Exploration of the polymorphic solid-state landscape of an amide-linked organic cage using computation and automation

C. E. Shields, T. Fellowes, A. G. Slater, A. I. Cooper, K. G. Andrews* and F. T. Szczypiński*

Supporting Information

1.   Experimental methodology

1.1. Robot configuration

Semi-automated solubility and crystallisation screening was performed on a Chemspeed SWING ISYNTH robotic platform. Solid handling was carried out using the gravimetric solid dispensing tool (SDU). Liquid handling was carried out using the four needle overhead dispensing tool with 4 syringe pumps. Needles were washed using the system solvent between dispenses to avoid contamination. Solid and liquid handling was performed at RT in a closed system (**Figure S1**). Solubility and crystallisation screening was performed in standard 8 mL vials with pre-slit septum-sealed caps.
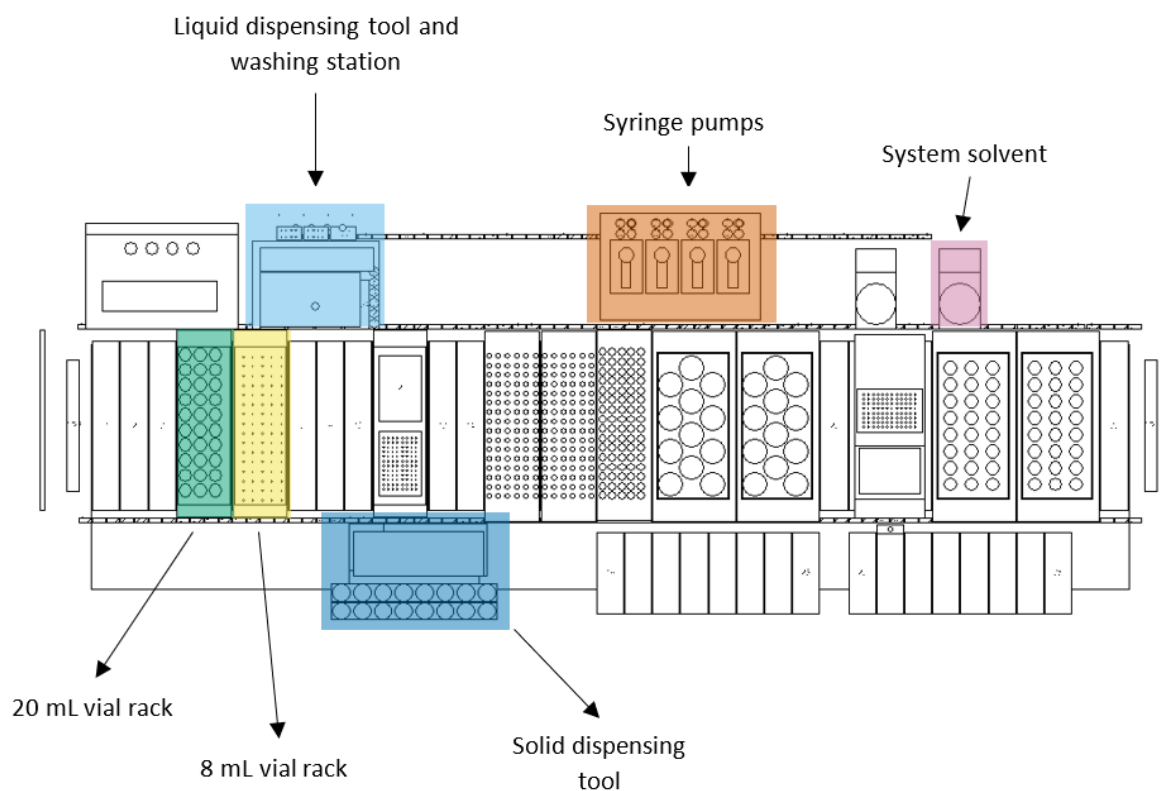
Figure S1: Chemspeed SWING ISYNTH robotic platform configuration for semi-automated solubility and crystallisation screening.

1.2. Liquid dispensing parameters

Liquid dispensing parameters were adjusted to improve layer formation between the stock solution and antisolvent. Using the standard dispense conditions (needle height: 25 mm from top of vial, dispense speed: 10 mL/min), some layer formation was observed, but the majority of the solution was mixed immediately. By bringing the needle closer to the level of the stock solution in the vial (35 mm from top of vial) and reducing the dispense speed (5 mL/min), antisolvent addition caused less immediate mixing. The improvement in layer formation was demonstrated by dispensing an IPA solution containing green dye onto a solution of chloroform (**Figure S2**).
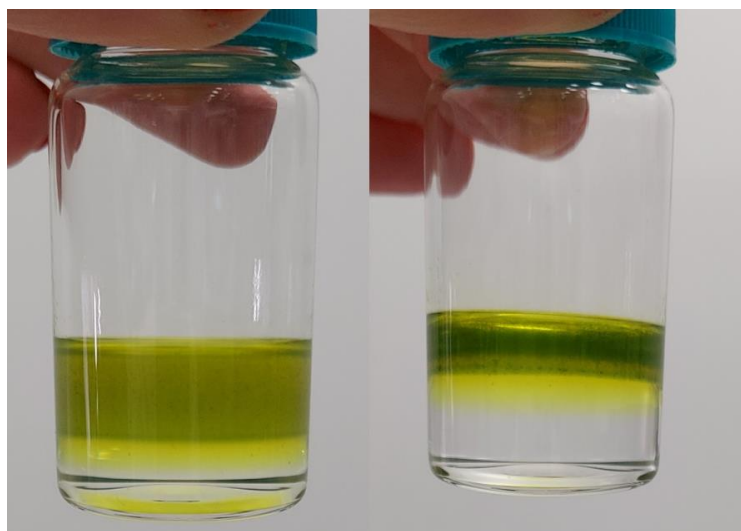
Figure S2: The difference in solvent-antisolvent layering with default dispense conditions (left) and adjusted dispense conditions (right).

1.3. Powder X-ray diffraction (PXRD)

High-throughput PXRD patterns were collected in vertical transmission mode from loose powder samples held on Mylar film in aluminium well plates, using a Panalytical Empyrean equipped with a high throughput screening XYZ stage, X-ray focusing mirror and PIXcel detector with Cu-Kα radiation (λ = 1.541 Å). PXRD patterns were recorded at room temperature.

1.4. Single crystal X-ray diffraction

SCXRD data sets were measured on a Rigaku AFC12K-007 HF rotating anode diffractometer (Mo-Kα radiation, λ = 0.71073 Å, Kappa 4-circle goniometer, HyPix-6000HE detector), or at beamline I19, Diamond Light Source, Didcot, UK using silicon double crystal monochromated synchrotron radiation (λ = 0.6889 Å, Pilatus 2M detector). Solvated single crystals were isolated from the crystallization solvent, immersed in a protective oil, mounted on a MiTeGen loop, and flash-cooled to 100 K under a dry N_2 gas flow unless stated otherwise. For synchrotron data collected at Diamond, data reduction and absorption corrections were performed with xia2. For data collected in-house using the Rigaku instrument, reduction was performed using the CrysAlisPro software. Structures were solved with SHELXT or SHELXD and refined by full-matrix least-squares on $|F|^2$ by SHELXL,[1,2] interfaced through the programme OLEX2.[3] All non-H atoms were refined anisotropically, and all H-atoms were fixed in geometrically estimated positions and refined using the riding model unless stated otherwise. Where structures were found to contain disordered solvent molecules, the SQUEEZE routine of PLATON was used to remove scattering caused by disordered guests.[4,5]

1.5. Semi-automated crystallisation

37 organic solvents were used for the solvent library across two semi-automated crystallisation screens (**Table S1**). Initially, the solubility of cage **1** in each solvent was determined, then solvent-antisolvent crystallisation experiments were performed.

