Systematic construction of progressively larger capsules from a 5-fold linking subcomponent

Kai Wu[1], Tanya K. Ronson[1], Pingru Su[3], Zhi Chen[3], Leonard Goh[1], Andrew W. Heard[1,2], Xiaopeng Li[3], Fabian Klautzsch[4], Christoph A. Schalley[4], Mladen Vinković[2] & Jonathan R. Nitschke[1]*

[1]Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.

[2]Astex Pharmaceuticals, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, U.K.

[3]College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen, Guangdong 518055, China.

[4]Institut für Chemie und Biochemie der Freien Universität Berlin, Arnimallee 20, 14195 Berlin, Germany.

*e-mail: jrn34@cam.ac.uk

Biological encapsulants, such as viral capsids and ferritin protein cages, use many identical subunits to tile the surface of a polyhedron. Inspired by these natural systems, synthetic chemists have prepared an extensive series of artificial nanocages, with well-defined shapes and cavities. Rational control over the self-assembly of discrete, nanometre-scale, hollow coordination cages composed of simple components still poses considerable challenges as a result of the entropic costs associated with binding many subunits together, difficulties in the error-correction processes associated with assembly, and increasing surface energy as their size grows. Here we demonstrate the construction of a family of nanocages of increasing size derived from a single simple pentatopic pyrrole-based subcomponent. Reasoned shifts in the preferred coordination number of the metal ions employed, along with the denticity and steric hindrance of the ligands, enabled the generation of progressively larger cages, incorporating more subunits. These structural changes of the cages through these 'mutations' are reminiscent of differences in the folding of proteins caused by minor variations in their amino acid sequences; understanding how they impact capsule structure and thus cavity size may help to elucidate construction principles for still larger, more complex and functional capsules, capable of binding and carrying large biomolecules as cargoes.

Self-assembly is a powerful tool for the construction of complex, functional architectures from simple subunits in both biological and synthetic contexts. Self-assembled viral capsids, for example, enclose and protect the genetic material until its delivery. Instead of coding for a

single protein large enough to serve as a capsid, multiple copies of smaller subunits spontaneously self-assemble into capsids, which vary in size according to the cargo. *Pseudo*-spherical cavities with icosahedral symmetry[1,2] have the largest volume-to-surface-area ratio, thus minimizing the amount of protein required, following the principle of genetic economy[3]. Capsids with icosahedral symmetry are often described by a "triangulation number" ***T***. Evolution has not only optimised the symmetry of the capsid, it also developed capsids with increasing ***T*** numbers that reflect progressively larger cavity sizes, to enclose larger and more complex viral genomes (Fig. 1a). The quasi-equivalence principle[4] set forth by Caspar and Klug laid the foundation for understanding how viruses build increasingly larger capsids. While small capsids can maintain a high symmetry, larger capsids require the subunits to interact in a deformative manner. These subunits are no longer fully identical, with pentagonal and hexagonal subunits giving Goldberg structures[5] that are closer to being truly spherical. Capsids with 12 fixed pentagonal subunits as vertices thus increase the number of hexagonal subunits between them to expand their sizes. Mimicking icosahedral capsid assembly would require a complicated heteroleptic assembly strategy to adopt this quasi-equivalence. Instead, we develop an alternative strategy to assemble larger cages in a homoleptic way from a pentatopic subcomponent, by altering the key dihedral angle between the pentagonal faces of the cage (Fig. 1b). Systematic changes in this angle are analogous to 'mutations' that change the sizes of viral capsids, progressively forming larger synthetic capsules that parallel nature's containers.

Inspired by viral capsids and many other biological encapsulants, synthetic chemists have prepared artificial nanocages, with well-defined shapes and cavities[6–8]. Rational control over the self-assembly of discrete, large, hollow coordination cages[9] composed of simple components still poses considerable challenges. Large self-assembled polyhedral cages, with structures corresponding to Platonic and Archimedean solids, have been reported by Fujita[10,11], Stang[12], Kim[13], Ward[14], and other groups[15–18]. Building blocks with five-fold symmetry cannot pack regularly, as evidenced by the particular properties of Penrose tiling patterns in two dimensions (2D) and quasicrystal packing in three dimensions (3D). In contrast with the pentameric capsomeres that form the vertices of icosahedral viral capsids, synthetic building blocks with five-fold symmetry are rare[19,20] because of the synthetic difficulties in obtaining them. Few cages have been prepared using pentagonal building blocks, such as corannulene- or calixarene-based components[15,17], to define the five-fold symmetry axes of icosahedra or dodecahedra. As a simple pentatopic building block[21], the rigidity and planar geometry of the pyrrole backbone render it highly suitable for capping the pentagonal faces of large self-

