Electronic Supplementary Information

Orientational self-sorting in cuboctahedral Pd cages

Ru-Jin Li,[a] Andrew Tarzia,[b] Victor Posligua,[b] Kim E. Jelfs,[b] Nicolas Sanchez,[c] Adam Marcus,[c] Ananya Baski,[d] Guido H. Clever,[d] Farzaneh Fadaei-Tirani,[a] and Kay Severin*[a]

[a] Institut of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

[b] Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, White City Campus, 82 Wood Lane, London, W12 0BZ, UK.

[c] Institut of Mathematics, EPFL, 1015 Lausanne, Switzerland

[d] Fakultät für Chemie und Chemische Biologie, Technische Universität Dortmund, Universität Dortmund, 44227 Dortmund, Germany

Table of Contents

1 General

All chemicals were obtained from commercial sources and used without further purification unless stated otherwise.

NMR spectra were measured on a Bruker Avance III HD spectrometer (^{1}H: 400 MHz) equipped with a BBFO-Plus$_z$ 5 mm probe, a Bruker Avance III spectrometer (^{1}H: 400 MHz) equipped with a BBFO$_z$ 5 mm probe, and a Bruker Avance III spectrometer (^{1}H: 800 MHz) equipped with a Prodigy BBO 10 mm cryoprobe, a Bruker Avance II HD spectrometer (^{13}C: 201 MHz) equipped with a TBI$_{xyz}$ 5 mm probe. The chemical shifts are reported in parts per million (ppm) using the solvent residual signal as a reference.

High resolution mass spectrometry experiments were carried out using a hybrid ion trap-Orbitrap Fourier transform mass spectrometer, Orbitrap Elite (Thermo Scientific) equipped with a TriVersa Nanomate (Advion) nano-electrospray ionization source. Mass spectra were acquired with a minimum resolution setting of 120,000 at 400 m/z. To reduce the degree of analyte gas phase reactions leading to side products unrelated to solution phase, the transfer capillary temperature was lowered to 50 °C. Experimental parameters were controlled via standard and advanced data acquisition software.

Cryo spray ionization (CSI) mass spectrometry experiment was operated in the positive ion mode and performed on a Bruker timsTOF instrument with Cryo spray attachment. The temperature of the nebulizing and drying gases was set at –40 °C. The sample was introduced into the mass spectrometer ion source by direct infusion using a syringe pump (sample concentration: 18.75 μM, solvent: acetonitrile, flow rate: 180uL/h, capillary voltage: 2500 V, end plate offset voltage: 200 V, dry temperature: 233 K, funnel 1 RF: 150V, funnel 2 RF: 150 V, multiple RF: 150 V, deflection delta: 60 V, quadrupole ion energy: 3 V, collision energy: 3 V, transfer time: 80 μs, prepulse storage: 15 μs).

2 Syntheses

2.1 Synthesis of the ligands
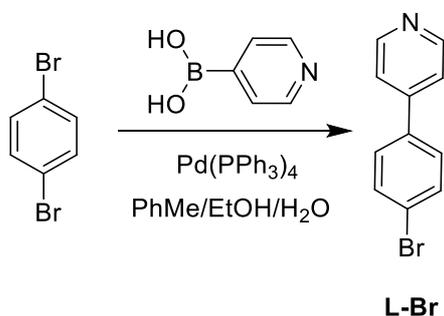
L-Br: 1,4-Dibromobenzene (472 mg, 2.0 mmol), pyridin-4-ylboronic acid (246 mg, 2.0 mmol), Pd(PPh3)4 (115 mg, 0.1 mmol), and K_2CO_3 (0.6 g, 4 mmol) were combined in a 50 mL Schlenk tube. After vacuum/backfilling with N_2 for three times, 20 mL degassed toluene/EtOH/H_2O (2:1:1) was added with a syringe. The mixture was heated at 85 °C overnight. After cooling to RT, and reaction was quenched with water. The product was extracted with DCM (2 x 50 mL). The combined organic phase was washed with water (2 x 50 mL) and brine (50 mL), and dried over $MgSO_4$. The solvent was removed under reduced pressure, and the residue was purified by column chromatography on silica to yield **L-Br** as a white powder (300 mg, 1.28 mmol, 64%). ^{1}H NMR (400 MHz, CDCl3) δ 8.72 – 8.66 (m, 2H), 7.69 – 7.63 (m, 2H), 7.60 – 7.56 (m, 2H), 7.56 – 7.50 (m, 2H). ^{13}C NMR (101 MHz, CDCl3) δ 148.70, 136.44, 132.70, 128.82, 124.52, 122.03. ESI-MS: *m/z* calculated for C11H9NBr [M+H]$^+$ 235.99, found 235.99.
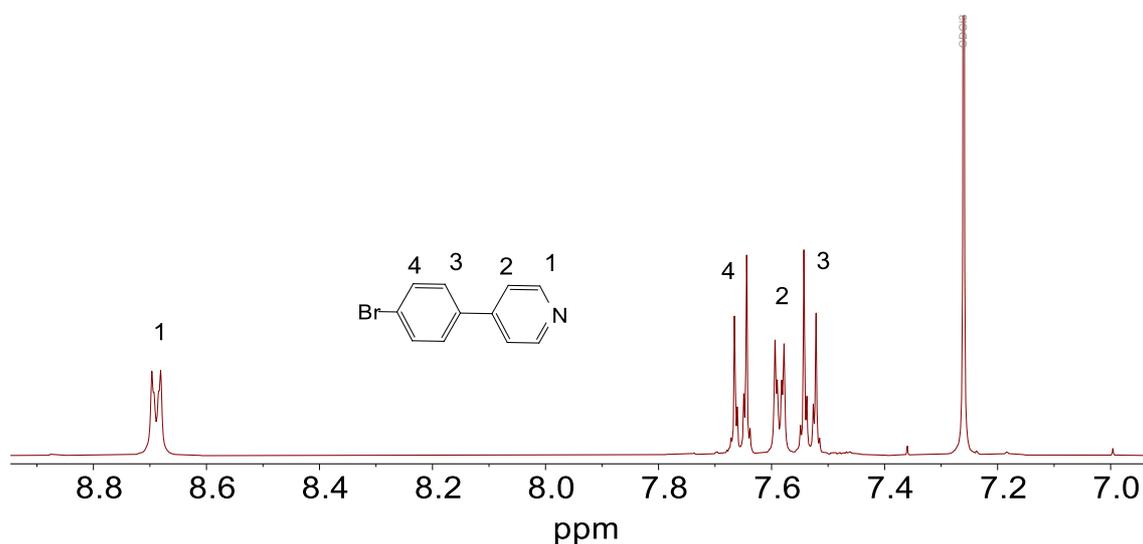
Figure S1. ^{1}H NMR (400 MHz, CDCl3) spectrum of **L-Br**.
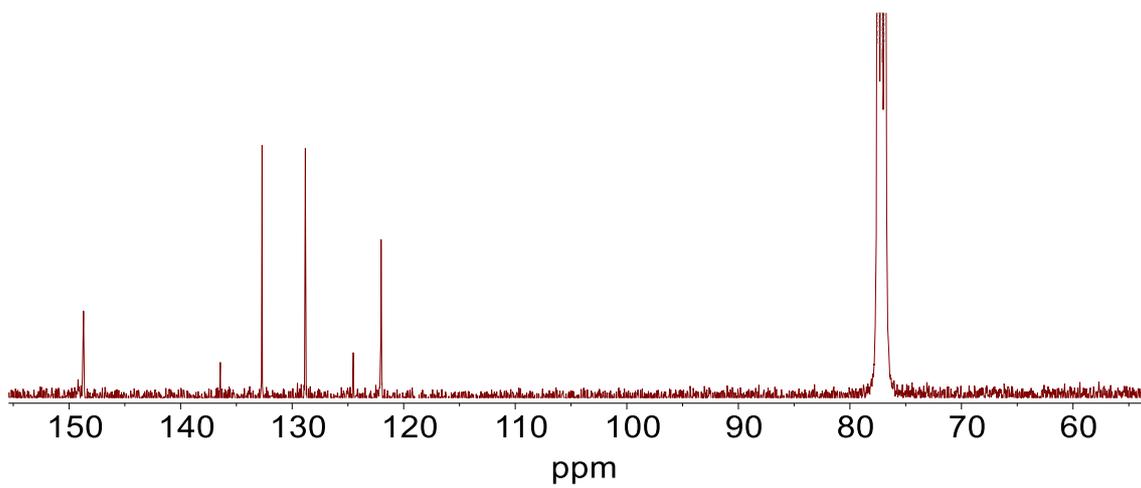
Figure S2. ^{13}C NMR (101 MHz, CDCl$_3$) spectrum of **L-Br**.
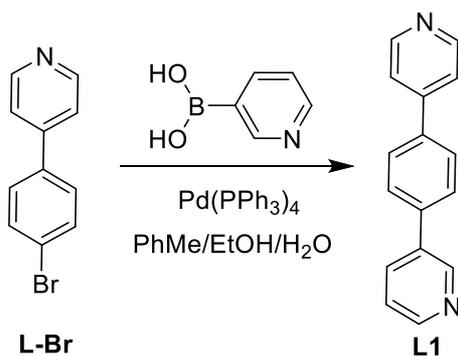
L1: Pyridin-3-ylboronic acid (185 mg, 1.5 mmol), **L-Br** (300 mg, 1.28 mmol), Pd(PPh$_3$)$_4$ (81 mg, 0.07 mmol), and K$_2$CO$_3$ (0.6 g, 4 mmol) were combined in a 50 mL Schlenk tube. After vacuum/backfilling with N$_2$ for three times, 20 mL degassed toluene/EtOH/H$_2$O (2:1:1) was added with a syringe. The mixture was heated at 85 °C overnight. After cooling to RT, the reaction was quenched with water, and the product was extracted with DCM (2 x 50 mL). The combined organic phase was washed with water (2 x 50 mL) and brine (50 mL), and dried over MgSO$_4$. The solvent was removed under reduced pressure, and the residue was purified by column chromatography on silica to yield **L1** as a brown powder (140 mg, 0.60 mmol, 47%). ^{1}H NMR (400 MHz, CD$_3$CN) δ 8.92 (dd, J = 2.4, 0.8 Hz, 1H), 8.68 – 8.62 (m, 2H), 8.59 (dd, J = 4.8, 1.6 Hz, 1H), 8.06 (ddd, J = 8.0, 2.4, 1.6 Hz, 1H), 7.91 – 7.86 (m, 2H), 7.85 – 7.80 (m, 2H), 7.70 – 7.65 (m, 2H), 7.46 (ddd, J = 7.9, 4.8, 0.9 Hz, 1H). ^{13}C NMR (101 MHz, CD$_3$CN) δ 151.27, 149.83, 148.97, 147.98, 139.45, 138.48, 136.33, 135.18, 128.74, 128.61, 124.69, 122.30. ESI-MS: m/z calculated for C16H13N2 [M+H]$^+$ 233.11, found 233.11.
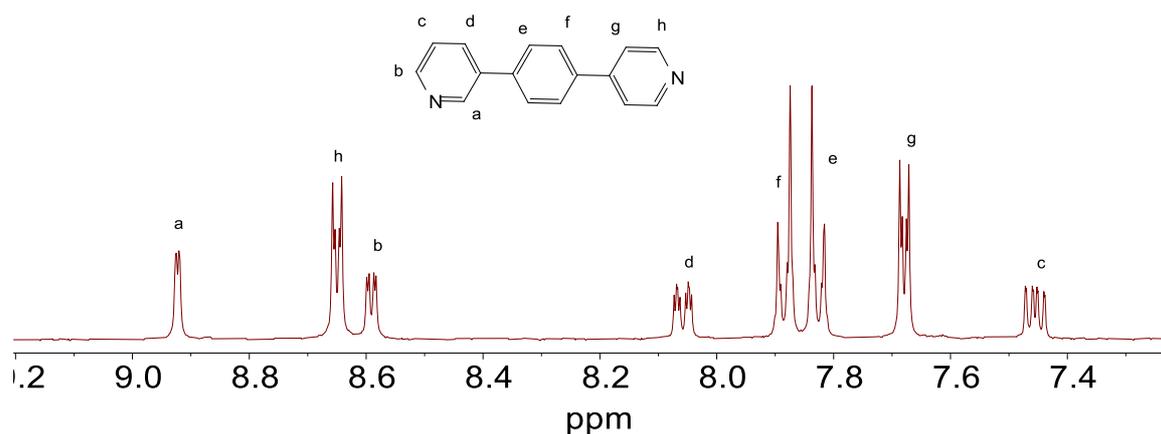
Figure S3. ^{1}H NMR (400 MHz, CD$_3$CN) spectrum of ligand **L1**.
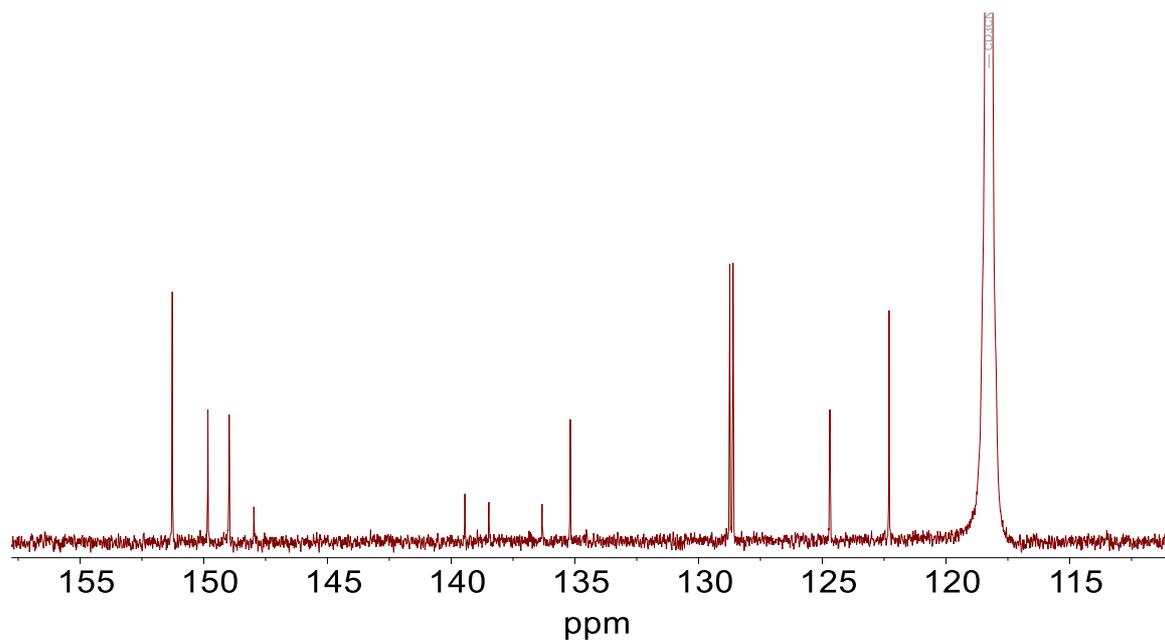
Figure S4. ^{13}C NMR (101 MHz, CD$_3$CN) spectrum of ligand **L1**.
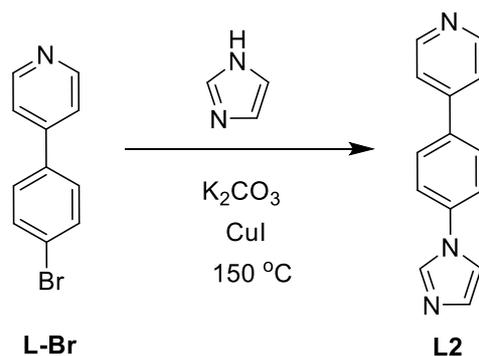
L2: Excess imidazole (870 mg, 12.8 mmol), **L-Br** (300 mg, 1.28 mmol), CuI (25 mg, 0.13 mmol), and K_2CO_3 (0.3 g, 2 mmol) were combined in a 25 mL Schlenk tube. After vacuum/backfilling with N_2 for three times, the mixture was heated at 150 °C overnight without any solvent. After cooling to RT, DCM (150 mL) was added. The organic phase was washed with water (2 x 100 mL) and brine (100 mL), and dried over $MgSO_4$. The solvent was removed under reduced pressure, and the residue was purified by column chromatography on silica to yield **L2** as a brown powder (246 mg, 1.11 mmol, 87%).

