 Hz); **^{13}C NMR** (126 MHz, $CDCl_3$) δ_C 145.2, 132.8, 28.7, 22.9, 15.8. Data in accordance with literature values.[5]

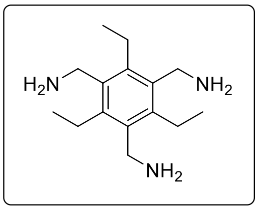
2,2',2''-((2,4,6-Triethylbenzene-1,3,5-triyl)tris(methylene))tris(isoindoline-1,3-dione), S4



A modification of the procedure of Grawe *et al*. was used for this reaction.[3] To a solution of 1,3,5-tris(bromomethyl)-2,4,6-triethylbenzene **S3** (19.2 g, 43.53 mmol, 1.0 eq.) and 18-crown-6 (3.45 g, 13.06 mmol, 0.3 eq.) in toluene (555 mL) was added potassium phthalamide (29.02 g, 156.72 mmol, 3.6 eq.). The mixture was heated at 100 °C under N_2 for 20 h before being allowed to cool to room temperature. The mixture was concentrated *in vacuo* and the resulting solid suspended in water (400 mL) and collected by filtration. The resulting solid was further washed with water (2 × 400 mL) and MeOH (400 mL) before being dried *in vacuo* to afford the desired product **S4** as a colourless solid which was used without further purification (26.79 g, 41.88 mmol, 96%).

^{1}H NMR (500 MHz, $CDCl_3$) δ_H 7.80 (6H, dd, *J* = 5.5, 3.0 Hz), 7.68 (6H, dd, *J* = 5.4, 3.1 Hz), 4.94 (6H, s), 3.10 (6H, q, *J* = 7.5 Hz), 0.97 (9H, t, *J* = 7.5 Hz); **^{13}C NMR** (101 MHz, $CDCl_3$) δ_C 168.4, 145.7, 134.0, 132.2, 129.6, 123.4, 37.6, 23.5, 15.9. Data in accordance to literature values.[5]

(2,4,6-Triethylbenzene-1,3,5-triyl)trimethanamine, 16



A modification of the procedure of Grawe *et al*. was used for this reaction.[3] To a suspension of 2,2',2''-((2,4,6-triethylbenzene-1,3,5-triyl)tris(methylene))tris(isoindoline-1,3-dione) **S4** (26.79 g, 41.88 mmol, 1.0 eq.) in a mixture of toluene (330 mL) and EtOH (670 mL) was added hydrazine hydrate in a single portion (15.6 mL, 50 wt% solution in water, 251.27 mmol, 6.0 eq.). The resulting mixture was heated at 90 °C for 44 h, at which point a large amount of solid had precipitated, before being allowed to cool to room temperature. The reaction mixture was concentrated *in vacuo* (not to dryness) and partitioned between an aqueous KOH solution (150 mL, 40 wt%) and $CHCl_3$ (400 mL). The aqueous layer was further extracted with $CHCl_3$ (2 × 200 mL) before the combined organic layers were dried (Na_2SO_4) and concentrated *in vacuo* to afford the desired triamine **16** as a pale yellow solid which was used without further purification (9.34 g, 37.45 mmol, 89%).

^{1}H NMR (500 MHz, $CDCl_3$) δ_H 3.87 (6H, s), 2.82 (6H, q, J = 7.6 Hz), 1.43 (6H, br s), 1.23 (9H, t, J = 7.5 Hz); **^{13}C NMR** (126 MHz, $CDCl_3$) δ_C 140.5, 137.6, 39.8, 22.7, 16.9. Data in accordance with literature values.[5]

3. High-throughput screening protocol and workflow

3.1. Synthetic screening

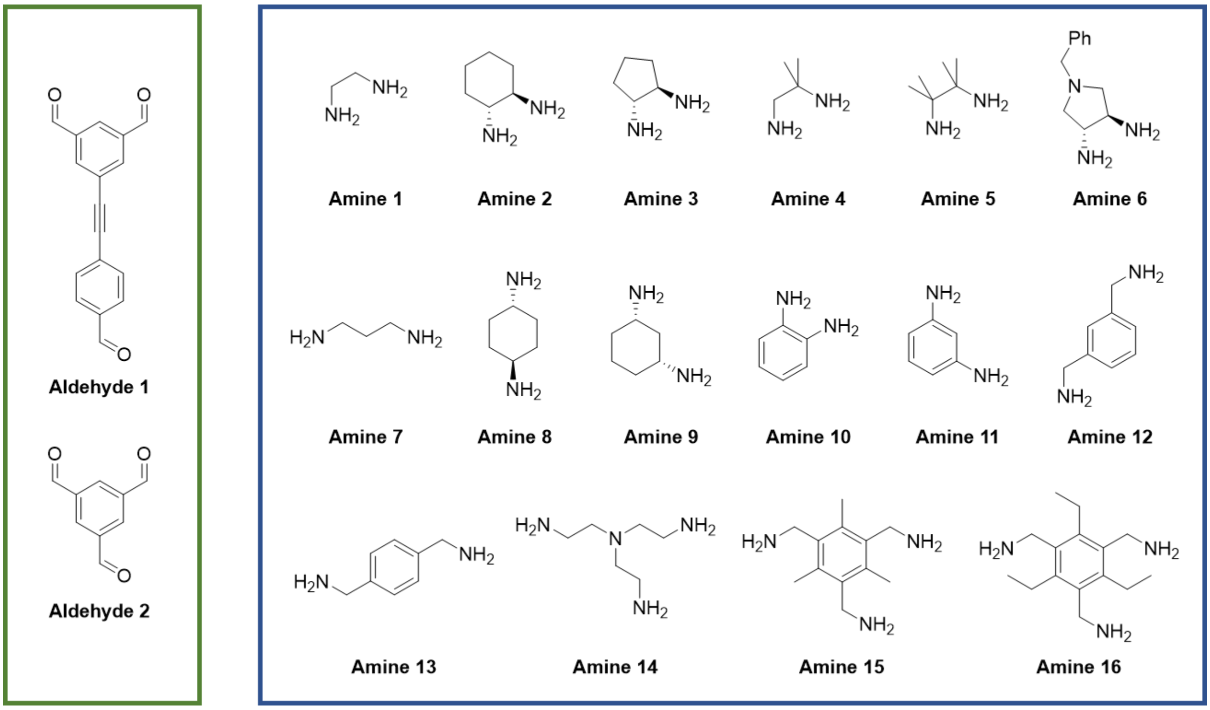


Figure S2: Summary of high-throughput cage screen: A range of ditopic (**1**–**13**) and tritopic (**14**–**16**) amines were selected to screen for the ideal partner for 5-((4-formylphenyl)ethynyl)isophthalaldehyde (**aldehyde 1**), leading to a total of 16 combinations for the high-throughput cage screen. A control reaction using the previously reported **CC3** was included to ensure a successful high-throughput screen.

Precursors: 1,3,5-Triformylbenzene (**aldehyde 2**), ethylenediamine (**amine 1**), (1*R*,2*R*)-cyclohexane-1,2-diamine (**amine 2**), (1*R*,2*R*)-cyclopentane-1,2-diamine dihydrochloride (**amine 3**), 1,2-diamino-2-methylpropane (**amine 4**), 2,3-dimethyl-2,3-butanediamine dihydrochloride (**amine 5**), (3*S*, 4*S*-(+)-3,4-diamino-1-benzylpyrrolidine (**amine 6**), 1,3-diaminopropane (**amine 7**), *trans*-1,4-cyclohexanediamine (**amine 8**), *cis*-1,3-cyclohexanediamine (**amine 9**), 1,2-phenylenediamine (**amine 10**), 1,3-phenylenediamine (**amine 11**), *m*-xylylenediamine (**amine 12**), *p*-xylylenediamine (**amine 13**) and tris(2-aminoethyl)amine (**amine 14**) were purchased from Sigma Aldrich, TCI UK, Apollo Scientific, or Manchester Organics. Amines (2,4,6-trimethylbenzene-1,3,5-triyl)trimethanamine (**amine 15**) and (2,4,6-triethylbenzene-1,3,5-triyl)trimethanamine (**amine 16**), and 5-((4-formylphenyl)ethynyl)isophthalaldehyde (**aldehyde 1**) were synthesised according to the procedures in section 2.

High-throughput screening method: All amine and aldehyde precursors (**Figure S2**) were dissolved in $CDCl_3$ using sonication to make stock solutions (5 mg/mL) for use in high-throughput screening (**Table S1**). On a Chemspeed Accelerator SLT-100 platform (**Figure S3**), the required volume of each amine stock solution, followed by the required volume of each aldehyde stock solution, was added to jacketed reactors (27 mL maximum volume) *via* liquid dispensing, followed by additional $CDCl_3$ to make each total volume up to 13 mL (**Table S2**). An excess of amine was used in each reaction (an additional 1 eq. relative to the cage stoichiometry being targeted) as it has previously been found to be well tolerated and promote further conversion to the targeted cages. The resulting solutions were vortexed at 800 rpm at room temperature for 3 days before 5 mL of each reaction was volumetrically transferred into vials for analysis. The remainder of the solutions were heated to 65 °C for 3 days, before being allowed to cool to room temperature and analysed. All reactions removed from reactor vessels and filtered through a small cotton wool plug to remove any insoluble precipitate prior to analysis.

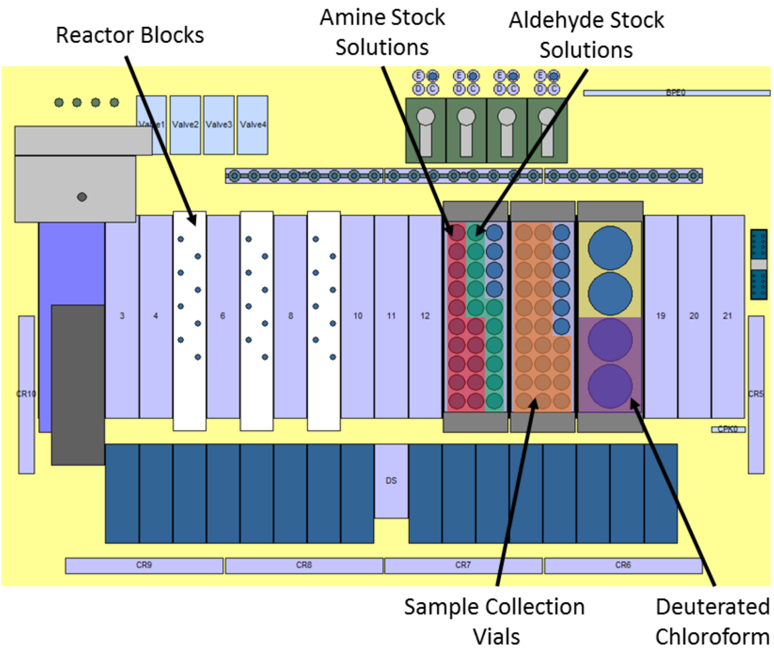


Figure S3: General Chemspeed Accelerator SLT-100 deck layout used for the high-throughput synthetic screen. Reactors blocks on the left contained the cage forming reactions after addition of the stock solutions of the amines and aldehydes, and top-up solvent (deuterated chloroform), by liquid dispensing from the right.

Table S1: Precursor stock solutions for high-throughput screening

Stock Solution Number	Reactant	MW (g/mol)	Stock Solution Concentration (mg/mL)	Stock Solution Concentration (mmol/mL)
1	Aldehyde 1	262.26	5	0.0191
2	Aldehyde 2	162.14	5	0.0308
3	Amine 1	60.10	5	0.0832
4	Amine 2	114.19	5	0.0438
5*	Amine 3	173.08	5	0.0289
6	Amine 4	88.15	5	0.0567
7*	Amine 5	189.12	5	0.0264
8	Amine 6	191.28	5	0.0261
9	Amine 7	74.13	5	0.0674
10	Amine 8	114.19	5	0.0438
11	Amine 9	114.19	5	0.0438
12	Amine 10	108.14	5	0.0462
13	Amine 11	108.14	5	0.0462
14	Amine 12	136.20	5	0.0367
15	Amine 13	136.20	5	0.0367
16	Amine 14	146.24	5	0.0342
17	Amine 15	207.32	5	0.0241
18	Amine 16	249.39	5	0.0200

*Diamine hydrochloride salts used, therefore triethylamine (0.1 mL) and methanol (1 mL) added to the stock solutions.

Table S2: Reaction volumes used in each reaction on the Chemspeed Accelerator SLT-100

Reaction Number	Aldehyde Reactant	Volume Aldehyde Stock Solution (mL)	Amount of Aldehyde (mmol)	Amine Reactant	Ratio of equiv. used	Amount Amine required (mmol)	Volume Amine Stock Solution (mL)	Additional Amount of $CDCl_3$ (mL) added (Total Volume = 13 mL)
1	1	2.53	0.0482	1	4:7	0.0603	0.72	9.75
2	1	2.53	0.0482	2	4:7	0.0603	1.38	9.09
3	1	2.53	0.0482	3	4:7	0.0603	2.09	8.38
4	1	2.53	0.0482	4	4:7	0.0603	1.06	9.41
5	1	2.53	0.0482	5	4:7	0.0603	2.28	8.19
6	1	2.53	0.0482	6	4:7	0.0603	2.31	8.16
7	1	2.53	0.0482	7	4:7	0.0603	0.89	9.58
8	1	2.53	0.0482	8	4:7	0.0603	1.38	9.09
9	1	2.53	0.0482	9	4:7	0.0603	1.38	9.09
10	1	2.53	0.0482	10	4:7	0.0603	1.30	9.17
11	1	2.53	0.0482	11	4:7	0.0603	1.30	9.17
12	1	2.53	0.0482	12	4:7	0.0603	1.64	8.83
13	1	2.53	0.0482	13	4:7	0.0603	1.64	8.83
14	1	2.53	0.0482	14	4:5	0.0603	1.76	8.71
15	1	2.53	0.0482	15	4:5	0.0603	2.50	7.97
16	1	2.53	0.0482	16	4:5	0.0603	3.01	7.46
Control	2	1.12	0.0345	2	4:7	0.0604	1.38	10.50

NB. Reaction number corresponds to the amine used in the reaction (see Figure S2), with the exception of the control reaction.

3.2. Summary of high-throughput screen

To determine if any cage had formed during the high-throughput screen, crude reaction aliquots were taken after 3 days at room temperature, and after a further 3 days at 65 °C, for both ^{1}H NMR and HRMS analysis. Overall, it was unclear if any cage had formed based on ^{1}H NMR analysis, with either a large amount of starting material remaining, or a complex mixture produced, presumably due to the likelihood of the cage products being unsymmetrical if formed. Therefore, hits from the screen were determined by HRMS based on the indication of a clear mass ion corresponding to the formation of a cage (Table S3).

The most promising partner amine was **Amine 2,** which at room temperature and after heating at 65 °C for 3 days showed mass ions corresponding to a [4+6] cage in the mass spectrum (entry 2 and Figure S5), and was therefore investigated further. Other vicinal diamines (**amines 1**, **3**, **4**, and **6**) also gave an indication of forming [4+6] cages, while **amines 7** and **12** indicated the formation of [2+3] cages (which would likely not contain a cavity), but were not further investigated.

Table S3: Summary of HRMS characterisation data from high-throughput cage screen on crude reaction solutions. The most promising partner, **Amine 2**, is highlighted in green.

Reaction Number	Conditions	HRMS (direct injection- - Cage Formation?)			
		Cage Formation?	Mass ions indicating cage formation?	Main m/z ion?	Size Cage Indicated?
1	RT, 3 days	✓	$[M+2H]^{2+}$ 597.2837; $[M+3H]^{3+}$ 398.5257	✗	**[4+6]**
	65 °C, 3 days	✓	$[M+2H]^{2+}$ 597.2811; $[M+3H]^{3+}$ 398.5232	✗	
2	RT, 3 days	✓	$[M+2H]^{2+}$ 759.9285; $[M+3H]^{3+}$ 506.9561	✗	**[4+6]**
	65 °C, 3 days	✓	$[M+2H]^{2+}$ 759.9246; $[M+3H]^{3+}$ 506.9535	✓	
3	RT, 3 days	✓	$[M+2H]^{2+}$ 717.8832; $[M+3H]^{3+}$ 478.9233	✗	**[4+6]**
	65 °C, 3 days	✓	$[M+2H]^{2+}$ 717.8759; $[M+3H]^{3+}$ 478.9206	✗	
4	RT, 3 days	✓	$[M+2H]^{2+}$ 681.3779; $[M+3H]^{3+}$ 454.5889	✗	**[4+6]**
	65 °C, 3 days	✗	✗	-	
5	RT, 3 days	✗	✗	-	-
	65 °C, 3 days	✗	✗	-	
6	RT, 3 days	✓	$[M+3H]^{3+}$ 661.0105; $[M+4H]^{4+}$ 496.0102	✗	**[4+6]**
	65 °C, 3 days	✓	$[M+3H]^{3+}$ 661.0076; $[M+4H]^{4+}$ 496.0082	✗	
7	RT, 3 days	✓	$[M+H]^{+}$ 639.3314; $[M+2H]^{2+}$ 320.1698	✗	**[2+3]**
	65 °C, 3 days	✓	$[M+H]^{+}$ 639.3292; $[M+2H]^{2+}$ 320.1684	✗	
8	RT, 3 days	✗	✗	-	-
	65 °C, 3 days	✗	✗	-	
9	RT, 3 days	✗	✗	-	-
	65 °C, 3 days	✗	✗	-	
10	RT, 3 days	✗	✗	-	-
	65 °C, 3 days	✗	✗	-	
11	RT, 3 days	✗	✗	-	-
	65 °C, 3 days	✗	✗	-	
12	RT, 3 days	✓	$[M+H]^{+}$ 825.3793; $[M+2H]^{2+}$ 413.1956	✓	**[2+3]**
	65 °C, 3 days	✓	$[M+H]^{+}$ 825.3777; $[M+2H]^{2+}$ 413.1932	✓	
13	RT, 3 days	✗	✗	-	-
	65 °C, 3 days	✗	✗	-	
14	RT, 3 days	✗	✗	-	-
	65 °C, 3 days	✗	✗	-	
15	RT, 3 days	✗	✗	-	-
	65 °C, 3 days	✗	✗	-	
16	RT, 3 days	✗	✗	-	-
	65 °C, 3 days	✗	✗	-	
Control	RT, 3 days	✓	$[M+2H]^{2+}$ 559.3618		**[4+6]**
	65 °C, 3 days	✓	$[M+2H]^{2+}$ 559.3612		

3.3. HRMS spectra of combinations showing cage formation

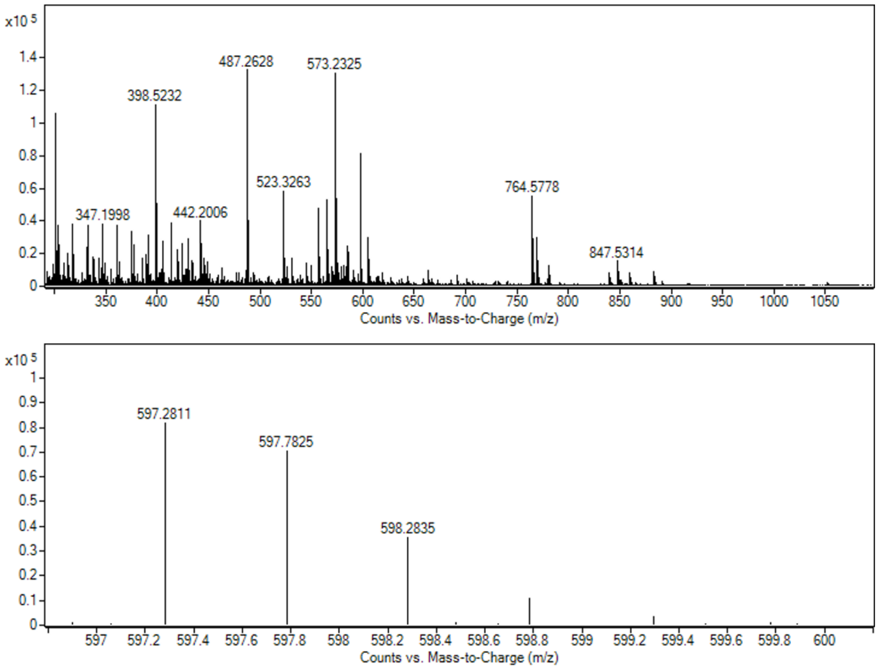


Figure S4: HRMS spectra for crude reaction **1**, taken directly from the high-throughput screen after heating at 65 °C for 3 days, indicating some formation of a [4+6] cage — calc. for $C_{80}H_{64}N_{12}$ 1192.5377, found $[M+2H]^{2+}$ 597.2811 and $[M+3H]^{3+}$ 398.5232.

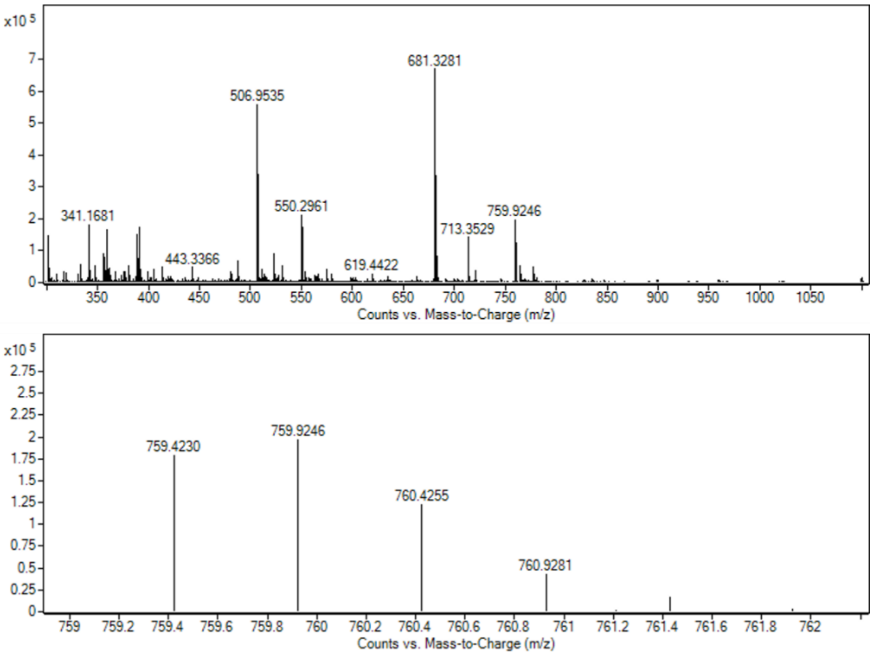


Figure S5: HRMS spectra for crude reaction **2**, taken directly from the high-throughput screen after heating at 65 °C for 3 days, indicating clear formation of a [4+6] cage — calc. for $C_{104}H_{100}N_{12}$ 1517.8227, found $[M+2H]^{2+}$ 759.9246 and $[M+3H]^{3+}$ 506.9535.

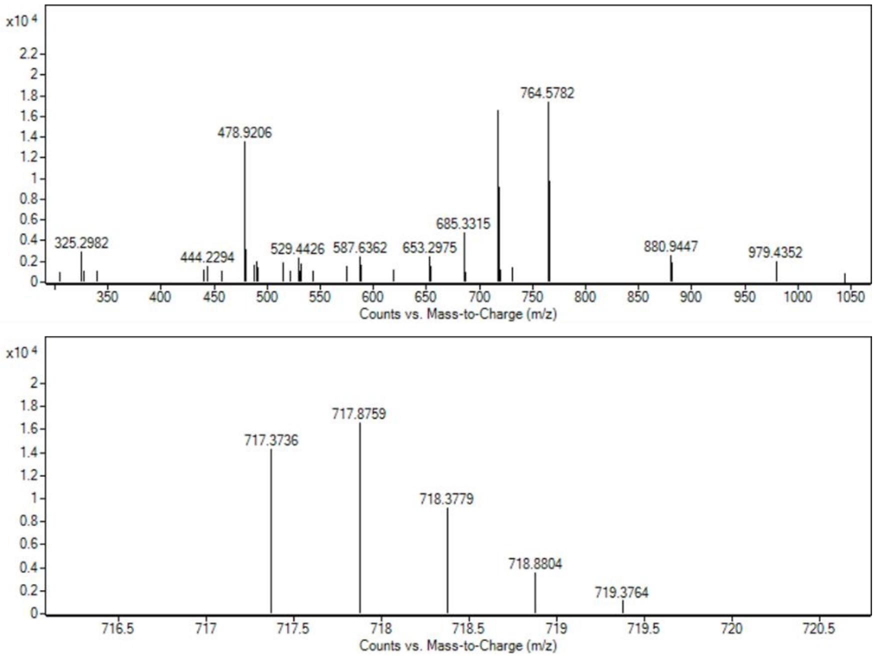


Figure S6: HRMS spectra for crude reaction **3**, taken directly from the high-throughput screen after heating at 65 °C for 3 days, indicating formation of a [4+6] cage — calc. for $C_{98}H_{88}N_{12}$ 1433.7286, found $[M+2H]^{2+}$ 717.8759 and $[M+3H]^{3+}$ 478.9206.