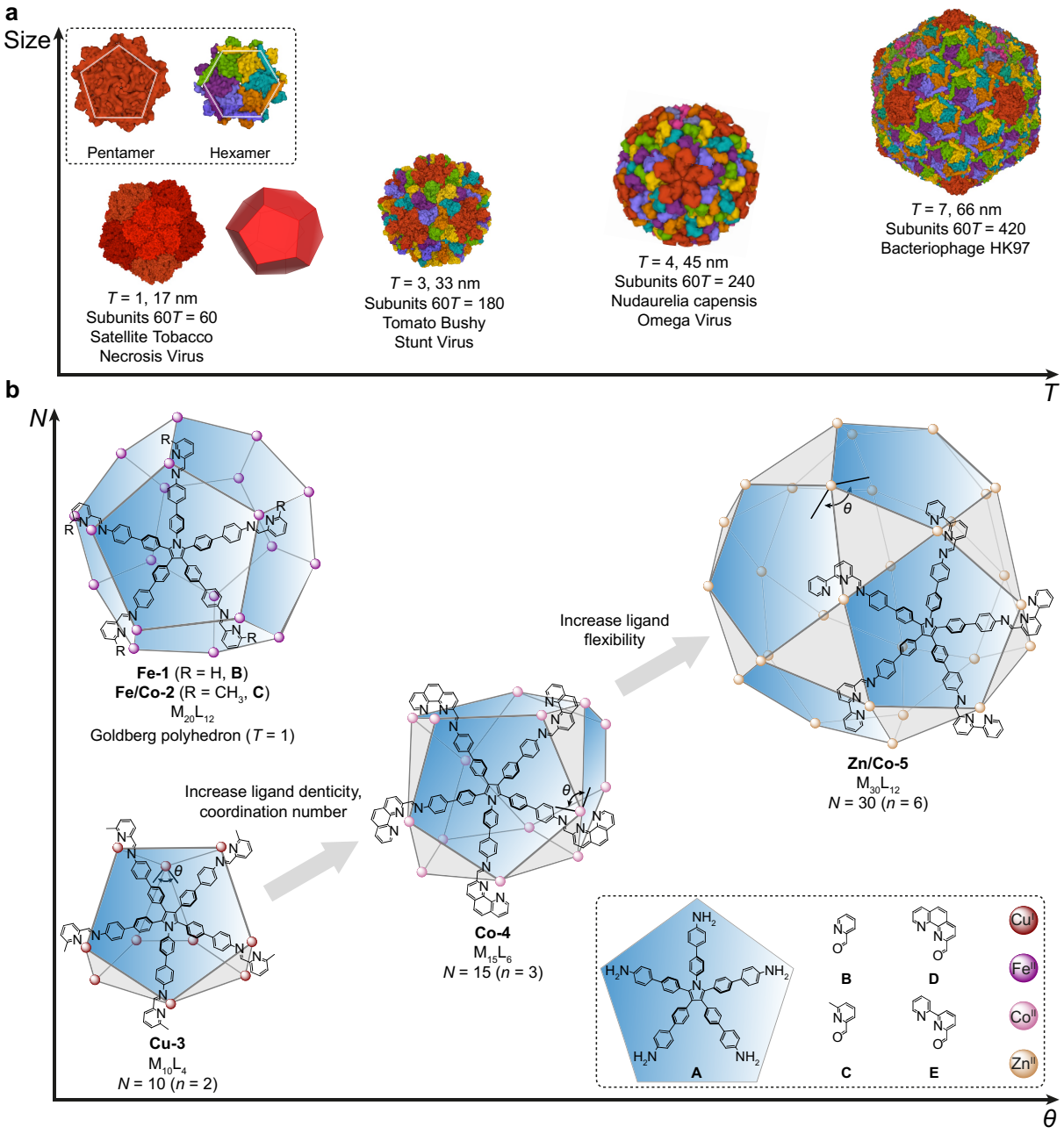


Fig. 1 | Schematic representation of the cages assembled from pentatopic pyrrole-based subcomponent A. a, Progressively larger viral capsids with icosahedral symmetry[24–27] are constructed from pentagonal and hexagonal subunits by increasing the "triangulation number" ***T***. Inset shows the pentamer and hexamer of virus with ***T*** = 7. **b,** Progressive increase in the number of metal vertices, i.e. the nuclearity (***N***), of the $M_{5n}L_{2n}$ cages assembled from subcomponent **A** as the dihedral angle between pentagonal faces at vertices (***θ***) increases; grey arrows highlight changes to the system that brought about an increase in both ***θ*** and ***N***. Dodecahedra **1** and **2** correspond to the Goldberg structure with ***T*** = 1. The inset box shows organic subcomponents **A**–**E** and the metal ions used to construct the cages shown.

assembled structures, we thus set out to explore the subcomponent self-assembly[22] of nanocages using pentatopic pyrrole-based building block **A** (Fig. 1b).

Pyrrole-centred pentakis(biphenyleneamine) subcomponent **A** was prepared through one step of Suzuki-Miyaura cross-coupling (Extended Data Fig. 1, Figs. S1–6). We envisaged that penta-amine **A** would condense with formylpyridines **B** or **C** to generate a pentakis(bidentate) ligand, which would combine with metal ions with an octahedral coordination geometry to form structures with $M_{5n}L_{3n}$ formulae, or metal ions with a tetrahedral coordination geometry, to form $M_{5n}L_{2n}$ structures. Alternatively, **A** could combine with subcomponents **D** or **E** to generate a pentakis(tridentate) ligand, which could further coordinate with metal ions with an octahedral coordination geometry to produce $M_{5n}L_{2n}$ structures. In all cases the composition would be driven by the principle of maximum site occupancy[23], whereby all metal ions are coordinatively saturated, and all ligand nitrogen atoms bound to a metal.

For structures with $M_{5n}L_{3n}$ formulae, the dodecahedron ($\boldsymbol{N} = 20$, $n = 4$, Fig. 1b) is the smallest of the Goldberg structure series. Unlike viruses, which increase capsid size by adding hexagonal subunits between pentagons while maintaining a constant dihedral angle between pentamers, here we develop an alternative strategy of increasing capsule size by modifying this key dihedral angle. For the $M_{5n}L_{2n}$ structures, as cage nuclearity ($\boldsymbol{N}$) – the number of metal centres per cage – increases, so do dihedral angles $\boldsymbol{\theta}$ between pentagonal faces sharing the same vertex, from the smaller truncated tetrahedron (**3** in Fig. 1b) to the largest icosidodecahedron (**5**), as shown schematically in Fig. 1b. In light of this correlation, we developed a distinct strategy to increase $\boldsymbol{\theta}$ in order to increase cage size of the new $M_{5n}L_{2n}$ structure series, for example by switching from tetrahedral to octahedral metal coordination, and by increasing steric hindrance, ligand denticity and flexibility (Fig. 1b), as detailed below.

Cage structures assembled from **A**, 2-formylpyridine **B** and first-row transition metals that favour an octahedral configuration should adopt an $M_{5n}L_{3n}$ formula ($\boldsymbol{N} = 20$, $n = 4$, Fig. 1b) with tris(pyridylimine) metal vertices. We hypothesised that the tighter coordination spheres of such metals would preclude the assembly of smaller structures with lower $\boldsymbol{\theta}$ values. Although the reaction of **A** and **B** with zinc(II) bis(trifluoromethanesulfonyl)imide (triflimide, NTf_2^-) or $Co(NTf_2)_2$ did not lead to isolable cages, the use of $Fe(NTf_2)_2$ resulted in the formation of an $Fe^{II}_{20}L_{12}$ dodecahedral cage, **Fe-1**, as confirmed by electrospray ionization-mass spectrometry (ESI-MS). The low intensities of the ESI-MS signals and very broad ^{1}H nuclear magnetic resonance (^{1}H NMR) signals precluded further analysis (Figs. S13–19).

There is also a mismatch between the geometry required to form a dodecahedron and the regular octahedral coordination sphere of low-spin Fe^{II}, which may destabilise **Fe-1**. As shown in Extended Data Fig.2, the optimal angles between pyridylimine ligands bound to low-spin

Fe^{II} are narrower than optimal for dodecahedral **Fe-1**, instead adapting readily to the smaller optimum angles of a tetrahedral or a cubic assembly. To stabilise the required geometry, we sought to modulate the second coordination sphere of Fe^{II} through the introduction of steric hindrance[28] by incorporating the bulky methyl group of subcomponent **C** (Fig. 1b). This steric bulk destabilises the low-spin state of iron(II)[29], instead favouring a high-spin complex with ^{1}H NMR signals over a wider frequency range. We infer that the more flexible coordination sphere of high-spin Fe^{II} might enable the metal vertex to open to the wider angles required by dodecahedral **Fe-1** with a lower enthalpic penalty than low-spin Fe^{II} would impose (Extended Data Fig. 2).

Indeed, the reaction of penta-amine **A** (12 equiv), methylformylpyridine **C** (60 equiv) and $Fe(NTf_2)_2$ (20 equiv) did produce a product with ^{1}H NMR and ESI-MS spectra corresponding to $Fe^{II}_{20}L_{12}$ dodecahedral cage **Fe-2** (Figs. S20, S23). Travelling wave ion mobility-mass spectrometry (TWIM-MS) for each charge state showed a single band, ruling out the formation of other isomers (Fig. S24). All ^{1}H NMR signals were assigned using T_1 relaxation time measurements, which correlated the distances between hydrogen atoms and paramagnetic Fe^{II} centres according to the Solomon equation[30] (Figs. S20, S22, Supplementary Table 1).

When $Co(NTf_2)_2$ was used instead of the iron salt, dodecahedral **Co-2** was formed. We infer **Co-2** and **Fe-2** to be isostructural on the basis of their NMR and ESI-MS characterisation data (Fig. 2, Figs. S25, S29), as expected due to the similar coordination properties[31] of high-spin Fe^{II} and Co^{II}. Although numerous attempts to obtain X-ray quality crystals of **Fe-2** were unsuccessful, **Co-2** proved more amenable. Its single-crystal X-ray structure is shown in Fig. 2d.

Twelve ligands span the faces of **Co-2**, with 20 Co^{II} centres serving as its vertices. The structure thus possesses pseudo-*I* symmetry, with rotationally disordered orientations of the pyrrole-N atoms. All Co^{II} centres within each cage have the same Δ or Λ handedness, while every pentaphenylene pyrrole core adopts a propeller-like configuration. Co^{II}–Co^{II} distances average 14.7 Å between neighbouring metal centres and 41.2 Å for the farthest-spaced metal centres. The overall diameter of the structure is ca. 5 nm, as measured by the distance between the outermost hydrogen atoms. The cavity volume was calculated to be 13132 $Å^3$ using the MoloVol program[32] (Fig. 4a, c), and the cavity is fairly enclosed, with a pore aperture of ca. 6 Å, compared to other edge-directed giant cages with large open pore sizes of ca. 11–22 Å[9,10,11,16].