Table S1: Solvent library.

First screen				
Dimethyl sulfoxide (DMSO)	1, 4-Dioxane	Tetrahydrofuran (THF)	Pyridine	Acetone
Pentane	Diethyl ether (Et$_2$O)	1-Propanol	1-Octanol	Acetonitrile (MeCN)
1-Butanol	Methanol (MeOH)	Cyclohexane	Mesitylene	Xylene
Ethyl acetate	2-Butanol	Propan-2-ol (IPA)	Dichloromethane (DCM)	Chloroform (CHCl$_3$)
Second screen				
N,N-dimethyl formamide (DMF)	1,3-Dioxolane	*N*-methyl-2-pyrrolidone (NMP)	*N,N*-dimethyl acetamide (DMAc)	1,3-Dimethoxy-benzene
Methyl *t*-butyl ether (MTBE)	2,2,2-Triflouroethanol (TFE)	Propylene carbonate (PC)	1,2-Dichlorobenzene	Methyl acetate
Benzyl alcohol	Toluene	Triethyl orthoformate (TEOF)	Cyclohexanone	Tetrahydropyran (THP)
Methyl benzoate	2-Butanone			

1.6. Solubility screening

10 mg of cage **1** was dispensed into 8 mL vials, followed by 1 mL solvent. The vials were then removed from the Chemspeed platform, lightly shaken, then examined by eye. The solvents which dissolved the cage under these conditions were labelled as good solvents and used to prepare stock solutions of the cage. Across the two solubility screens, DMSO, 1,4-dioxane, THF, pyridine, DMF, 1,3-dioxolane, NMP, and DMAc were identified as good solvents. The remaining solvents which did not dissolve the cage were used as antisolvents.

1.7. Crystallisation screening

Stock solutions of cage **1** at 10 mg/mL were prepared manually at RT using the good solvents identified in the solubility screen. Due to the limited amount of cage available, only three good solvents were chosen per screen (Screen 1: DMSO, 1,4-dioxane, THF. Screen 2: DMF, 1,3-dioxolane, NMP). On the Chemspeed platform, 1 mL of each stock solution was dispensed into 8 mL sample vials. 1 mL of antisolvent was then carefully layered on top. Vials were moved into a fume hood and left at room temperature to crystallise.

1.8. XRD analysis

After two weeks, samples were visually inspected for crystal formation. Vials containing material suitable for single crystal analysis were set aside. For the remaining samples containing precipitate, excess crystallisation solvents were removed by pipette, then the samples were left to dry (either in air or under vacuum at 25 °C if required) and analysed by PXRD.

Table S2: Summary of PXRD data from crystallisation screens. Colours show how the PXRD patterns were grouped based on similarity. Asterisks (*) denote which samples were dried under vacuum. Dashes (-) indicate no precipitate was formed over the screening period.

	DMSO	1,4-Dioxane	THF
Acetone	Crystalline*	Amorphous	Amorphous
Pentane	Crystalline*	Amorphous	Amorphous
Et$_2$O	Crystalline*	Amorphous	Amorphous
1-Propanol	Crystalline*	Crystalline	Amorphous
1-Octanol	-	Single crystals	Single crystals
MeCN	Crystalline*	Amorphous	Crystalline
1-Butanol	Crystalline*	Crystalline*	Amorphous
MeOH	Crystalline*	Crystalline	Crystalline
Cyclohexane	Crystalline*	Amorphous	Amorphous
Mesitylene	Crystalline*	Amorphous*	Crystalline*
Xylene	Crystalline*	Amorphous	Amorphous
Ethyl Acetate	Crystalline*	Amorphous	Amorphous
2-Butanol	Crystalline*	Crystalline	Amorphous
IPA	Crystalline*	Amorphous	Amorphous
DCM	Crystalline*	Amorphous	Amorphous
CHCl$_3$	Crystalline*	Amorphous	Amorphous
	DMF	1,3-Dioxolane	NMP
1,3-Dimethoxybenzene	-	-	-
MTBE	Single crystals	Crystalline	Crystalline
TFE	Single crystals	Crystalline	-
PC	-	Single crystals	-
1,2-Dichlorobenzene	-	-	-
Methyl acetate	Single crystals	Amorphous	-
Benzyl alcohol	-	-	-
Toluene	Crystalline	Amorphous	-
Triethyl orthoformate	Single crystals	Crystalline	-
Cyclohexanone	-	-	-
THP	Single crystals	Amorphous	-
Methyl benzoate	-	-	-
2-Butanone	Single crystals	Crystalline	-

PXRD patterns were compared and tentatively grouped based on similarity across both crystallisation screens. Each group of similar patterns were considered to represent a different solvate or polymorph. A summary can be found in Table S2. Crystallisation experiments which yielded unique patterns were then repeated manually in an attempt to grow single crystals. In cases where many different solvent combinations resulted in a similar pattern, a few conditions were selected at random to try manually. In the first crystallisation screen, we found that all samples dried under vacuum had almost identical PXRD patterns, suggesting a phase transformation upon desolvation that may not be indicative of the solvated structure. Vacuum drying was therefore avoided in the second crystallisation screen. A representative example of PXRD patterns from the first crystallisation screen is displayed in **Figure S3**.
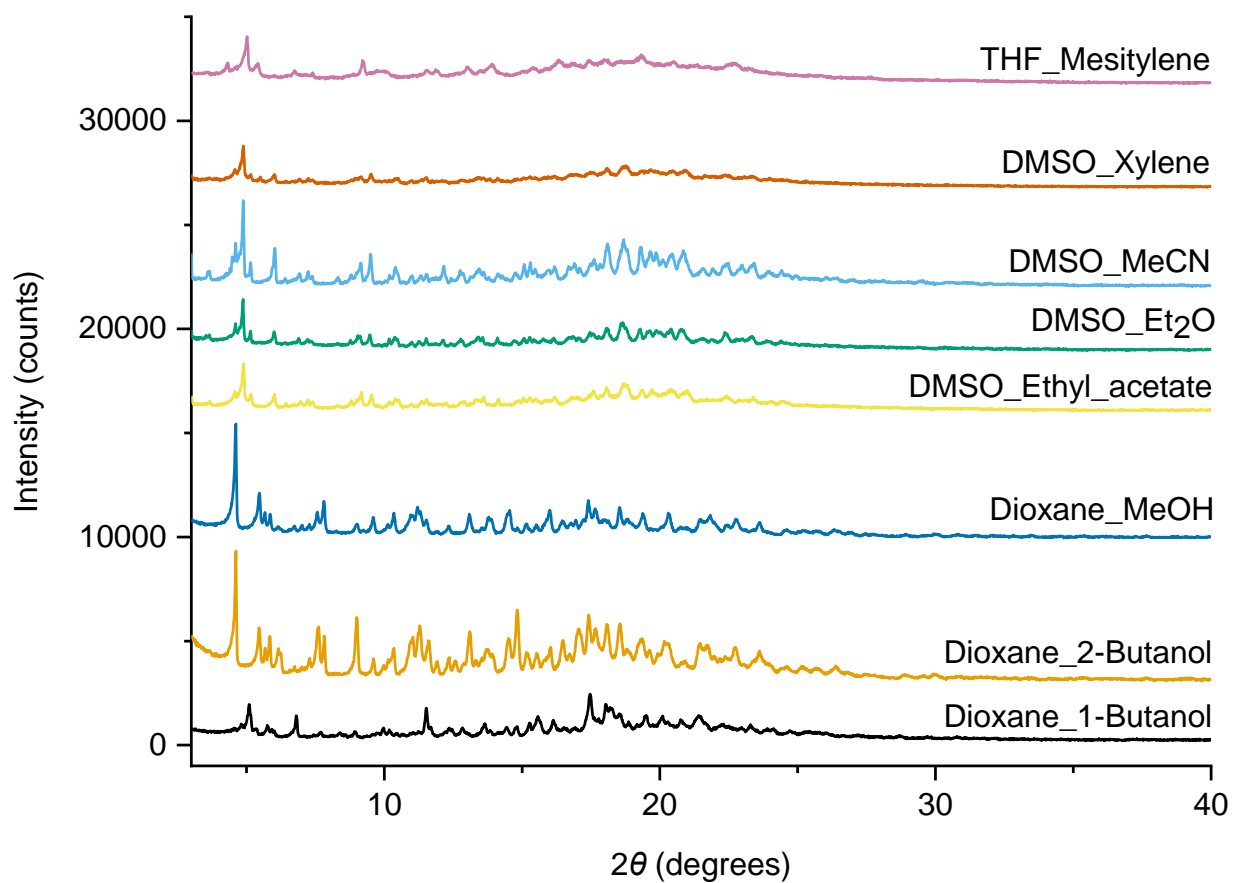
Figure S3: A representative selection of PXRD patterns from the first crystallisation screen.