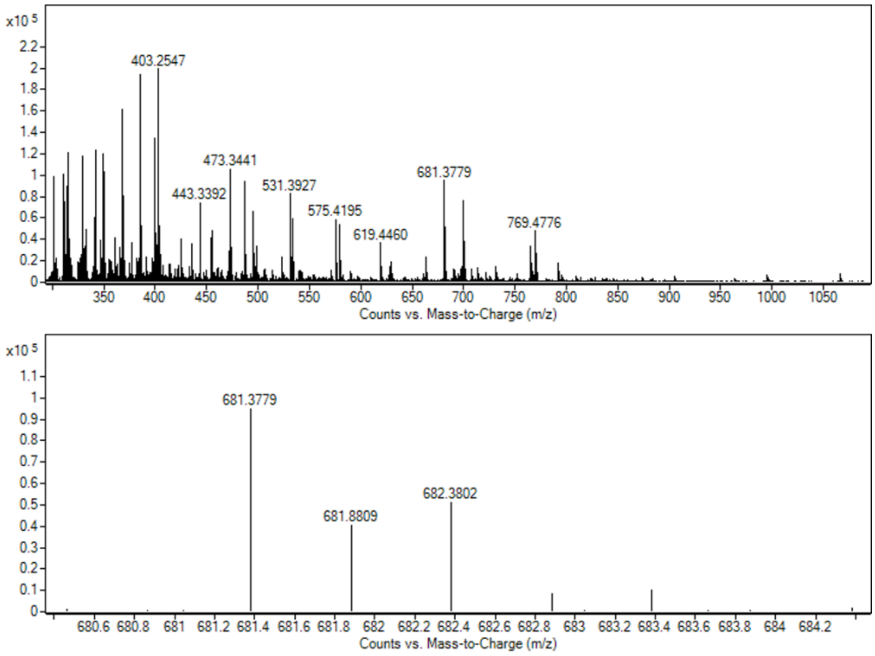


Figure S7: HRMS spectra for crude reaction **4**, taken directly from the high-throughput screen after stirring at room temperature for 3 days, indicating some formation of a [4+6] cage — calc. for $C_{92}H_{88}N_{12}$ 1361.7286, found $[M+2H]^{2+}$ 681.3779 and $[M+3H]^{3+}$ 454.5889.

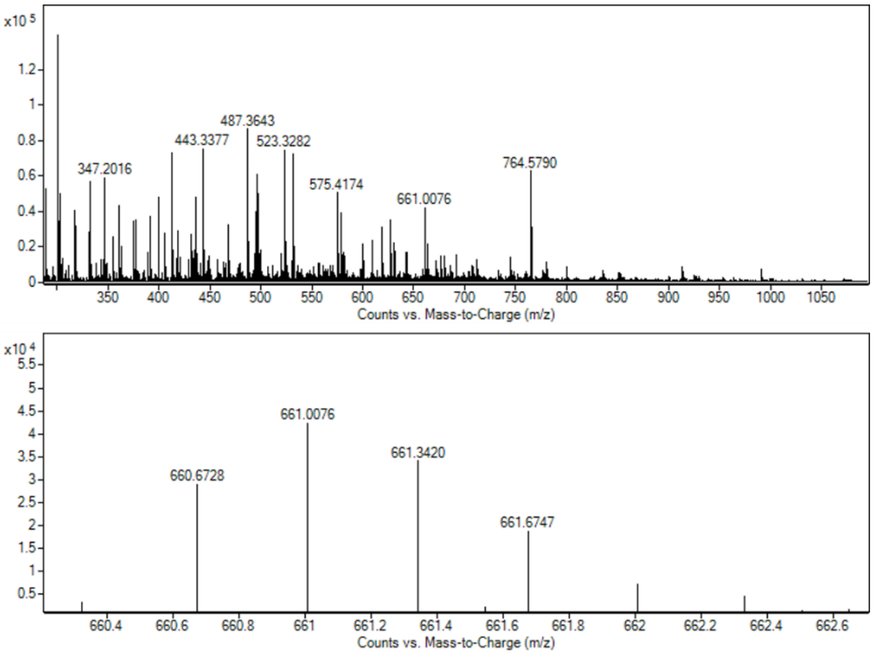


Figure S8: HRMS spectra for crude reaction **6**, taken directly from the high-throughput screen after heating at 65 °C for 3 days, indicating some formation of a [4+6] cage — calc. for $C_{134}H_{118}N_{18}$ 1979.9818, found $[M+3H]^{3+}$ 661.0076 and $[M+4H]^{4+}$ 496.0082.

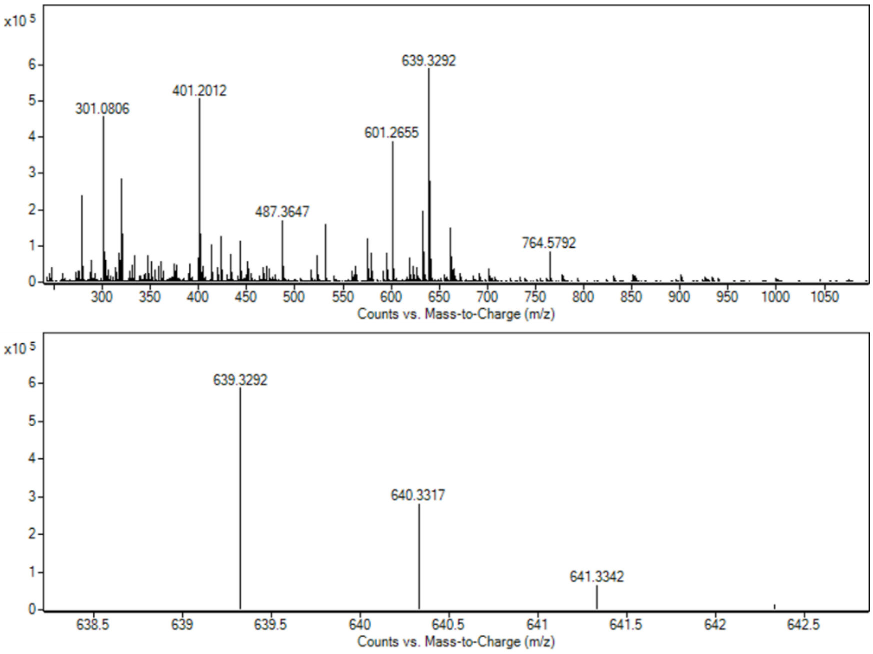


Figure S9: HRMS spectra for crude reaction **7**, taken directly from the high-throughput screen after heating at 65 °C for 3 days, indicating formation of a [2+3] cage — calc. for $C_{43}H_{38}N_6$ 638.3158, found $[M+H]^+$ 639.3292 and $[M+2H]^{2+}$ 320.1684.

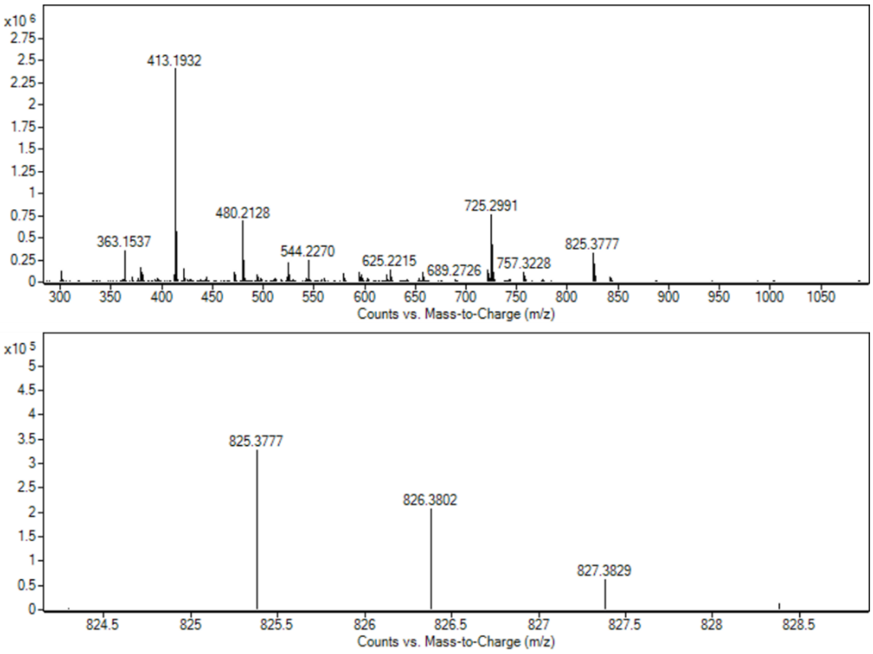
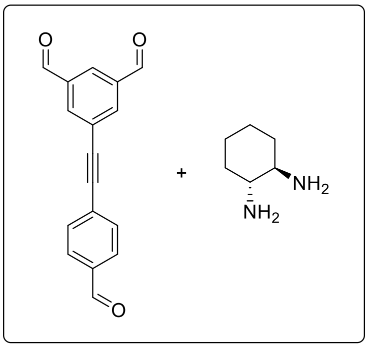


Figure S10: HRMS spectra for crude reaction **12**, taken directly from the high-throughput screen after heating at 65 °C for 3 days, indicating some formation of a [2+3] cage — calc. for $C_{58}H_{44}N_6$ 824.3627, found $[M+H]^+$ 825.3777 and $[M+2H]^{2+}$ 413.1932.

4. Scale-up and characterisation of unsymmetrical cage



Method A: A solution of 5-((4-formylphenyl)ethynyl)-isophthalaldehyde **1** (1.0 g, 3.81 mmol, 4.0 eq.) and (1*R*,2*R*)-cyclohexane-1,2-diamine (0.65 g, 5.72 mmol, 6.0 eq.) in DCM (1.23 L) was stirred at room temperature for 5 days before the addition of hexane (500 mL). The DCM was removed *in vacuo* and the pale yellow solid collected by filtration and dried *in vacuo* to afford the cage product as a pale yellow solid (1.4 g, 0.92 mmol, 96% mass recovery, 91% purity by HPLC).

Method B (pure cage product): A solution of 5-((4-formylphenyl)ethynyl)isophthalaldehyde **1** (300 mg, 1.14 mmol, 4.0 eq.) and (1*R*,2*R*)-cyclohexane-1,2-diamine (196 mg, 1.71 mmol, 6.0 eq.) in DCM (360 mL) was stirred at room temperature for 5 days before concentration *in vacuo to* ~10 mL. Methanol (10 mL) was added to the crude solution and the sample was purified by preparative HPLC. The collected product fractions were combined and carefully concentrated *in vacuo* (20 °C). The resulting solid was re-dissolved in DCM (50 mL) and the solution dried (Na_2SO_4), before the addition of hexane (100 mL) and removal of DCM *in vacuo*. The resulting precipitate was collected by filtration and dried under vacuum to afford the cage product as a very pale yellow solid (104 mg, 0.068 mmol, 24% yield, >98% purity by HPLC).

mpt > 190 °C, no decomposition observed; **IR** (ν_{max}/cm^{-1}) 2931, 2858, 2209, 1642, 1604, 1556, 1448, 1372, 1343, 1300, 1158, 1139, 1087; **^{1}H NMR** (500 MHz, $CDCl_3$) δ_H 8.289 (1H, s), 8.244 (1H, s), 8.236 (1H, s), 8.204 (1H, s), 8.202 (1H, s), 8.196 (1H,s), 8.172 (3H, m), 8.163 (1H, s), 8.131 (1H, s), 8.093 (1H, t, *J* = 1.6 Hz), 8.087 (1H, s), 8.024 (1H, t, *J* = 1.5 Hz), 7.998 (1H, t, *J* = 1.5 Hz), 7.958 (1H, t, *J* = 1.5 Hz), 7.914 (3H, m), 7.823 (1H, t, *J* = 1.5 Hz), 7.571–7.520 (8H, m), 7.462 (1H, t, *J* = 1.6 Hz), 7.446 (1H, s), 7.435 (1H, t, *J* = 1.6 Hz), 7.429 (1H, s), 7.417 (2H, m), 7.403 (3H, m), 7.385 (1H, s), 7.368 (1H, s), 7.350 (1H, t, *J* = 1.5 Hz), 3.518–3.289 (12H, m), 1.919–1.647 (36H, m), 1.538–1.419 (12H, m); **^{13}C NMR** (126 MHz, $CDCl_3$) δ_C 160.61, 160.54, 159.97, 159.85, 159.69, 159.38, 159.34, 159.28, 159.10, 159.08, 158.87, 158.30, 137.18, 136.89, 136.83, 136.80, 136.74, 136.71, 136.56, 136.49, 136.17, 136.12, 136.06, 136.02, 135.68, 135.48, 135.00, 131.91, 131.89, 131.86, 131.82, 130.60, 130.47, 130.30, 129.97, 128.03, 127.96, 127.90, 127.87, 127.67, 127.53, 127.46, 127.19, 125.14, 125.06, 125.03, 124.99, 123.98, 123.91, 123.86, 123.67, 90.31, 90.29, 90.24, 90.08, 90.03, 89.97, 89.90, 75.47, 75.26, 75.22, 75.12, 74.78, 74.74, 74.51, 74.41, 74.38, 74.35, 74.14, 73.65, 33.70, 33.51, 33.40, 33.38, 33.16, 33.13, 33.10, 32.96, 32.82, 32.77, 32.62, 24.68, 24.65, 24.58, 24.50, 24.45, 24.44; **HRMS** (ES+) calc. for $C_{104}H_{100}N_{12}$ 1517.8227, found $[M+2H]^{2+}$ 759.9211 and $[M+3H]^{3+}$ 506.9507; **CHN Analysis** calc. for $C_{104}H_{100}N_{12}$: C, 82.29; H, 6.64; N, 11.07; found: C, 79.72; H, 6.53; N, 10.83.

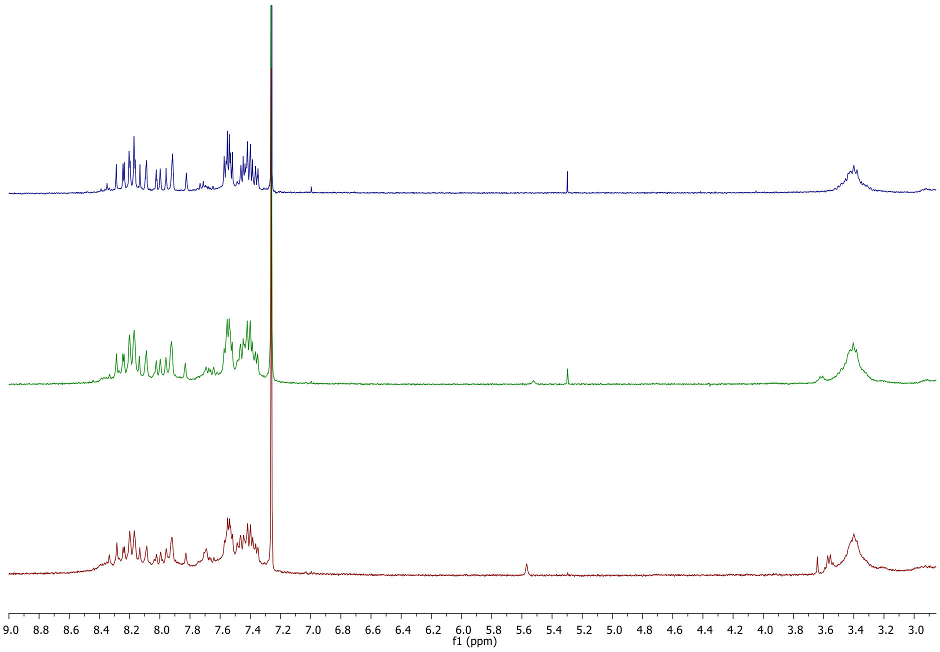


Figure S11: Stacked ^{1}H NMR spectra of different batches of unsymmetrical cage, prior to purification by prep-HPLC, demonstrating synthetic reproducibility in the formation of the main cage species, with slight variation in the amount of side-products formed, likely due to inaccuracies in the reaction stoichiometry.

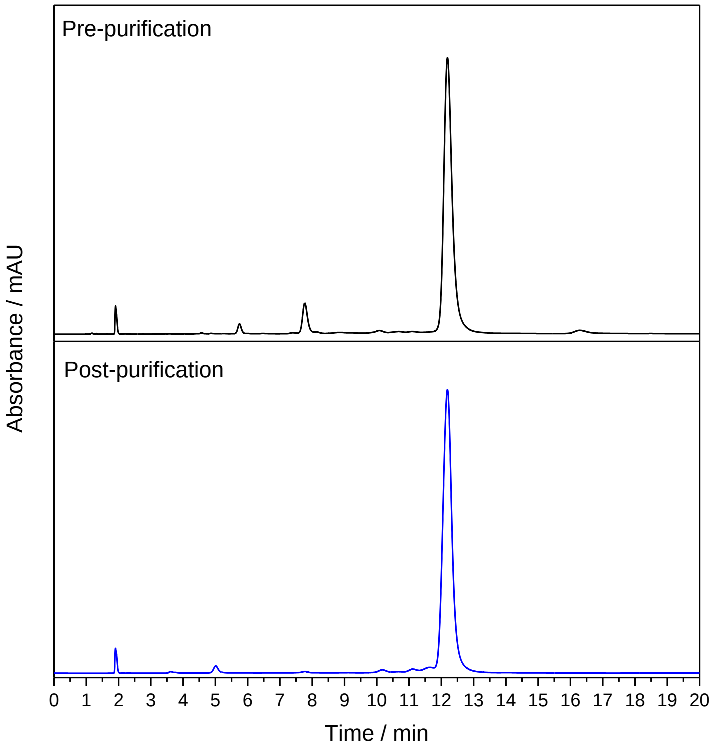


Figure S12: Stacked HPLC traces showing unsymmetrical cage at 12.2 min pre- (black trace), and post-purification (blue trace) by prep-HPLC, showing an improvement in purity from 91% to >98% (a/a). The peak at ~2 min is the solvent front.

HPLC analysis was carried out using previously optimised conditions for the separation and analysis of organic cages (Thermo-Scientific Syncronis C8 column, 150x4.6 mm, 3 µm (SN 10136940, Lot 12459, mobile phase = isocratic MeOH, flow rate = 1 mL/min). This column and solvent system has successfully separated the different cage species in scrambled cage mixtures where the exterior vertices differ slightly in which alkyl functionality is present,[6,7] and **CC14**, which is analogous to **CC3** but with a single 1,3,5-triformylbenzene replaced with 1,3,5-triacetylbenzene.[8] Based on this, the HPLC trace suggests the formation of a single cage species.

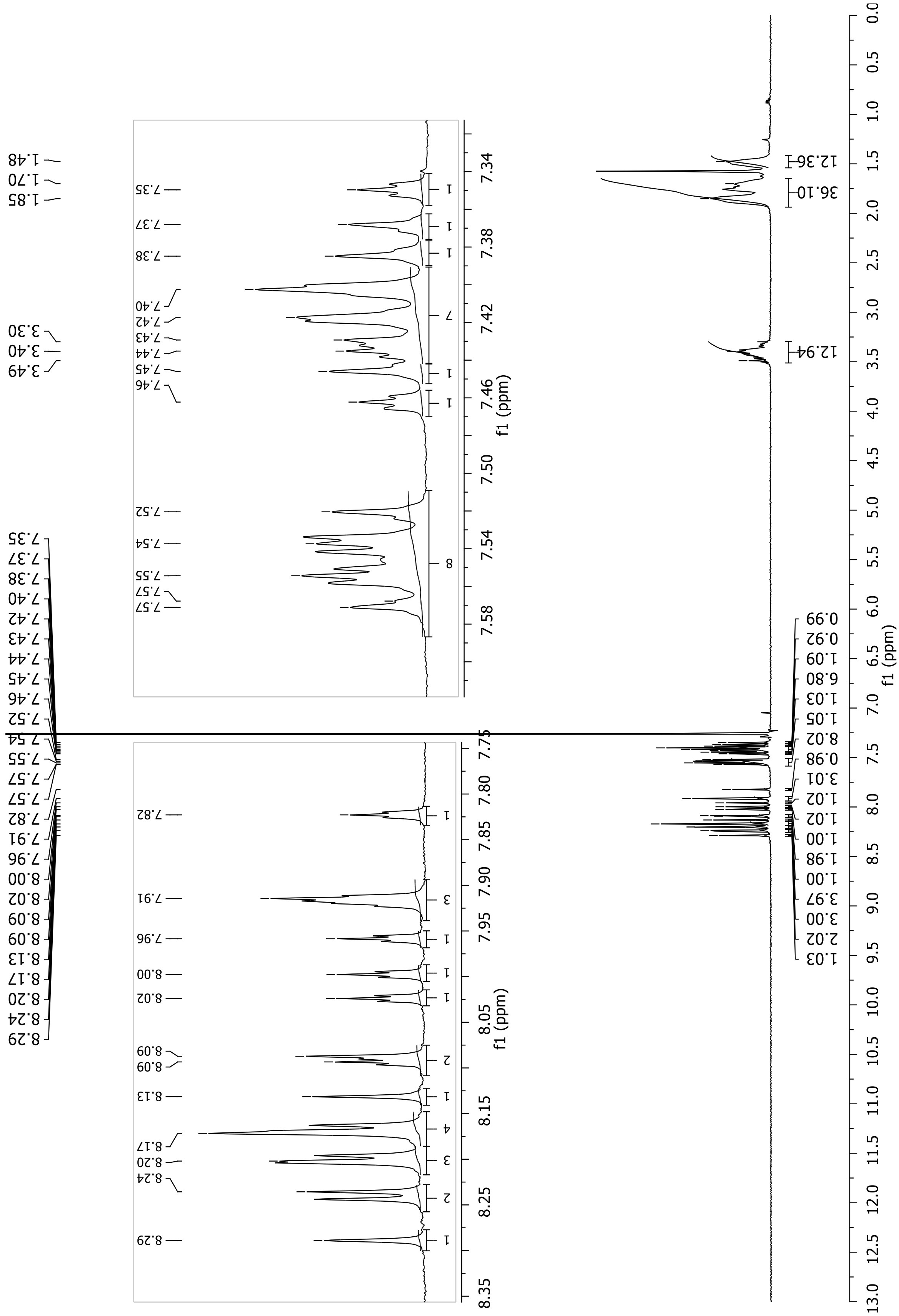


Figure S13: ^{1}H NMR ($CDCl_3$) for unsymmetrical cage after purification by prep-HPLC.