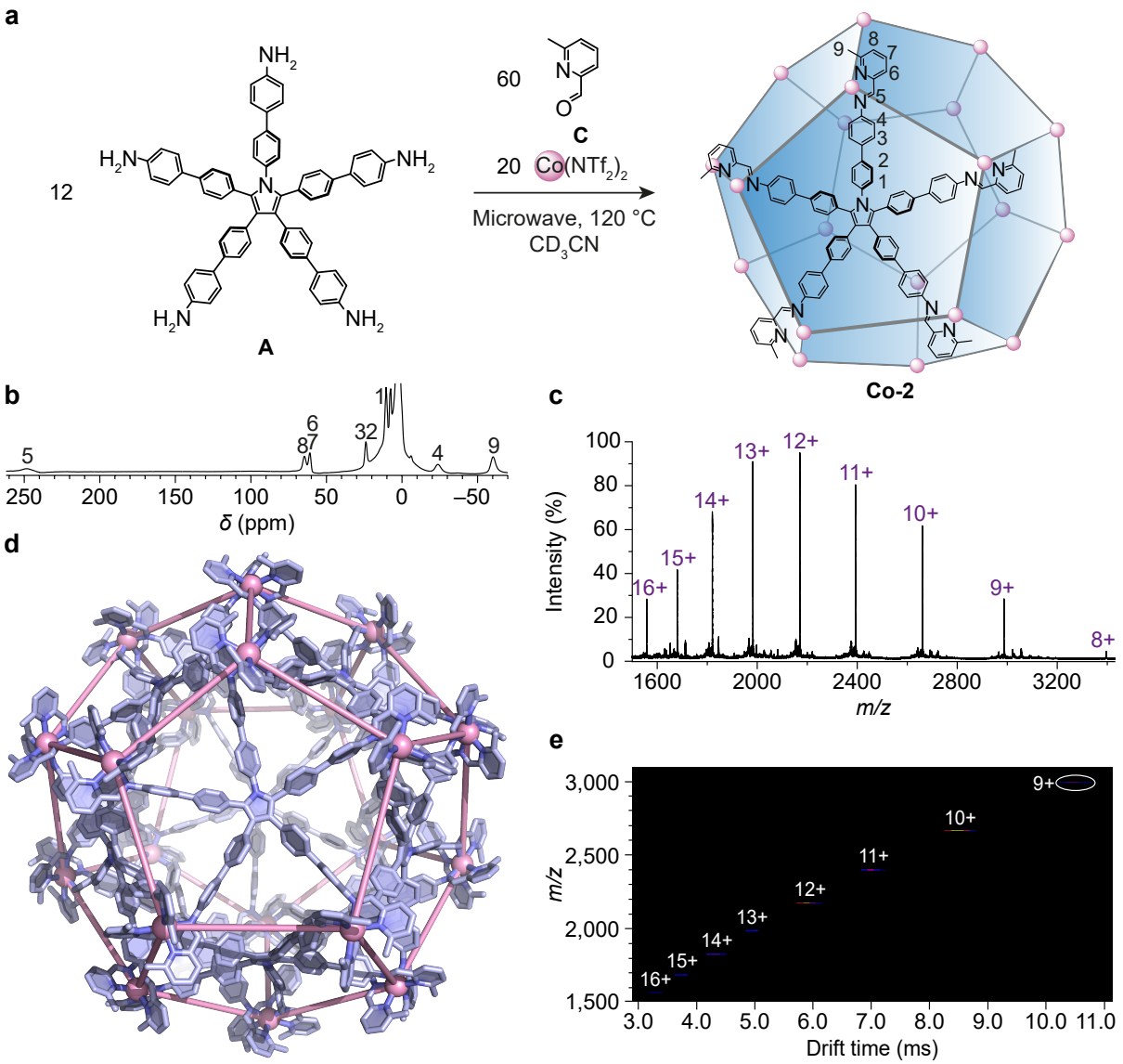


Fig. 2 | Self-assembly and characterisation of dodecahedral cage Co-2. **a,** Subcomponent self-assembly of **Co-2**. **b,** Wide sweep ^{1}H NMR spectrum (400 MHz, CD_3CN, 298 K) of **Co-2**. **c,** ESI-MS of **Co-2**. **d,** Crystal structure of **Co-2** shown in stick mode with Co^{II} in space-filling mode (The Λ_{20} enantiomer is shown. Hydrogens of the ligands, anions and cations are omitted for clarity). **e,** 2D ESI-TWIM-MS plot of **Co-2** (*m/z* vs. drift time).

Ultrahigh-vacuum, low-temperature scanning tunnelling microscopy (UHV-LT-STM) was also employed to characterise **Co-2**. Individual cage molecules were observed in STM images as bright spots with a width of ca. 5 nm (Fig. S28).

We reasoned that the use of Cu^{I}, which favours a tetrahedral configuration, might lead to the formation of a $Cu^{I}_{30}L_{12}$ icosidodecahedral structure with pentagonal faces derived from **A**. Instead, subcomponents **A** and 2-formyl-6-methylpyridine **C** reacted with $[Cu(MeCN)_4](BF_4)$ in DMSO-d_6 to form the $Cu^{I}_{10}L_4$ truncated tetrahedron **Cu-3** (***N*** = 10, *n* = 2, Fig. 1b), as

confirmed by ESI-MS (Fig. S37). We infer that this smaller assembly formed because tetrahedral Cu^{I} centres are able to tolerate a narrow θ angle between pentagon vertices, thus disfavouring the larger icosidodecahedron and avoiding the entropic cost of forming a larger structure.

In seeking a higher-nuclearity icosidodecahedral structure based upon **A**, we next turned to metal ions with an octahedral coordination geometry together with tridentate, instead of bidentate, ligand motifs to also target structure with the formulae $M_{5n}L_{2n}$ (Fig. 1b). To this end, 2-formyl-phenanthroline **D**, which is able to form tridentate ligand arms with penta-amine **A**, was employed in place of **B** for cage assembly. The assembly of **A** and **D** with $Co(NTf_2)_2$ resulted in a species **Co-4** with a composition of $Co^{II}_{15}L_6$ (Fig. 1b), as elucidated by ESI-MS (Fig. S43). TWIM-MS showed a single band for each charge state, indicating the formation of one discrete structure (Fig. S44). We infer the structure of **Co-4** to be a truncated rhombohedron, resembling a reported organic cage[21], with a θ value between pentagon vertices of ~80° (Fig. 4). **Co-4** is inferred to have approximate D_3 symmetry, giving rise to three types of symmetry-inequivalent Co^{II} vertices and five distinct ligand arm environments (Figs. S45–46). This lower symmetry results in a more complex ^{1}H NMR spectrum than **Co-3**, with pseudo-*I* symmetry.

Intriguingly, the reaction of penta-amine **A**, formylphenanthroline **D**, and $Zn(NTf_2)_2$ did not yield a discrete assembly. The Zn^{II}–N coordination bond strength may not be sufficient to compensate for either the strain necessarily incorporated into the irregular structure of a putative **Zn-4** framework, with its different metal–metal distances, or the entropic cost of forming larger, more regular structures that would lack this strain.

We thus sought next to add ligand flexibility by incorporating a more flexible subcomponent **E** and returning to the more strongly coordinating Co^{II}. We reasoned that the less flexible phenanthroline rings incorporated into **D** might favor the narrower θ values of **Co-4**, whereas more flexible bipyridines could enable expansion of these angles to the 117° required for an icosidodecahedron, whose more regular structure would eliminate any enthalpic penalty associated with the formation of the less regular **Co-4**.