2. Summary of single crystal diffraction data

Table S3: Summary of single crystal data.

Name	Crystal habit	Lattice	Unit cell parameters	Space group	Z'	Conformers observed
THF-MeOH-C9	Needles	aP	a=22, b=26, c=29, α=73, β=71, γ=72	$P1$	1	**C9** (previous work)[6]
THF-octanol-C9	Needles	mP	a=23, b=27, c=32, α=90, β=93, γ=90	$P2_1/n$	1	**C9**
THF-mesitylene	Needles	-	a=20, b=22, c=37, α=90, β=99, γ=90	-	-	Poorly diffracting. Not collected.
Dioxane-octanol-C9	Needles	mP	a=22, b=27, c=32, α=90, β=90, γ=90	$P2_1/n$	1	**C9**
Dioxane-MeOH	Needles	-	a=23, b=27, c=32, α=90, β=93, γ=90	-	-	Not collected – assumed to be **C9**.
Dioxane-1-butanol	Needles	mP	a=38, b=27, c=39, α=90, β=93, γ=90	-	-	Poorly diffracting. Not collected.
Dioxane-2-butanol	Needles	mP	a=52, b=54, c=23, α=90, β=104, γ=90	-	-	Poorly diffracting. Not collected.
DMSO-MeOH-C5-C9	Needles	mI	a=52, b=22, c=57, α=90, β=90, γ=90	$I2$	4	**C5** and **C9**
DMSO-MeOH-C13	Blocks	hP	a=20, b=20, c=33, α=90, β=90, γ=120	P6₃22	1/6	**C13**
DMSO-MeCN-C10-C9	Needles	aP	a=21, b=26, c=31, α=74, β=72, γ=68	$P1^-$	2	**C10** and **C9**
DMSO-1-propanol	Needles	-	a=22, b=26, c=32, α=74, β=72, γ=68	-	-	Not collected – assumed to be **C9/C10**.
DMF-MTBE	Needles	-	a=21, b=26, c=30, α=73, β=71, γ=72	-	-	Not collected – assumed to be **C9**.
DMF-methyl-acetate	Needles	-	a=22, b=26, c=31, α=90, β=90, γ=90	-	-	Not collected – assumed to be **C9**.
DMF-TEOF	Needles	-	a=21, b=26, c=29, α=94, β=108, γ=111	-	-	Not collected – assumed to be **C9**.
DMF-TFE-C9	Needles	aP	a=22, b=27, c=28, α=94, β=111, γ=110	$P1^-$	2	**C9**
DMF-THP-C10-C12	Needles	aP	a=21, b=27, c=29, α=94, β=108, γ=113	$P1^-$	2	**C10** and **C12**
DMF-2-butanone	Needles	-	a=21, b=26, c=29, α=94, β=108, γ=111	-	-	Not collected – assumed to be **C9**
Dioxolane-PC-C9	Needles	mC	a=39, b=28, c=35, α=90, β=109, γ=90	$C2/c$	1	**C9**

IDENTIFICATION CODE	KA1_DIOX_OCT_C9	KA1_DMF_TFE	KA1_DMF_THP	KA1_DMSO_MECN
EMPIRICAL FORMULA	$C_{198}H_{251}N_6O_{46}$	$C_{154.5}H_{176.5}N_{15.5}O_{19.5}$	$C_{342}H_{393}N_{12}O_{80.01}$	$C_{152.46}H_{189.38}N_6O_{23.23}S_{13.23}$
FORMULA WEIGHT	3451.03	2562.6	5951.82	2901.85
TEMPERATURE/K	100.15	100.15	100.15	100.15
CRYSTAL SYSTEM	monoclinic	triclinic	triclinic	triclinic
SPACE GROUP	$P2_1/n$	P-1	P-1	P-1
A/Å	22.4005(12)	21.60129(19)	21.4803(10)	21.1953(10)
B/Å	26.9673(7)	27.1502(2)	27.0745(12)	26.1285(12)
C/Å	32.2242(12)	28.3933(3)	29.3858(17)	31.3285(14)
A/°	90	93.8642(8)	94.133(4)	73.564(4)
B/°	90.916(4)	111.2557(9)	108.263(4)	72.242(4)
Γ/°	90	110.4407(7)	112.968(4)	68.045(4)
VOLUME/Å³	19463.5(14)	14178.7(2)	14568.6(11)	15042.0(13)
Z	4	4	2	4
$P_{CALC}G/CM^3$	1.178	1.2	1.357	1.281
M/MM⁻¹	0.083	0.075	0.09	0.26
F(000)	7396	5472	6338	6175
CRYSTAL SIZE/MM³	0.1 × 0.025 × 0.007	0.207 × 0.04 × 0.013	0.13 × 0.012 × 0.008	0.365 × 0.09 × 0.054
RADIATION	MoKα (λ = 0.71073)	synchrotron (λ = 0.6889)	synchrotron (λ = 0.6889)	MoKα (λ = 0.71073)
2Θ RANGE FOR DATA COLLECTION/°	3.44 to 43.934	1.592 to 40.296	1.624 to 40.296	3.054 to 41.632
INDEX RANGES	-23 ≤ h ≤ 23, -28 ≤ k ≤ 28, -33 ≤ l ≤ 33	-21 ≤ h ≤ 21, -27 ≤ k ≤ 27, -28 ≤ l ≤ 28	-21 ≤ h ≤ 21, -27 ≤ k ≤ 27, -29 ≤ l ≤ 29	-21 ≤ h ≤ 21, -26 ≤ k ≤ 26, -31 ≤ l ≤ 31
REFLECTIONS COLLECTED	145511	122084	116593	177974
INDEPENDENT REFLECTIONS	23765 [R_{int} = 0.0658, R_{sigma} = 0.0447]	29682 [R_{int} = 0.0630, R_{sigma} = 0.0716]	30461 [R_{int} = 0.1178, R_{sigma} = 0.1014]	31480 [R_{int} = 0.1131, R_{sigma} = 0.0702]
DATA/RESTRAINTS/PARAMETERS	23765/257/1281	29682/503/2598	30461/542/2569	31480/3008/2907
GOODNESS-OF-FIT ON F²	1.474	1.404	1.406	1.985
FINAL R INDEXES [I>=2Σ (I)]	R_1 = 0.1175, wR_2 = 0.3649	R_1 = 0.1060, wR_2 = 0.3365	R_1 = 0.1324, wR_2 = 0.3782	R_1 = 0.1777, wR_2 = 0.4696
FINAL R INDEXES [ALL DATA]	R_1 = 0.1475, wR_2 = 0.3900	R_1 = 0.1281, wR_2 = 0.3610	R_1 = 0.1763, wR_2 = 0.4083	R_1 = 0.2007, wR_2 = 0.4857
LARGEST DIFF. PEAK/HOLE / E Å⁻³	0.79/-0.46	1.09/-0.40	0.73/-0.46	1.63/-1.25
CSD DEPOSITION CODE	2343720	2343784	2343788	2343734