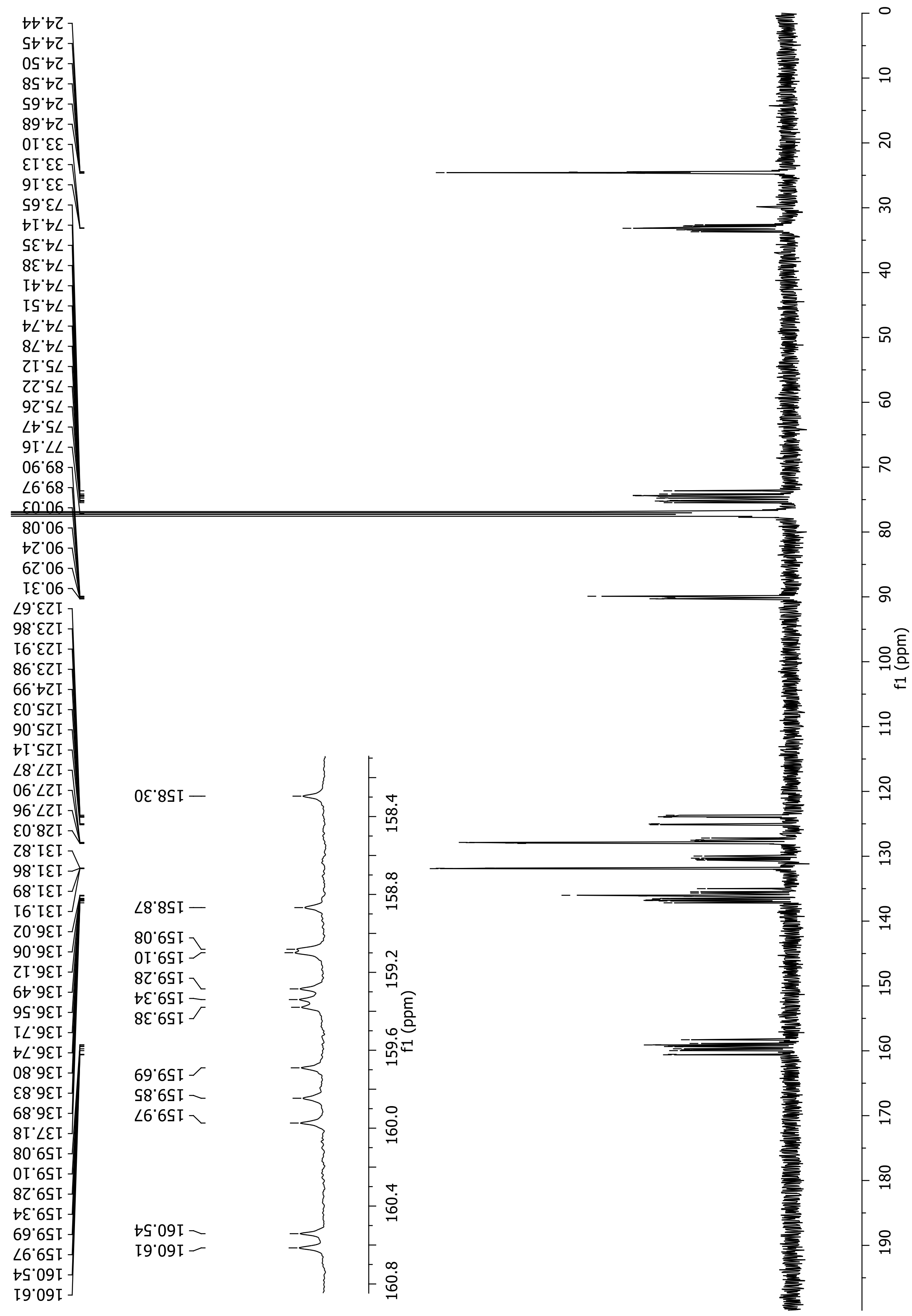


Figure S14: ^{13}C NMR (CDCl$_3$) for unsymmetrical cage after purification by prep-HPLC.

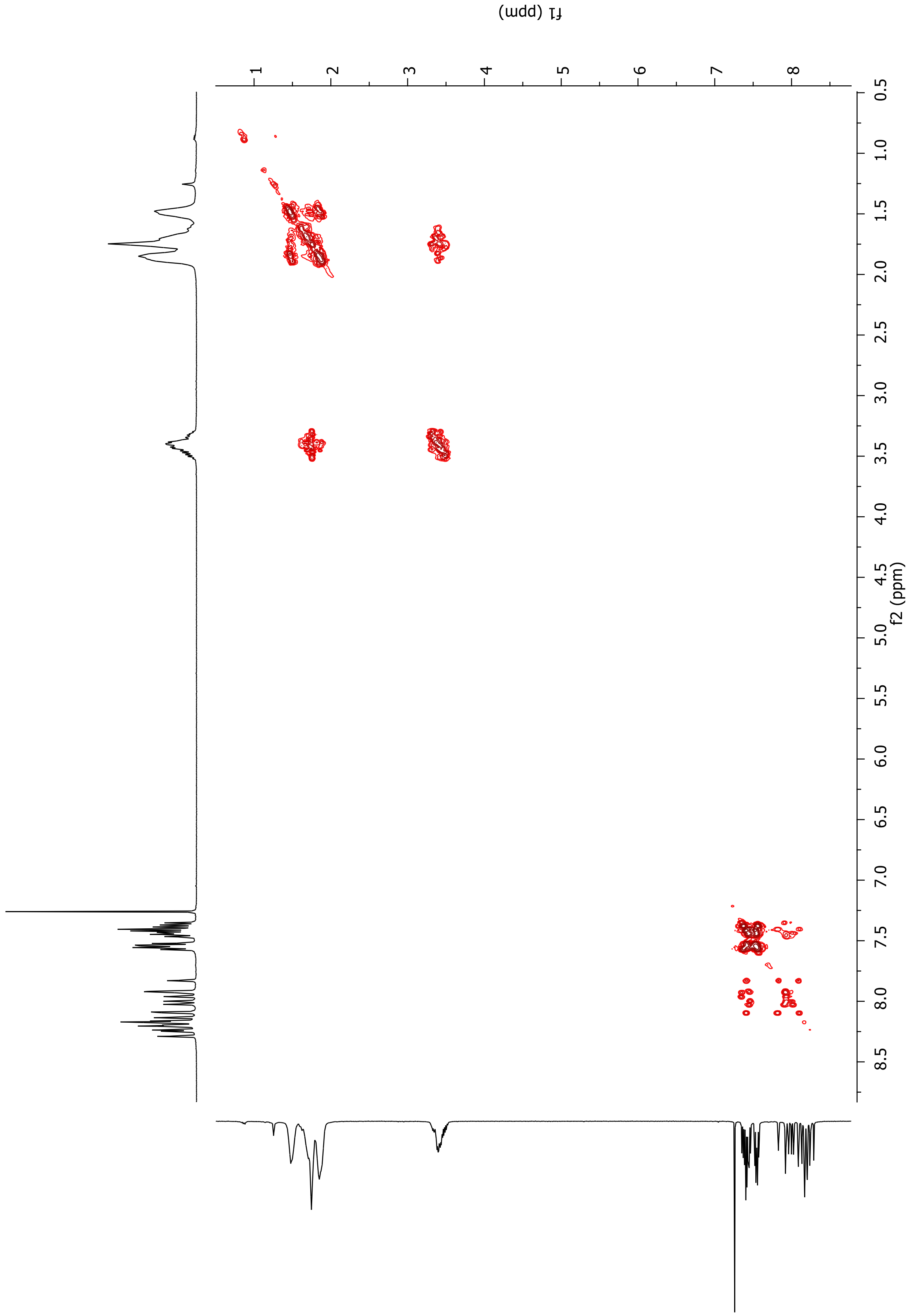


Figure S15: COSY ($CDCl_3$) for unsymmetrical cage after purification by prep-HPLC.

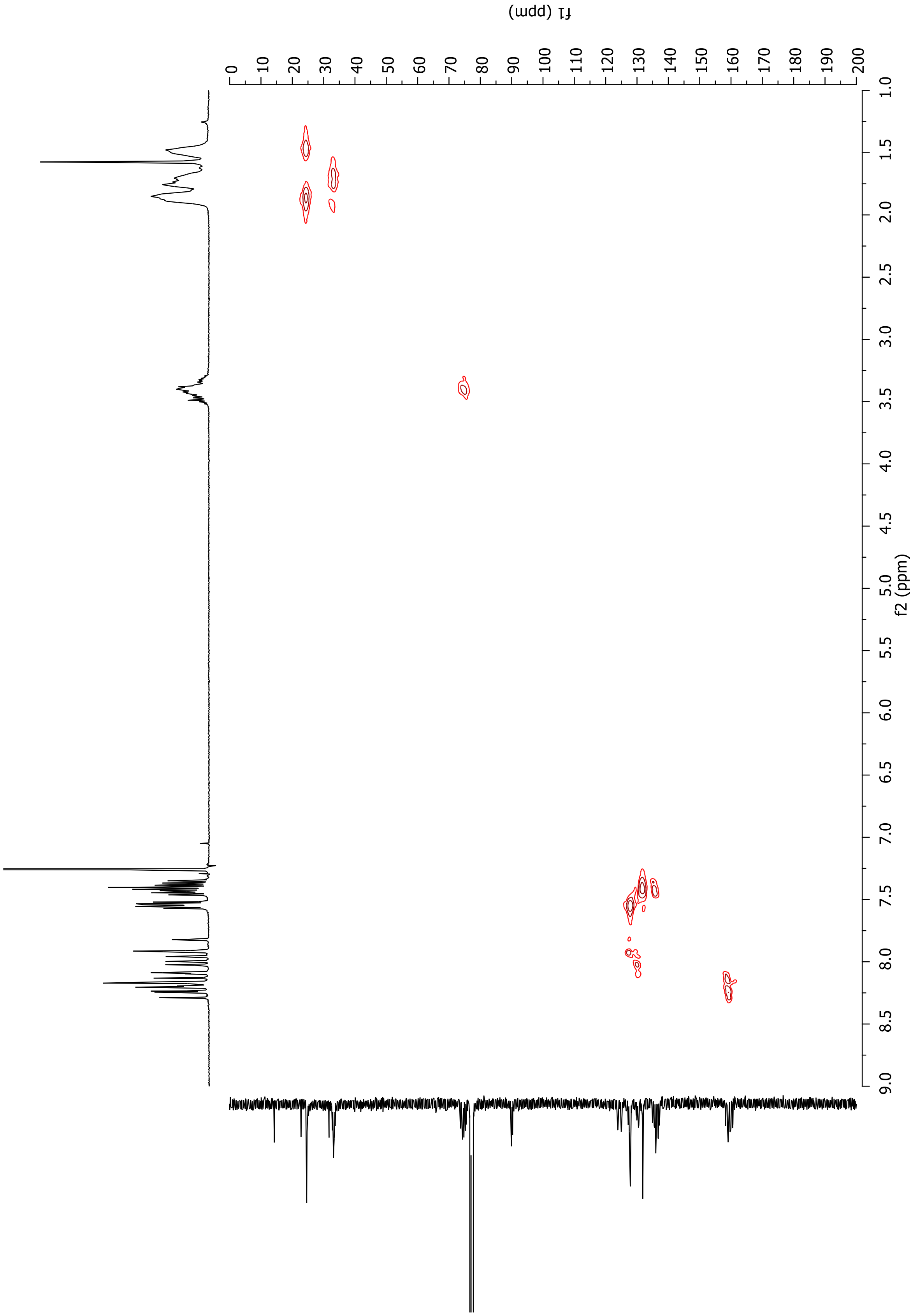


Figure S16: HSQC ($CDCl_3$) for unsymmetrical cage after purification by prep-HPLC.

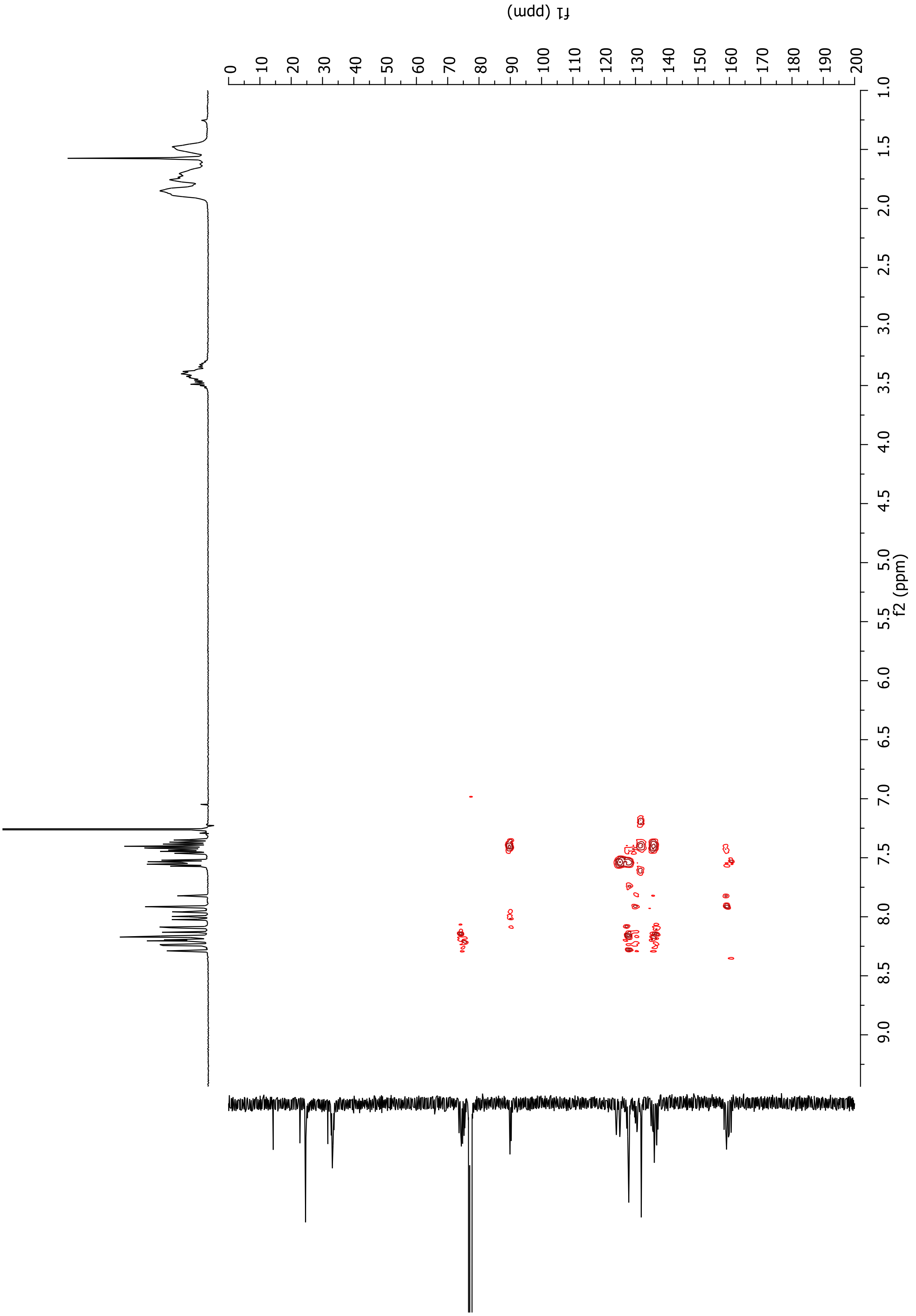


Figure S17: HMBC ($CDCl_3$) for unsymmetrical cage after purification by prep-HPLC.

Interpretation of NMR Data:

The presence of discrete peaks for the different protons and carbons, with whole number integrals in the ^{1}H NMR spectrum, suggests the formation of a single cage species. This is in contrast to previously reported cage mixtures, such as scrambled cages where broad NMR signals were observed,[6,7] and more symmetrical species where single chemical shifts are observed for chemically equivalent environments (see Figures S45-S46).[9,10]

Detailed analysis, interpretation and assignment of the 1D (^{1}H, ^{13}C) and 2D NMRs (COSY, HSQC, HMBC) are shown in Figures S18-S21. In particular, this indicates the presence of 4 discrete faces (I-IV) in the unsymmetrical cage.

A 2D-NOESY spectrum was also measured in an attempt to confirm connectivity and the existence of a single species (Figure S22). However, it was difficult to interpret specific interactions due to the close proximity of some of the peaks. Therefore, a series of 1D-NOESY experiments were carried out across all peaks in the aromatic region of the ^{1}H NMR spectrum (>7.26 ppm). This enabled the assignment of the 12 imine protons present in the spectra to either the para- or meta-substituted segments of the four tri-imine faces identified, with those on the meta-substituted end appearing to adopt an ‘up’/’down’ configuration on each face as seen previously in other cage systems (see Figure S46), as shown in Figures S23-S27.

For comparison of the configurations of the meta-substituted segments to those in the final four computationally modelled candidates see Fig. S48.

Finally, by disregarding any ^{1}H-^{1}H interactions seen in the COSY spectrum (Figure S19) from the 1D-NOEs, there is evidence of connectivity between the 4 faces, which provides further support for the presence of a single unsymmetrical species (Figures S28).”

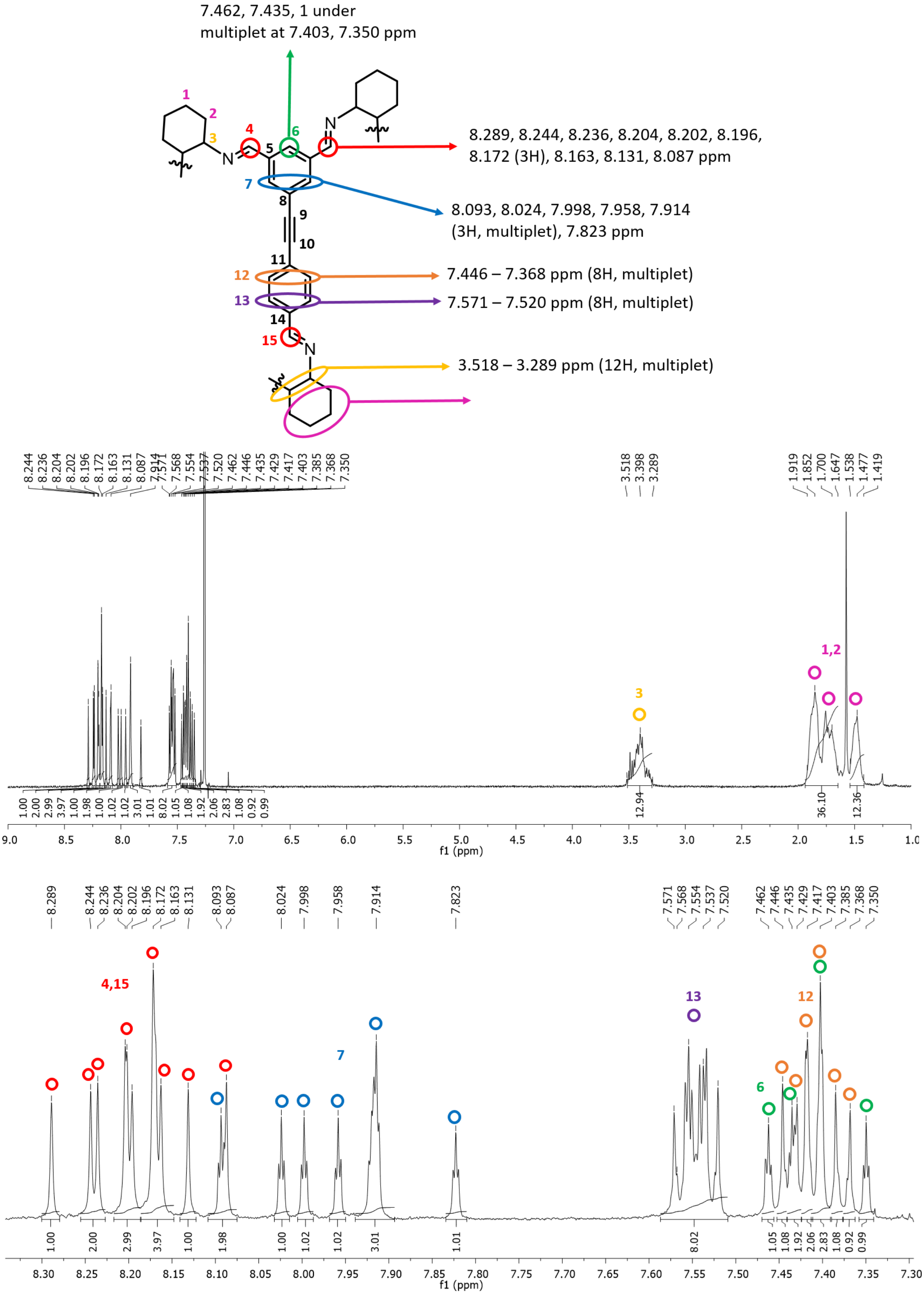


Figure S18: Interpretation of ^{1}H NMR spectra – assigned using COSY and HSQC interactions.

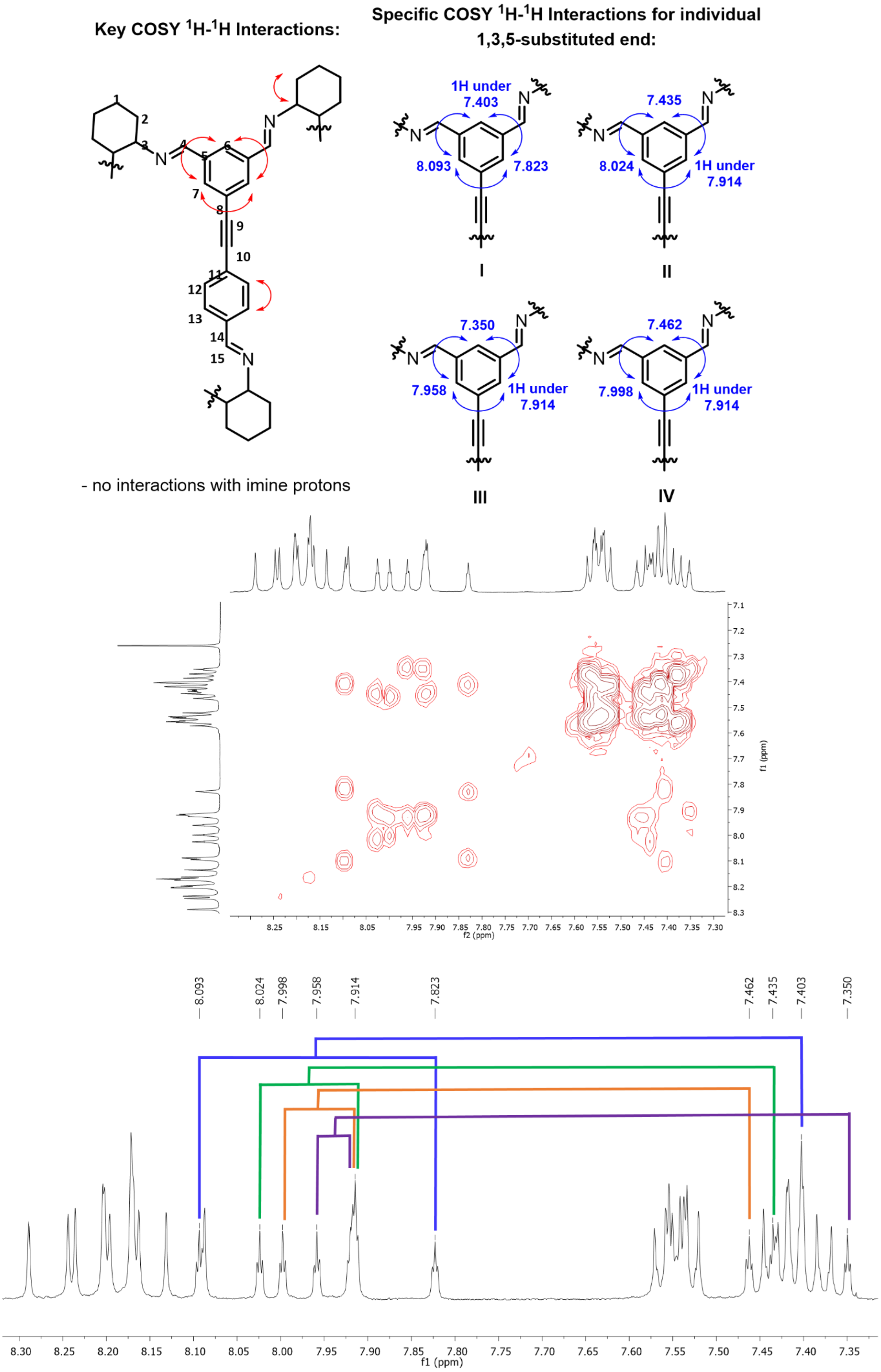


Figure S19: Interpretation of COSY spectra, with key and specific interactions shown, and the identification of 4 individual faces.