This minor modification in subcomponent structure – the replacement of rigid subcomponent **D** with the more flexible **E** in the self-assembly reaction – thus resulted in the formation of a $Co^{II}_{30}L_{12}$ structure, **Co-5**. We infer **Co-5** (Fig. 1b) to have an icosidodecahedral framework, as designed.

Although the $Co^{II}_{30}L_{12}$ formulation of **Co-5** is clear from ESI-MS (Fig. S49), its broad and complex ^{1}H NMR spectra rendered further analysis very challenging. The broadness of these spectra may be a consequence of the presence of both Δ and Λ Co^{II} centres within **Co-5**, with the flexibility of **E** leading to minimal energetic differentiation between the two handednesses,

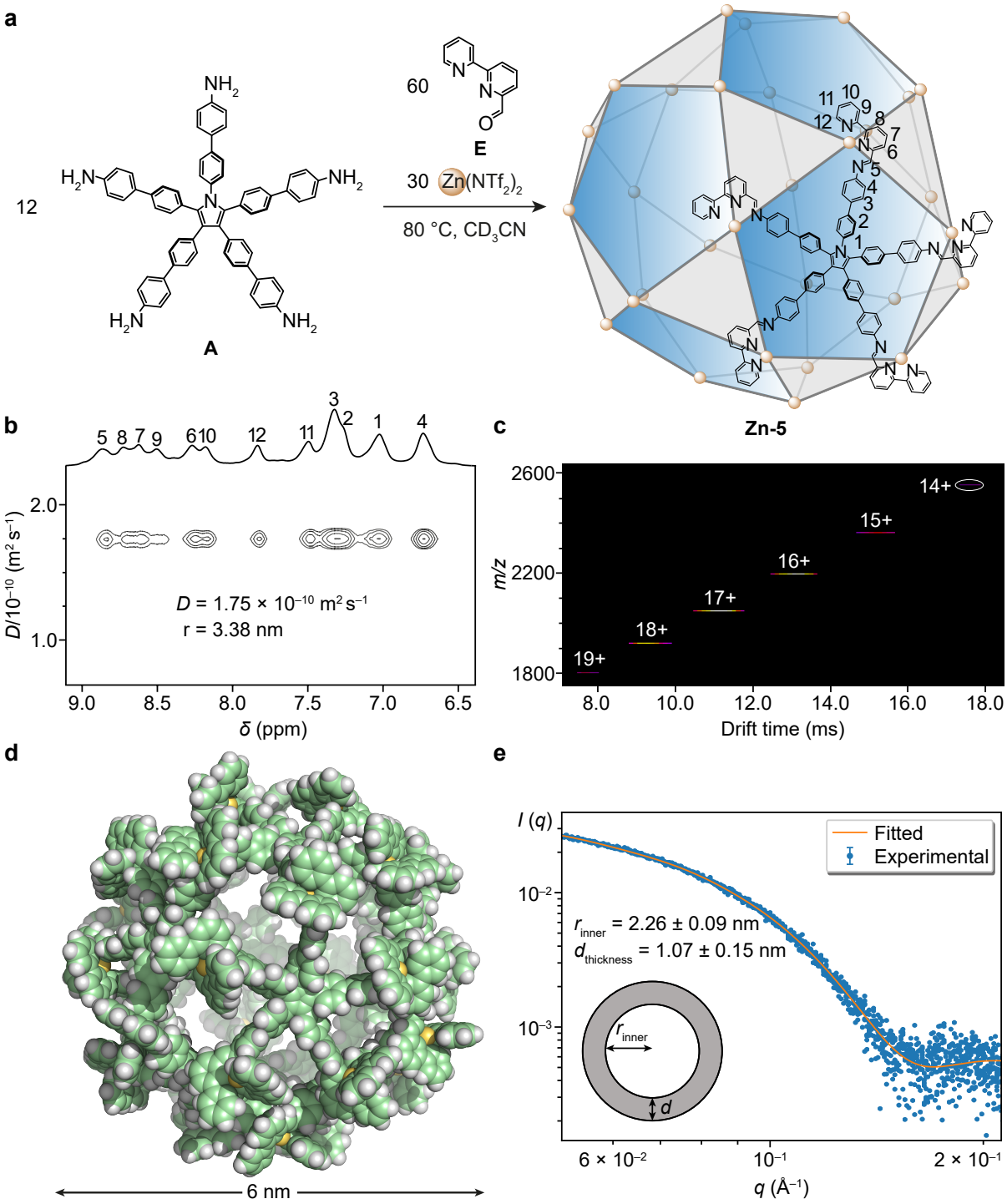


Fig. 3 | Schematic representation and characterisation of icosidodecahedral cage Zn-5. a, Subcomponent self-assembly of **Zn-5**. **b,** ^{1}H DOSY spectrum (400 MHz, CD_3CN, 298 K) of **Zn-5**, and **c,** 2D ESI-TWIM-MS plot of **Zn-5** (*m/z* vs. drift time), both consistent with a single product. **d,** Space-filling view of an MM3 model[33] of the all-Δ stereoisomer of **Zn-5**. Colour code: Zn^{II} yellow, C and N green, H white. **e,** SAXS profile of **Zn-5** (blue circles) and the fitted result (orange line) based on a core-shell sphere model. The error bars represent uncertainties in the scattering intensity.

thus favouring the formation of many diastereomers. We thus explored the use of other metal ions to optimise the formation of the icosidodecahedral cage framework **5**. The assembly of penta-amine **A** (12 equiv), flexible **E** (60 equiv), and $Zn(NTf_2)_2$ (30 equiv) in CD_3CN gave a relatively broad, yet well-resolved, ^1H NMR spectrum with one set of signals, indicating the formation of a discrete assembly (Fig. 3a, b). The broad ^1H NMR signals may reflect slow tumbling on the NMR relaxation time scale, as reported for other giant assemblies[9,10,11], and the Zn^{II} centres may also be racemizing at an intermediate rate on the NMR time scale. ESI-MS confirmed the exclusive formation of cage **Zn-5** (Fig. 1b) with a $Zn^{II}_{30}L_{12}$ composition (Fig. S58). TWIM-MS also showed narrow signals, consistent with the presence of a single product (Fig. 3c). ^1H DOSY provided further evidence for the icosidodecahedral assembly, showing only a single band for all signals, from which a solvodynamic radius of 3.38 nm could be calculated (Fig. 3b).

Despite numerous attempts, the diffraction of the single crystals of **Zn-5** obtained was not strong enough to obtain the crystal structure. Instead, an MM3 model[33] provided a view of **Zn-5** that was consistent with all of our experimental data (Fig. 3d). Twelve pentagonal ligands span the faces of the icosidodecahedron, leaving 20 triangular faces open. Each of the 30 Zn^{II} vertices is bound by two tridentate ligand arms, while the pentaphenylene pyrrole cores adopt propeller-like configurations, giving rise to pseudo-*I* symmetry with an overall diameter of ca. 6.0 nm (Fig. 3d).