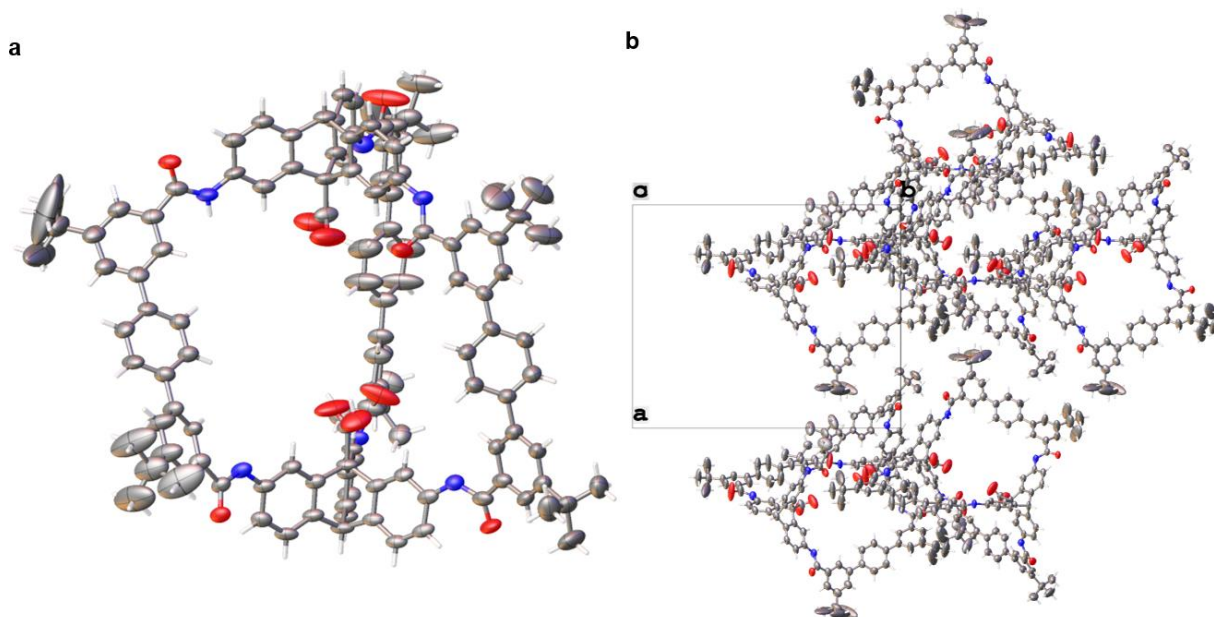| IDENTIFICATION CODE	KA1_DMSO_MEOH	KA1_DMSO_MEOH_C5	KA1_DMSO_MEOH_C13	KA1_PC_DIOX
EMPIRICAL FORMULA	$C_{138}H_{148}N_6O_{16}S_6$	$C_{132}H_{130}N_6O_{13}S_3$	$C_{174}H_{256}N_6O_{34}S_{24}$	$C_{190}H_{208}N_6O_{58}$
FORMULA WEIGHT	2338.98	2104.59	3745.27	3503.61
TEMPERATURE/K	100.1	100.15	100.15	100.15
CRYSTAL SYSTEM	triclinic	monoclinic	hexagonal	monoclinic
SPACE GROUP	P-1	I2	$P6_322$	C2/c
A/Å	21.6989(3)	51.7627(6)	19.90656(19)	39.4383(4)
B/Å	27.1612(3)	21.7575(2)	19.90656(19)	27.5839(2)
C/Å	27.6258(3)	57.0058(7)	33.4511(5)	35.2015(4)
A/°	94.1928(10)	90	90	90
B/°	111.5642(11)	90.4430(10)	90	109.9980(9)
Γ/°	109.9034(11)	90	120	90
VOLUME/Å³	13867.0(3)	64199.6(12)	11479.8(2)	35985.4(6)
Z	4	16	2	8
$P_{CALC}G/CM^3$	1.12	0.871	1.084	1.293
M/MM⁻¹	0.148	0.085	0.26	0.09
F(000)	4968	17856	3996	14832
CRYSTAL SIZE/MM³	0.23 × 0.18 × 0.15	0.075 × 0.055 × 0.01	0.01 × 0.008 × 0.008	0.299 × 0.034 × 0.006
RADIATION	synchrotron (λ = 0.6889)	synchrotron (λ = 0.6889)	synchrotron (λ = 0.6889)	synchrotron (λ = 0.6889)
2Θ RANGE FOR DATA COLLECTION/°	2.924 to 40.296	2.92 to 54.646	2.576 to 33.35	2.13 to 40.296
INDEX RANGES	-21 ≤ h ≤ 21, -27 ≤ k ≤ 27, -27 ≤ l ≤ 27	-68 ≤ h ≤ 68, -28 ≤ k ≤ 28, -75 ≤ l ≤ 75	-16 ≤ h ≤ 16, -16 ≤ k ≤ 16, -27 ≤ l ≤ 27	-39 ≤ h ≤ 39, -27 ≤ k ≤ 27, -35 ≤ l ≤ 35
REFLECTIONS COLLECTED	115133	492725	61628	62103
INDEPENDENT REFLECTIONS	29012 [R_{int} = 0.0614, R_{sigma} = 0.0715]	157670 [R_{int} = 0.0728, R_{sigma} = 0.0647]	2332 [R_{int} = 0.0950, R_{sigma} = 0.0603]	18806 [R_{int} = 0.0693, R_{sigma} = 0.0677]
DATA/RESTRAINTS/PARAMETERS	29012/221/2714	157670/565/5678	2332/261/217	18806/477/1703
GOODNESS-OF-FIT ON F²	1.079	0.947	1.203	1.085
FINAL R INDEXES [I>=2σ (I)]	R_1 = 0.0867, wR_2 = 0.2694	R_1 = 0.0661, wR_2 = 0.1894	R_1 = 0.1451, wR_2 = 0.3485	R_1 = 0.1539, wR_2 = 0.4287
FINAL R INDEXES [ALL DATA]	R_1 = 0.1161, wR_2 = 0.2935	R_1 = 0.0970, wR_2 = 0.2115	R_1 = 0.1987, wR_2 = 0.3888	R_1 = 0.1886, wR_2 = 0.4705
LARGEST DIFF. PEAK/HOLE / E Å⁻³	1.08/-0.67	0.43/-0.27	0.46/-0.45	1.48/-0.87
FLACK PARAMETER		0.15(2)	-0.1(10)	
CSD DEPOSITION CODE	2343766	2343777	2343792	2343780