^{1}H NMR (400 MHz, CD$_3$CN) δ 8.69 – 8.61 (m, 2H), 8.02 (t, J = 1.1 Hz, 1H), 7.92 – 7.83 (m, 2H), 7.70 – 7.62 (m, 4H), 7.53 (t, J = 1.4 Hz, 1H), 7.14 (t, J = 1.2 Hz, 1H). ^{13}C NMR (101 MHz, CD$_3$CN) δ 151.33, 147.42, 138.94, 137.46, 136.55, 131.16, 129.40, 122.35, 122.23, 118.96. ESI-MS: *m/z* calculated for C14H12N3 [M+H]$^+$ 222.10, found 222.10.
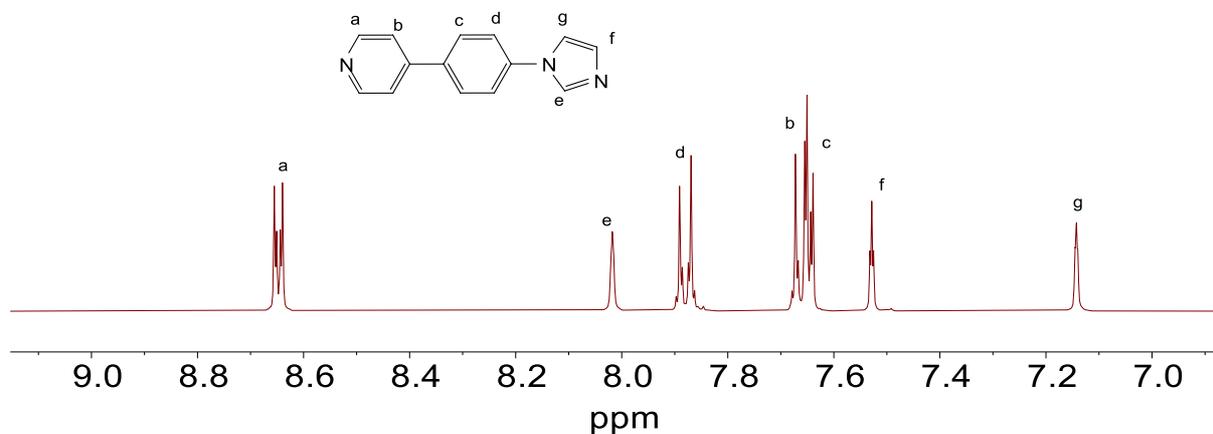
Figure S5. ^{1}H NMR (400 MHz, CD$_3$CN) spectrum of ligand **L2**.
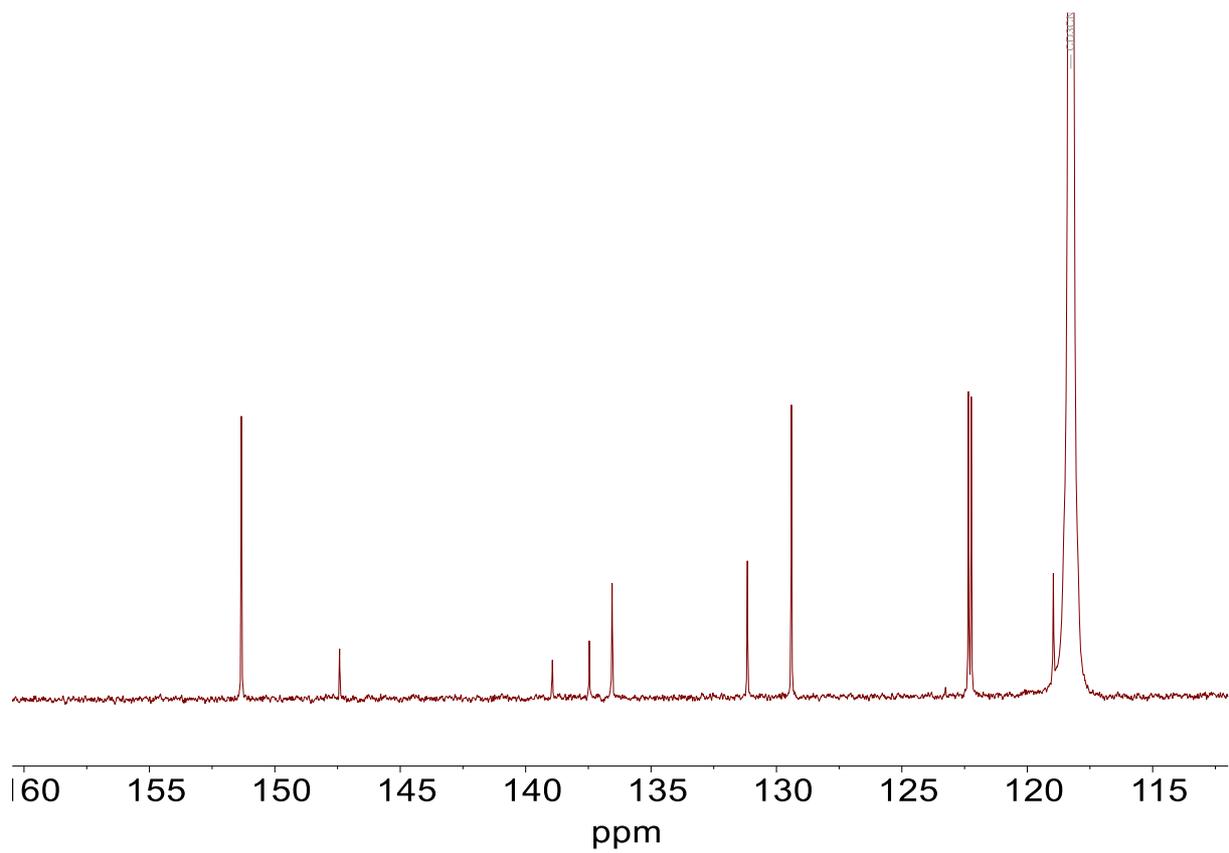
Figure S6. ^{13}C NMR (101 MHz, CD$_3$CN) spectrum of ligand **L2**.