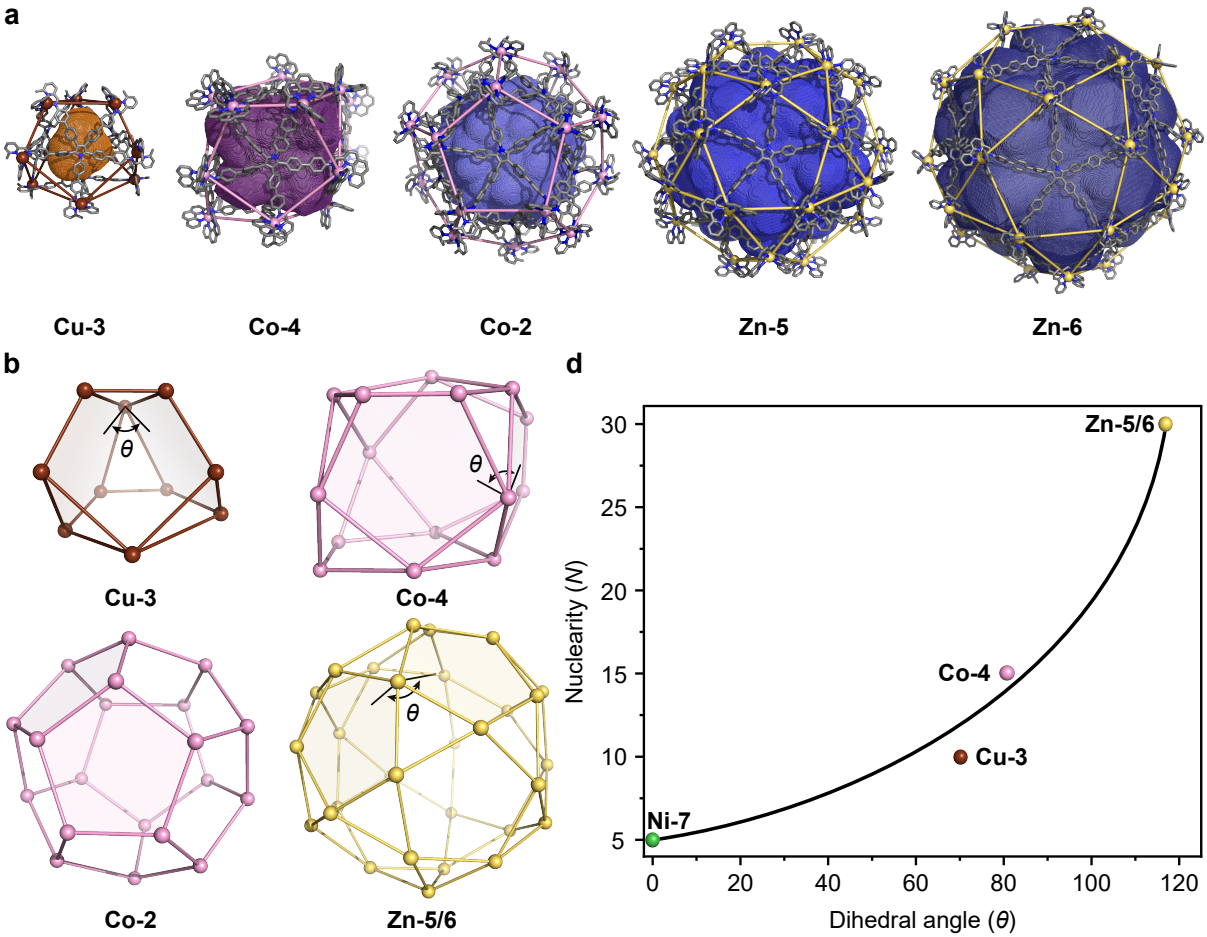
Further structural information for **Zn-5** in CD_3CN was obtained through small-angle X-ray scattering (SAXS) analysis. The SAXS profile fits well to the density function of a monodisperse core-shell sphere model, corresponding to a hollow-sphere cage with an inner radius of 2.26 nm and a shell thickness of 1.07 nm (Fig. 3e), consistent with both the ^1H DOSY measurement and the model. The cavity volume of our model of **Zn-5** was calculated to be 36415 $Å^3$ using the MoloVol program[32] (Fig. 4a, c).

To demonstrate the possibility of further enlarging the internal cavities of these cages, we synthesised a new subcomponent **A'** by introducing one additional phenylene to each arm of pentatopic component (Figs. S7–12). Self-assembly of **A'** with **E** and Zn^{II} produced a $Zn^{II}_{30}L_{12}$ cage, **Zn-6**, isostructural to **Zn-5**, confirmed by ESI-MS and DOSY NMR (Figs. S69–70). The larger solvodynamic radius of **Zn-6**, 3.63 nm from ^1H DOSY, led to a much greater cavity volume of 92173 $Å^3$ (Fig. 4c), calculated using MoloVol[32], than in the case of **Zn-5**. **Zn-6** thus represents the largest architecture prepared to date through subcomponent self-assembly, coming together from 102 discrete organic subcomponents and metal ions in a precisely defined fashion. Its formation illustrates the use of rational design principles to push the boundaries of large cage assembly.

Nickel(II), with its relatively small ionic radius and less adaptable coordination sphere, was also screened for cage assembly with penta-amine **A** and **D** or **E**. Although no discrete assembly was obtained with rigid formylphenanthroline **D**, the more flexible formylbipyridine **E** led to the formation of the small $Ni^{II}_5L_2$ sandwich structure **Ni-7** (Extended Data Fig. 3), confirmed by ESI-MS (Figs. S73–74).

The cavity sizes of models (X-ray structure in the case of **Co-2**) of the cages obeying the $M_{5n}L_{2n}$ or $M_{5n}L_{3n}$ formulae, with metal nuclearities ***N*** ranging from 10 to 30, were calculated using MoloVol and are visualised in Fig. 4a. To gain further insight into the factors that govern these diverse structural outcomes, we considered the geometries of these architectures. The angles $\boldsymbol{\theta}$ between pentagonal faces of each idealised polyhedron were calculated, as shown in Fig. 4d. **Ni-7** (n = 1), with a sandwich structure, has the smallest $\boldsymbol{\theta}$ of 0°, whereas truncated tetrahedron **Cu-3** (n = 2), with more open vertices, has the same $\boldsymbol{\theta}$ of 70° as a regular tetrahedron. For the larger structures based upon hexacoordinate metal ions, $\boldsymbol{\theta}$ increases progressively to 117°, as ***N*** increases from 15 for **Co-4** (n = 3) through to 30 for the largest icosidodecahedron (**5/6**) (n = 6).

Progressive design alterations thus allowed for the increase of $\boldsymbol{\theta}$, together with the cage nuclearity and size, along the series of polyhedral structures shown in Fig. 1b. This series culminates in a remarkably large icosidodecahedron **Zn-6**, having an enclosed volume greater than 92 nm^3, representing the largest ligand-enclosed inner cavity volume prepared to date; larger reported cages have more open ligand frameworks. The formation of this diverse array of well-defined large cages marks the five-fold pyrrole backbone as a versatile building block for subcomponent self-assembly. The voluminous internal cavities of these cages may be able to serve as a platform for the binding of large biomolecules, such as hydrophobic membrane proteins or proteases *via* hydrophobic effects, charge complementarity, hydrogen bonding and electrostatics. These specific interactions may be designed in through modification of either the aldehyde subcomponent or pyrrole backbone. The higher degree of cooperativity linking these assemblies together[34], associated with the five-fold connectivity of subcomponent **A**, may lend them greater robustness compared to structures built from two-fold, three-fold and four-fold subcomponents.



c

$M_{5n}L_{2n}/M_{5n}L_{3n}$	M_5L_2	$M_{10}L_4$	$M_{15}L_6$	$^*M_{20}L_{12}$	$M_{30}L_{12}$	$M_{30}L_{12}$
Cage framework	**Ni-7**	**Cu-3**	**Co-4**	**Co-2**	**Zn-5**	**Zn-6**
Metal source	Ni^{II}	Cu^{I}	Co^{II}	Fe^{II}/Co^{II}	Co^{II}/Zn^{II}	Zn^{II}
Aldehyde	**E**	**C**	**D**	**C**	**E**	**E**
Nuclearity (*N*)	5	10	15	20	30	30
Dihedral angle (*θ*) /°	0	70.5	80.2	-	116.6	116.6
Cavity volume/Å^3	-	1620	8508	13132	36415	92173

Fig. 4 | Comparison of cage cavity volumes and geometrical analysis of their frameworks. **a,** Increase in cavity volume as the nuclearity (*N*) of the cages increases. **b,** Stick view of the frameworks of the different cages, with the dihedral angles between pentagonal faces (*θ*) of the polyhedral faces shown for the $M_{5n}L_{2n}$ structure series; the *θ* at the edges of **Co-4** is mathematically equivalent to the *θ* at the vertices of **Zn-5/6**. **c,** Summary of parameters of the polyhedral cages ordered by *N*. **d,** Trendline illustrating the relationship between *θ* and *N*. *denotes the dodecahedral topology of **2** that obeys an $M_{5n}L_{3n}$ formula. Self-assembly of **Ni-7** is shown in Extended Data Fig. 3.