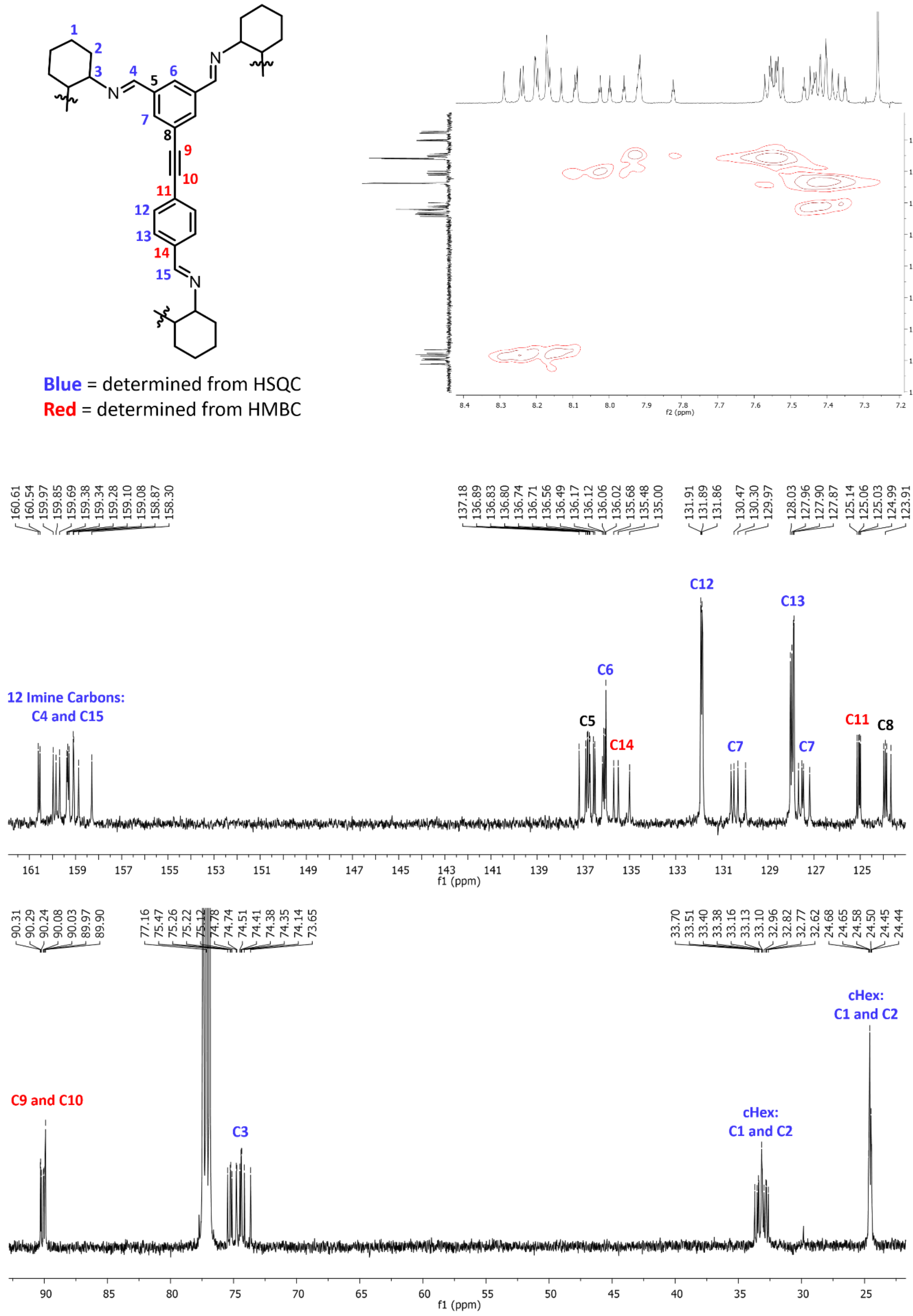


Figure S20: Interpretation of ^{13}C NMR spectra – assigned using HSQC (shown top right) and HMBC interactions (see Figure S21).

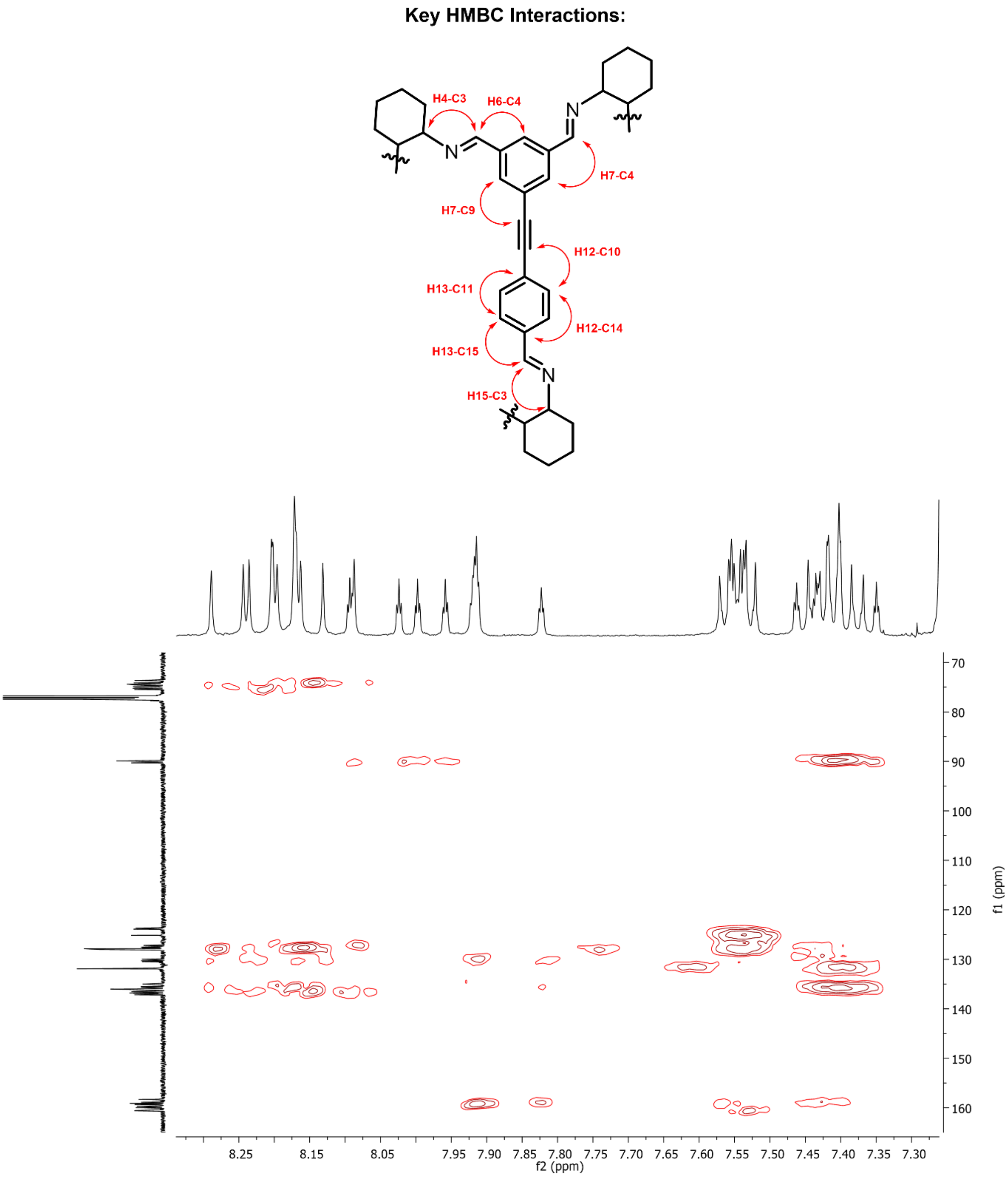


Figure S21: Key HMBC Interactions used to assign ^{13}C NMR spectra

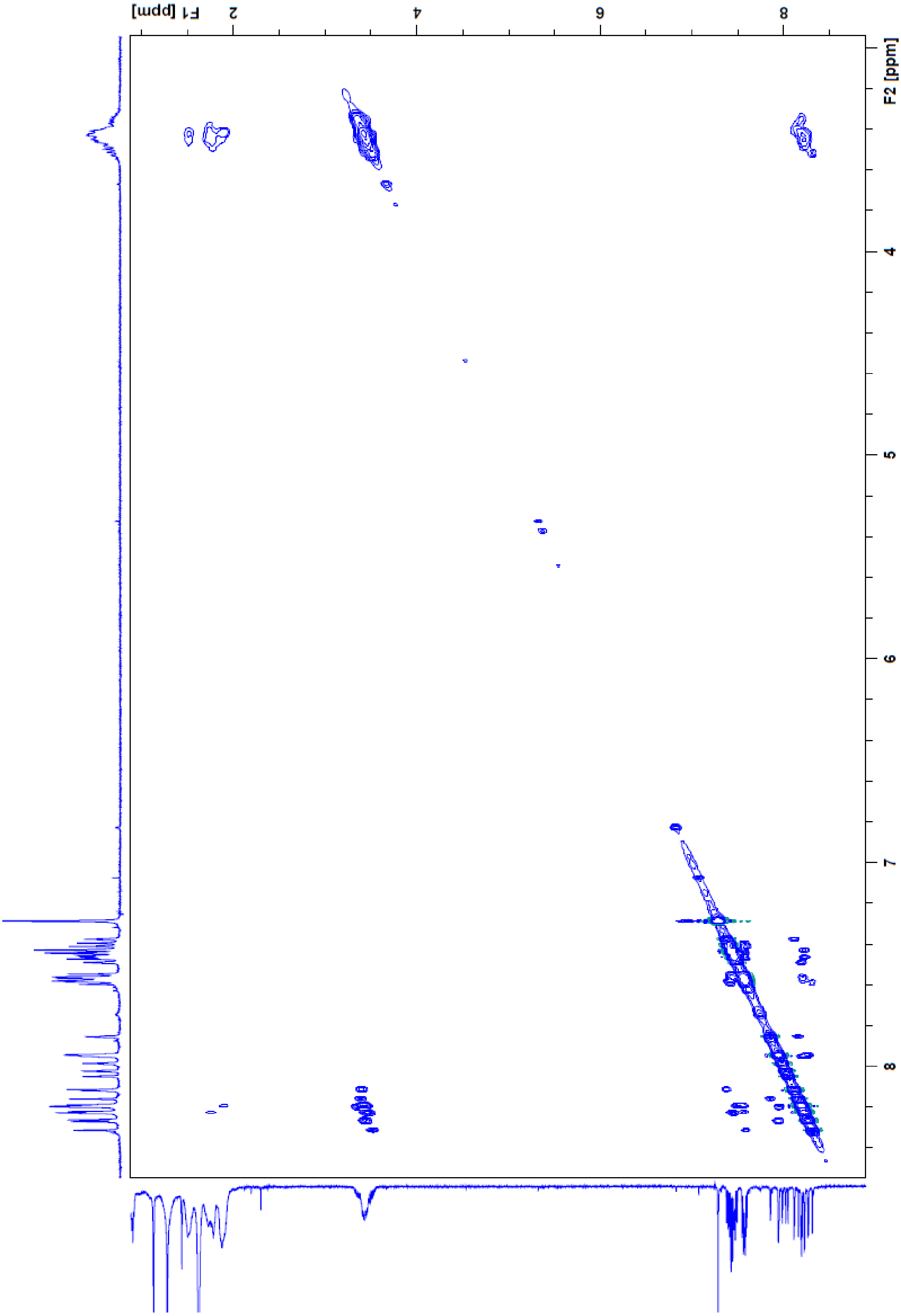


Figure S22: 2D-NOESY ($CDCl_3$) for unsymmetrical cage after purification by prep-HPLC

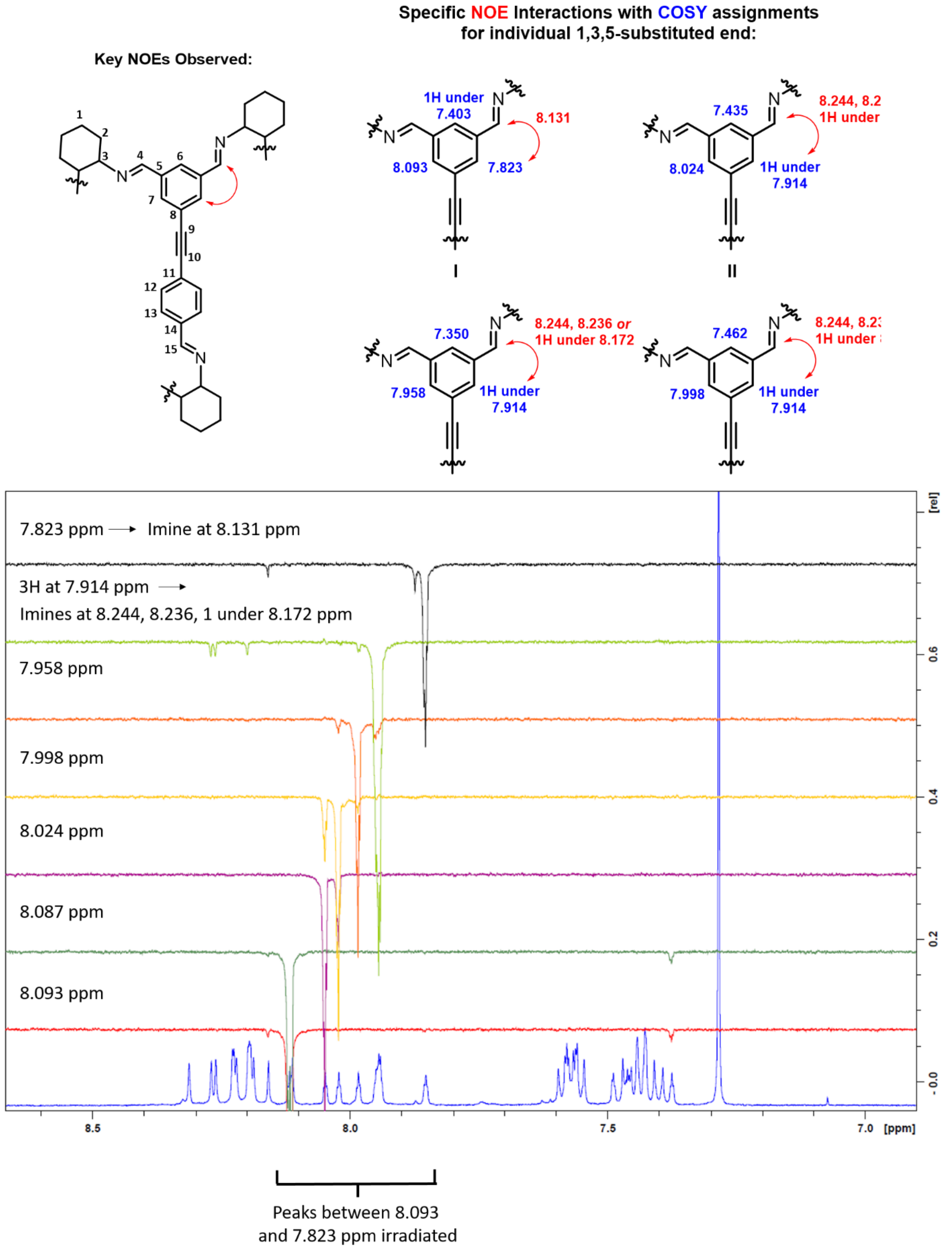


Figure S23: Summary of the 1D-NOESYs carried out on peaks in the region 8.093 ppm to 7.823 ppm, allowing determination of the four imine environments seen by the aromatic protons 7 as shown. The shifts on the left of the stacked 1D-NOE spectra show which peaks were selectively irradiated, with the ^{1}H NMR shown at the bottom for comparison.

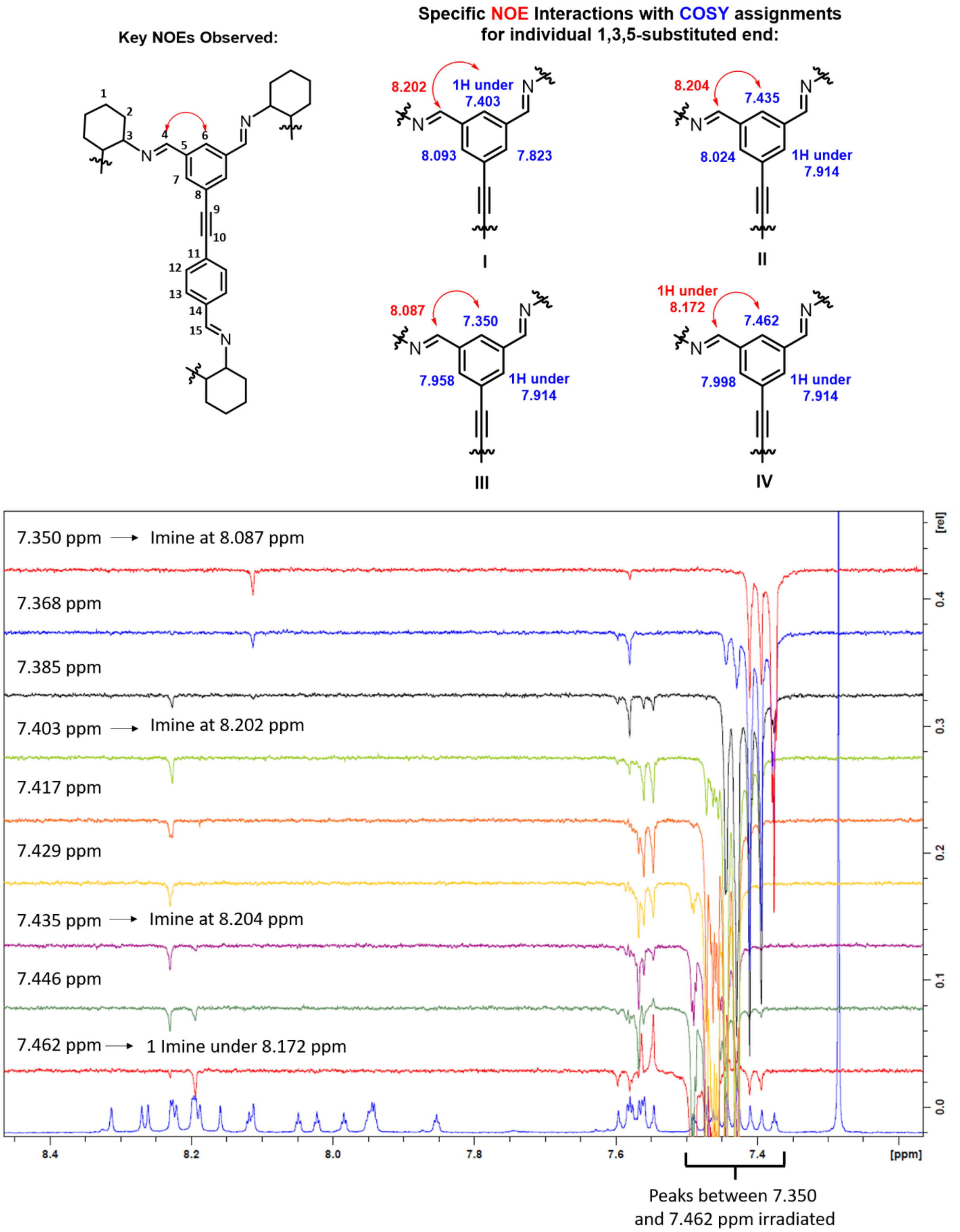


Figure S24: Summary of the 1D-NOESYs carried out on peaks in the region 7.462 ppm to 7.350 ppm, allowing determination of the four imine environments seen by the aromatic protons 6 as shown. The shifts on the left of the stacked 1D-NOE spectra show which peaks were selectively irradiated, with the ^{1}H NMR shown at the bottom for comparison.

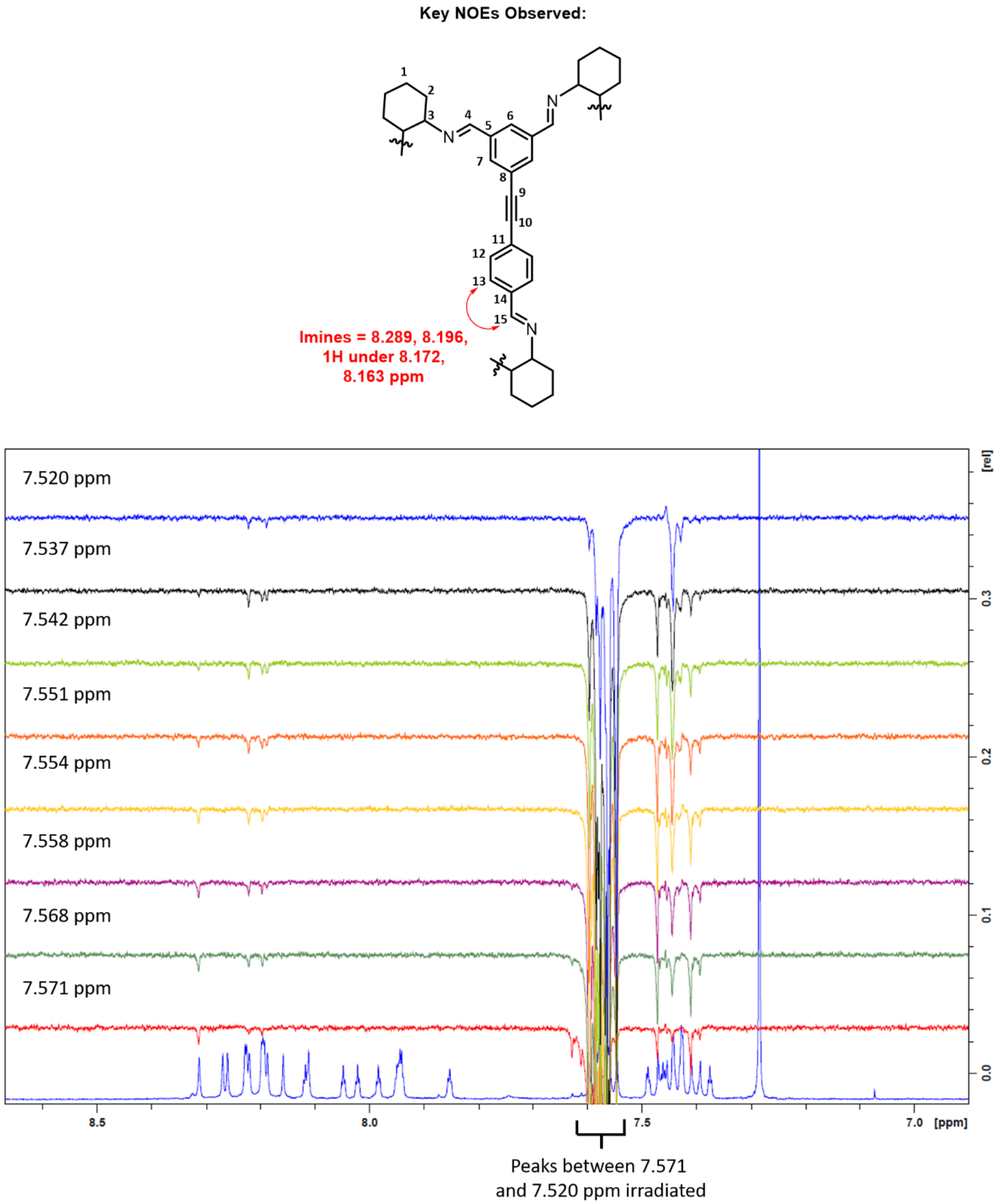


Figure S25: Summary of the 1D-NOESYs carried out on peaks in the region 7.571 ppm to 7.520 ppm, allowing determination of the four imine environments corresponding to the para-substituted end of the tri-imine (aromatic protons 13) as shown. The shifts on the left of the stacked 1D-NOE spectra show which peaks were selectively irradiated, with the ^{1}H NMR shown at the bottom for comparison.