2.2 Syntheses of the cages

2.2.1 [Pd$_{12}$(L1)$_{24}$](BF$_4$)$_{24}$
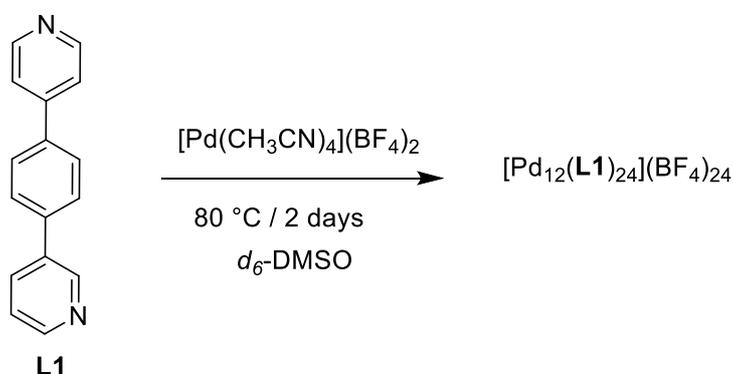
The cage [Pd$_{12}$(L1)$_{24}$](BF$_4$)$_{24}$ was synthesized by stirring a mixture of ligand L1 (9 µmol, 200 µL of a 45 mM stock solution of L1 in d_6-DMSO) and [Pd(CH$_3$CN)$_4$](BF$_4$)$_2$ (4.5 µmol, 150 µL of a 30 mM stock solution in d_6-DMSO) in 650 µL d_6-DMSO at 80 °C for 2 days to give 1000 µL of a 0.375 mM solution of cage [Pd$_{12}$(L1)$_{24}$](BF$_4$)$_{24}$. ^{1}H NMR (400 MHz, d_6-DMSO) δ 9.87 (s, 8H), 9.74 (s, 8H), 9.65 (b, 8H), 9.55 (b, 16H), 9.51 (s, 8H), 9.38 (b, 16H), 9.36 – 9.30 (b, 16H), 9.22 (b, 16H), 8.48 (b, 24H), 8.20 (b, 48H), 8.11 – 7.86 (b, 120H). ^{13}C NMR (201 MHz, d_6-DMSO) δ 150.93, 150.00 (b), 149.14 (b), 138.67 (b), 137.11, 136.39, 135.14, 128.08 (b), 124.25 (b).
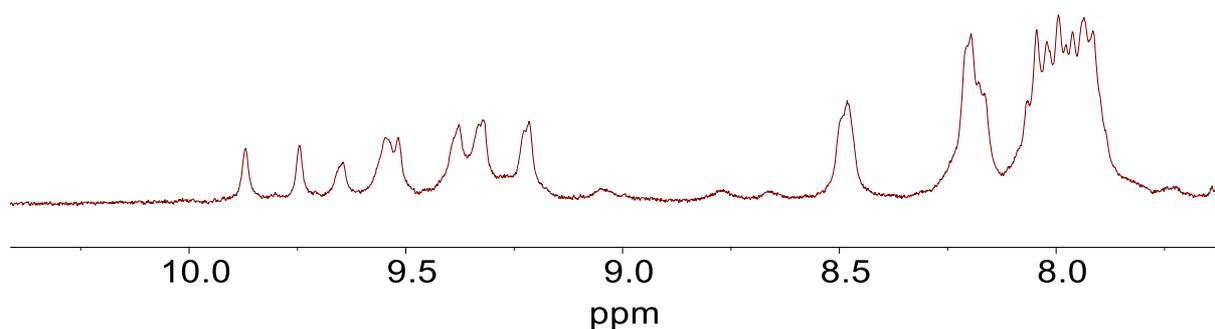
Figure S7. ^{1}H NMR spectrum (400 MHz, d_6-DMSO) of [Pd$_{12}$(L1)$_{24}$](BF$_4$)$_{24}$.
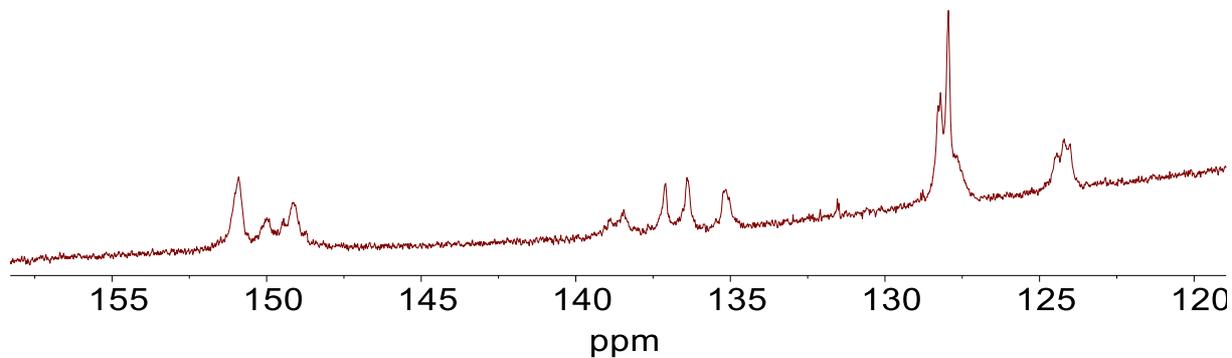
Figure S8. ^{13}C NMR spectrum (201 MHz, d_6-DMSO) of [Pd$_{12}$(**L1**)$_{24}$](BF$_4$)$_{24}$.
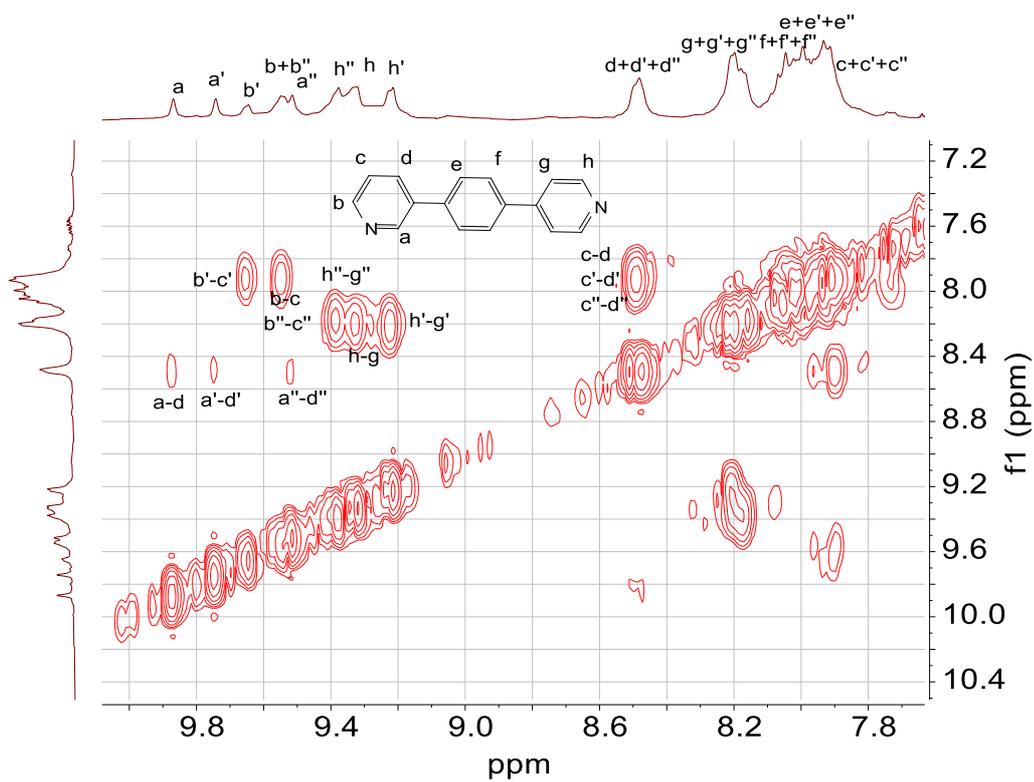
Figure S9. ^1H-^1H COSY NMR spectrum (400 MHz, d_6-DMSO) of [Pd$_{12}$(**L1**)$_{24}$](BF$_4$)$_{24}$.
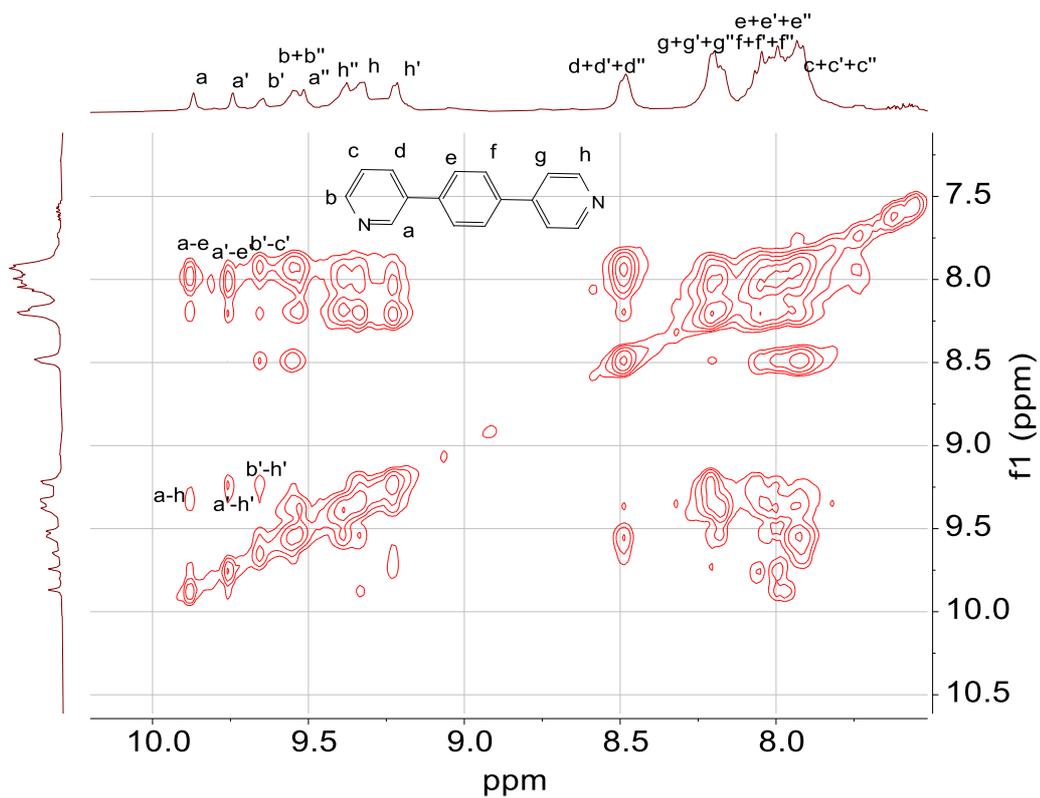
Figure S10. ^{1}H-^{1}H NOESY NMR spectrum (400 MHz, d_6-DMSO) of [Pd$_{12}$(**L1**)$_{24}$](BF$_4$)$_{24}$.
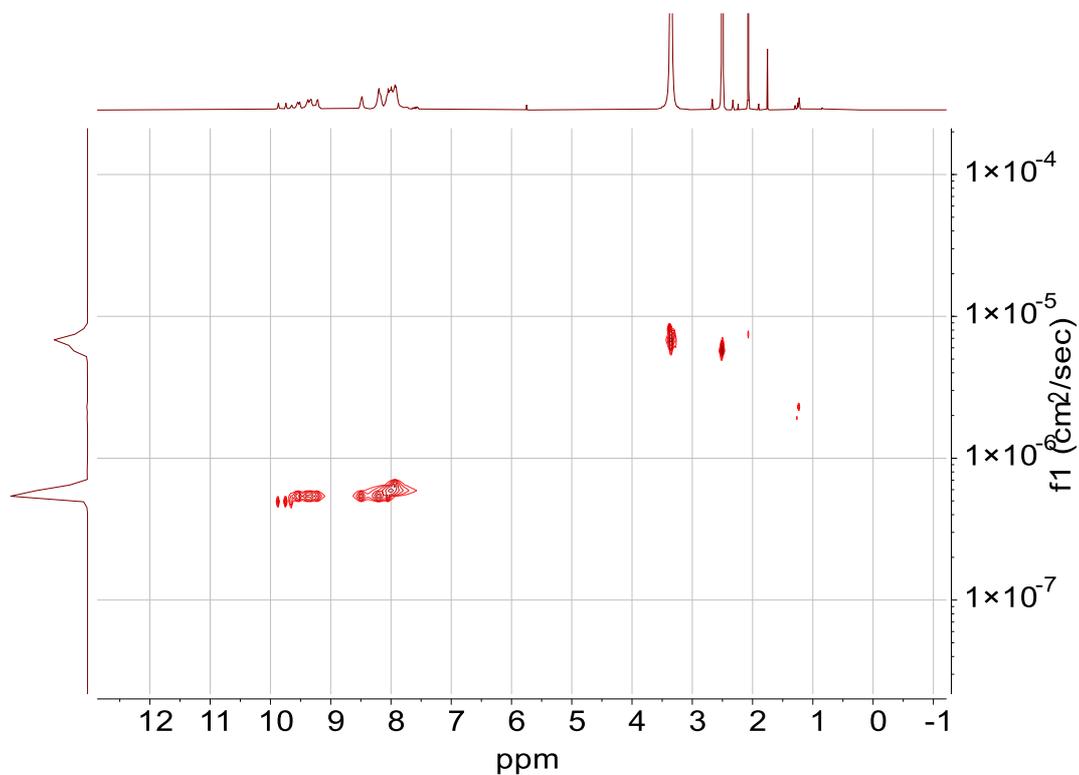
Figure S11. ^{1}H-^{1}H DOSY NMR spectrum (400 MHz, d_6-DMSO) of [Pd$_{12}$(**L1**)$_{24}$](BF$_4$)$_{24}$

(D = 5.44 x 10^{-11} m^2 s^{-1}, r_{dosy} = 20 Å).
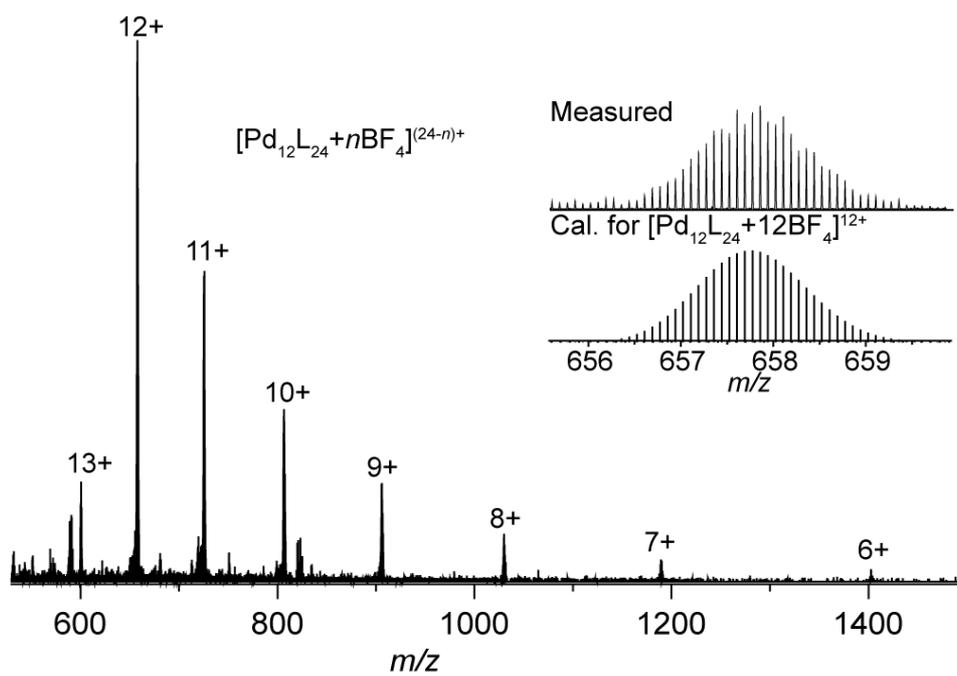
Figure S12. High-resolution ESI mass spectrum of $[Pd_{12}(\textbf{L1})_{24}+nBF_4]^{(24-n)+}$ (n = 11–18).