Figure S4: Crystal structure obtained from 1,4-dioxane and 1-octanol (**Dioxane-octanol-C9**). The asymmetric unit consists of a single cage **1** molecule with one amide carbonyl on the top face and one amide carbonyl on the bottom face pointing into the cavity, which corresponds to predicted conformation **C9** (**a**). Packing of cage molecules is shown along the crystallographic *c* axis (**b**). Ellipsoids are displayed at 50% probability. Solvent molecules were not explicitly modelled due to disorder.
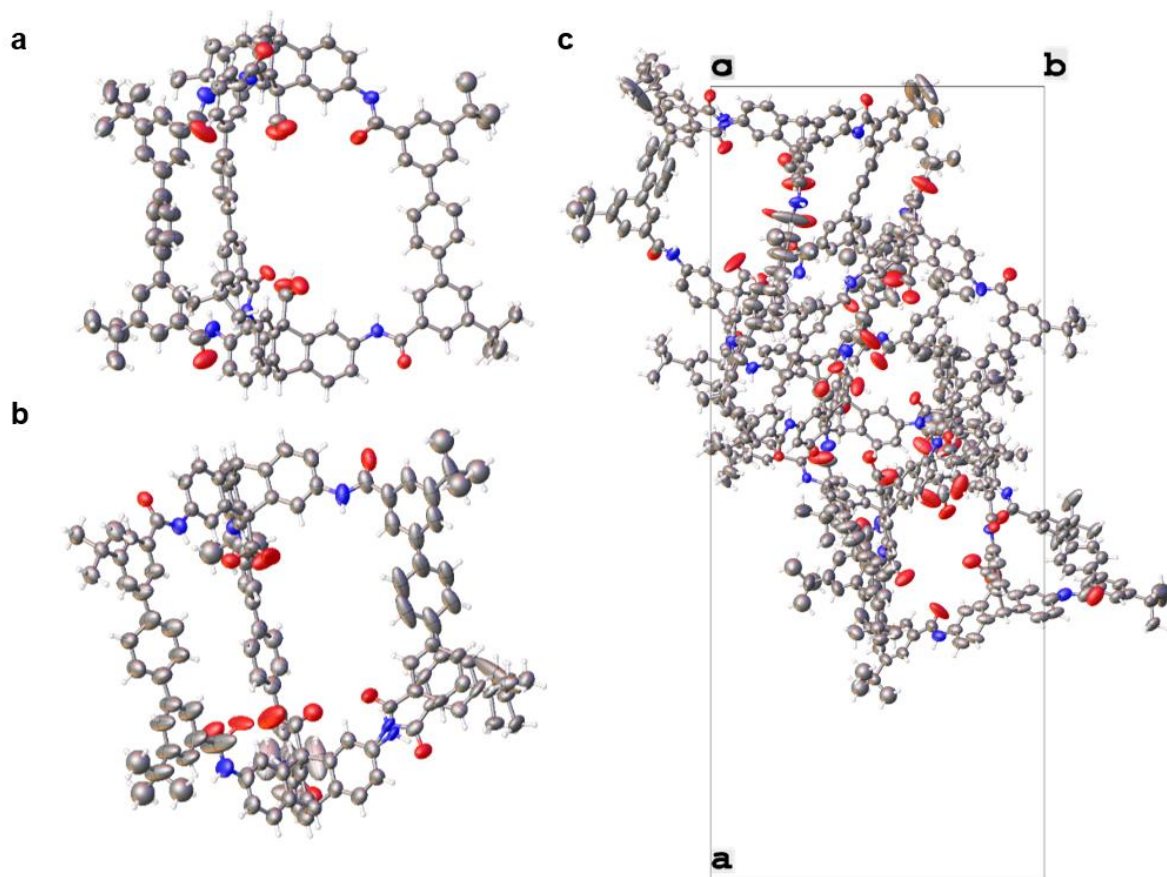
Figure S5: Crystal structure obtained from DMSO and MeOH (**DMSO-MeOH-C5-C9**). The asymmetric unit consists of four cage **1** molecules. In three of these cage molecules, two amide carbonyls on the top face and one on the bottom face of the cage are pointing into the cavity, corresponding to the predicted conformer **C5** (**a**). The fourth cage molecule has one amide carbonyl on each face pointing into the cavity, and a third carbonyl that is disordered so that it points into the cavity with an occupancy of 0.48, and out of the cavity with an occupancy of 0.52, which changes the arrangement from **C5** to **C9** respectively (**b**). Packing of the cage molecules is shown along the crystallographic *c* axis (**c**). Ellipsoids displayed at 50% probability. Solvent molecules have been omitted for clarity.
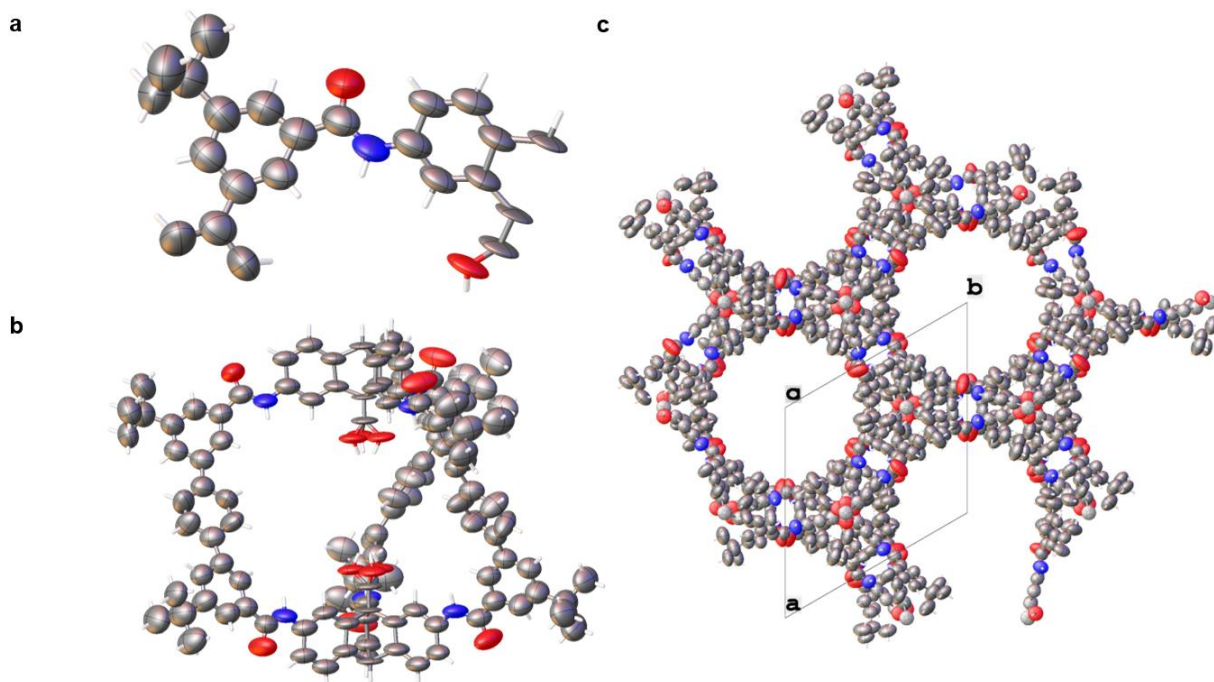
Figure S6: Crystal structure obtained from DMSO and MeOH (**DMSO-MeOH-C13**). The asymmetric unit (**a**) consists of one sixth of a single cage molecule (z' = 6) where all amide carbonyls are pointing out of the cavity, corresponding to the predicted conformer **C13** (**b**). The cage molecules pack hexagonally in the extended structure (**c**). Ellipsoids displayed at 50% probability. Solvent molecules were not explicitly modelled due to disorder.
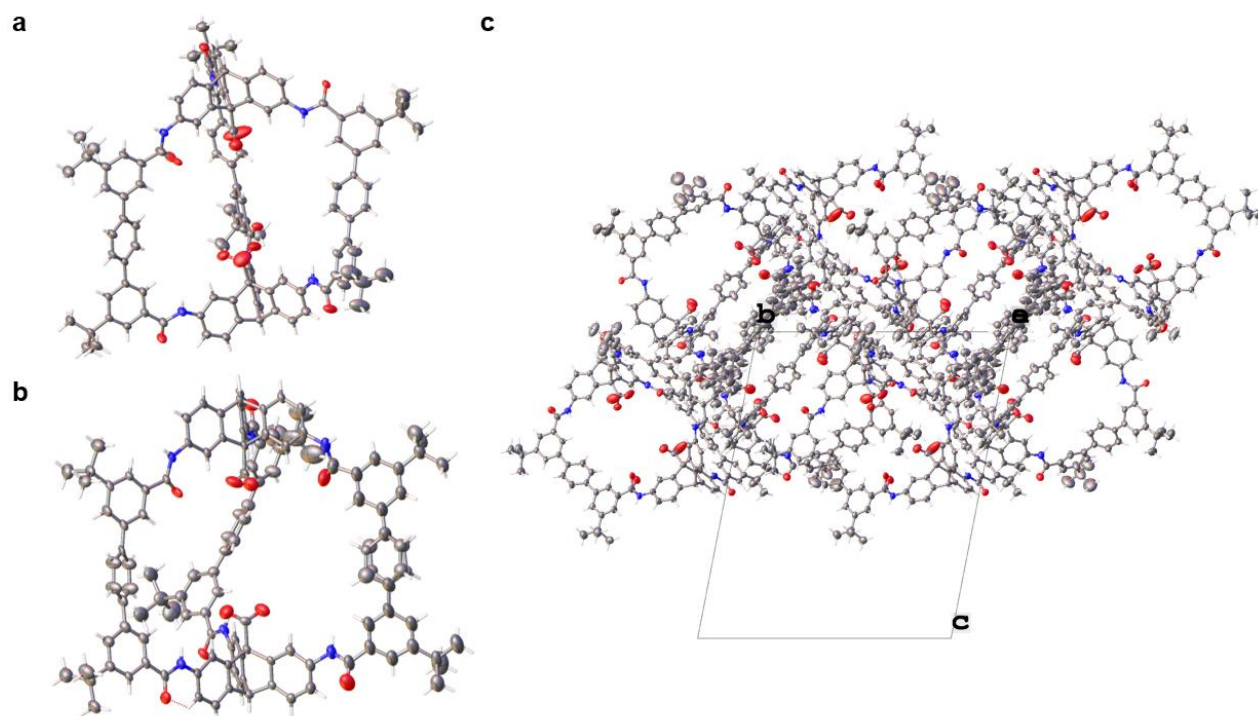