References and Notes:

1. Li, S., Roy, P., Travesset, A. & Zandi, R. Why large icosahedral viruses need scaffolding proteins. *Proc. Natl. Acad. Sci. USA* **115**, 10971–10976 (2018).
2. Zandi, R., Reguera, D., Bruinsma, R. F., Gelbart, W. M. & Rudnick, J. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. USA* **101**, 15556–15560 (2004).
3. Crick, F. H. C. & Watson, J. D. Structure of small viruses. *Nature* **177**, 473–475 (1956).
4. Caspar, D. L. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
5. Goldberg, M. A class of multi-symmetric polyhedra. *Tohoku Math. J.* **43**, 104–108 (1937).
6. Chakrabarty, R., Mukherjee, P. S. & Stang, P. J. Supramolecular coordination: self-assembly of finite two- and three-dimensional ensembles. *Chem. Rev.* **111**, 6810–6918 (2011).
7. Smulders, M. M. J., Riddell, I. A., Browne, C. & Nitschke, J. R. Building on architectural principles for three-dimensional metallosupramolecular construction. *Chem. Soc. Rev.* **42**, 1728–1754 (2013).
8. Zhang, G. & Mastalerz, M. Organic cage compounds – from shape-persistency to function. *Chem. Soc. Rev.* **43**, 1934–1947 (2014).
9. Fujita, D. et al. Self-assembly of tetravalent Goldberg polyhedra from 144 small components. *Nature* **540**, 563–566 (2016).
10. Sun, Q.-F. et al. Self-assembled $M_{24}L_{48}$ polyhedra and their sharp structural switch upon subtle ligand variation. *Science* **328**, 1144–1147 (2010).
11. Fujita, D. et al. Self-assembly of $M_{30}L_{60}$ icosidodecahedron. *Chem* **1**, 91–101 (2016).
12. Olenyuk, B., Levin, M. D., Whiteford, J. A., Shield, J. E. & Stang, P. J. Self-assembly of nanoscopic dodecahedra from 50 predesigned components. *J. Am. Chem. Soc.* **121**, 10434–10435 (1999).
13. Koo, J. et al. Gigantic porphyrinic cages. *Chem* **6**, 3374–3384 (2020).
14. Liu, Y. Z., Hu, C. H., Comotti, A. & Ward, M. D. Supramolecular Archimedean cages assembled with 72 hydrogen bonds. *Science* **333**, 436–440 (2011).
15. Pasquale, S., Sattin, S., Escudero-Adán, E. C., Martínez-Belmonte, M. & de Mendoza, J. Giant regular polyhedra from calixarene carboxylates and uranyl. *Nat. Commun.* **3**, 785 (2012).
16. Wang, H. et al. Hierarchical self-assembly of nanowires on the surface by metallo-supramolecular truncated cuboctahedra. *J. Am. Chem. Soc.* **143**, 5826–5835 (2021).

17. Chen, Y.-S. et al. Chemical mimicry of viral capsid self-assembly via corannulene-based pentatopic tectons. *Nat. Commun.* **10**, 3443 (2019).
18. Zhang, D. et al. Templation and concentration drive conversion between a $Fe^{II}_{12}L_{12}$ pseudoicosahedron, a $Fe^{II}_{4}L_{4}$ tetrahedron, and a $Fe^{II}_{2}L_{3}$ helicate. *J. Am. Chem. Soc.* **144**, 1106–1112 (2022).
19. Lee, S., Chen, C.-H. & Flood, A. H. A pentagonal cyanostar macrocycle with cyanostilbene CH donors binds anions and forms dialkylphosphate [3]rotaxanes. *Nat. Chem.* **5**, 704–710 (2013).
20. Guieu, S., Crane, A. K. & MacLachlan, M. J. Campestarenes: novel shape-persistent Schiff base macrocycles with 5-fold symmetry. *Chem. Commun.* **47**, 1169–1171 (2011).
21. Qu, H. et al. Truncated face-rotating polyhedra constructed from pentagonal pentaphenylpyrrole through graph theory. *J. Am. Chem. Soc.* **142**, 16223–16228 (2020).
22. Zhang, D., Ronson, T. K. & Nitschke, J. R. Functional capsules via subcomponent self-assembly. *Acc. Chem. Res.* **51**, 2423–2436 (2018).
23. Kramer, R., Lehn, J. M. & Marquis-Rigault, A. Self-recognition in helicate self-assembly: spontaneous formation of helical metal complexes from mixtures of ligands and metal ions. *Proc. Natl. Acad. Sci. USA* **90**, 5394–5398 (1993).
24. Jones, T. A. & Liljas, L. Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution. *J. Mol. Biol.* **177**, 735–767 (1984).
25. Hopper, P., Harrison, S. C. & Sauer, R. T. Structure of tomato bushy stunt virus: V. Coat protein sequence determination and its structural implications. *J. Mol. Biol.* **177**, 701–713 (1984).
26. Helgstrand, C., Munshi, S., Johnson, J. E. & Liljas, L. The refined structure of *Nudaurelia capensis* ω Virus reveals control elements for a $T = 4$ capsid maturation. *Virology* **318**, 192–203 (2004).
27. Wikoff, W. R. et al. Topologically Linked Protein Rings in the Bacteriophage HK97 Capsid. *Science* **289**, 2129–2133 (2000).
28. Luo, D., Wang, X. Z., Yang, C., Zhou, X. P. & Li, D. Self-assembly of chiral metal-organic tetartoid. *J. Am. Chem. Soc.* **140**, 118–121 (2018).
29. McConnell, A. J., Aitchison, C. M., Grommet, A. B. & Nitschke, J. R. Subcomponent exchange transforms an $Fe^{II}_{4}L_{4}$ cage from high- to low-spin, switching guest release in a two-cage system. *J. Am. Chem. Soc.* **139**, 6294–6297 (2017).
30. Solomon, I. Relaxation processes in a system of two spins. *Phys. Rev.* **99**, 559–565 (1955).