2.2.2 [Pd$_{12}$(**L1**)$_{24}$](NO$_3$)$_{24}$
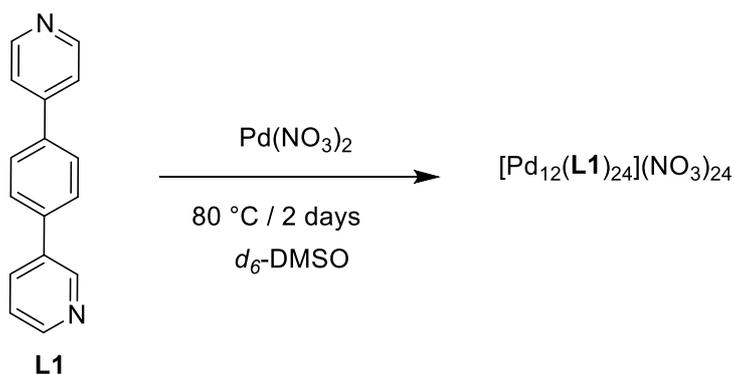
The cage [Pd$_{12}$(**L1**)$_{24}$](NO$_3$)$_{24}$ was synthesized by stirring a mixture of ligand **L1** (9 µmol, 200 µL of a 45 mM stock solution of **L1** in d_6-DMSO) and Pd(NO$_3$)$_2$ (4.5 µmol, 150 µL of a 30 mM stock solution in d_6-DMSO) in 650 µL d_6-DMSO at 80 °C for 2 days to give 1000 µL of a 0.375 mM solution of cage [Pd$_{12}$(**L1**)$_{24}$](NO$_3$)$_{24}$. ^{1}H NMR (400 MHz, d_6-DMSO) δ 10.01 (s, 8H), 9.89 (s, 8H), 9.70 (s, 8H), 9.66 (b, 8H), 9.57 (b, 16H), 9.47 (b, 16H), 9.43 – 9.36 (b, 16H), 9.33 (b, 16H), 8.49 (b, 24H), 8.23 (b, 48H), 8.05 (b, 96H), 7.90 (b, 24H). ^{13}C NMR (201 MHz, d_6-DMSO) δ 151.04, 150.03 (m), 149.12 (m), 138.39 (m), 137.11, 136.23, 135.01, 128.10 (b), 124.23 (m).
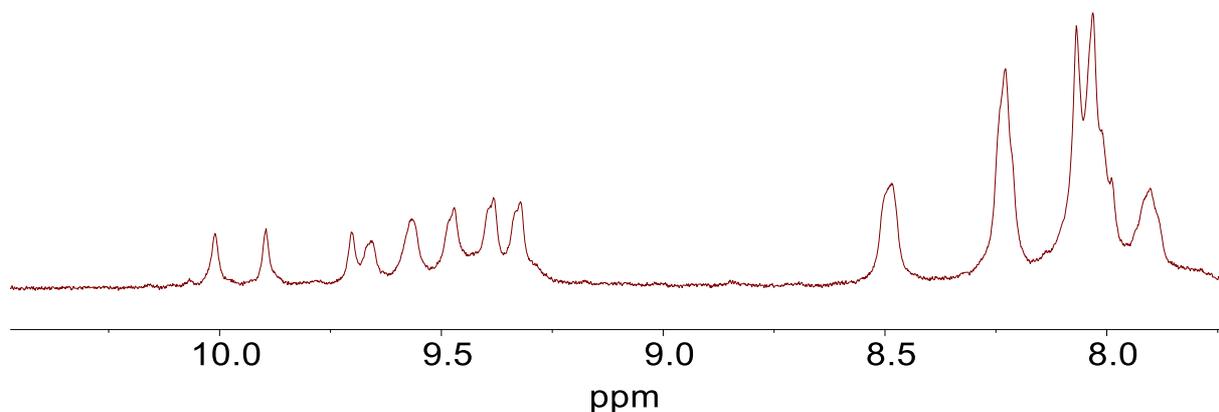
Figure S13. ^{1}H NMR spectrum (400 MHz, d_6-DMSO) of [Pd$_{12}$(**L1**)$_{24}$](NO$_3$)$_{24}$.
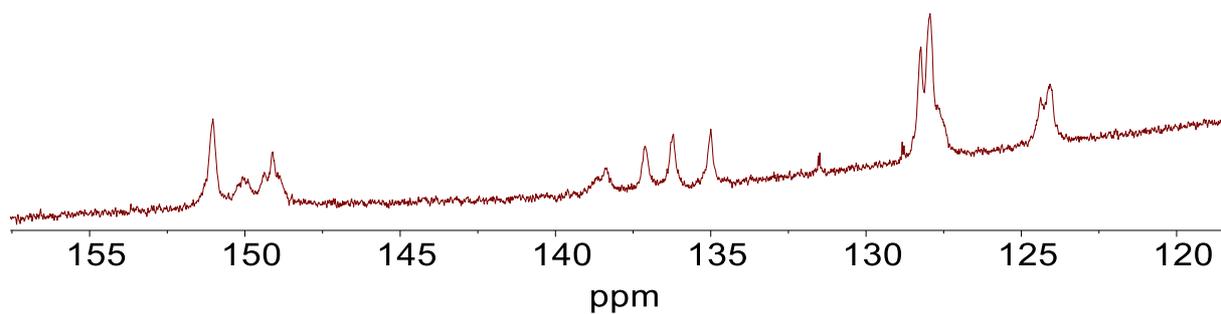
Figure S14. ^{13}C NMR spectrum (201 MHz, d_6-DMSO) of [Pd$_{12}$(**L1**)$_{24}$](NO$_3$)$_{24}$.
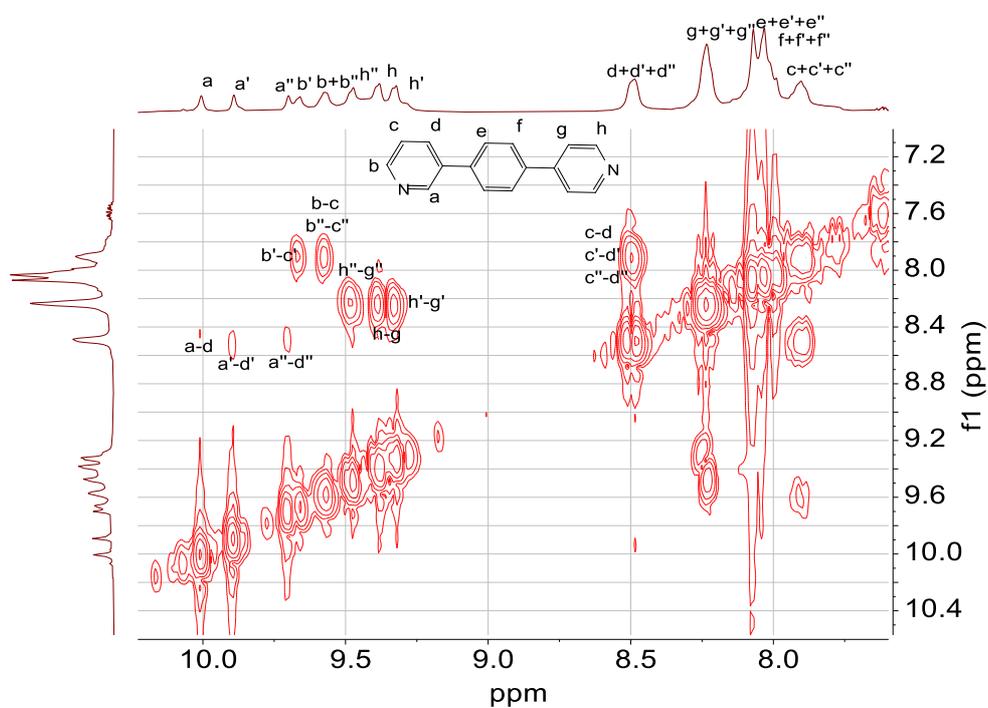
Figure S15. ^{1}H-^{1}H COSY NMR spectrum (400 MHz, d_6-DMSO) of [Pd$_{12}$(**L1**)$_{24}$](NO$_3$)$_{24}$.
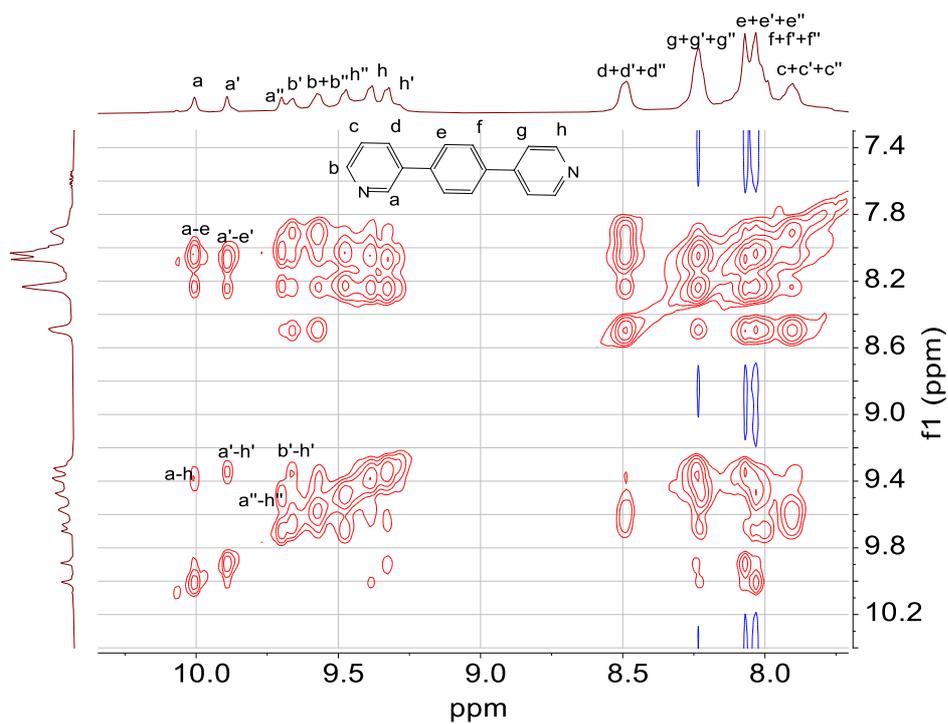
Figure S16. ^{1}H-^{1}H NOESY NMR spectrum (400 MHz, d_6-DMSO) of $[Pd_{12}(\textbf{L1})_{24}](NO_3)_{24}$.
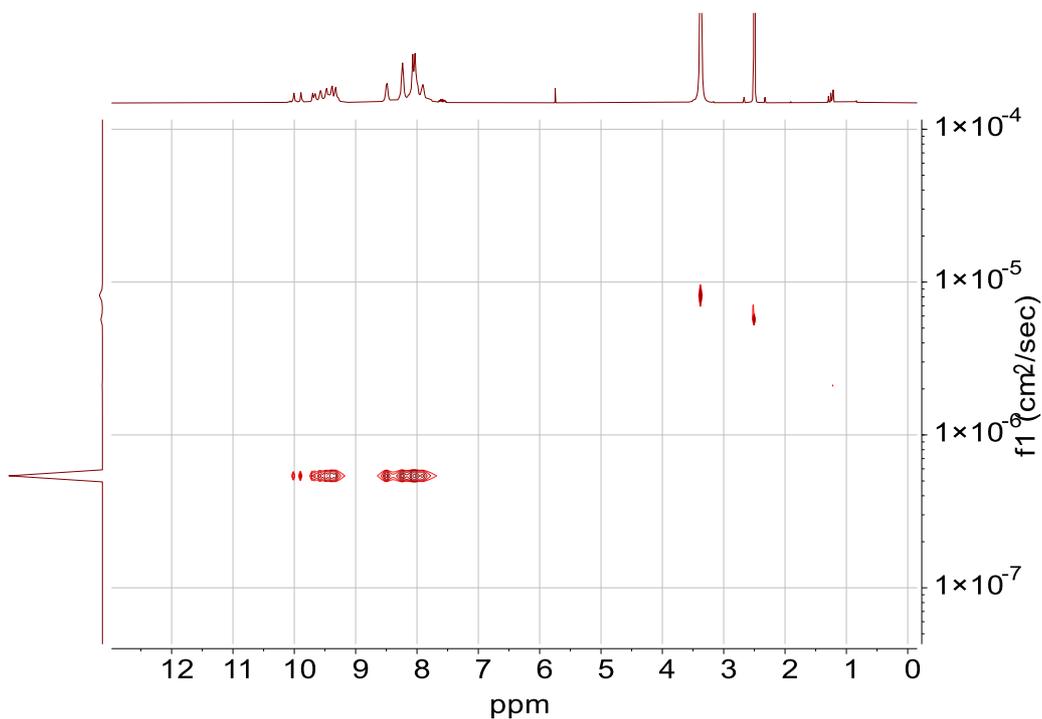
Figure S17. ^{1}H-^{1}H DOSY NMR spectrum (400 MHz, d_6-DMSO) of $[Pd_{12}(\textbf{L1})_{24}](NO_3)_{24}$
($D = 5.40 \times 10^{-11}$ m^2 s^{-1}, $r_{dosy} = 20$ Å).
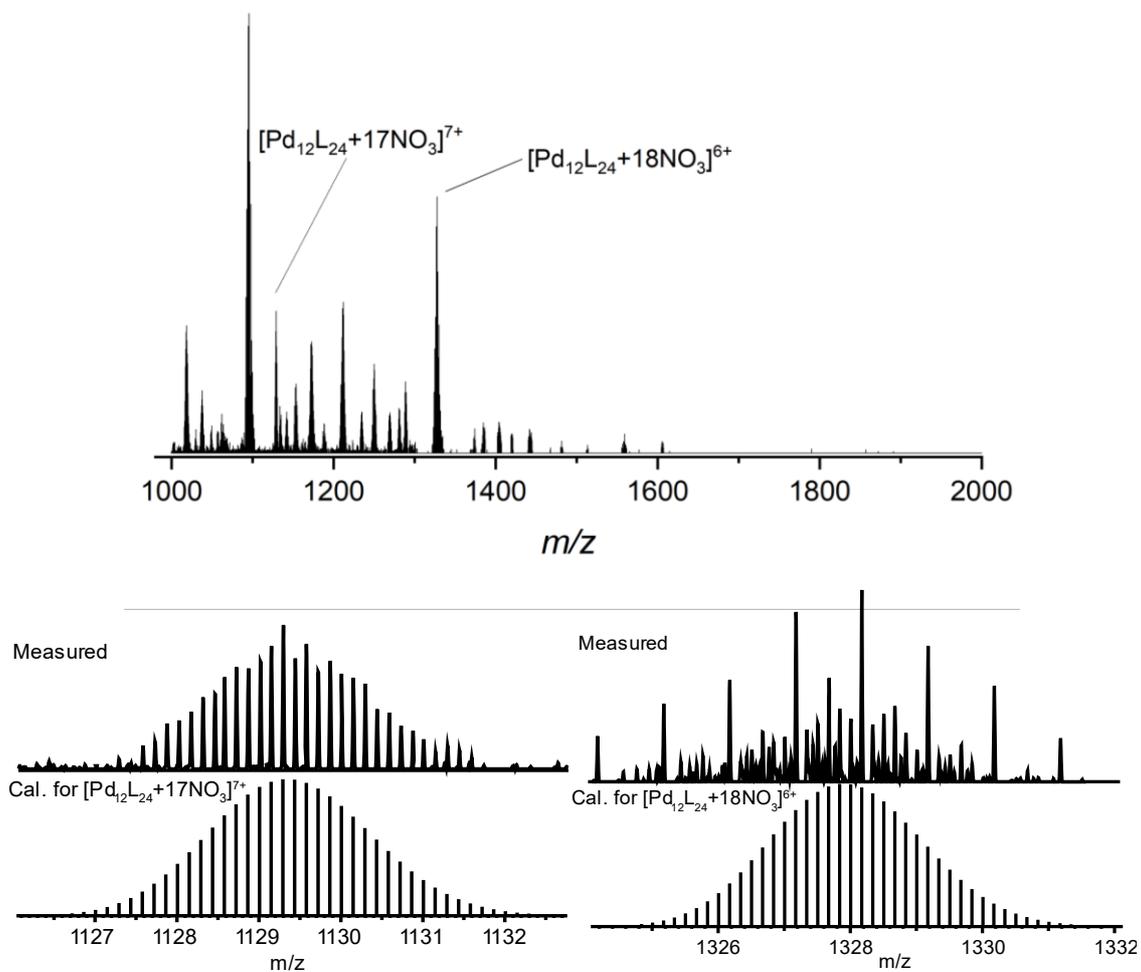
Figure S18. High-resolution ESI mass spectrum of $[Pd_{12}(\textbf{L1})_{24}+nNO_3]^{(24-n)+}$ (n = 17 or 18).