Figure S7: Crystal structure obtained from DMSO and MeCN (**DMSO-MeCN-C9-C10**). The asymmetric unit consists of two cage **1** molecules, one in the **C9** conformation (**a**) and one in the **C10** (**b**) conformation, where two amide carbonyls on the top face are pointing into the cavity. Packing of cage molecules is shown along the crystallographic *a* axis (**c**). Ellipsoids displayed at 50% probability. Explicitly modelled solvent molecules were omitted for clarity.
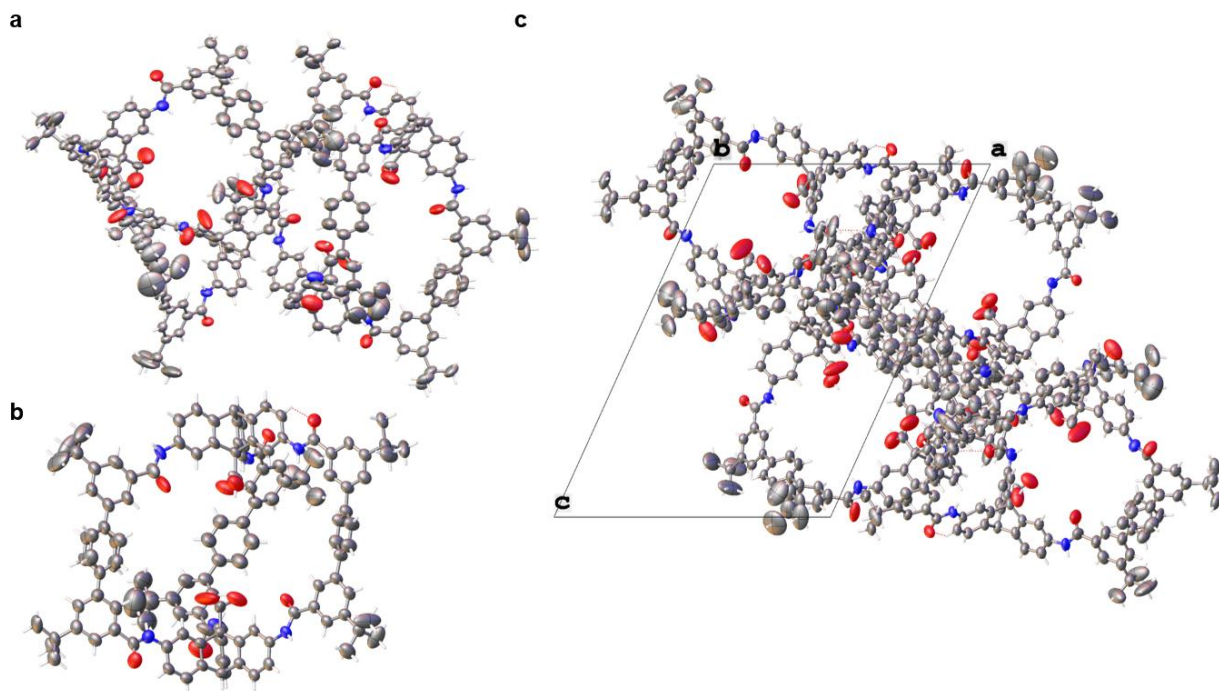
12

Figure S8: Crystal structure obtained from DMF and TFE (**DMF-TFE-C9**). The asymmetric unit (**a**) consists of two cage **1** molecules in the **C9** conformation (**b**). Packing of the cage molecules is shown along the crystallographic *b* axis (**c**). Ellipsoids displayed at 50% probability. Solvent molecules were not explicitly modelled due to disorder.