31. Schweinfurth, D. et al. Spin crossover in Fe(II) and Co(II) complexes with the same click-derived tripodal ligand. *Inorg. Chem.* **53**, 8203–8212 (2014).
32. Maglic, J. B. & Lavendomme, R. MoloVol: an easy-to-use program for analyzing cavities, volumes and surface areas of chemical structures. *J. Appl. Crystallogr.* **55**, 1033–1044 (2022).
33. Geometry optimised structures were modelled using the MM3 force field on SCIGRESS software (Fujitsu Limited, Tokyo, Japan, 2013) version FJ 2.6 (EU 3.1.9) Build 5996.8255.20141202.
34. von Krbek, L. K. S., Schalley, C. A. & Thordarson, P. Assessing cooperativity in supramolecular systems. *Chem. Soc. Rev.* **46**, 2622–2637 (2017).
35. Kieffer, M., Bilbeisi, R. A., Thoburn, J. D., Clegg, J. K. & Nitschke, J. R. Guest binding drives host redistribution in libraries of $Co^{II}_4L_4$ cages. *Angew. Chem., Int. Ed.* **59**, 11369–11373 (2020).
36. Miller, T. F. et al. Small structural variations have large effects on the assembly properties and spin state of room temperature high spin Fe(II) iminopyridine cages. *Inorg. Chem.* **57**, 13386–13396 (2018).
37. Percástegui, E. G. et al. Waterproof architectures through subcomponent self-assembly. *Chem. Sci.* **10**, 2006–2018 (2019).

Methods:

Synthesis of pyrrole-based penta-amine (Subcomponent A)

To a Schlenk flask were added 1,2,3,4,5-pentakis(4-bromophenyl)-1H-pyrrole (**A1**) (337 mg, 0.400 mmol, 1 equiv), 4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)aniline (788 mg, 3.60 mmol, 9 equiv), and K_3PO_4 (980 mg, 4.32 mmol, 10.8 equiv). Then 32 mL dioxane and 8 mL water were added to the mixture to give a brownish suspension. After two freeze-pump-thaw cycles, $Pd(PPh_3)_4$ (231 mg, 0.200 mmol, 0.5 equiv) powder was carefully added while the mixture was frozen and the reaction system was degassed again. The resulting mixture was then heated at 95 °C under nitrogen for 48 h. The solvent was removed under reduced pressure, and then CH_2Cl_2 and water were added. The aqueous layer was extracted three times with CH_2Cl_2 and then washed with brine. The organic phases were combined and dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The crude product was purified by silica gel column chromatography with CH_2Cl_2/CH_3OH (v/v 40:1 to 25:1) to give **A** as a brown solid (201 mg, 56%).

^{1}H NMR (500 MHz, DMSO-d_6) δ 7.44–7.39 (m, 2H), 7.37–7.24 (m, 18H), 7.09–7.05 (m, 2H), 7.01–6.92 (m, 8H), 6.59–6.50 (m, 10H), 5.23 (s, 2H), 5.19 (s, 4H), 5.15 (s, 4H).

^{13}C NMR (125 MHz, DMSO-d_6) δ 148.6, 148.4, 148.1, 138.9, 138.3, 137.3, 135.9, 132.6, 131.4, 131.4, 130.9, 129.5, 129.2, 126.96, 126.8, 126.7, 126.7, 126.2, 125.6, 124.6, 124.3, 124.1, 122.0, 114.2, 114.1.

HR-ESI-MS (CH_3CN) *m/z*: $[M+H]^+$ calc. for $C_{64}H_{51}N_6$, 903.4169; found 903.4194; $[M+2H]^{2+}$ calc. for $C_{64}H_{52}N_6$, 452.2120; found 452.2105.

Self-assembly of Dodecahedron Fe-2

Subcomponent **A** (0.9 mg, 1 μmol), $Fe(NTf_2)_2 \cdot 4.5H_2O$ (1.2 mg, 1.7 μmol), 2-formyl-6-methyl pyridine **C** (0.6 mg, 5 μmol) and CD_3CN (0.4 ml) were added to a J-Young NMR tube, and the mixture was sonicated for 1 min. After two cycles of freeze-pump-thaw, the mixture was heated at 75 °C overnight. The brown solution was concentrated by blowing with N_2 and then Et_2O was added. The resulting solid was collected by centrifugation, washed three times with additional Et_2O and then vacuum dried to give dodecahedron **$[Fe_{20}L_{12}](NTf_2)_{40}$** as a brown solid (2.2 mg, 0.075 μmol, 90%).

^{1}H NMR (400 MHz, CD_3CN) δ 205.6–190.3 (br), 58.3 (br), 56.2 (br), 12.0 (br), 9.8 (br), 7.6 (br), 4.6 (br), –15.1 (br), –30.1 (br).

^{19}F NMR (376 MHz, CD_3CN) δ –79.19 (s, CF_3).

ESI-MS (CH_3CN): *m/z*: 1054.2 $[M–22NTf_2^-]^{22+}$, 1117.8 $[M–21NTf_2^-]^{21+}$, 1187.7 $[M–20NTf_2^-]^{20+}$, 1264.9 $[M–19NTf_2^-]^{19+}$, 1350.7 $[M–18NTf_2^-]^{18+}$, 1446.7 $[M–17NTf_2^-]^{17+}$, 1554.6 $[M–16NTf_2^-]^{16+}$, 1676.9 $[M–15NTf_2^-]^{15+}$, 1816.1 $[M–14NTf_2^-]^{14+}$, 1977.4 $[M–13NTf_2^-]^{13+}$, 2165.6 $[M–12NTf_2^-]^{12+}$, 2387.8 $[M–11NTf_2^-]^{11+}$, 2654.6 $[M–10NTf_2^-]^{10+}$, 2980.7 $[M–9NTf_2^-]^{9+}$.

Data availability

All data needed to evaluate the conclusions given in the paper are present in the paper and Supplementary Information. Any additional data related to this paper may be requested from the authors. Crystallographic data for the structure reported in this paper has been deposited at the Cambridge Crystallographic Data Centre (deposition number 2193949) and can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif.

Acknowledgements This study was supported by the European Research Council (695009), the UK Engineering and Physical Sciences Research Council (EPSRC, EP/P027067/1) and Astex Therapeutics Ltd. (RG73357, Sustaining Innovation Postdoctoral Training Program –

AH). We thank Diamond Light Source for providing time on Beamline I24. The SAXS data collection was performed with the help of Dr. Claire Pizzey using beamline B21 of the UK Diamond Light Source, a dedicated beamline for small angle X-ray scattering. This work benefited from the use of the SasView application, originally developed under NSF award DMR-0520547. SasView contains code developed with funding from the European Union's Horizon 2020 research and innovation programme under the SINE2020 project, grant agreement No 654000. The authors thank the University of Cambridge Chemistry Department NMR service for performing some NMR experiments. We thank the ThermoFisher scientific company and the mass team of the University of Cambridge Chemistry Department for help with some mass spectrometry measurements. We also thank Jack D. Davis, Dr. Xiang Sun, Dr. Weichao Xue, Dr. Carles Fuertes-Espinosa, Dr. Hua-kui Liu for helpful discussion and assistance.

Author Contributions:

K.W. and J.R.N. designed the work and wrote the paper. K.W. carried out the research, grew the crystals and analyzed data. T.K.R., and M.V. contributed to X-ray crystallographic data collection and analysis, A.W.H helped in data analysis. M.V. contributed to SAXS data analysis. P.S. and X.L. conducted some of the ESI-MS and the TWIM-MS measurements. L.G. helped with ligand synthesis and provided helpful discussion. F. K. and C. A. S. measured the mass spectrum of **Co-5**. Z.C. performed STM measurements. J.R.N. is the principal investigator. All authors discussed the results and commented on the manuscript.