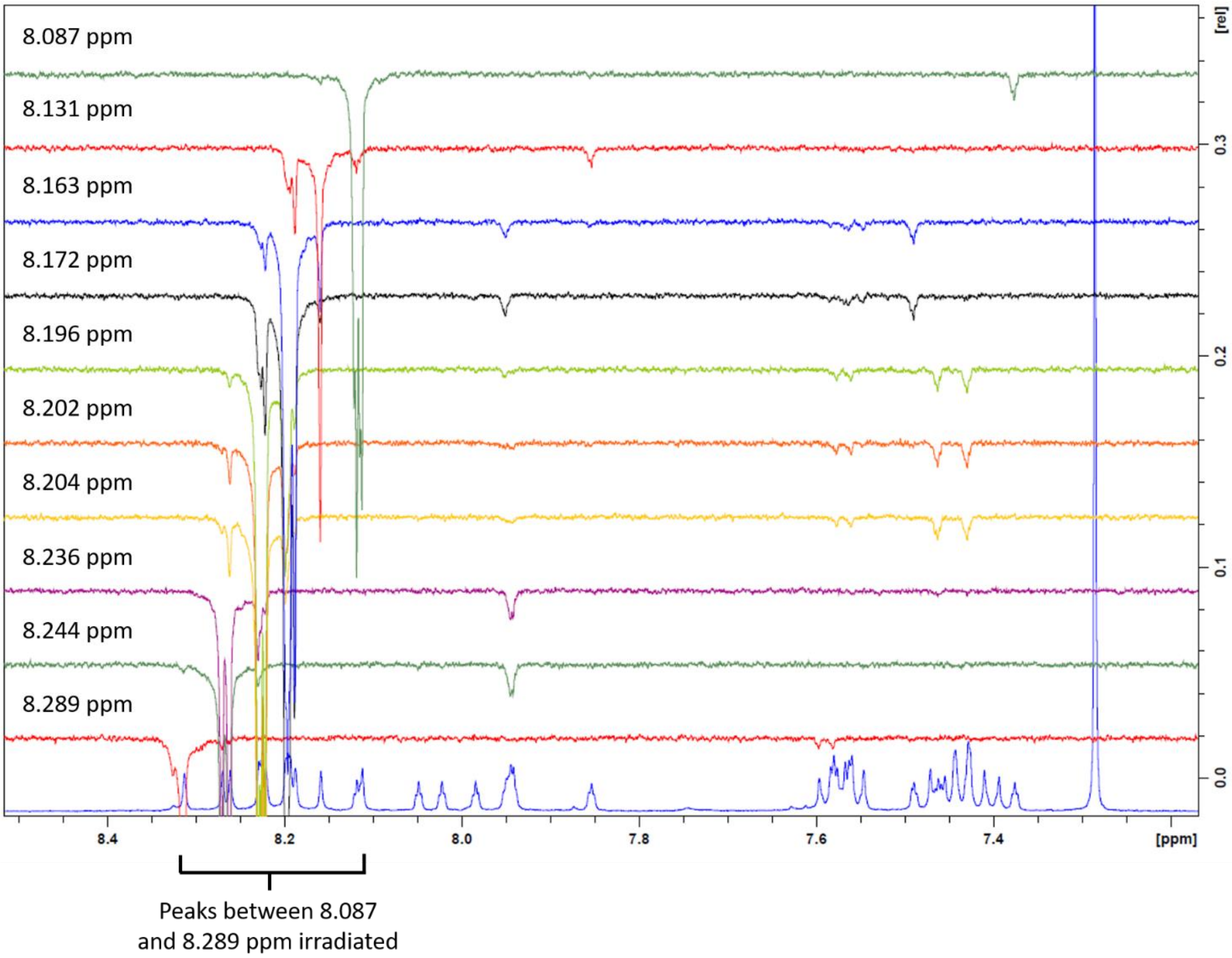


Figure S26: Summary of the 1D-NOESYs carried out on imine peaks in the region of 8.289 ppm to 8.087 ppm, showing the interactions back to the aromatic species initially irradiated in Figure S23-S25.8.087

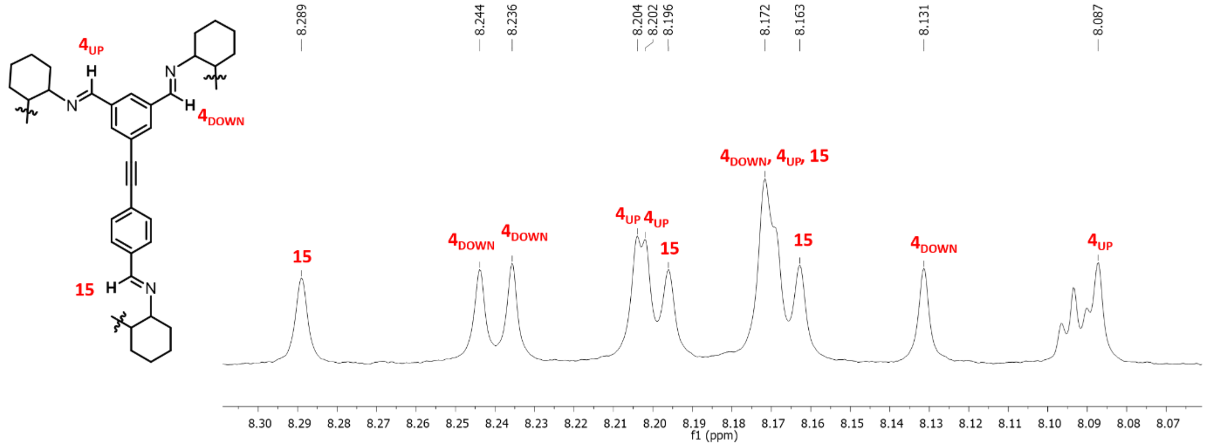


Figure S27: Overall imine assignments from 1D-NOEs as shown in Figures S23-S26.

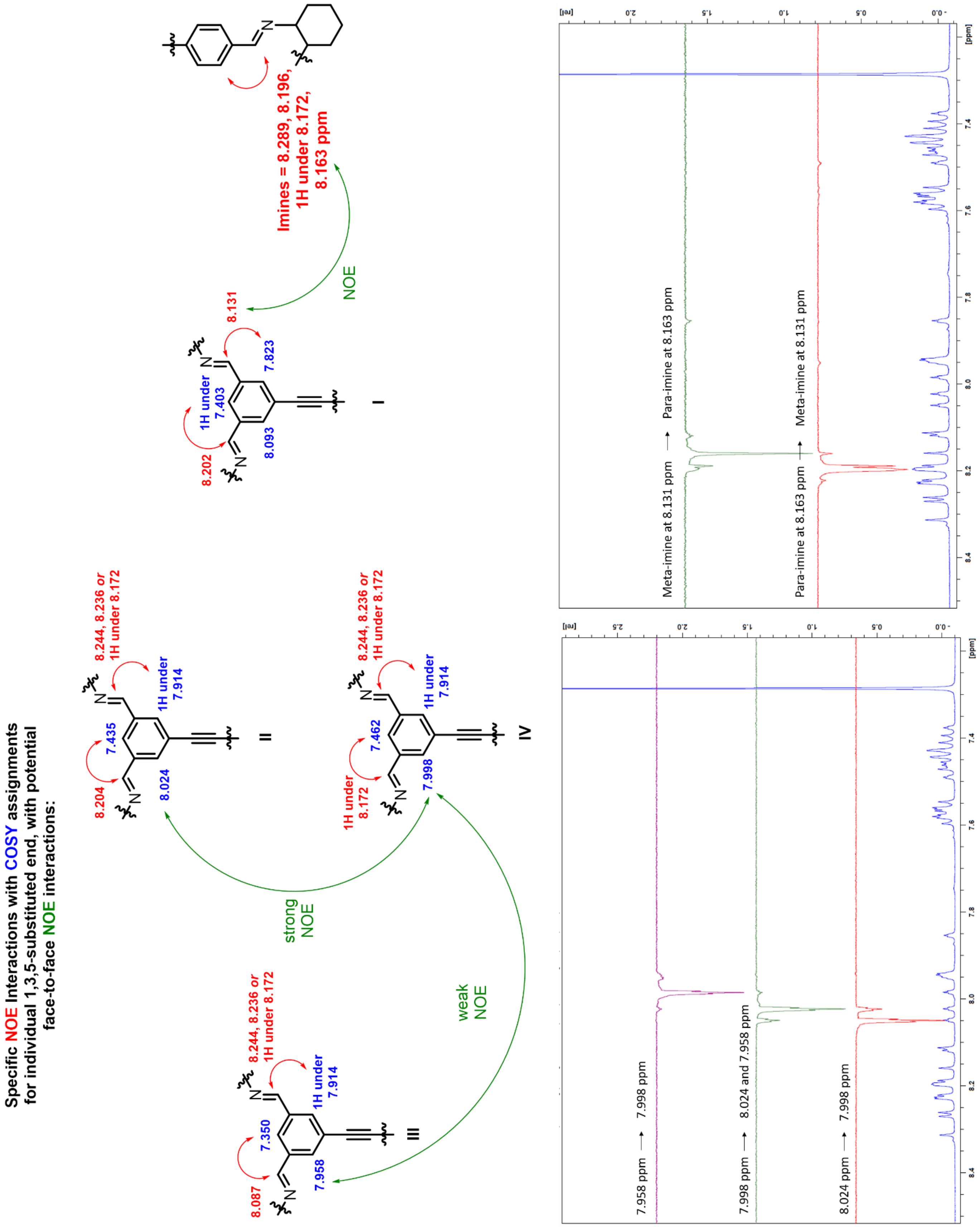


Figure S28: Key NOEs showing through space face-to-face interactions providing support for the formation of a single unsymmetrical cage species.

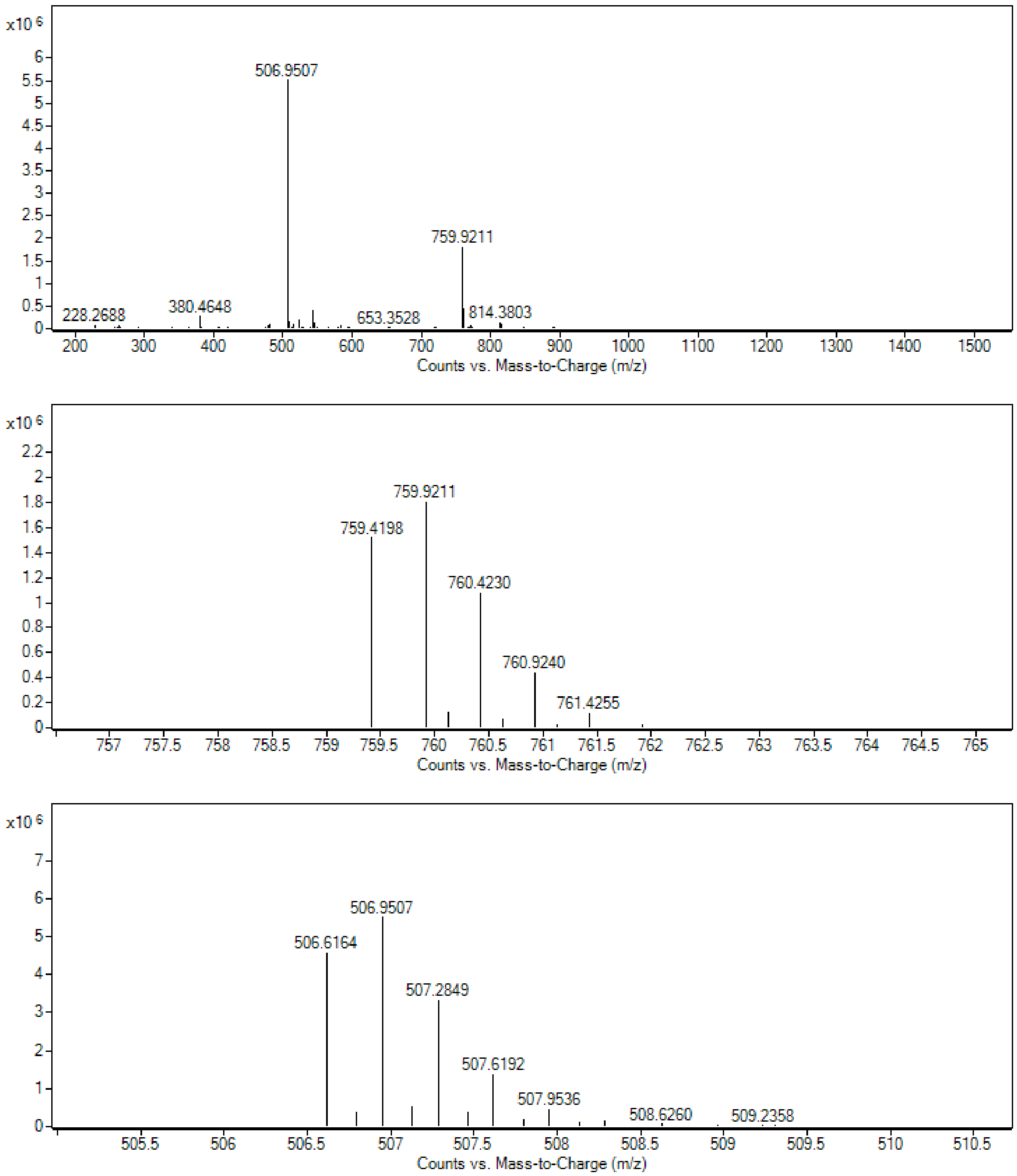


Figure S29: HRMS of unsymmetrical cage after purification by prep-HPLC. From top to bottom — full mass spectra showing purity of cage; mass ion corresponding to doubly charged unsymmetrical [4+6] cage; mass ion corresponding to triply charged unsymmetrical [4+6] cage.

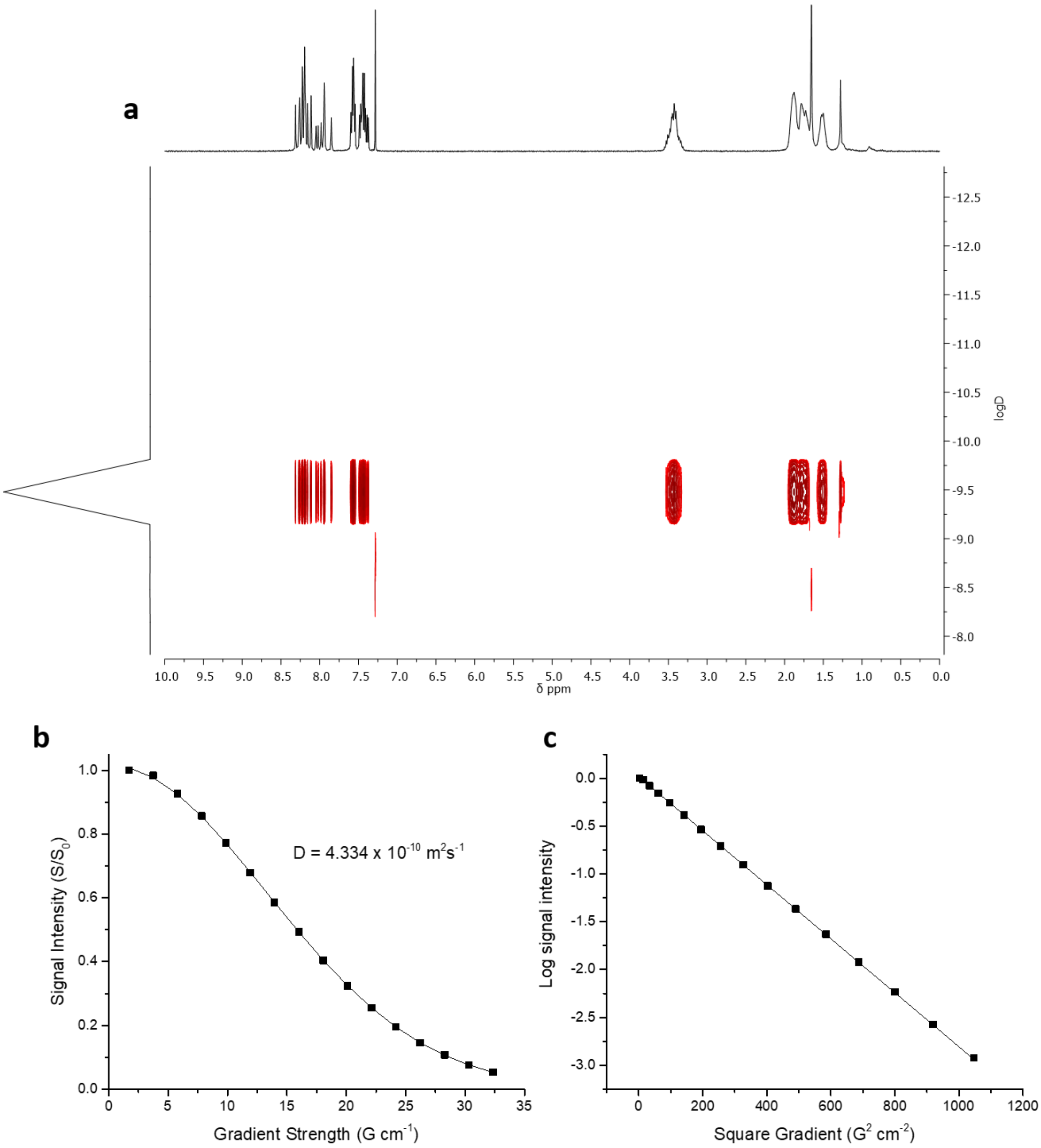


Figure S30: Diffusion NMR of unsymmetrical cage: (a) ^{1}H DOSY spectra (with the least attenuated spectrum [top]); (b) Attenuation of aromatic ^{1}H NMR signal (8.340–8.084 ppm) with increasing gradient strength; (c) Straight-line Stejskal-Tanner PFG-NMR response curve. A linear fit was used to calculate the diffusion co-efficients, with all peaks having the same diffusion co-efficients (see Table S4), suggesting the formation of a single size species.

Table S4: Summary of measured diffusion co-efficients for the unsymmetrical cage.

	Peak Range (ppm)	**Diffusion Coefficient (D, 10^{-10} $m^2.s^{-1}$)**
1	8.340 to 8.084	4.334
2	8.068 to 7.917	4.344
3	7.869 to 7.829	4.262
4	7.617 to 7.528	4.317
5	7.501 to 7.362	4.312
6	3.568 to 3.296	4.305
7	1.595 to 1.402	4.375
Average		**4.321**
Standard Deviation		**0.033**

Table S5: Calculated solvodynamic radii (R_S) for the unsymmetrical cage, using the Stokes-Einstein equation with the measured diffusion co-efficients and the viscosity of chloroform (0.542 cP at 25 °C).

	Peak Range (ppm)	**Solvodynamic Radius (R_S, nm)**
1	8.340 to 8.084	0.929
2	8.068 to 7.917	0.927
3	7.869 to 7.829	0.944
4	7.617 to 7.528	0.932
5	7.501 to 7.362	0.934
6	3.568 to 3.296	0.935
7	1.595 to 1.402	0.920
Average		0.932
Standard Deviation		0.007

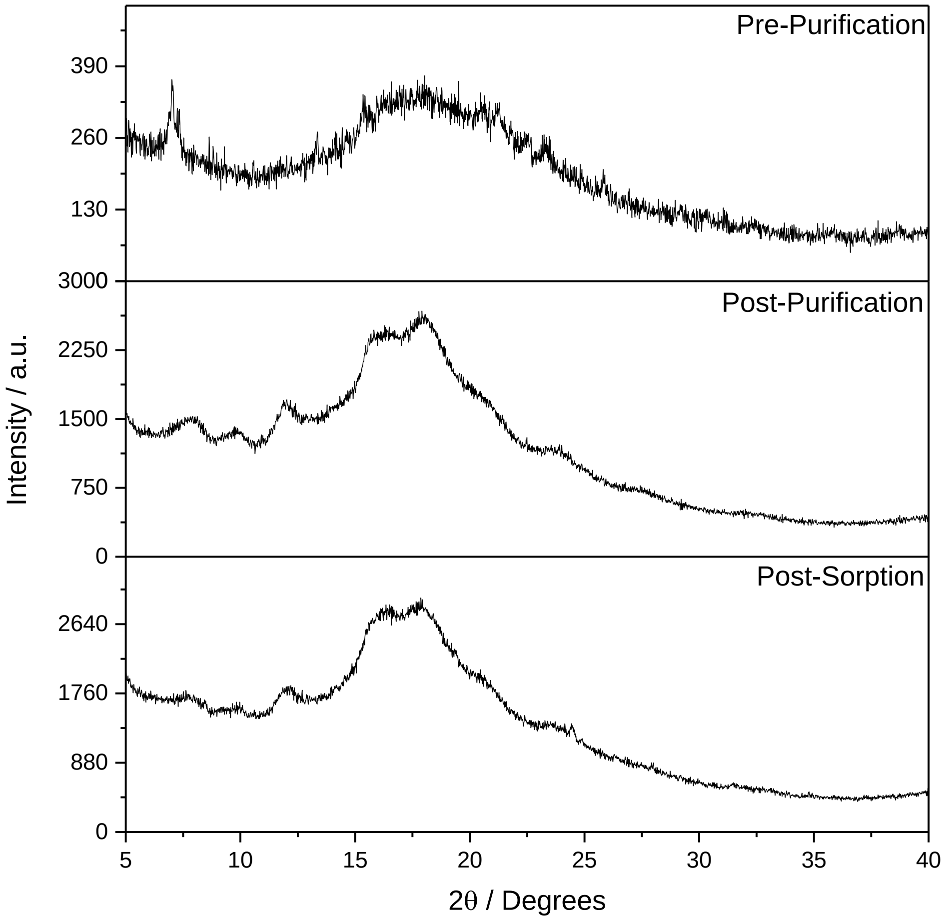


Figure S31: Stacked PXRD analysis of unsymmetrical cage pre-purification (top), post-purification (middle) by prep-HPLC (pre-sorption), and post-sorption (bottom), showing amorphous nature of cage.

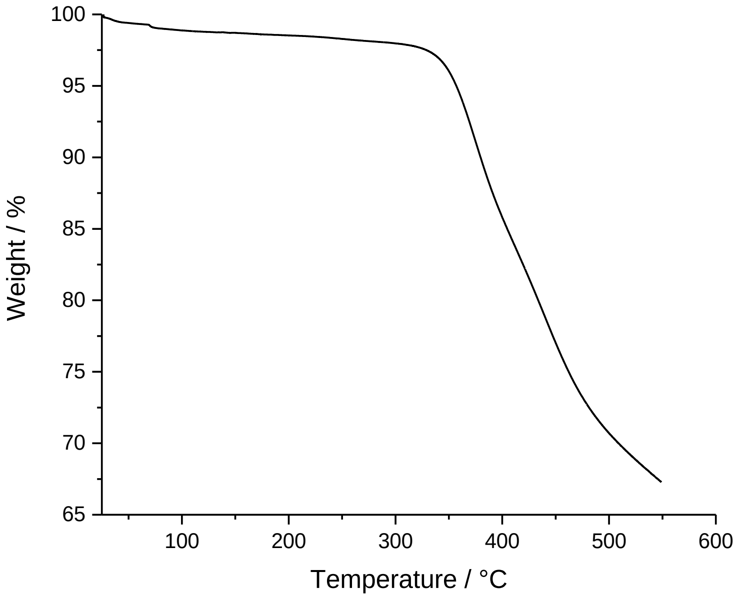


Figure S32: TGA analysis of the unsymmetrical cage after purification by prep-HPLC — ramp rate of 10 °C/min under Ar (35 mL/min) to 550 °C.

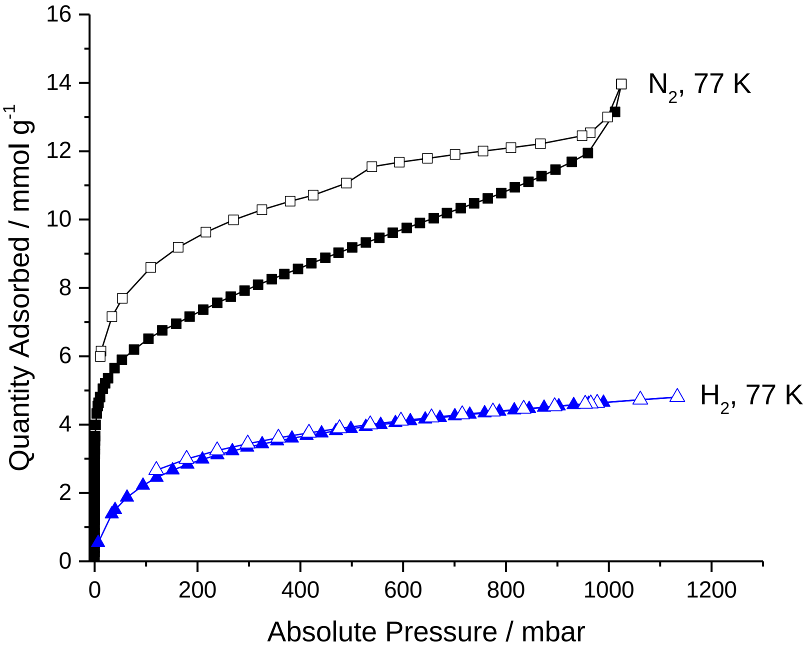


Figure S33: Gas sorption isotherms for N_2 (black, squares) and H_2 (blue, triangles) at 77 K for the unsymmetrical cage after purification by prep-HPLC. Filled symbols represent adsorption, and open symbols represent desorption.

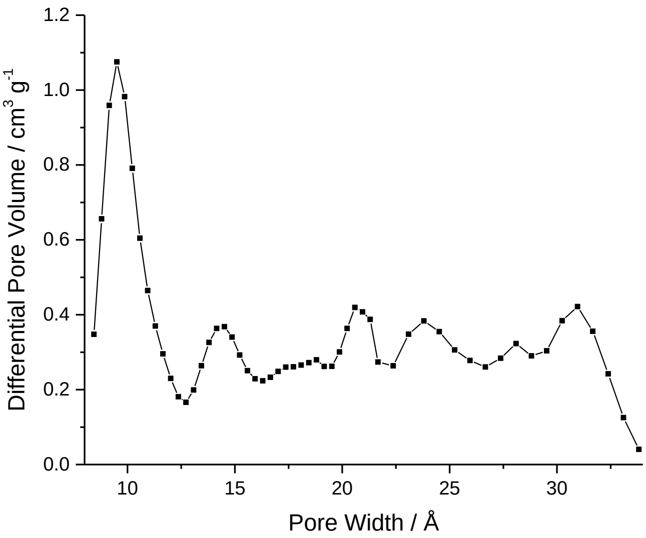


Figure S34: Pore-size distribution of unsymmetrical cage after purification by prep-HPLC calculated based on the experimental N_2 sorption isotherm at 77 K using a DFT model in the Micromeritics software.