Figure S9: Crystal structure obtained from DMF and THP (**DMF-THP-C10-C12**). The asymmetric unit (**a**) consists of two cage **1** molecules, one in the **C12** conformation, where a single amide carbonyl is pointing into the cavity (**b**), and one in the **C10** conformation (**c**). Packing of the cage molecules is shown along the crystallographic *b* axis (**d**). Ellipsoids displayed at 50% probability. Solvent molecules were not explicitly modelled due to disorder.
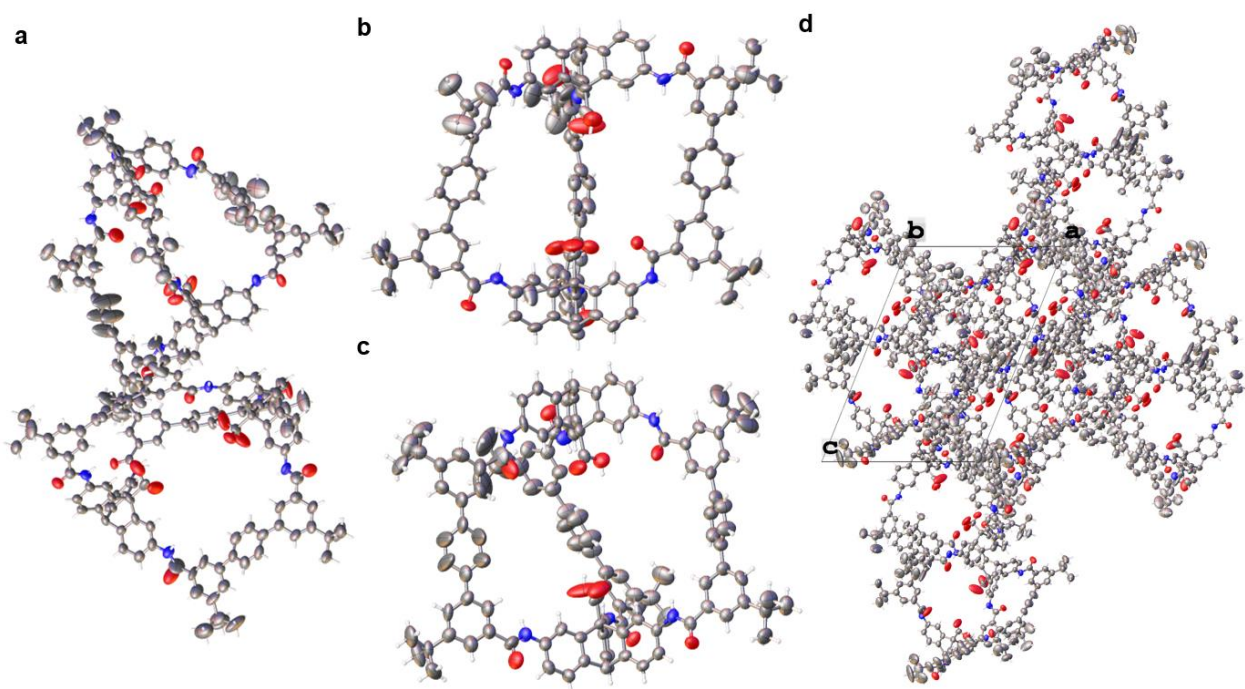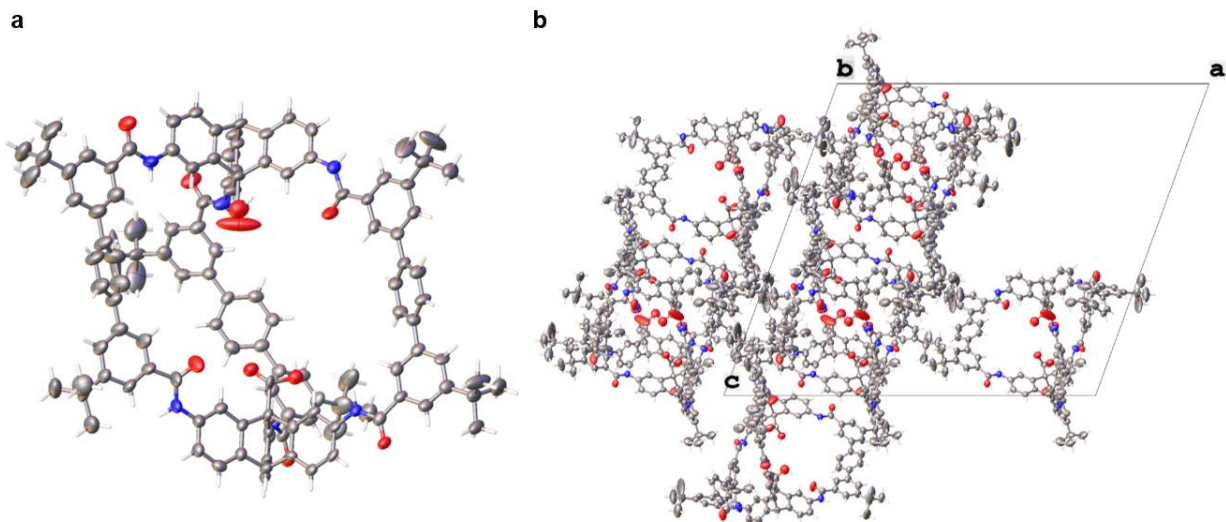
Figure S10: Crystal structure obtained from 1,3-dioxolane and propylene carbonate (**Dioxolane-PC-C9**). The asymmetric unit consists of a single cage **1** molecule in the **C9** conformation (**a**). Packing of the cage molecules is shown along the crystallographic *b* axis (**b**). Ellipsoids displayed at 50% probability. Solvent molecules have been omitted for clarity.
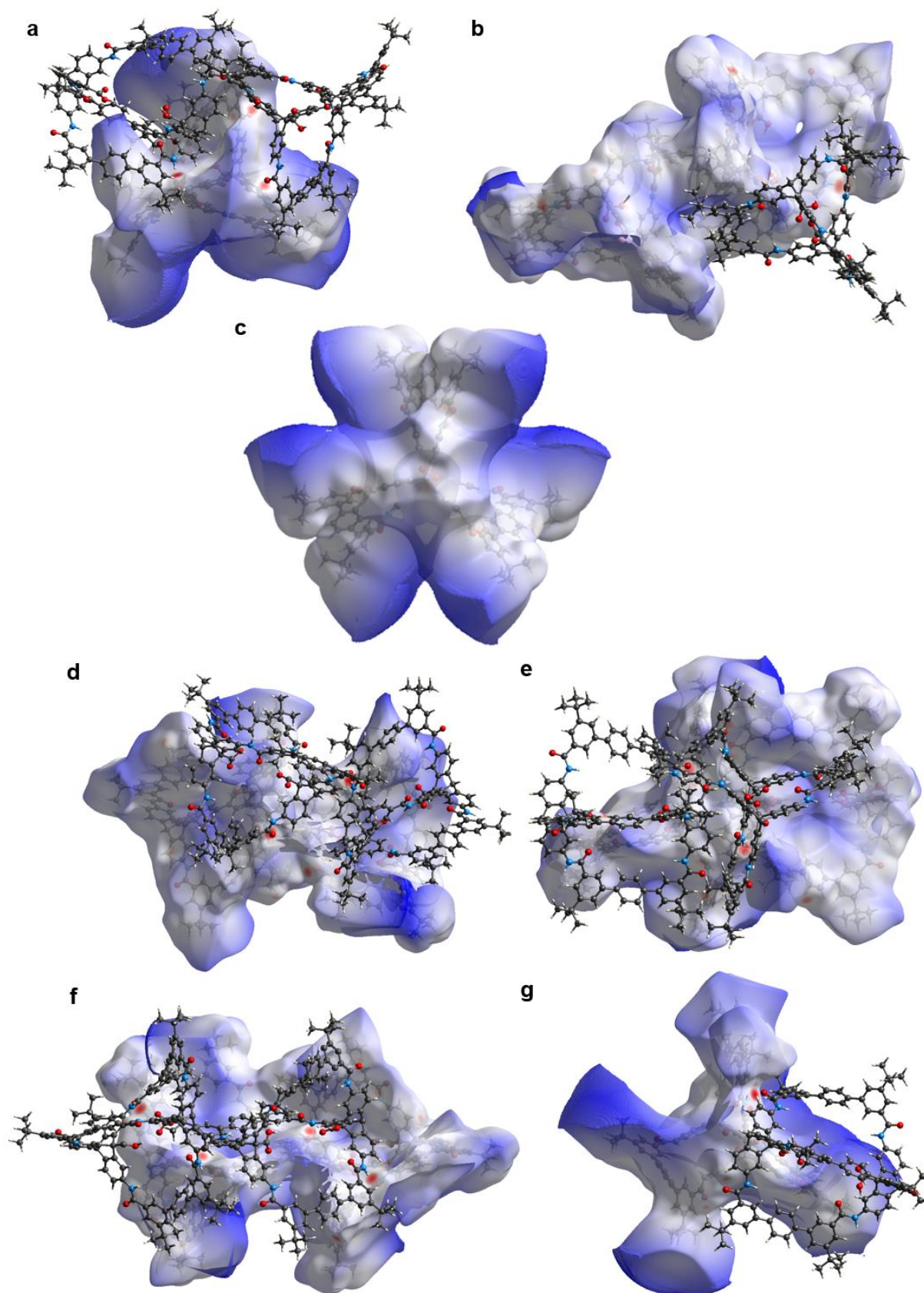
Figure S11: Hirshfeld surface analysis for **Dioxane-octanol-C9 (a)**, **DMSO-MeOH-C5-C9 (b)**, **DMSO-MeOH-C13 (c)**, **DMSO-MeCN-C9-C10 (d)**, **DMF-TFE-C9 (e)**, **DMF-THP-C10-C12 (f)**, and **Dioxolane-PC-C9 (g)**, showing intermolecular interactions between neighbouring cage **1** molecules.