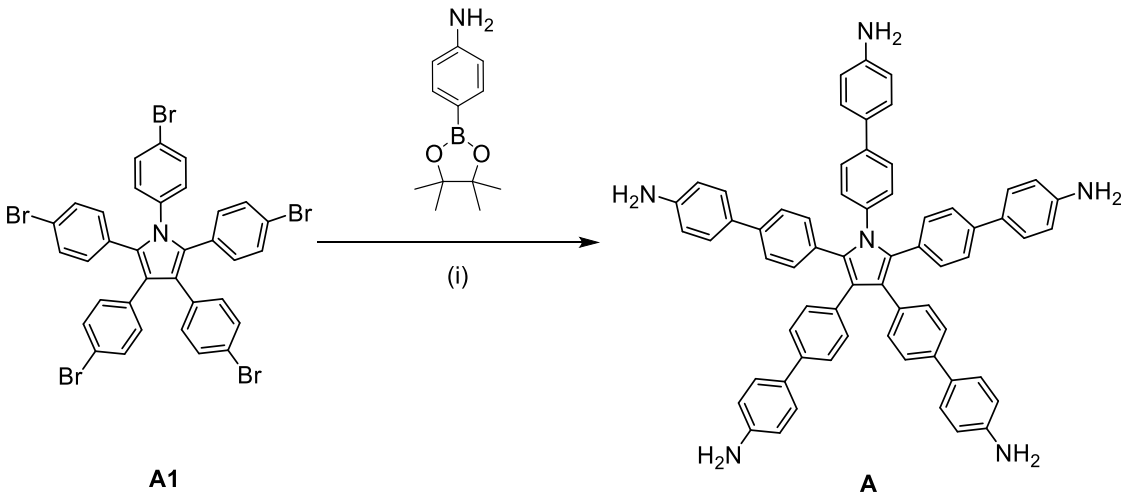
Competing interests: The authors declare no competing interests.

Additional information

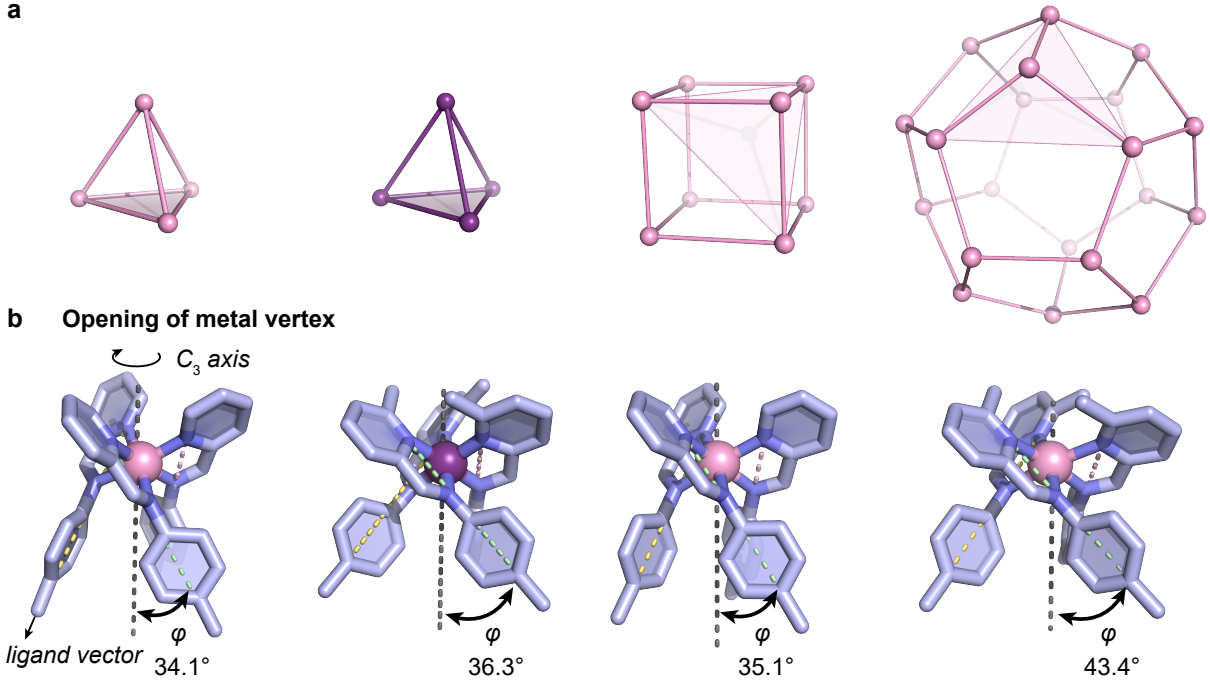
Supplementary information is available for this paper at [DOI link].

Correspondence and requests for materials should be addressed to J.R.N.

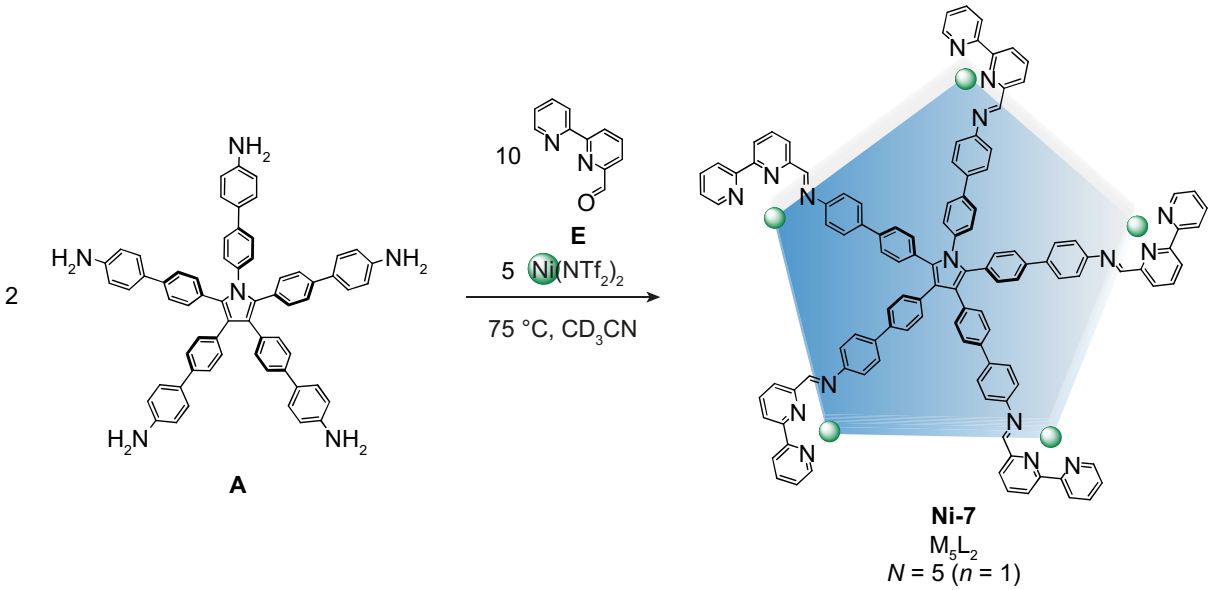
Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | Preparation of penta-amine A from A1. Reagents and conditions: (i) $Pd(PPh_3)_4$, K_3PO_4, dioxane:H_2O = 4:1, 95 °C, 48 h.



Extended Data Fig. 2 | Schematic representation of the metal vertex opening angles of tetrahedral, cubic and dodecahedral frameworks. a, Stick view of metal ion positions in the different cage frameworks. **b**, Opening angles (***φ***) of the metal vertex of the corresponding tetrahedra[35,36], cube[37], or dodecahedron shown above, between the metal centre C_3 axis (grey dashed line) and the ligand vector defined by the two nitrogen atoms of the chelate plane, measured from their crystal structures. ***φ*** was calculated as the complementary angle of the dihedral angle between the five-membered chelate plane and the plane formed by the three neighbouring metal centres (highlighted). Colour code: Co^{II} pink, Fe^{II} purple, C light blue, N blue). The presence of bulky methyl group and the expansion of frameworks are inferred to both increase ***φ***.



Extended Data Fig. 3 | Schematic representation of subcomponent self-assembly of Ni-7. Only one ligand is shown for clarity.