2.2.3 [Pd$_{24}$(**L2**)$_{48}$](BF$_4$)$_{48}$
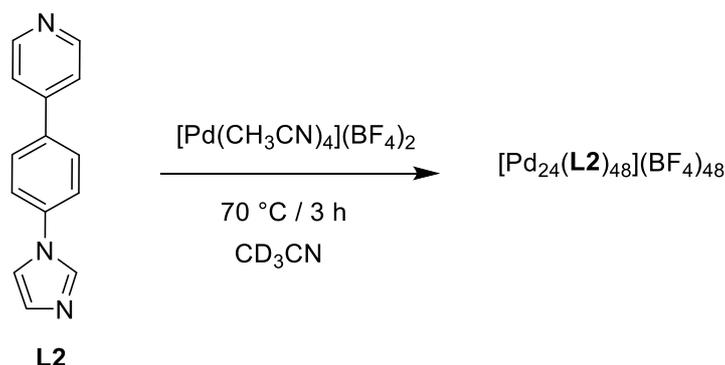
The cage [Pd$_{24}$(**L2**)$_{48}$](BF$_4$)$_{48}$ was synthesized by stirring a mixture of ligand **L2** (9 µmol, 200 µL of a 45 mM stock solution of **L2** in CD$_3$CN) and [Pd(CH$_3$CN)$_4$](BF$_4$)$_2$ (4.5 µmol, 150 µL of a 30 mM stock solution in CD$_3$CN) in 650 µL d_6-DMSO at 70 °C for 3 h to give 1000 µL of a 0.188 mM solution of cage [Pd$_{24}$(**L2**)$_{48}$](BF$_4$)$_{48}$.

The ^{1}H NMR spectrum of the reaction mixture showed broad peaks in the aromatic region (Fig. S19). It is worth noting that line broadening of the NMR signals is expected for assemblies of this size. According to the DOSY NMR spectrum, the peaks belong to species with a uniform diffusion coefficient (Fig. S20). The MS data confirmed that a [Pd$_{24}$(**L2**)$_{48}$]$^{48+}$ assembly had formed (Fig. S21). Unfortunately, the available analytical data do not allow us to draw conclusions about the degree of orientational self-sorting for this system.
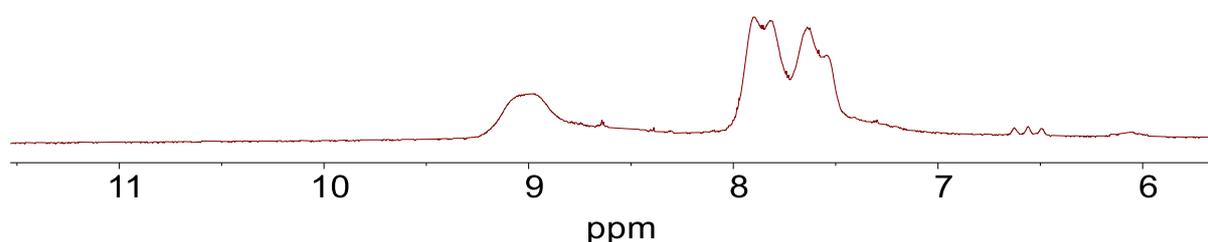
Figure S19. ^{1}H NMR spectrum (800 MHz, CD$_3$CN) of [Pd$_{24}$(**L2**)$_{48}$](BF$_4$)$_{48}$.
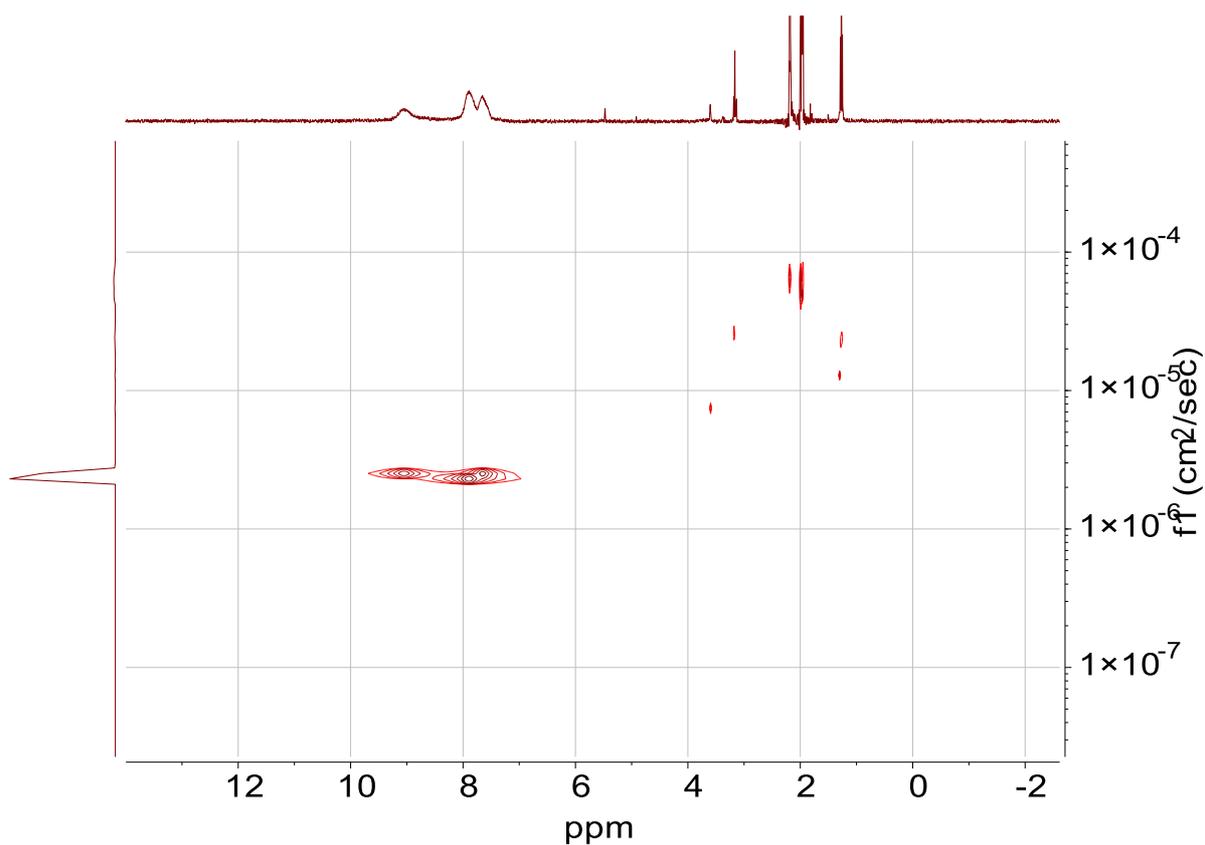
Figure S20. ^{1}H-^{1}H DOSY NMR spectrum (400 MHz, CD$_3$CN) of [Pd$_{24}$(**L2**)$_{48}$](BF$_4$)$_{48}$ (D = 2.31 x 10^{-10} m^2 s^{-1}, r_{dosy} = 28 Å).
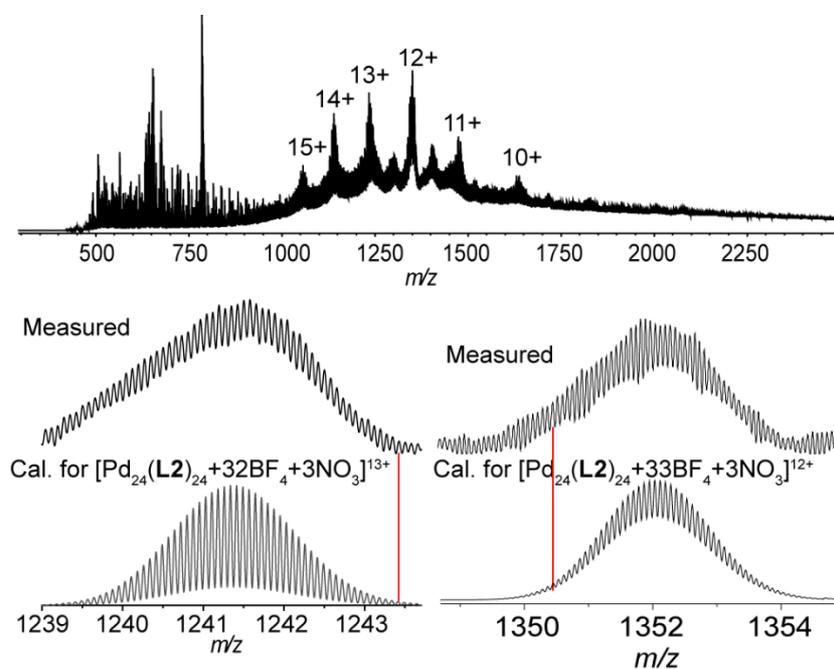
Figure S21. Cryo spray ionization (CSI) mass spectrum of [Pd$_{24}$(**L2**)$_{48}$+nBF$_4$]$^{(48-n)+}$ (n = 33–38).