3. Computational methodology

3.1. Conformer search

CREST was used with GFN2-xTB semi-empirical functional and the built-in analytical linearized Poisson-Boltzmann (ALPB) solvation model for tetrahydrofuran to perform conformational searches starting from CCDC structure XIVMAU (`original_structure.xyz`).[7,8] Energy window of 12 kcal mol^{-1} was used for conformer structures and no genetic crossing was used, resulting in around 6000 unique structures being generated. One of the thirteen pre-defined amide configurations was assigned to each structure based on the comparison of the distances from the centre of mass of the structure to the positions of the amide oxygen and nitrogen atoms. Conformer **C1** was above the energy threshold and thus was not found probably due to high strain in the structure. The lowest energy structure for each conformer was optimised in Orca 5.0.4. using the B97-3c functional with tight optimisation criteria and the universal solvent model based on density (SMD) for implicit treatment of chloroform (see `B97-3c.inp` for an example Orca input file).[9,10] Resulting structures (see files `C2.xyz` to `C13.xyz` in `B97-3c structures`) were used for further calculations.

3.2. Single-point calculations

Final energetics were obtained with single point energies at a DFT level of theory in Orca 5.0.4. For all reported calculations, we compared def2-TZVP, def2-TZVPP and def2-QZVP basis sets with the def2/J auxiliary basis.[11,12] Results were compared between the PBE,[13] PBE0,[14] B3LYP,[15] M06-2X,[16] ωB97M-V,[17] and ωB97X-D3.[18] The calculations used atom-pairwise dispersion correction with the Becke-Johnson damping scheme (D3ZERO for M06-2X and D3BJ for other functionals),[19] or using the non-local VV10 correction (in the case of ωB97M-V).[10] Final energies were calculated with SMD tetrahydrofuran solvation and are reported in Table S4 (see `PBE0-def2-QZVP.inp` for an example Orca input file).

Table S4: Summary of the single-point calculations. Energies are quoted in kJ mol^{-1} relative to the lowest energy identified for each method (combination of a functional and a basis set).

Conformer ID	Relative single-point energy (SMD=THF) / kJ mol^{-1}								
	PBE-D3(BJ)			PBE0-D3(BJ)			B3LYP-D3(BJ)		
	TZVP	TZVPP	QZVP	TZVP	TZVPP	QZVP	TZVP	TZVPP	QZVP
2	25.86	25.80	25.81	25.54	25.49	25.96	27.34	27.29	27.87
3	8.84	8.84	8.93	8.41	8.38	8.86	9.52	9.53	10.11
4	20.00	19.94	19.79	19.52	19.48	19.75	20.76	20.73	21.10
5	0.00	0.00	0.25	0.44	0.40	1.08	1.29	1.36	2.02
6	24.04	23.79	23.64	23.64	23.41	23.70	25.27	25.07	25.38
7	34.61	34.52	34.00	34.35	34.32	34.38	35.72	35.62	35.75
8	17.08	17.11	16.65	16.35	16.42	16.52	16.97	16.99	17.17
9	0.57	0.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	5.68	5.66	5.48	4.99	4.95	5.25	5.89	5.90	6.21
11	17.32	17.43	16.59	16.06	16.19	16.00	16.27	16.35	16.11
12	1.66	1.74	1.02	0.48	0.59	0.49	0.40	0.50	0.33
13	4.31	4.48	2.92	2.51	2.79	2.07	1.00	1.17	0.42

Conformer ID	M06-2X-D3(0)			ωB97X-D3			ωB97M-V		
	TZVP	TZVPP	QZVP	TZVP	TZVPP	QZVP	TZVP	TZVPP	QZVP
2	28.02	27.43	27.47	33.49	33.18	34.49	29.10	29.08	30.36
3	12.73	12.03	11.74	17.03	16.73	18.03	13.95	13.92	15.20
4	20.95	20.39	20.36	26.82	26.53	27.63	23.27	23.25	24.34
5	6.27	5.50	5.50	11.29	11.01	12.41	9.18	9.19	10.51
6	25.58	24.67	24.41	32.10	31.61	32.68	28.92	28.74	29.76
7	35.61	35.12	34.97	39.37	39.08	40.01	35.42	35.35	36.29
8	17.19	16.74	16.68	20.94	20.74	21.58	18.34	18.34	19.21
9	2.42	1.82	1.62	5.60	5.33	6.11	4.26	4.20	4.88
10	6.88	6.21	6.16	11.99	11.68	12.71	9.92	9.84	10.94
11	16.56	16.23	16.15	18.64	18.49	19.08	16.35	16.35	16.99
12	1.84	1.43	0.96	4.42	4.26	4.89	3.18	3.17	3.81
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

3.3. Relaxed coordinate scans

Each identified conformer optimised using B97-3c was further optimised at the PBE/def2-TZVP level of theory (see files `C2.xyz` to `C13.xyz` in `PBE-def2-TZVP structures`). All energies (in Hartrees) and acid-acid distances are reported in `opt-summary.json` and summarised here in Table S5.

Table S5: Summary of the acid-acid distances calculated with structures optimised with PBE/def2-TZVP.

Conformer ID	2	3	4	5	6	7	8	9	10	11	12	13
Acid-acid distance / Å	9.88	9.57	9.67	9.33	9.41	9.63	9.21	8.84	8.87	8.68	8.33	7.73

We then performed a relaxed coordinate scan using at the PBE/de2-TZVP level of theory. Based on the single-point energy calculations, we concluded that a triple-zeta functional is necessary for reliable energetics and decided to use the reliable generalised-gradient approximation PBE functional for accessible computational cost while maintaining reasonable chemical accuracy (i.e., the previously observed conformer **C9** identified as the lowest energy conformer when using the def2-QZVP basis set that was assumed to be complete). Potential energy scans were performed for acid-acid separations ranging from 6 to 12 Å in 0.25 Å increments. See `coordinate-scan-down.inp` and `coordinate-scan-up.inp` for example Orca input files for the coordinate scans.

Only separations between 7 to 10 Å were successfully optimised for all conformers and hence they are reported in the manuscript. Energies (in Hartrees) for all acid-acid separations that successfully optimised for the entire range of 6 to 12 Å for each conformer are reported in `scan-summary.json`.

4. References

1   G. M. Sheldrick, *Acta Cryst. Sect. Found. Adv.*, 2015, **71**, 3–8.
2   G. M. Sheldrick, *Acta Cryst. Sect. C Struct. Chem.*, 2015, **71**, 3–8.
3   O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. a. K. Howard and H. Puschmann, *J. Appl. Cryst.*, 2009, **42**, 339–341.
4   A. L. Spek, *Acta Cryst. Sect. C Struct. Chem.*, 2015, **71**, 9–18.
5   A. L. Spek, *J. Appl. Cryst.*, 2003, **36**, 7–13.
6   K. G. Andrews and K. E. Christensen, *Chem. – Eur. J.*, 2023, **29**, e202300063.
7   P. Pracht, F. Bohle and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 7169–7192.
8   C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
9   F. Neese, *WIREs Comput. Mol. Sci.*, 2022, **12**, e1606.
10 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
11 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
12 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.
13 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
14 J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.
15 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
16 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
17 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2016, **144**, 214110.
18 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
19 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.