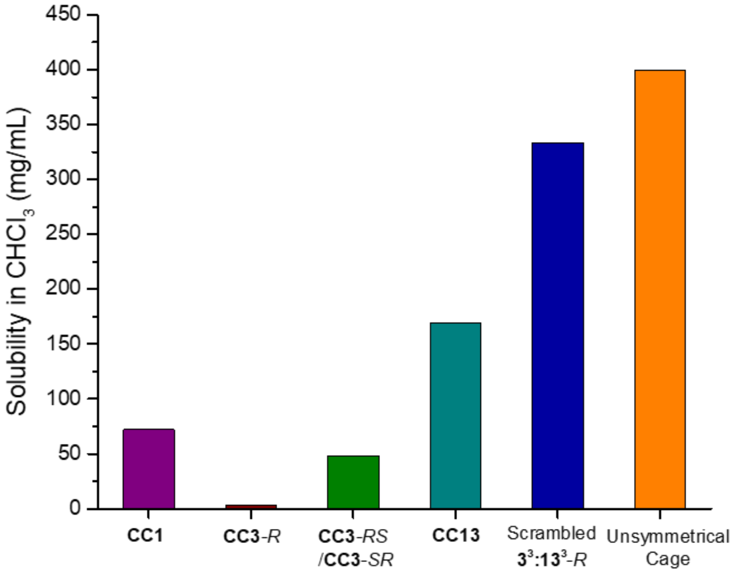


Figure S35: Solubility of the formed unsymmetrical cage in chloroform, compared to a range of other symmetric (**CC1**, **CC3**-*R*, **CC13**), dissymmetric (**CC3**-*RS*/**CC3**-*SR*), and scrambled cages (**3^3:13^3**-*R*) – the solubility is >400 mg/mL, exceeding the solubility of symmetrical cages reported previously by over two orders of magnitude, as well as the solubility of scrambled cages which were engineered as porous liquids with high solubility.[7,11]

Summary of crystallisation conditions attempted:

Prior to purification, the crude cage (2 x 15 mg, ~90% purity by HPLC, a/a) was dissolved in both DCM (15 mL) and chloroform (15 mL), and the solutions (1 mg/mL) split into 30 small vials which were capped with lids containing holes. For each of the solvents, vial-in-vial slow diffusion crystallisation studies were carried out with a range of anti-solvents: methanol, *n*-hexane, tetrahydrofuran, acetone, ethyl acetate, acetonitrile, isopropanol, *m*-xylene, *o*-xylene, *p*-xylene, ethanol, petroleum ether (40-60), toluene and pentane.

After purification by prep-HPLC, the cage (20 mg) was dissolved in chloroform (20 mL), and the solution (1 mg/mL) split into 20 small vials which were capped with lids containing holes. One sample was left to slowly evaporate, three had either trifluoroethanol, NMP or mesitylene layered onto them, and the remainder were used in vial-in-vial slow diffusion crystallisation studies with a range of anti-solvents: acetone, acetonitrile, diethyl ether, ethanol, ethyl acetate, methanol, *n*-hexane, pentane, petroleum ether (40-60), isopropanol, tetrahydrofuran, toluene, *p*-xylene, *o*-xylene, *m*-xylene and trifluoroethanol.

5. Computational modelling of the unsymmetrical cage

5.1. Cage topology nomenclature

We employ the nomenclature for cage topologies defined by Santolini *et al.*[12] This nomenclature was introduced in order to avoid confusion when labelling the specific topology, as the use of polyhedra (e.g. tetrahedron, cube, dodecahedron) can lead to ambiguous cases. For example, various cages with equivalent topology can present different shapes and therefore they can be associated to different polyhedra.

In this work each structure is labelled as:

$$\mathbf{X_p^m Y^n}$$

Where **X** and **Y** define the topicity (number of functional groups) of the two different molecular precursors that constitute the cage. **X** and **Y** are **Tri** and **Di** when the cage is built with a tritopic (three reactive functionalities) and ditopic (two reactive functionalities) building block, respectively. By convention, **X** represents the building block with the highest number of reactive end groups (unless **X** = **Y**) and when building a cage, if the topology relates to a polyhedron, then **X** is placed on the vertices. **Y**, the second building block, can have a number of functionalities, which is less or equal to **X**. If **X** = **Y**, then the allocation of the labels is arbitrary. The superscripts m and n define the numbers of precursors embedded into the topology for **X** and **Y**, respectively. The subscript **p** describes the number of connections between precursor pairs within a topology. Generally, **X**-type building blocks are linked to other **X**-type precursors through only one **Y**-type molecule; if this is the case, then no subscript is given, thus **p** is omitted. However, if two **X**-type building blocks are connected through links with two distinct **Y**-type precursors, then **p** = 2.

5.2. Cage assembly and screening

The 3D coordinates of the trialdehyde and diamine molecules where generated through the use of the ETKDG[13] approach as implemented in RDKit.[14] To build a molecular cage, the trialdehydes are positioned on the vertices of each topology and the diamines are positioned on the edges. The aldehyde C atom is then connected to the nearest amine N atom, and in order to simulate the imine condensation reaction, a C=N bond is created between those two atoms, while the 2 amine H and the aldehyde O are deleted. The newly created molecular cage contains 12 imine bonds.

During the screening of the Reaxys database, cages were assembled into the $\mathbf{Tri^4Di^6}$ topology (Figure 1b) and low energy conformers were searched for, as described later in Section 5.4, this involved sampling molecular dynamics simulations for low energy structures. The precursors selected from Reaxys consisted of diamines, triamines, dialdehydes and trialdehydes that can undergo a [4+6] imine condensation reaction to form an imine cage. We removed all precursors that were charged systems, radical systems, contained heavy elements and any duplicates prior to assembly into cage molecules.

From the resulting lowest energy conformer for each precursor pair (assembled without consideration of low symmetry of building blocks, thus in an arbitrary connectivity), we used pyWINDOW (see section 5.5) to calculate the largest sphere that could fit in the cage's cavity and the size of the four windows. If the molecule contained a cavity that was large enough to accommodate a nitrogen molecule (kinetic diameter 3.84 Å) and had four windows (spherical diameter greater than 2.8 Å, the kinetic diameter of H_2), then it was deemed to be "shape persistent".

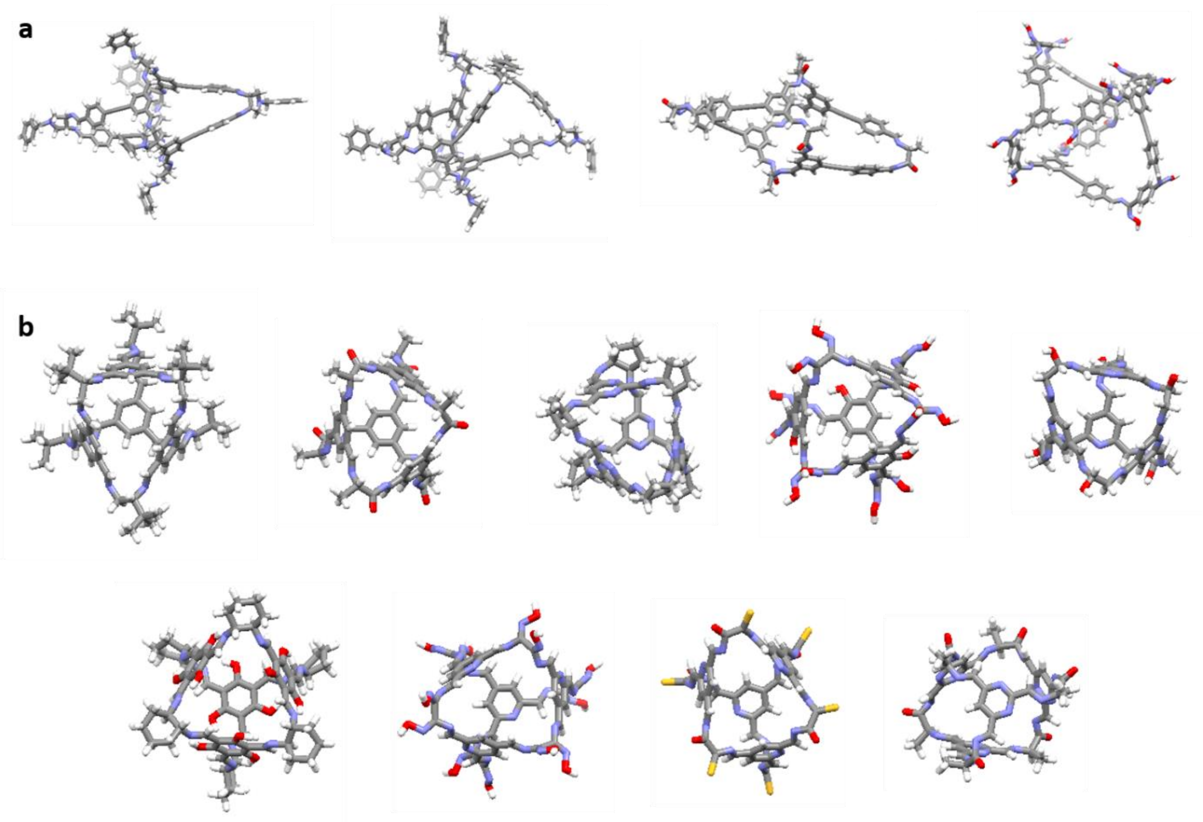


Figure S36: Examples of some of the computationally discovered shape persistent porous organic molecules from the screening of the Reaxys database. (a) Examples of shape-persistent cages involving the C_{2v} trialdehyde, showing the potential for the formation of an unsymmetrical cage. (b) Examples of other shape-persistent $\mathbf{Tri^4Di^6}$ cages from the screen, the majority of which resemble previously synthesised cages.

5.3. Cage identification

For the generation of the unsymmetrical cage, we considered 2 possible topologies that include 4 trialdehydes and 6 diamines in total, leading to the formation of 12 imine bonds in the final molecular cage. The first topology corresponds to that typically expected for a [4+6] molecule, a $\mathbf{Tri^4Di^6}$ topology, as defined in our recent work,[12] whereas the second is the $\mathbf{Tri^4_2Di^6}$ topology.[12] Both topologies are shown in Figure 3 in the main text.

Depending on how the molecular precursors are oriented on the topology, a specific structural isomer of the cage is generated. By rotating the precursors across their highest rotation axis centred at the centre of geometry of the molecule's functional groups (*e.g.*, C_3 for the trialdehydes and C_2 for the diamines), different structural isomers can be generated. However, if the trialdehyde and the diamine chosen are characterised by C_{3h} and C_{2v} point group symmetry respectively, all the structural isomers generated via rotation of the precursors during the assembly step reduce to a single structural isomer.

The selected trialdehyde **1** has a point group symmetry of C_{2v}, therefore the rotation of each trialdehyde will generate a different structural isomer. During the assembly, each functional group of the trialdehyde can connect to any of the 3 adjacent amine groups from the neighbouring diamines, located on the edges of the topology, through a rotation of 120º around the axis passing through the centre of geometry of its 3 functional groups. Each of the 4 trialdehydes has 3 possible orientations, this leads to a total number of $3^4 = 81$ structural isomers per topology. Two structural isomers obtained for the same topology differ at a minimum by a 120° rotation of at least one trialdehyde.

However, these 81 isomers are not all unique. A rotation of 120° of a single trialdehyde, if followed by a complementary rotation in a second trialdehyde, could lead to the restoration of a non-unique connectivity.

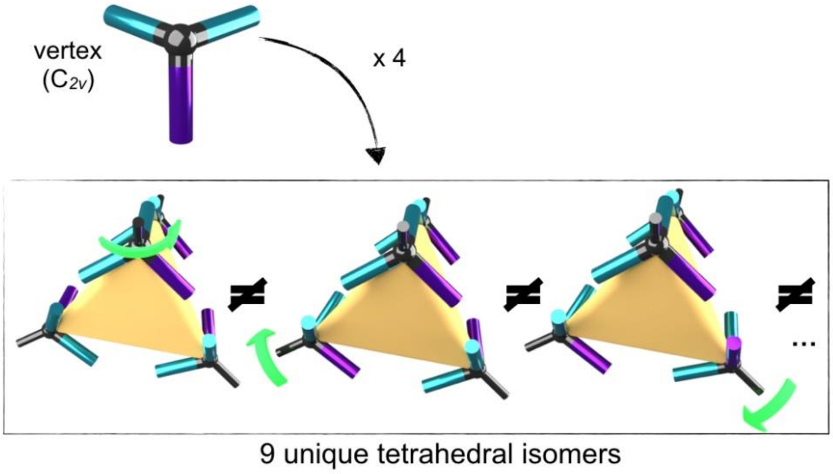


Figure S37: Effect of using a lower symmetry precursor or vertex element (C_{2v} symmetry) for the assembly of a tetrahedral structure (including 4 vertex elements). When a lower symmetry vertex element is employed, a different structural isomer will be generated during the assembly by the rotation of each of the C_3 symmetry axis of the tetrahedron. In the case of a C_{2v} vertex element, the number of unique structural isomers amounts to 9 for a tetrahedral topology.

In order to determine if a structural isomer was unique, we flagged all the functional groups in the molecular precursors and kept track of their orientations on an ideal topology. We then compared all the structural isomers and checked if, by applying the topology's symmetry operations, any 2 of them would become equivalent. We then deleted one of the non-unique isomers.

After this process, we obtained a total number of 36 unique structural isomers for the two topologies (9 unique $\mathbf{Tri^{4}Di^{6}}$ cages and 27 unique $\mathbf{Tri}_{2}^{4}\mathbf{Di}^{6}$ cages). Cages labelled 0 – 8 were generated from the use of the $\mathbf{Tri^{4}Di^{6}}$ topology, whereas all the $\mathbf{Tri}_{2}^{4}\mathbf{Di}^{6}$ cages were labelled with numbers 9 – 35.

In order to evaluate the level of symmetry observed in the final assembled isomers, we calculated the number of distinct orientations obtained from each isomer by applying all rotational symmetry operations of the base topology. We defined this number as the number of global unique environments present in each cage isomer (Table S6), as the number of distinct orientations generated by simple rotations strictly depends on the overall symmetry of the molecule. Isomers that only show a single unique orientation are more symmetric, as all the newly generated imine bonds (12 in total) are equivalent. Isomers with 12 unique environments are completely unsymmetrical, where each imine bond is symmetrically different from the other 11. The isomers with lowest symmetry (*i.e.* 12 unique imine environments) are numbers 0, 2, 4, 6, and 8. All of these isomers are defined by the base topology of $\mathbf{Tri^{4}Di^{6}}$, whereas the isomers generated from the $\mathbf{Tri}_{2}^{4}\mathbf{Di}^{6}$ generally tended to be higher in symmetry.

Table S6: Symmetry analysis for the 36 assembled cages.

Original isomer identifier	Number of unique imine environments
0	12
1	6
2	12
3	6
4	12
5	3
6	12
7	6
8	12
9	1
10	4
11	4
12	4
13	2
14	2
15	4
16	2
17	2
18	4
19	4
20	4
21	4
22	4
23	2
24	4
25	4
26	2
27	4
28	4
29	2
30	4
31	4
32	2
33	2
34	1
35	1

5.4. Cage optimization

The 36 assembled molecular cages were then optimised through a 3-step process. In the first step, the atoms involved in the imine bonds generated during the cage assembly are relaxed through a constrained optimisation (all the other bond distances are kept fixed) by using the OPLS3 force field.[15] In the second step, in order to ensure that the lowest energy conformer of each cage is selected, we use a simulated annealing approach to probe the molecule's conformational space (e.g. multiple conformational isomers are explored for each structural isomer). For this step a molecular dynamics (MD) simulation is performed with the OPLS3 force field at 700 K for 20000 ps, the time step is 0.5 fs and the equilibration time is 500 ps. 1000 random geometries are sampled along the MD trajectory and each geometry is then fully optimised with the OPLS3 force field. The lowest energy conformer among the 1000 sampled for each cage is then selected for the next step. In the final step, the selected cage conformers are re-optimised with GAUSSIAN16[16] at DFT level (B3LYP-D3/def2-TZVP) in the gas phase.

5.5. Cage analysis

The 36 fully optimised cages were then screened for a series of properties, and only the most promising candidates were kept at each step of the analysis. Figure 3 summarises the screening pipeline, which we employed to select the final 4 isomers from the initial 36 candidates. In order to obtain more accurate DFT energies, M06-2X/def2-TZVP single point calculations were performed in the presence of a polarizable continuum model on all the 36 B3LYP-D3/def2-TZVP optimised cages. The implicit solvent employed is CH_2Cl_2, where all the parameters were the ones selected by default within GAUSSIAN 16.[16]

Structural properties such as pore size and weighted average diameter and the largest sphere that can pass through a window (ie the window diameter) were then calculated for each cage with the pywindow[17] code. The pore size, which gives an approximate value of the porosity of the cage, is calculated as two times the distance between the centre of mass of the cage and its closest atom. Whenever the distance between the centre of mass of the cage and its closest atom is smaller than the atom's van der Waals radius, the value that pywindow outputs is negative. For all those occurrences we assumed the cage to be collapsed and defined its pore size equal to 0.0 Å. The weighted average diameter strongly correlates with the solvodynamic radius measured from diffusion NMR and is calculated by considering the weighted average distribution of atoms around the centre of mass of the cage. Isomers 13 and 14, which after the assembly corresponded to different stereoisomers, resulted in the same geometry after the high temperature MD runs and DFT optimization. The window diameters are calculated along the vector path passing through the window centre and connecting the centre of mass of a cage and the cage's exterior. Spheres with various dimensions are placed along this vector and the overlap with the neighbouring atoms van der

Waals spheres is calculated. The largest sphere that can pass through a window, that is also the smallest sphere along the sampled vector path (*ie*. the window is treated as the narrowest point of a channel connecting cage's internal void and the external environment) is the calculated window diameter.

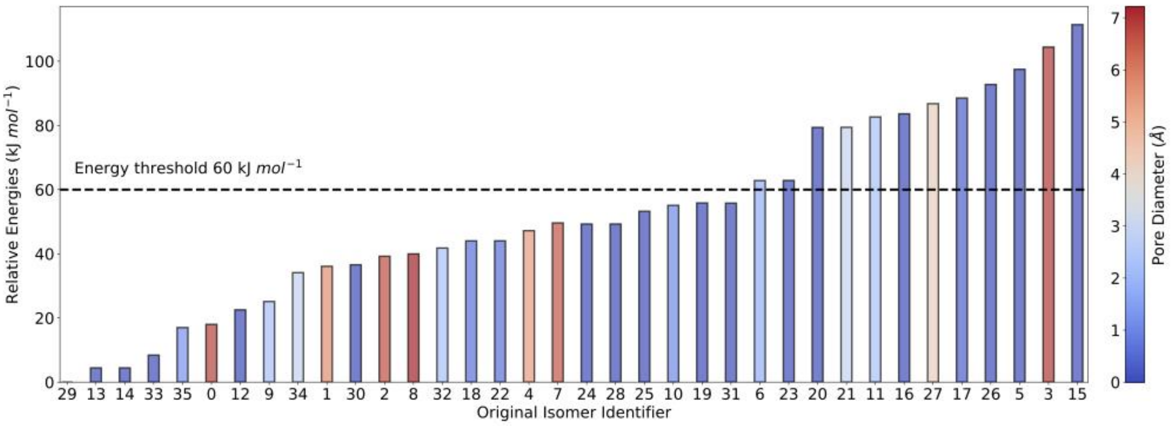


Figure S38: The isomer identifier of each cage is plotted against the M06-2X/def2-TZVP relative energy (kJ mol^{-1}) and the colour of each bar corresponds to the pore size of the cage (Å). The dashed line defines the energy threshold (60 kJ mol^{-1}) that we employed for the selection of the best hypothetical isomers. All the relevant values can be found in Table S7.

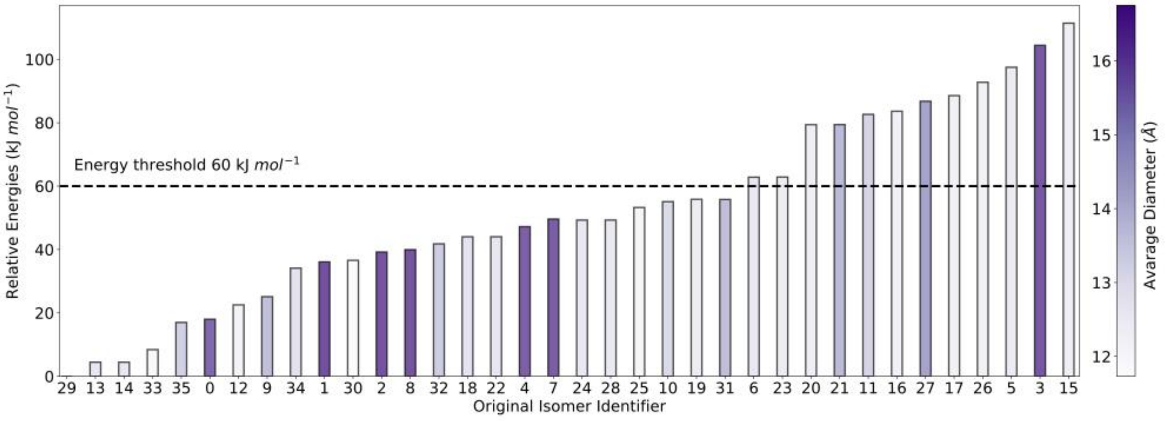


Figure S39: The isomer identifier of each cage is plotted against the M06-2X/def2-TZVP relative energy (kJ mol^{-1}) and the colour of each bar corresponds to the average weighted diameter (Å). The dashed line defines the energy threshold (60 kJ mol^{-1}) that we employed for the selection of the best hypothetical isomers. All the relevant values can be found on Table S7.

Table S7: For each isomer, the pore diameter, the average weighted diameters, the relative and absolute DFT energies are given. The isomers highlighted in orange are the structures selected as the 4 most promising candidates (relative DFT energies are compared with the lowest energy conformer, number 29).

Original isomer identifier	Pore diameter (Å)	Avg. weighted diameter (Å)	Relative DFT energy (kJ mol^{-1})	Absolute DFT energy a.u. (Hartree)	MAE ^{1}H NMR (ppm)
0	6.8	16.1	18	-4679.851288	0.085
1	5.7	16.7	36	-4679.844331	0.089
2	6.7	16.7	39	-4679.843129	0.113
3	6.9	16.5	104	-4679.818053	0.116
4	5.4	16.3	47	-4679.840059	0.116
5	0.0	12.7	98	-4679.820717	0.198
6	1.8	12.8	63	-4679.834053	0.169
7	6.5	16.4	50	-4679.839130	0.072
8	7.2	16.6	40	-4679.842845	0.104
9	2.2	14.0	25	-4679.848552	0.302
10	1.1	13.3	55	-4679.837024	0.19
11	2.3	13.4	83	-4679.826432	0.291
12	0.0	12.3	23	-4679.849545	0.27
13	0.0	12.8	4	-4679.856505	0.238
14	0.0	12.8	4	-4679.856505	0.238
15	0.0	12.5	111	-4679.815365	0.234
16	0.0	12.4	84	-4679.826054	0.216
17	0.3	12.4	89	-4679.824159	0.204
18	0.6	12.9	44	-4679.841284	0.218
19	0.0	12.6	56	-4679.836737	0.254
20	0.0	12.5	79	-4679.827680	0.14
21	2.9	14.2	79	-4679.827669	0.168
22	0.6	12.9	44	-4679.841280	0.218
23	0.0	12.6	63	-4679.834038	0.185
24	0.0	12.8	49	-4679.839260	0.186
25	0.0	12.1	53	-4679.837731	0.209
26	0.0	12.4	93	-4679.822538	0.281
27	4.2	14.8	87	-4679.824842	0.165
28	0.0	12.8	49	-4679.839258	0.186
29	4.8	14.4	0	-4679.858215	0.182
30	0.0	11.7	37	-4679.844145	0.173
31	0.0	14.1	56	-4679.836760	0.143
32	2.3	13.7	42	-4679.842141	0.137
33	0.0	12.0	8	-4679.854976	0.227
34	2.9	13.0	34	-4679.845093	0.217
35	1.3	13.6	17	-4679.851676	0.347

From the 36 original isomers, we selected the most promising structures by screening on pore diameter (> 3.64 Å) and DFT relative energies (within 60 kJ mol^{-1} in energy from the lowest energy isomer). A minimum pore diameter of 3.64 Å was chosen as this is the kinetic diameter of N_2, the molecule for which the porosity of the synthesized cage was tested. A 60 kJ mol^{-1} energy threshold was selected as we have previously found alternative reaction outcomes within that energetic range from a global minimum to be experimentally observable,[12,18] for instance due to the influence of solvent directing the reaction and perturbing the potential energy landscape of the isolated molecules. Among the 7 isomers that matched both criteria, we then checked how many unique imine environments were present in each system (Table S6) and only selected the structures that contained 12 unique imine environments (systems with the lowest symmetry possible). This led to the final selection of 4 cage isomers, which correspond to the isomers 0, 2, 8, and 4. The selected isomers where then renumbered depending on their DFT relative energy (as defined in Table S8).