3 Geometric analyses of potential isomers

3.1 Analysis of $[Pd_{12}L_{24}]^{24+}$
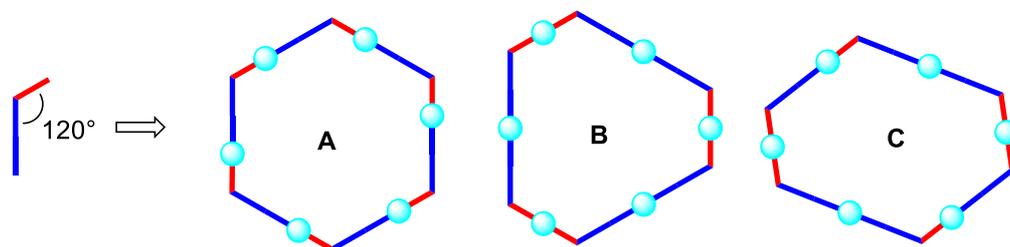
Figure S22. When combined with a linear connector M, a low-symmetry ligand L with a 120° angle between the coordinate vectors can form three different M_6L_6 hexagons (**A**–**C**).

cuboctahedra based on A
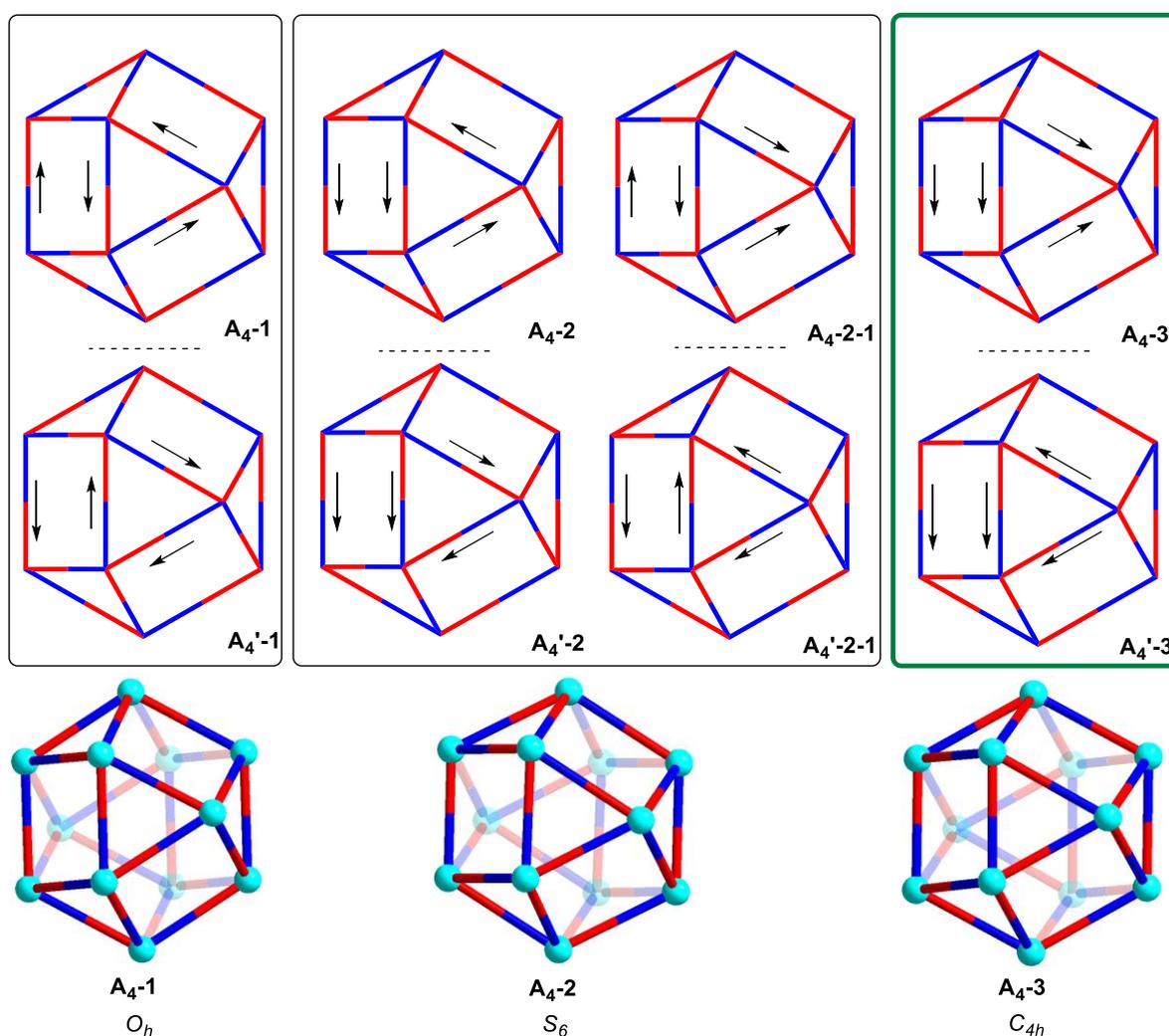
Figure S23. Re-construction of a cuboctahedron from hexagon **A** gives three $M_{12}L_{24}$ isomers. The bottom view shows the relative orientation of the ligands and the point group of the assembly.

cuboctahedra based on C
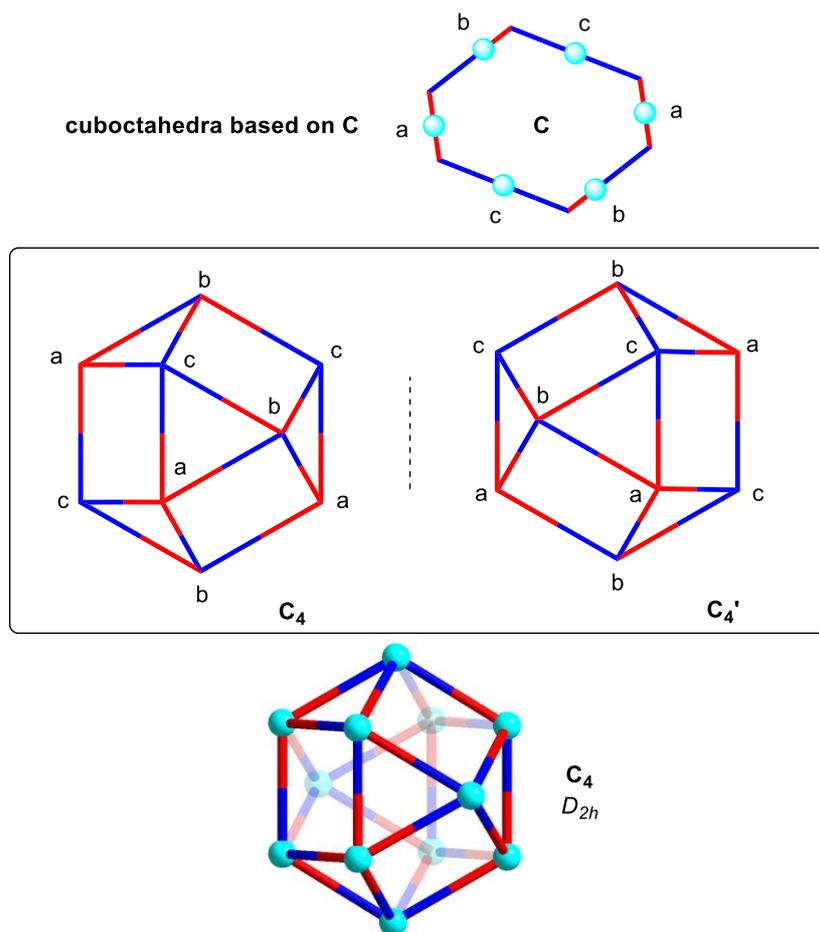
Figure S24. Re-construction of a (distorted) cuboctahedron from hexagon **C** gives one $M_{12}L_{24}$ isomer. The bottom view shows the relative orientation of the ligands and the point group of the assembly (D_{2h}).

chemically distinct ligands:

1	5	3	3
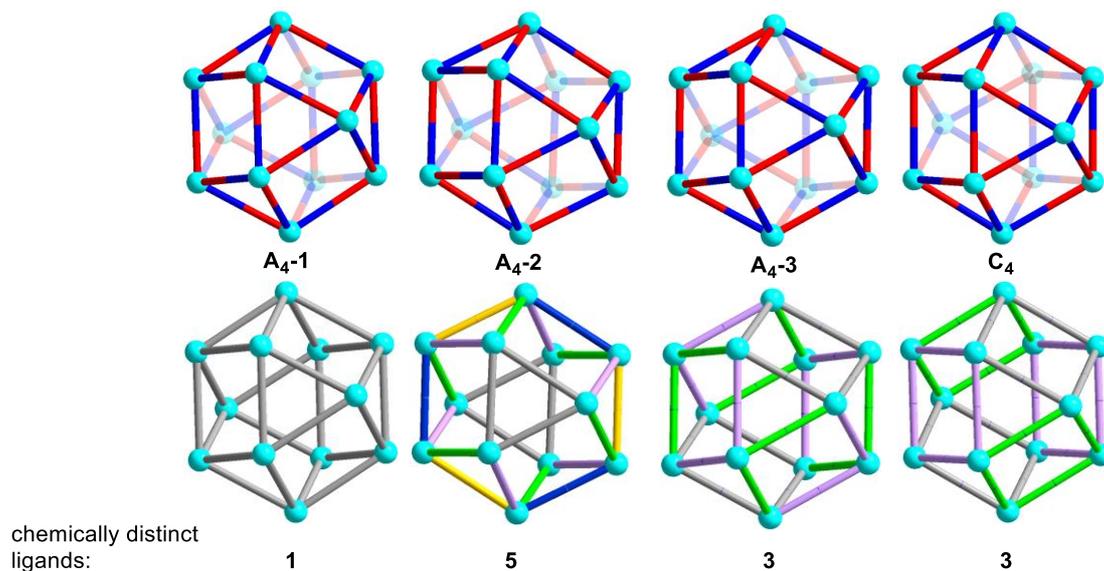
Figure S25. Graphic representations of the structures **A₄-1**, **A₄-2**, **A₄-3**, and **C₄** (top row) and the number of chemically distinct ligands for these assemblies (bottom row). The color coding shows the symmetry-related ligands.

3.2 Analysis of $[Pd_6L_{12}]^{12+}$

Assemblies of type M_6L_{12} can be deconstructed into three orthogonal M_4L_4 tetragons, with the metal centers having an occupancy of 50% (Fig. S26a). Let's consider the case, which is schematically depicted in Fig. S26b. The assemblies are based on ditopic ligands having the following general characteristics: the ligand is perfectly rigid, the donor groups are chemically distinct (red/blue), the coordinate vectors form an angle of 90°, and there is an equal distance between the donor atoms and the intersection of the coordinate vectors ($a = b$). When combined with metal ions that act as linear connectors (e.g. Pd^{2+}), these ligands can form four different M_4L_4 tetragons (**D–G**). The final M_6L_{12} complex can be obtained from all possible **D/E/F/G** combinations. Consequently, there is no intrinsic bias for a particular isomer.
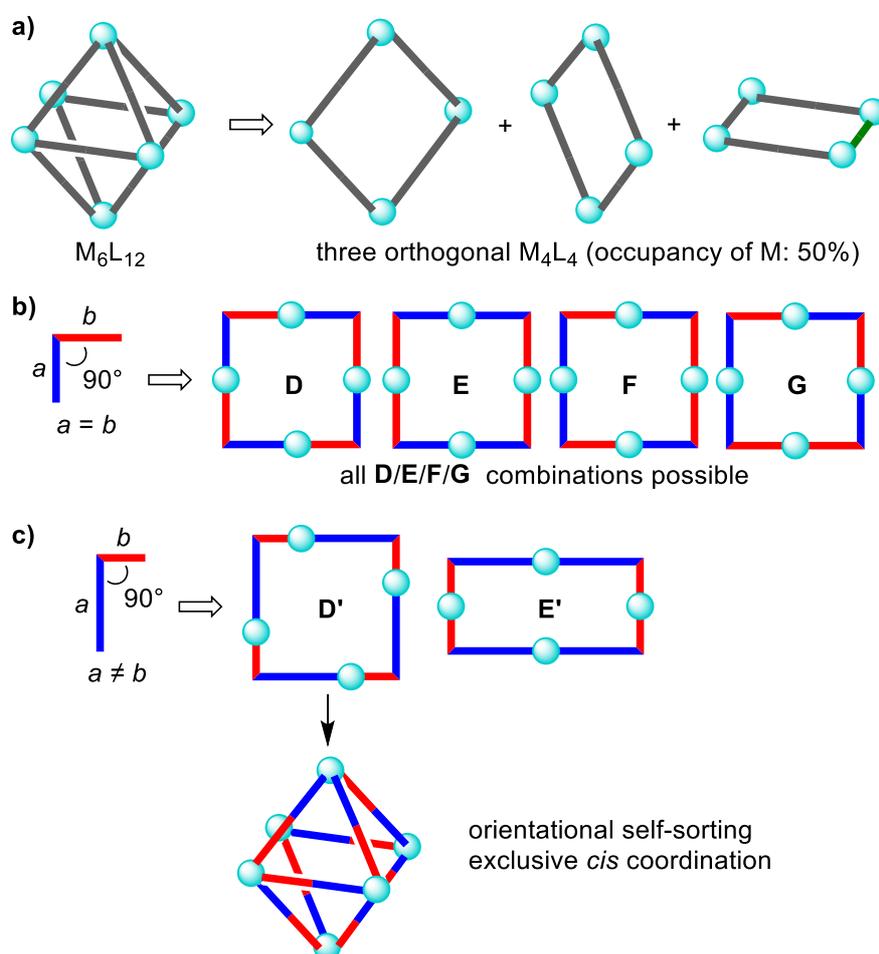
Figure S26. a) Deconstruction of a M_6L_{12} assembly into three tetragons with a metal occupancy of 50%; **b)** A ligand with two distinct donor groups (blue/red), a 90° angle between the coordinate vectors, and equal lengths a and b can form four different M_4L_4 tetragons (**D–G**). All possible combinations of tetragons allow the construction of an M_6L_{12} assembly; **c)** A ligand with two distinct donor groups (blue/red), a 90° angle between the coordinate vectors, and non-equal lengths a and b can form two different M_4L_4 tetragons, **D'** and **E'**. The construction of an M_6L_{12} assembly can only be achieved from **D'**, leading to a complex with exclusive *cis* coordination at the metal centers (for simplicity, the $M_{12}L_{24}$ complex is depicted with linear ligands).