Table S8: New isomers identifiers for the selected candidates, as used in the main paper.

Old isomer identifier	New isomer identifier
0	1
2	2
8	3
4	4

The average weighted diameter (Figure S39) was not used as a criterion for the screening of the final selected isomers since all the calculated values are slightly lower in size than the experimental solvodynamic diameter (18.6 Å). The discrepancy between the predicted values and the experimental solvodynamic diameters could be attributed to multiple factors: (i) the diffusion NMR measurements are performed in the presence of solvent, whereas our cage models are gas-phase optimized; (ii) the Stokes-Einstein equation employed for calculating the solvodynamic radius approximates the cage to a sphere, which has been shown to be a reasonable approximation for symmetrical cages, but arguably is not the most accurate representation for these molecules. Although, as these differences fall within the range obtained for similar porous cages assessed using this method (see Table S11), it was used to confirm that our computational models represent a good guess of the actual molecule.

5.6. NMR chemical shift predictions

NMR chemical shifts were calculated at DFT level for the 4 final selected isomers. For the computation of the ^{1}H shielding tensors we employed the WP04/DGTZVP level of theory, which was recently suggested by Benassi[19] to perform exceptionally well on a set of test molecules. All the calculations were performed with the Gauge-Independent Atomic Orbital (GIAO) approach and solvent effects (chloroform) were taken into account using the integral equation formalism variant (IEF-PCM) of the implicit polarizable continuum model. The IEF-PCM calculation used the radii and non-electrostatic terms of Truhlar's SMD solvation model,[20] for which the standard default values implemented in GAUSSIAN 16[16] were employed.

Due to molecular size limitations (each isomer contains 216 atoms), all the isomers were considered in only a single conformation, rather than with conformational averaging. The calculated shielding tensors were converted to chemical shifts by applying scaling factors (slope and intercept, respectively), which are obtained from linear regression analysis of a test set of molecules to each of the predicted tensor values. The scaling factors we employed were provided by Benassi[19] with a slope of -1.0420 and an intercept of 32.3165 for the WP04/DGTZVP ^{1}H shielding tensors.[21] The DFT predicted chemical shifts were then compared to the experimental chemical shifts and in order to determine the goodness of the fit for each isomer, we calculated the mean absolute error (MAE).[21] The calculated MAE for all the structures are collected in Table S9.

Table S9: MAE calculated for each isomer between the predicted ^{1}H chemical shifts (WP04/DGTZVP) compared to the experimental values.

New isomer identifier	Pore size (Å)	Average Weighted Diameter (Å)	M06-2X/Def2-TZVP relative Energy (kJ mol^{-1})	Wp04/DGTZVP MAE ^{1}H NMR (ppm)
1	6.8	16.1	18.0	0.085
2	6.7	16.7	39.2	0.113
3	7.2	16.6	40.0	0.104
4	5.4	16.3	47.2	0.116

All the MAE values obtained from this work are generally small, suggesting that all the structures show a reasonable match with the experimental results. The isomer with the lowest MAE corresponds to isomer **1**, which also displays the lowest relative energy, second largest pore size (6.8 Å) and the smallest weighted average diameter (16.1 Å) within the set of the 4 selected isomers. The predicted ^{1}H WP04/DGTZVP chemical shifts for the isomer **1** are shown in Figure S40.

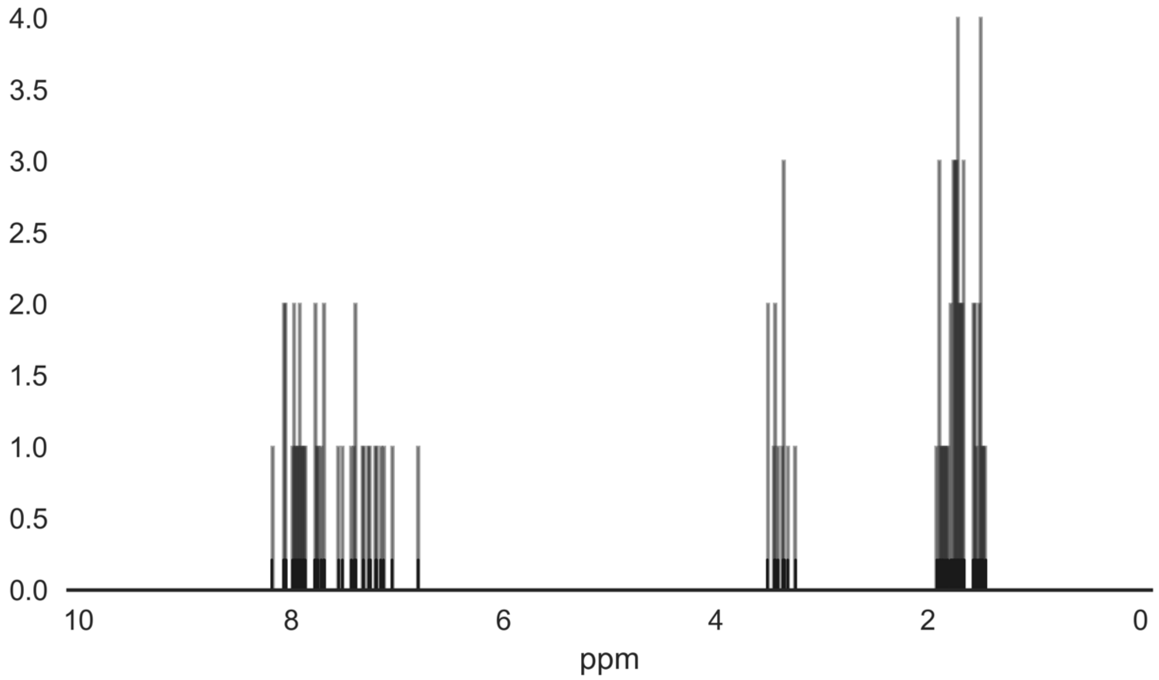


Figure S40: Predicted ^{1}H WP04/DGTZVP chemical shifts for isomer **1**. There is a MAE of 0.085 ppm compared to the experimental results shown in Fig. 2b.

5.7. Structural properties of the 4 final isomers

For the final four candidates, structural properties, pore and window size, were investigated through the use of pywindow. Figure 4 and Figures S28 – S30 show the structure of the isomers, with their cavities and four windows highlighted and labelled with size.

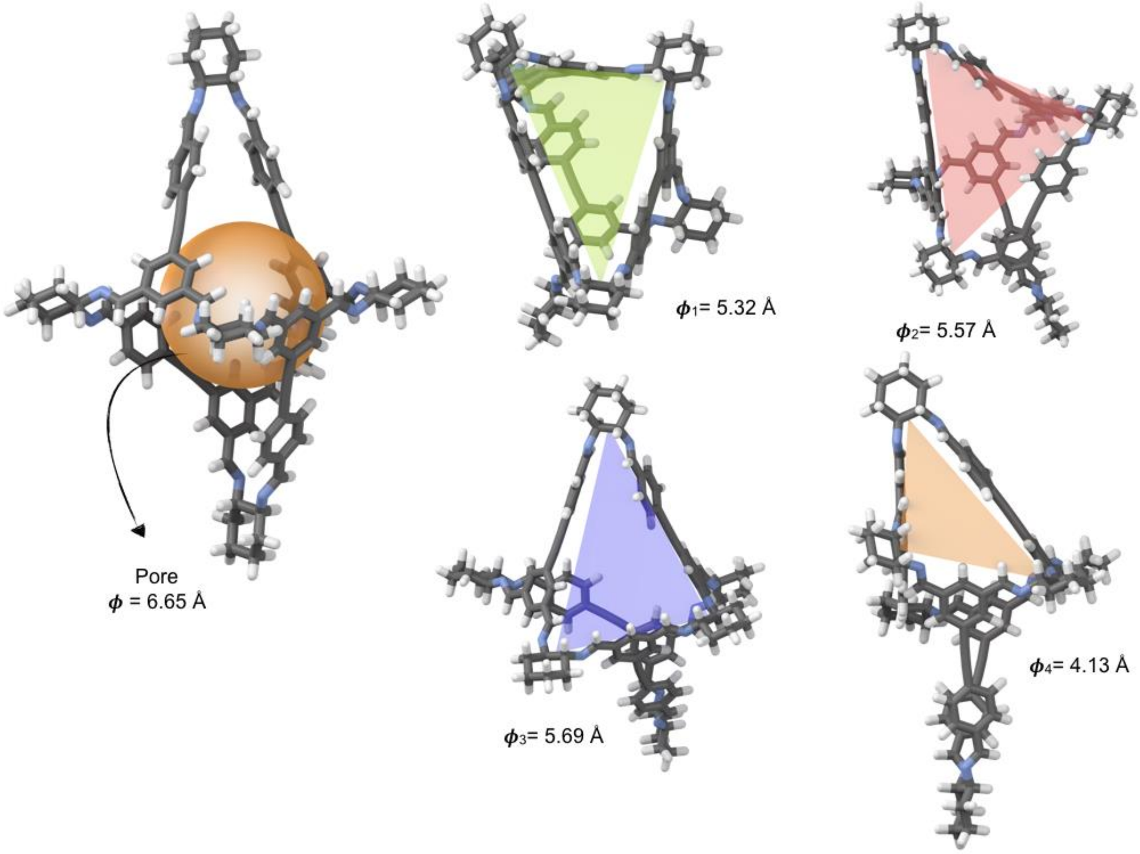


Figure S41: The structure of isomer **2**, with the 6.65 Å diameter pore shown as an orange sphere on the left. On the right, the diameter size of each of the 4 windows are shown. Transparent triangles have been used to facilitate the visualization of each window.

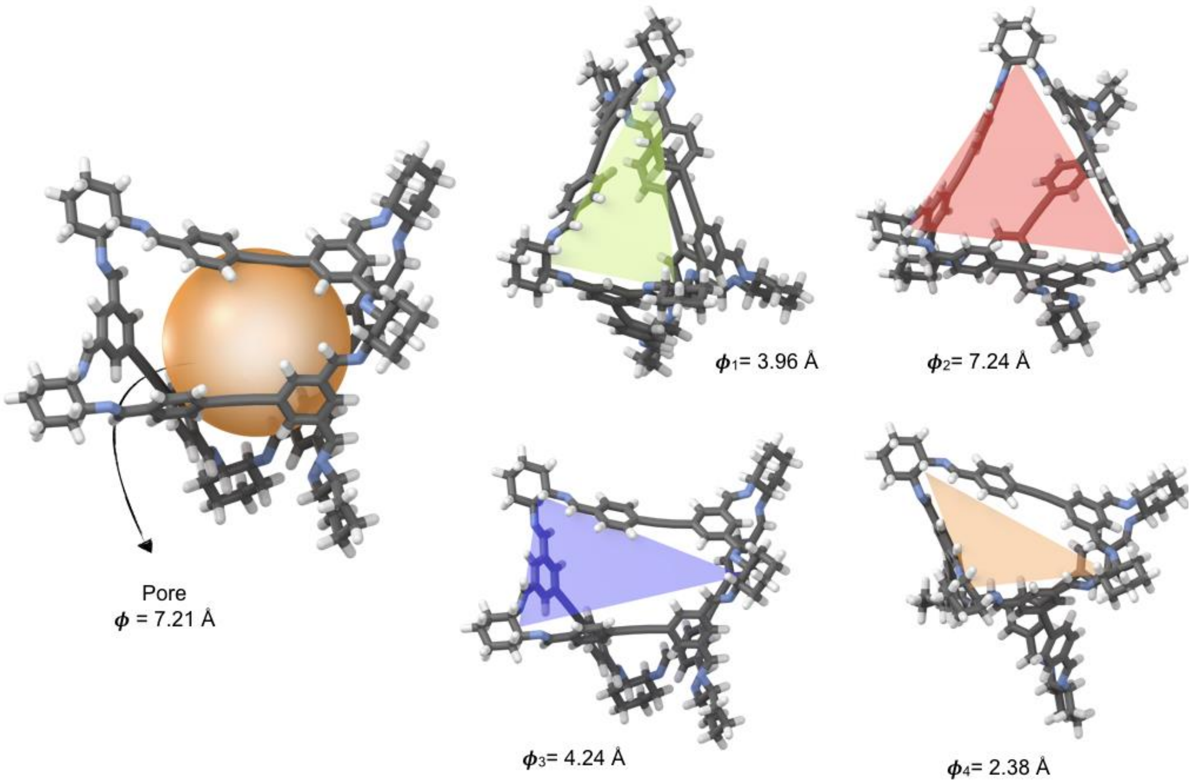


Figure S42: The structure of isomer **3**, with the 7.21 Å diameter pore shown as an orange sphere on the left. On the right, the diameter size of each of the 4 windows are shown. Transparent triangles have been used to facilitate the visualization of each window.

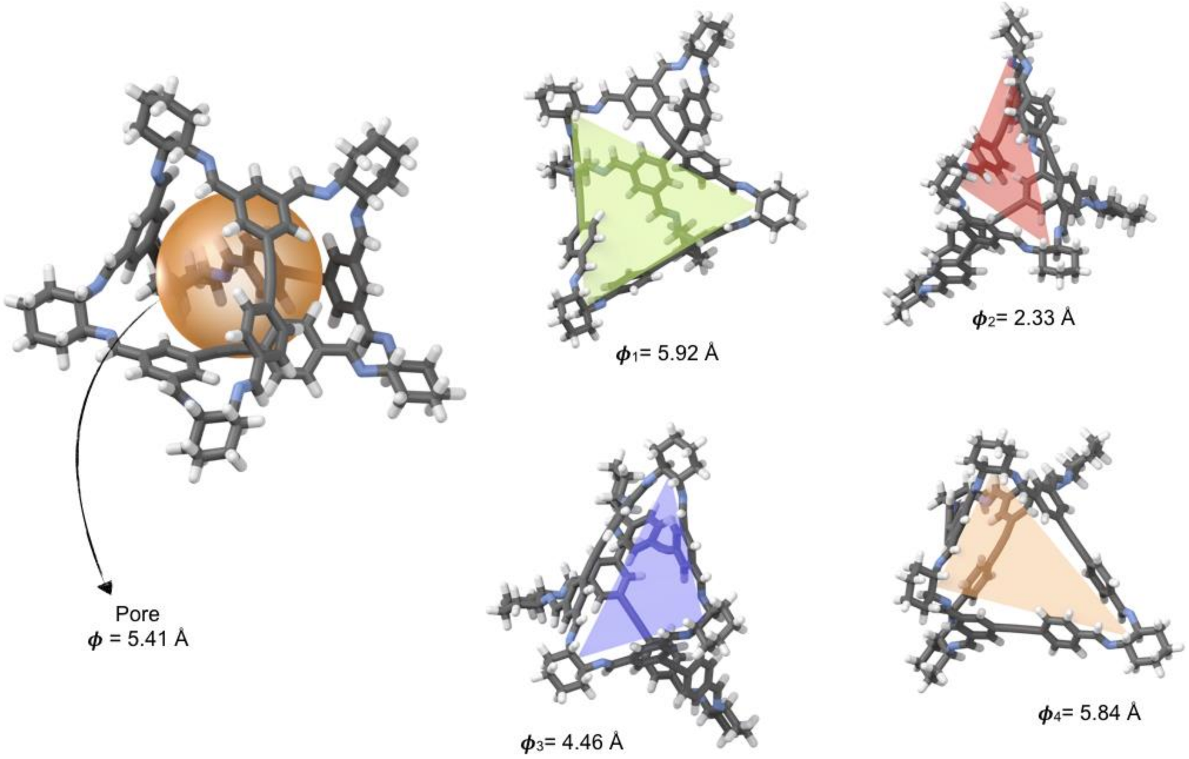


Figure S43: The structure of isomer **4**, with the 5.41 Å diameter pore shown as an orange sphere on the left. On the right, the diameter size of each of the 4 windows are shown. Transparent triangles have been used to facilitate the visualization of each window.

5.8. Amorphous model

For the generation of the amorphous model, we employed a multistep process. In the loading stage, the simulation cells were packed with 100 cage structures of isomer **1** at a density of 0.1 g cm^{-3} using the pack module from the Polymatic package.[22,23] All the cage structures were described using the cage specific force field (CSFF),[24] which was specifically developed for the description of imine cage systems. We were unable to use OPLS3[15] for solid state calculations, as OPLS3 is available only as part of the commercial software, Schrödinger PLC's Macromodel, which can only perform molecular calculations. Partial charges were calculated for the cage molecules by fitting atomic charges from the output of Gaussian16[16] single point calculations at the HF/6-31G* level of theory (on the B3LYP-D3/Def2-TZVP optimized geometry). The LAMMPS package (http://lammps.sandia.gov)[25] was then employed for a 21-step relaxation procedure.[23] This procedure has previously been shown to be suitable for producing representative structural models of POC systems.[26,27] Similarly, our previous experience suggests there is only minor deviation in properties across a series of amorphous models generated in this type of approach across the same system, therefore we only needed to produce a single model to represent the amorphous isomer **1** system.[27,28]

The final density of the amorphous model of isomer **1**, 0.87 g cm^{-3}, is in line with the densities of other amorphous models of different porous organic cages generated with similar computational approaches.[26] The final cell size for the 100 molecules was 66 x 66 x 66 Å. Zeo^{++}[29,30] was used for void analysis on the relaxed amorphous model. The void space, the largest cavity diameter, pore limiting diameter, pore size distribution and the interconnectivity of void space were calculated, the latter through a Voronoi decomposition. These values are given in Table S10. The pore limiting diameter (PLD), the largest probe that can diffuse from one side of the static model to the other, is 3.09 Å in the amorphous model, which is large enough for the kinetic radius of H_2 of 1.42 Å[31] to diffuse freely, but is a pore too small for either the van der Waals (1.55 Å) or kinetic radius (1.82 Å)[31] of N_2. However, our experience of dynamic porosity, whereby thermal fluctuations open up pores that are seemingly inaccessible in static models,[32,33] would suggest that this system is likely to be porous to both H_2 and N_2, as experimentally observed.

Table S10: Structural properties for the amorphous model of isomer **1**.

Density at 300 K, 1 bar (g cm^{-3})	Diameter of largest cavity (Å)	Pore limiting diameter, PLD (Å)	Total SA for 1.55 Å radius probe (m^2 g^{-1})	Total SA for 1.82 Å radius probe (m^2 g^{-1})
0.870	8.38	3.09	837	456

The surface area for the amorphous model was calculated with both the probe radii of 1.55 and 1.82 Å, equivalent to the van der Waals radius and the kinetic radius of N_2, respectively. In POC systems where the PLD is close to that of the guest of interest, as amorphous isomer **1** is to N_2, reducing the probe size from the kinetic radius to the van der Waals radius (representing the minimum cross-section of a N_2 molecule diffusing "end-on" through a pore) can assist in uncovering surface area that would be realised through dynamic porosity mechanisms.[33] For example, for **CC3**-*R*,[10] which has a PLD of 3.62 Å[32] and SA_{BET} of 409 ± 9 $m^2 g^{-1}$ in crystalline form,[34] we calculate a surface area in Zeo^{++} of 527 $m^2 g^{-1}$ with a probe of 1.55 Å and of 286 $m^2 g^{-1}$ with a probe of 1.82 Å. In other words, the experimental SA_{BET} sits between the values calculated with the two different sized probes to represent N_2.

For the amorphous isomer **1** model, the calculated surface areas range from 456 $m^2 g^{-1}$ to 837 $m^2 g^{-1}$, for probes of 1.82 Å and 1.55 Å respectively. As with **CC3**-*R*, this range would span the experimentally reported SA_{BET} of 578 $m^2 g^{-1}$. Thus isomer **1** generates an amorphous solid-state structure that can explain the experimentally observed properties.

It was of interest to determine to what extent the intrinsic porosity (voids within a cage cavity) was responsible for the void volume in the amorphous model, rather than extrinsic porosity (voids outside of the cage molecules resulting from inefficient packing). To uncover to what extent the intrinsic pores were contributing to the voids in the amorphous model, we used pywindow to compare the pore size of each isolated cage molecule extracted from the amorphous assembly against the actual pore size of the same cage within the packed system. As can be observed in Fig. S44, the isolated cages have pore diameters (when approximated as a sphere) that generally fall within the range of 6 – 8 Å, values that fit with the pore diameter of isomer **1** optimized in the gas phase at B3LYP-D3/Def2-TZVP level of theory (6.8 Å). Within the packed amorphous model, the cages show a much lower actual porosity, due to the interpenetration of adjacent cages within the large windows of isomer **1**, with the most frequent peak falling at around 0 Å. This analysis suggests that 83% of the potential void space is lost when isomer **1** is packed in the amorphous solid state and that the overall porosity exhibited by the amorphous model can therefore be mostly attributed to extrinsic porosity.

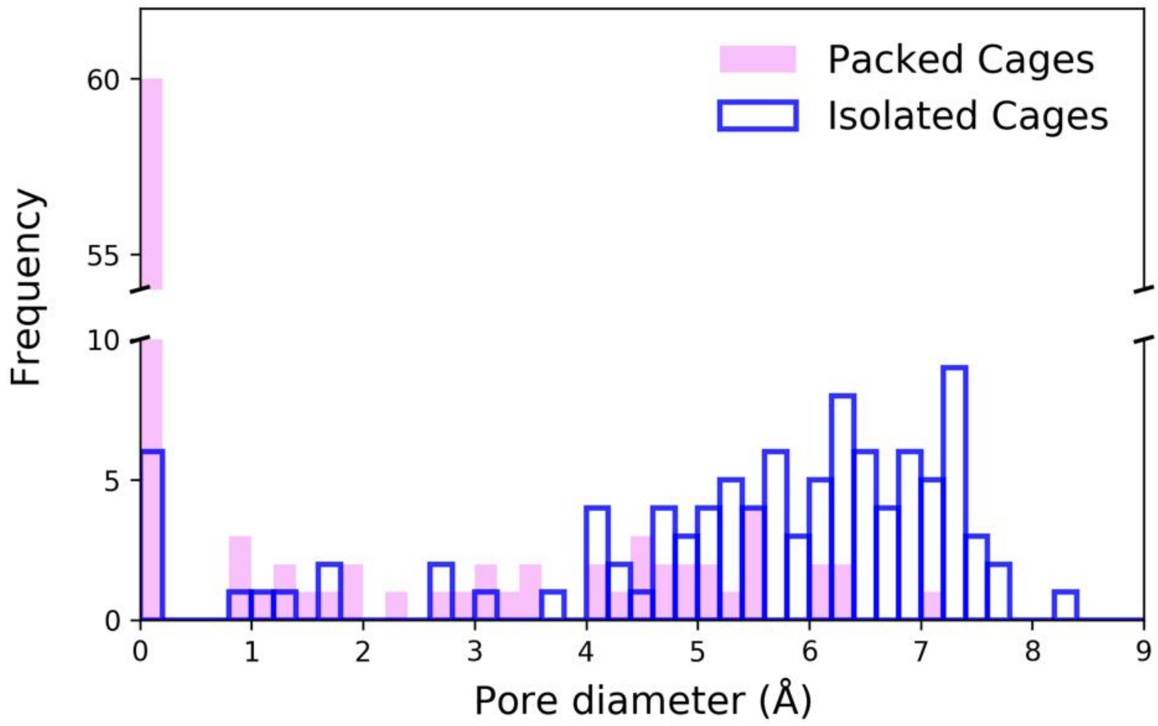


Figure S44: Comparison of the porosity for all the isolated cages extracted from the amorphous model (blue bars) *vs* the porosity of the cages within the packed assembly (pink bars).