A different situation is encountered when the distance between the donor atoms and the intersection of the coordinate vectors is not equal (Fig. S26c). A geometric match between the ligands and the linear metal (M) connectors is only found for two tetragons, **D'** and **E'**. An M_6L_{12} complex can be formed from three **D'**, but not from three **E'**, or from a mixture of **D'** and **E'** (the M···M distances are not compatible). There is only one way to construct an M_6L_{12} assembly from **D'**, and the resulting complex shows exclusively *cis* coordination at the six metal centers.

The conclusion from this geometrical analysis is the following: for rigid, ditopic ligands with a 90° angle between the coordinate vectors, the distances *a* and *b* between the donor atoms and the intersection of the coordinate vectors are crucial. For ligands with a large difference between *a* and *b*, orientational self-sorting is expected, and the preferred M_6L_{12} isomer has exclusive *cis* coordination at the six metal centers.

3.3 Analysis of $[Pd_{24}L_{48}]^{48+}$

A rhombicuboctahedron can be deconstructed into 8 octagons (Fig. S27). A low-symmetry ligand with a bend angle of 135° and distinct lengths *a* and *b* can form four different octagons (**H**–**K**), assuming that the metals connect the ligands in a perfectly linear fashion. Re-construction of a rhombicuboctahedron can be achieved from **H**, but not from **I**, **J** or **K**, or from mixtures of **H/I/J/K**. The combination of six octagons **H** allows forming six different $M_{24}L_{48}$ isomers. All six isomers display exclusive *cis* coordination at the metal centers. As for octahedral and cuboctahedral assemblies, a *cis* arrangement is favored from a purely geometrical point of view.

a)

$M_{24}L_{48}$ $\implies$ 6 x six M_8L_8 (occupancy of M: 50%)

b)

a
b 135° $\implies$ **H** **I** **J** **K**
$a \neq b$

six **H$_6$** isomers

c
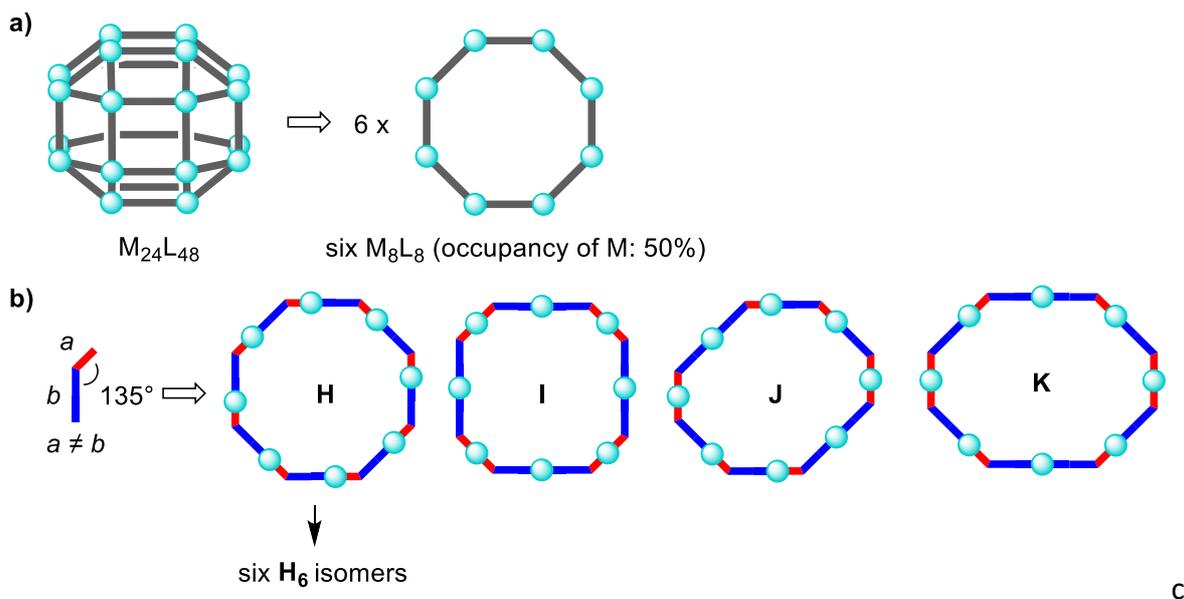
Figure S27. a) Deconstruction of an $M_{24}L_{48}$ assembly into six octagons with a metal occupancy of 50%; **b)** A ligand with two distinct donor groups (blue/red), a 135° angle between the coordinate vectors, and non-equal lengths *a* and *b* can form four different M_8L_8 hexagons (**H**–**K**). The construction of an $M_{24}L_{48}$ assembly can be achieved from hexagon **H** (six possible combinations), but not from **I**, **J** or **K**, or from mixtures of **H/I/J/K**.

Rhombicuboctahedron based on H

H₆-1
T

H₆-2
C₄

H₆-3
C₂

H₆-4
S₆

H₆-5
D₄

H₆-6
S₆
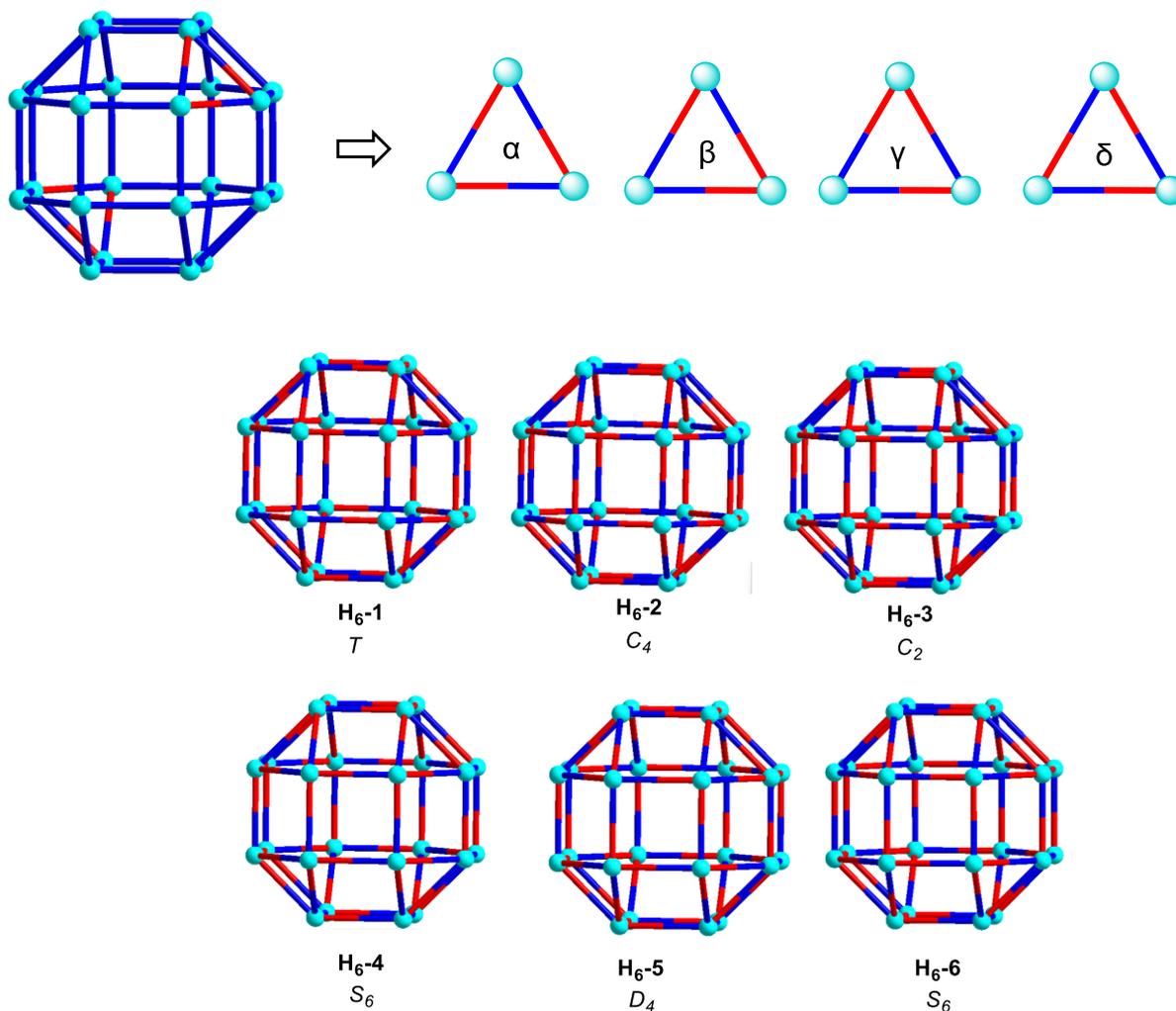
Figure S28. The two opposite triangles in the rhombicuboctahedron define the orientations of the six octagons **H**. There are four triangles with different orientations of the ligands (α, β, γ, δ). The combination of the triangles gives six $M_{24}L_{48}$ isomers (10 isomers in total, **H₆-1**, **H₆-2**, **H₆-3**, **H₆-5** are chiral).

4 Computational analysis of the potential isomers

The algorithm used to count distinct isomers ('configurations') is similar to the one introduced in our previous work,[1] and it is similar in spirit to the one used by Cronin and co-workers.[2] These previous algorithms, while sufficient for smaller assemblies (e.g. octahedral M_6L_{12} cages), are not suitable for larger $M_{12}L_{24}$ and $M_{24}L_{48}$ cages due to the explosion in the number of configurations. As a result, we were forced to modify the primary counting method in a way that could accommodate these larger cages (a modification which we attempt to explain here).

The primary idea of the modification is to count the number of distinct configurations indirectly using use the Orbit–Stabilizer Theorem[3] in a way that is referred to in the combinatorics literature as "Pólya enumeration." The general flow of the algorithms is shown in Figure S29 (the construction of the configurations with at least 1 symmetry) and Figure S30 (the counting of the distinct configurations).
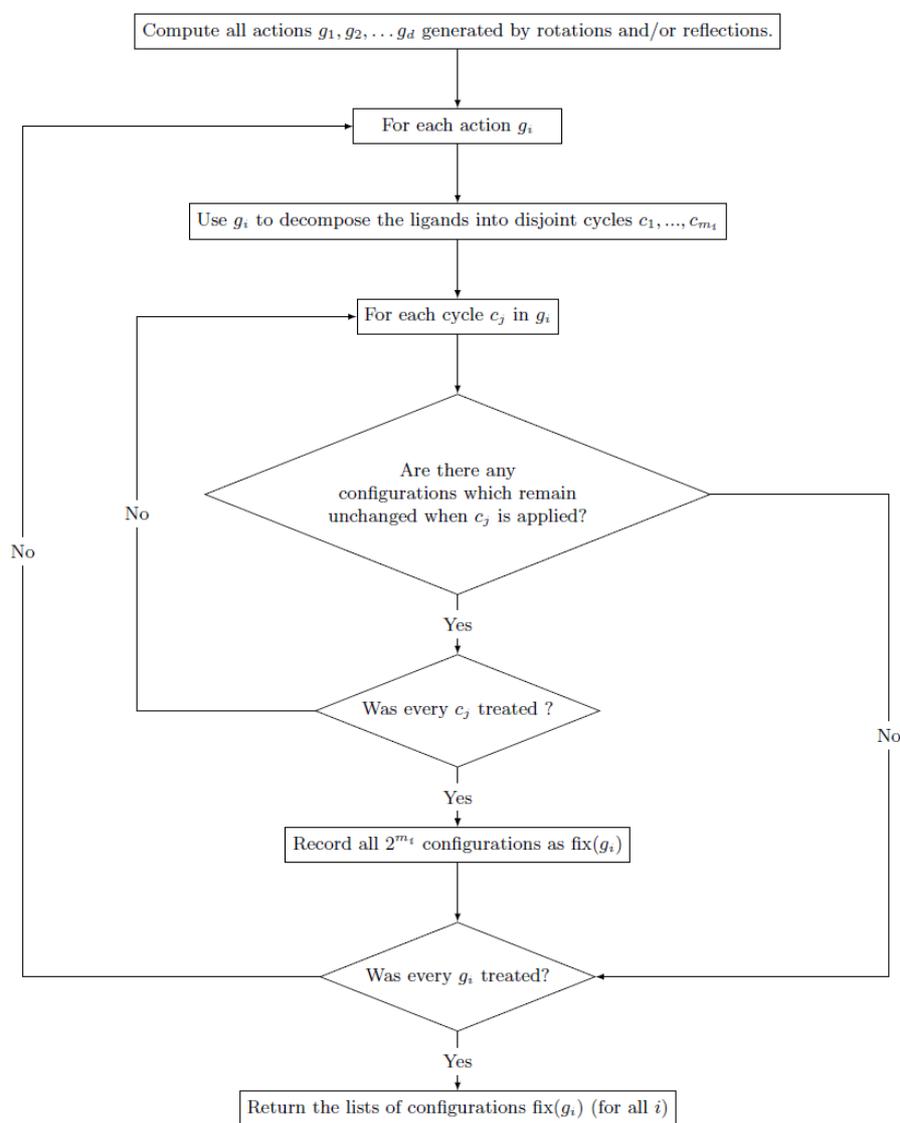
Figure S29. Construction of configurations with at least 1 symmetry.
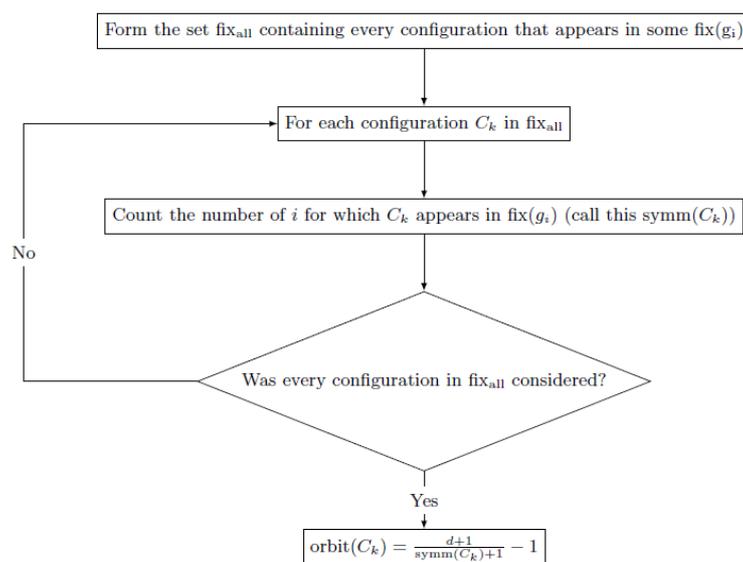
Figure S30. Counting the distinct configurations.