6. Comparison of experimental and computational data

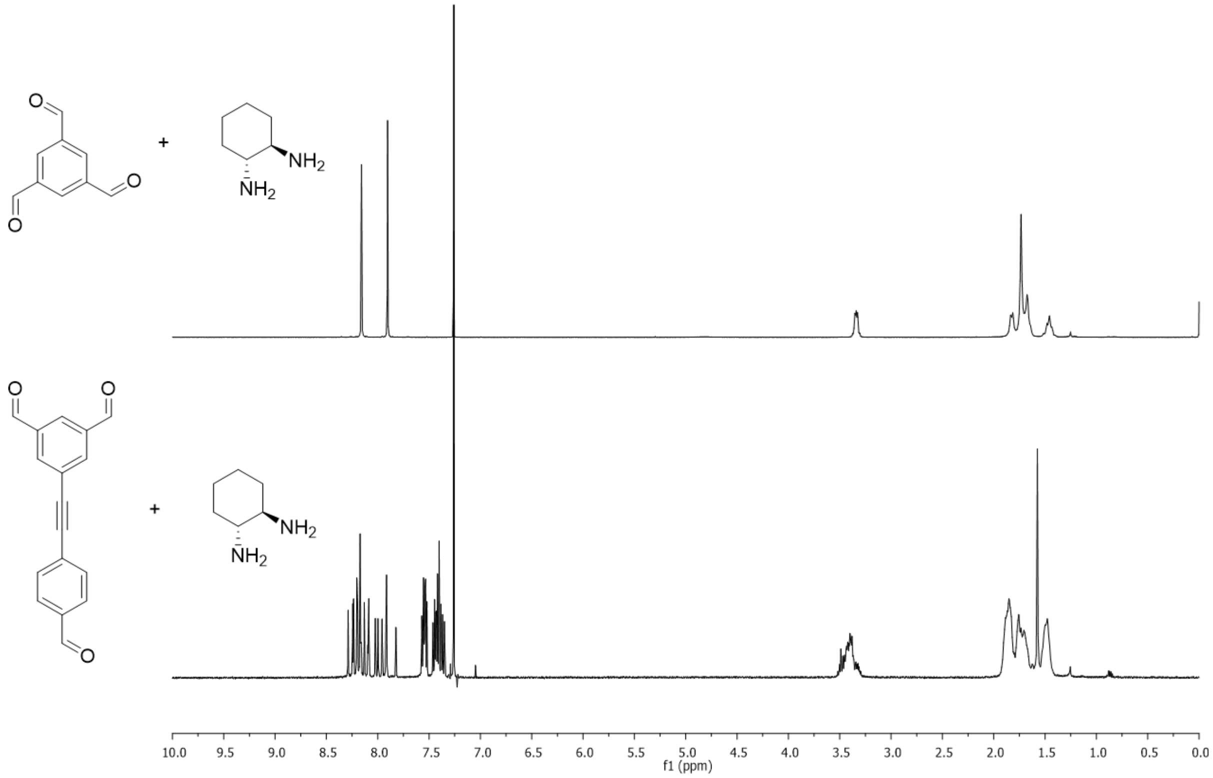


Figure S45: Stacked ^1H NMR spectra ($CDCl_3$) comparing symmetrical **CC3** (top),[10] formed using a D_{3h} symmetric trialdehyde (1,3,5-triformylbenzene), with the unsymmetrical cage formed (bottom), using a C_{2v} symmetric trialdehyde **1** (5-((4-formylphenyl)ethynyl)isophthalaldehyde).

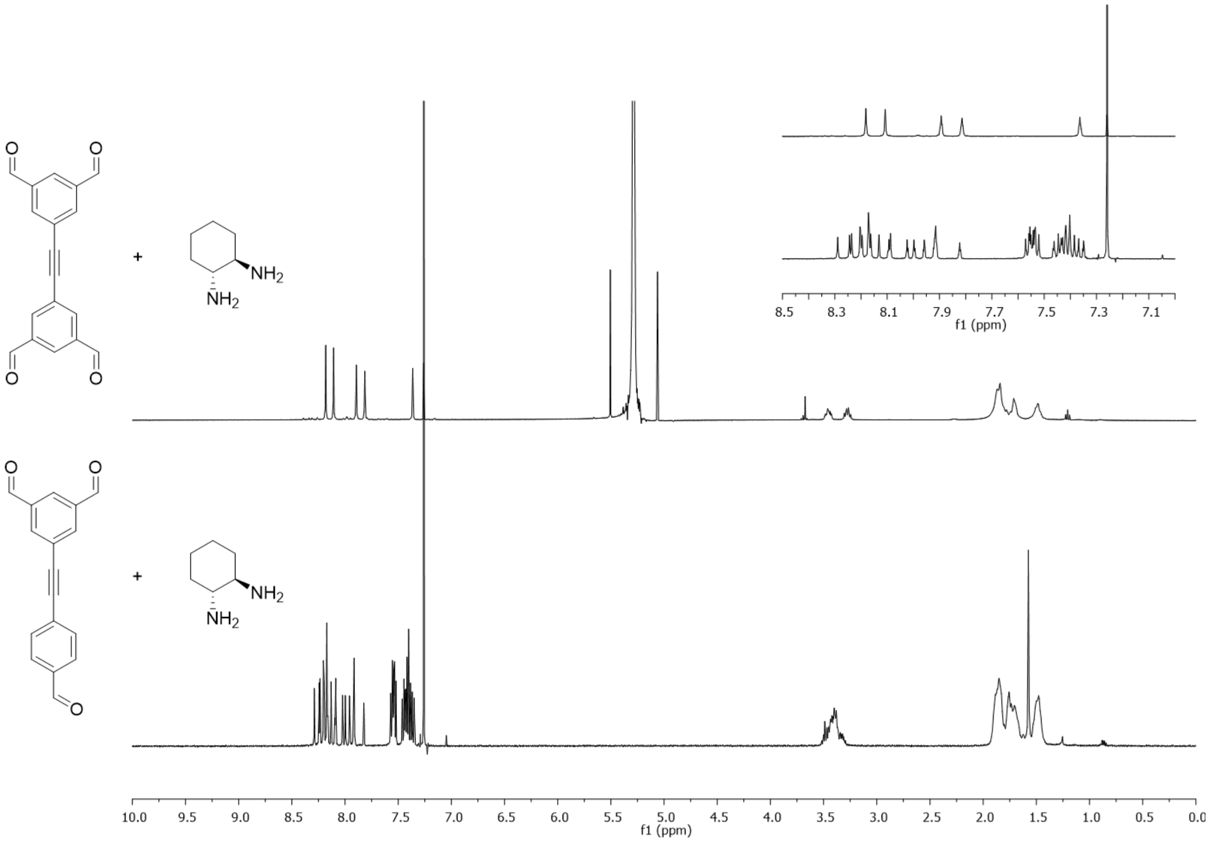
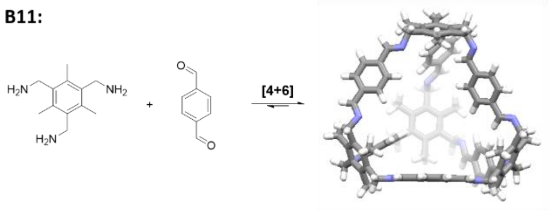


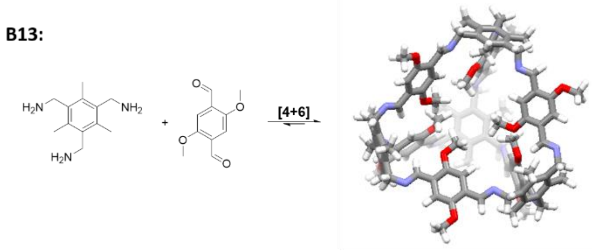
Figure S46: Stacked ^{1}H NMR spectra (CDCl$_3$) comparing symmetrical **TCC1**-*R* (top), formed using a D_{2h} symmetric tetraaldehyde (5,5′-(ethyne-1,2-diyl)diisophthalaldehyde), with the unsymmetrical cage formed (bottom), using a C_{2v} symmetric trialdehyde **1** (5-((4-formylphenyl)ethynyl)isophthalaldehyde). Inset (top right): unlike **CC3**, **TCC1** has 2 imine environments and 3 aromatic environments which have similar shifts to some of those in the unsymmetrical cage species.

Table S11: Comparison of previously calculated pore sizes and average diameters for a range of **Tri^{4}Di^6** cages, and their experimentally measured solvodynamic diameters,[18] with the unsymmetrical cage values (see Table S5).

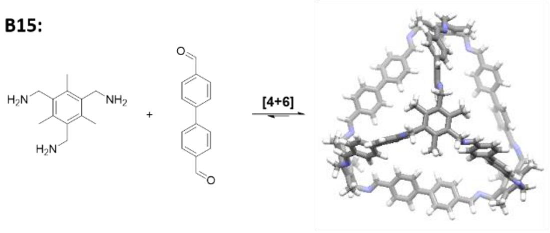
B11:



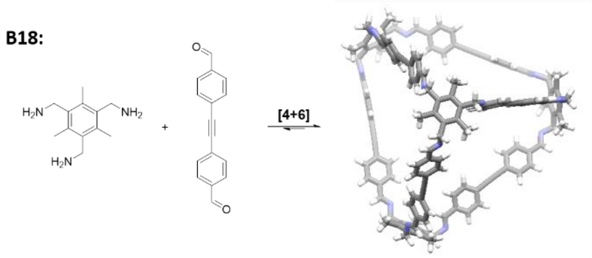
B13:



B15:



B18:



Calculated Sizes (Å)			Experimentally Measured Solvodynamic Diameter (Å)	Δ(CALC-EXP)
Cage	**Pore Size**	**Calculated Average Diameter**		
B11	6.3	18.5	16.7	1.8
B13	3.8	16.5	18.3	-1.8
B15	9.9	22.2	19.7	2.5
B18	12.1	24.6	20.6	4.0
Unsymmetrical Cage*	6.8	16.1	18.6	-2.5

*Data shown for isomer **1** (see Table S9)

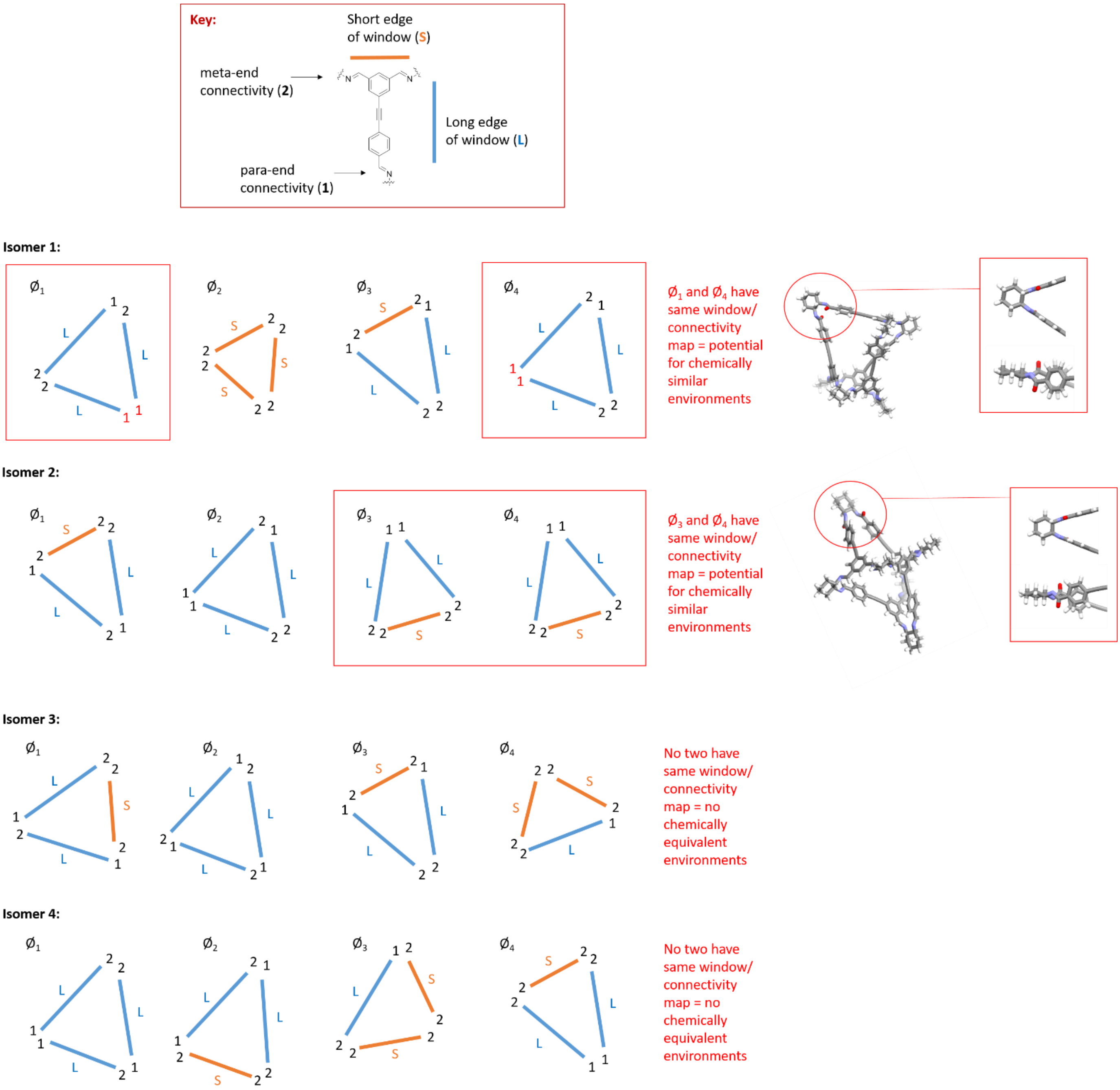


Figure S47: 2D Maps representing the different windows and connectivities of the isomers for the four potential candidates, confirming lack of symmetry. Whilst isomer **1** and **2** could potentially have 2 imines each in a chemically similar environment based on the mapping (highlighted in red on the corresponding models), this is still in agreement with the experimental NMR data with some of the protons and carbons being in similar environments based on the overlapping peak shifts.

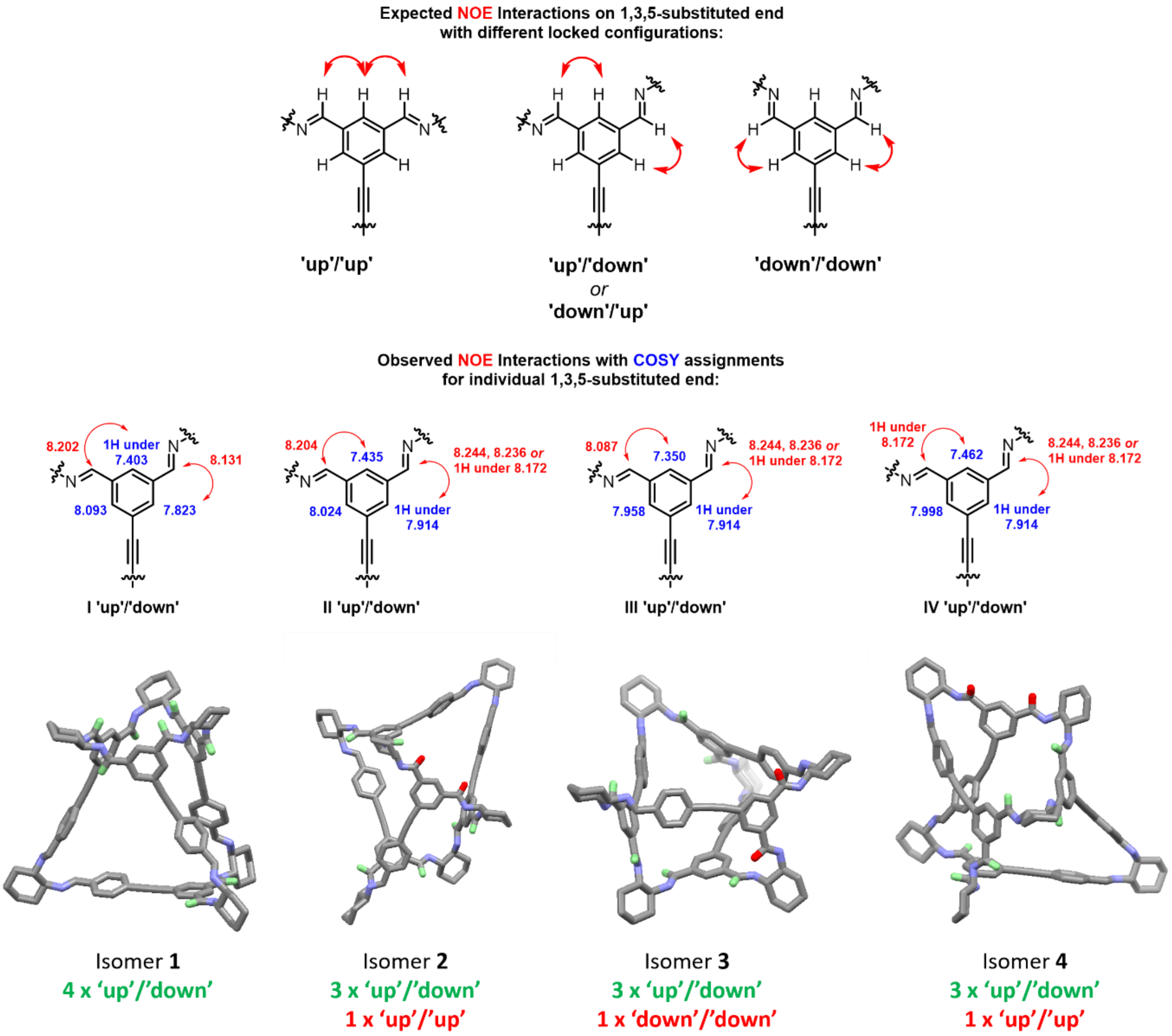


Figure S48: Comparison of expected NOEs for different imine orientations (top), compared to those seen experimentally (middle), confirming an 'up'/'down' configuration on all 4 faces. On comparison to the final 4 candidate isomers, only isomer **1** has the same configuration.

7. References

(1) Stejskal, E. O.; Tanner, J. E. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time Dependent Field Gradient. *J. Chem. Phys.* **1965**, *42* (1), 288–292.

(2) Blackburn, O. A.; Coe, B. J.; Helliwell, M.; Raftery, J. Syntheses, Structures, and Electronic and Optical Properties of platinum(II) Complexes of 1,3-Bis(imino)benzene-Derived Pincer Ligands. *Organometallics* **2012**, *31* (15), 5307–5320.

(3) Grawe, T.; Schrader, T.; Gurrath, M.; Kraft, A.; Osterod, F. Self-Organisation of Spheroidal Molecular Assemblies in Polar Solvents. *Org. Lett.* **2000**, *2* (1), 29–32.

(4) Roelens, S.; Vacca, A.; Francesconi, O.; Venturi, C. Ion-Pair Binding: Is Binding Both Binding Better? *Chem. Eur. J.* **2009**, *15* (33), 8296–8302.

(5) Vacca, A.; Nativi, C.; Cacciarini, M.; Pergoli, R.; Roelens, S. A New Tripodal Receptor for Molecular Recognition of Monosaccharides. A Paradigm for Assessing Glycoside Binding Affinities and Selectivities by 1H NMR Spectroscopy. *J. Am. Chem. Soc.* **2004**, *126* (50), 16456–16465.

(6) Jiang, S.; Jones, J. T. A.; Hasell, T.; Blythe, C. E.; Adams, D. J.; Trewin, A.; Cooper, A. I. Porous Organic Molecular Solids by Dynamic Covalent Scrambling. *Nat. Commun.* **2011**, *2*, 207.

(7) Greenaway, R. L.; Holden, D.; Eden, E. G. B.; Stephenson, A.; Yong, C. W.; Bennison, M. J.; Hasell, T.; Briggs, M. E.; James, S. L.; Cooper, A. I. Understanding Gas Capacity, Guest Selectivity, and Diffusion in Porous Liquids. *Chem. Sci.* **2017**, *8* (4), 2640–2651.

(8) Slater, A. G.; Reiss, P. S.; Pulido, A.; Little, M. A.; Holden, D. L.; Chen, L.; Chong, S. Y.; Alston, B. M.; Clowes, R.; Haranczyk, M.; et al. Computationally-Guided Synthetic Control over Pore Size in Isostructural Porous Organic Cages. *ACS Cent. Sci.* **2017**, *3* (7), 734–742.

(9) Slater, A. G.; Little, M. A.; Pulido, A.; Chong, S. Y.; Holden, D.; Chen, L.; Morgan, C.; Wu, X.; Cheng, G.; Clowes, R.; et al. Reticular Synthesis of Porous Molecular 1D Nanotubes and 3D Networks. *Nat. Chem.* **2016**, *9*, 17.

(10) Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; et al. Porous Organic Cages. *Nat. Mater.* **2009**, *8*, 973.

(11) Slater, A. G.; Little, M. A.; Briggs, M. E.; Jelfs, K. E.; Cooper, A. I. A Solution-Processable Dissymmetric Porous Organic Cage. *Mol. Sys. Des. Eng.* **2018**.

(12) Santolini, V.; Miklitz, M.; Berardo, E.; Jelfs, K. E. Topological Landscapes of Porous Organic Cages. *Nanoscale* **2017**, *9* (16), 5280–5298.

(13) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Theory Comput.* **2015**, *55* (12), 2562–2574.

(14) RDKit: Open-Source Cheminformatics, rdkit.org

(15) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.

(16) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. Gaussian 16 Rev. B.01. Wallingford, CT 2016.

(17) Miklitz, M.; Jelfs, K. E., pyWindow. *ChemRxiv*, **2018**, DOI: 10.26434/chemrxiv.6850109.

(18) Greenaway, R. L.; Santolini, V.; Bennison, M. J.; Alston, B. M.; Stackhouse, C.; Little, M. A.; Miklitz, M.; Eden, E. G. B.; Clowes, R.; Shakil, A.; et al. High-Throughput Discovery of Organic Cages and Catenanes Using Computational Screening Fused with Robotic Synthesis. *Nat. Commun.* No. 2018, 1–27.

(19) Benassi, E. Benchmarking of Density Functionals for a Soft but Accurate Prediction and Assignment of 1H and 13C NMR Chemical Shifts in Organic and Biological Molecules. *J. Comp. Chem.* **2017**, *38* (2), 87–92.

(20) Marenich, A. V; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378–6396.

(21) Willoughby, P. H.; Jansma, M. J.; Hoye, T. R. A Guide to Small-Molecule Structure Assignment through Computation of (1H and 13C) NMR Chemical Shifts. *Nat. Protoc.* **2014**, *9*, 643.

(22) Abbott, L. J. .; Colina, C. M. Polymatic: A Simulated Polymerization Algorithm. 2013.

(23) Abbott, L. J.; Hart, K. E.; Colina, C. M. Polymatic: A Generalized Simulated Polymerization Algorithm for Amorphous Polymers. *Theor. Chem. Acc.* **2013**, *132* (3), 1334.

(24) Holden, D.; Jelfs, K. E.; Cooper, A. I.; Trewin, A.; Willock, D. J. Bespoke Force Field for Simulating the Molecular Dynamics of Porous Organic Cages. *J. Phys. Chem. C* **2012**, *116* (31), 16639–16651.

(25) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.

(26) Evans, J. D.; Huang, D. M.; Hill, M. R.; Sumby, C. J.; Sholl, D. S.; Thornton, A. W.; Doonan, C. J. Molecular Design of Amorphous Porous Organic Cages for Enhanced Gas Storage. *J. Phys. Chem. C* **2015**, *119* (14), 7746–7754.

(27) Jiang, S.; Jelfs, K. E.; Holden, D.; Hasell, T.; Chong, S. Y.; Haranczyk, M.; Trewin, A.; Cooper, A. I. Molecular Dynamics Simulations of Gas Selectivity in Amorphous Porous Molecular Solids. *J. Am. Chem. Soc.* **2013**, *135* (47), 17818–17830.

(28) Jimenez-Solomon, M. F.; Song, Q.; Jelfs, K. E.; Munoz-Ibanez, M.; Livingston, A. G. Polymer Nanofilms with Enhanced Microporosity by Interfacial Polymerization. *Nat. Mater.* **2016**, *15*, 760.

(29) Pinheiro, M.; Martin, R. L.; Rycroft, C. H.; Haranczyk, M. High Accuracy Geometric Analysis of Crystalline Porous Materials. *CrystEngComm* **2013**, *15* (37), 7531–7538.

(30) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Micropor. Mesopor. Mat.* **2012**, *149* (1), 134–141.

(31) Robeson, L. M. Correlation of Separation Factor versus Permeability for Polymeric

Membranes. *J. Membr. Sci.* **1991**, *62* (2), 165–185.

(32) Holden, D.; Jelfs, K. E.; Trewin, A.; Willock, D. J.; Haranczyk, M.; Cooper, A. I. Gas Diffusion in a Porous Organic Cage: Analysis of Dynamic Pore Connectivity Using Molecular Dynamics Simulations. *J. Chem. Phys. C* **2014**, *118* (24), 12734–12743.

(33) Holden, D.; Chong, S. Y.; Chen, L.; Jelfs, K. E.; Hasell, T.; Cooper, A. I. Understanding Static, Dynamic and Cooperative Porosity in Molecular Materials. *Chem. Sci.* **2016**, *7* (8), 4875–4879.

(34) Hasell, T.; Chong, S. Y.; Jelfs, K. E.; Adams, D. J.; Cooper, A. I. Porous Organic Cage Nanocrystals by Solution Mixing. *J. Am. Chem. Soc.* **2012**, *134* (1), 588–598.