4.1 Counting with the Orbit–Stabilizer Theorem

The Orbit–Stabilizer Theorem[3] provides a relationship between the number of distinct configurations (with respect to a collection of symmetries) and the number of symmetries each configuration possesses. The advantage of this approach comes from the fact that there are a (very) large number of configurations with no symmetries at all. So rather than examining all configurations in an exhaustive manner (as is done previously[1,2]), we instead generate the much smaller set of configurations that have symmetries by iterating over all possible symmetries (and then examine these configurations in an exhaustive manner). The fact that all other configurations have no symmetries is sufficient to use the Orbit–Stablilizer Theorem without having to generate them explicitly. As the algorithm depends directly on our ability to generate all configurations with symmetries correctly and efficiently, we will attempt to explain this part in a way accessible to non-mathematicians. For those interested in learning more, we recommend the work by Dummet and Foote.[3]

We start by explaining the use of the Orbit–Stabilizer Theorem. For the purpose of this explanation, let us define an 'action' on a 3-dimensional cage to be any rigid transformation comprised of a sequence of rotations and/or reflections with the property that every ligand either

1. moves to a place where a different ligand was previously, or
2. stays in the same place but reverses direction.

The number of possible actions will depend on the number of ligands as well as the structure of the cage. This idea of 'actions' is intended to give us a way to count what we know (intuitively) as a 'symmetry' in a configuration. Precisely, a 'symmetry' exists in a configuration whenever there exists an action for which the orientation of ligands is the same before and after that action.

When it comes to the way actions affect ligands, there will be two natural things we will want to consider

- or each ligand *l*, we will consider the set of sites that this ligand visits over repeated application of an action *g* (which we will call its 'cycle under *g*).
- For each action *g*, we will consider the set of configurations for which the action *g* is a symmetry — that is, the configurations before and after applying *g* are the same. We will call such configurations "fixed points of g" and denote the set of these configurations as fix(*g*).

As an example, consider a cube sitting on a table under the action *g* = "spin the cube 90 degrees in such a way that the bottom face stays in contact with the table." Then Figure S31(a) shows the 3 cycles (blue, green, and red) caused by *g*, and (b) and (c) are two of the configurations in fix(*g*) for this *g*.
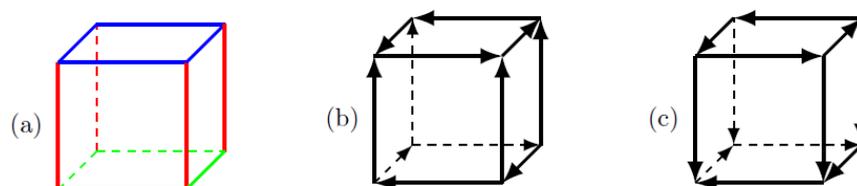
Figure S31. (a) The cycles (red, blue, green) formed by *g* and (b), (c) two elements of fix(*g*) (where *g* is the 90 degree spin that leaves the bottom face touching the table).

It is possible for a given configuration to be fixed with respect to multiple actions (say g_1 and g_2), in which case it will be listed in fix(g_1) and fix(g_2). Returning to the example above, each of the configurations shown in Figure S31 will appear in fix(*g*) for (at least) 3 different actions *g*:

1. the 90 degrees turn that keeps the bottom face in contact with the table
2. the same turn, but 180 degrees
3. the same turn, but 270 degrees

Note that if we do the same turn, but 360 degrees, then it takes us back to where we started (so is not a true "action") and larger degrees simply repeat the actions listed above (for example, 450 degrees is the same as 90 degrees as far as the ligands are concerned).

As mentioned previously, the Orbit–Stabilizer Theorem provides a relationship between the number of distinct configurations that one can obtain by applying a collection of actions and the number of symmetries it has, and the number of symmetries a configuration has can now be determined by counting the number of times a configuration appears in some fix(*g*). Specifically, if there are *d* distinct actions and we define

- orbit(*C*) to be the number of different configurations that can be achieved by applying all possible actions to the configuration *C*, and
- symm(*C*) to be the number of times the configuration *C* appears in some fix(*g*),

then[a]

---

[a] Note that is not true for any collection of actions — the actions must form what is known in mathematics as a "group" — but it will apply for collections of the type "all actions that correspond to rotations" and "all actions that correspond to combinations of rotations and reflections" which are the ones we will be interested in here.

$$(\text{orbit}(C) + 1) \times (\text{symm}(C) + 1) = d + 1$$

Note that the +1 appears each time because we need to include the original configuration in each count (not just the ones generated by actual actions).[b]

Another way to think about orbit(C) is as the set of configurations that we want to treat as being 'the same' with respect to these actions. So once we know the number of configurations in orbit(C) for every C, we know how many configurations to subtract from (or divide out of) the total.

4.2 Computational considerations

The main reason for using the approach above (as opposed to the brute force method) is computational. For $n > 20$, the ability to work with all of the 2^n different configurations becomes increasingly difficult. The method discussed above would (theoretically) allow us to work with a substantially smaller set of configurations and as a result could (theoretically) be applied to isomers with a larger number of ligands. Our use of the word "theoretically" at this point is due to the fact that there are still two things that will affect the running time of the algorithm:

- The number of configurations with at least 1 symmetry
- The construction of the sets fix(g) for each g (this was done in an exhaustive manner in our previous work,[1] which will be too slow)

Both issues can be resolved using the cycle decompositions mentioned earlier. A configuration remains fixed under a given action if and only if all of its cycles remain fixed under that action. Furthermore, it is easy to calculate which which ligand orientations that can remain fixed within a given cycle.

- If a cycle results in a ligand returning to the same location but with opposite orientation (like a reflection) than there will be 0 orientations that remain fixed
- Otherwise, there will be 2 orientations that remain fixed — start with one ligand (in either of two possible orientations) and then orient the other ligands in whatever way the action tells you to.

The set of configurations fixed by an action g will then consist of all possible ways of combining the fixed cycle orientations — a total of 0 (if any of the cycles has 0 orientations) or 2^m. Furthermore, we can record these configurations efficiently by simply recording the cycle orientations (and not building the full configurations explicitly).

Hence we can build each fix(g) quickly, and so it remains to see how many configurations will arise out of this procedure. For this, let us return to the definition of 'actions' and note that if the second possibility happens (some ligand stays in the same place but reverses direction), then fix(g) will have no configurations in it. So we only need to consider those actions g for which the first possibility (every ligand moves to a place where a different ligand was previously). But the fact that every ligand moves to a new place means that every cycle formed by such an action must have at least 2 ligands in it. Since the cycles are disjoint, this means an isomer with n ligands will be able to have at most $n/2$ cycles under

---

[b] Another way to accomplish this (and the way it is typically done in mathematics) is to allow for a 'do nothing' action — for this reason, formulas here will differ slightly from ones in mathematical texts by this factor of 1.

such an action g. Hence the maximum number of configurations that we will need to consider for each g will be $2^{n/2}$.

This gives us an indication as to which cages we expect that this algorithm will be able to handle (essentially double the number of ligands we were able to handle previously). This makes the cuboctahedron (24 ligands) easily accessible using this method but still puts the rhombicuboctahedron (48 ligands) in precarious territory. However, some small modifications can be applied that speed up the handling of the extreme case (when all cycles are size 2) and this speeds up the algorithm enough to allow it to handle the rhombicuboctahedron as well. Further improvements (and possibly a different approach) would be needed to handle isomers with more than 48 ligands.

4.3 Numerical results

Here we present the exact results for $M_{12}L_{24}$ (the cuboctahedron) and $M_{24}L_{48}$ (the rhombicuboctahedron). We will use the following notation: for an integer t, we will let

1. a_t = the number of configurations with $t-1$ reflection symmetries
2. b_t = the number of configurations with $t-1$ rotation/reflection symmetries
3. c_t = the number of chiral pairs in a_t
4. d_t = the number of non-chiral (reflection symmetric) configurations in a_t

Note that the $t-1$ comes from the fact that the counts in a_t and b_t include the original configurations as well. In both cases (cuboctahedron and rhombicuboctahedron), a_t and b_t were computed using the program above and then c_t and d_t derived using the identities

$$a_t = 2c_t + d_t \text{ and } b_{2t} = c_{2t} + d_t$$

Table S1. Results for $M_{12}L_{24}$ (the cuboctahedron).

t	1	2	3	4	6	8	12	24
a_t	697501	2982	119	63	14	6	1	2
b_t	348425	2124	56	63	14	12	0	2
c_t	348425	1473	56	27	7	3	0	1
d_t	651	36	7	9	0	0	1	0

Table S2. Results for $M_{24}L_{48}$ (the rhombicuboctahedron).

t	1	2	3	4	6	8	12	24
a_t	11728117728726	12577572	32634	4682	252	60	6	4
b_t	5864058514859	6987282	16253	3365	254	30	3	2
c_t	5864058514859	6288274	16253	2341	126	30	3	2
d_t	699008	1024	128	0	0	0	0	0

The code for the computations is publicly available at doi: <u>10.5281/zenodo.6614832</u>

5 Computational modelling of cage structures

5.1 Cage construction

Using *stk*,[4,5] we built a single conformer of ligand **L1**, then ran a CREST[6] conformer search in the gas phase using an energy window of 5 kcal mol^{-1} to get the lowest energy conformer, which was further optimized at the GFN2-xTB[7,8] level of theory with "extreme" convergence criteria (energy converged within 5E-8 E_h and gradient norm converged within 5E-5 $E_h$bohr^{-1}) and used to build all subsequent cages.

The cage construction process places palladium atoms on the four-connected nodes and the ligand on two-connected nodes of the $M_{12}L_{24}$ topology graph defined in *stk*. The ditopic ligands are aligned such that their pyridyl functional groups point toward their neighboring Pd atom. Different isomers are defined by modifying the relative orientation of each ligand. The default alignment in *stk* does not fit with a specific symmetry and is used here as a benchmark for a low-symmetry isomer ("**def**"). The different ligand orientations (termed "vertex alignment" in *stk*) required to generate the four expected isomers, "**A₄-1**", "**A₄-2**", "**A₄-3**" and "**C₄**" were determined by visual inspection. The maps are provided in the code https://github.com/andrewtarzia/big_unsymm/blob/main/build_cage.py.

Each cage has a charge of 24+ and undergoes an optimization sequence similar to recent work on unsymmetrical Pd_2L_4 cages:[9]

1. *stk* assembles structures based on predefined topology graphs with unphysical, long bonds between building blocks (nodes).

2. The cage structure is geometry optimized using UFF[10–12] in the General Utility Lattice Program (GULP; version 5.1),[13,14] with the conjugate gradient method and then the Newton-Raphson method. The atom typing is handled by a Python implementation of the "ForceFieldHelpers" module in RDKit,[15] except for the metal atoms, which are manually typed to match the target types in UFF (because RDKit does not handle the metal-atom typing). Palladium atoms are assigned the square planar atom type, "Pd^{2+}". The bonding used within GULP matches the bonding in the *stk* molecule.

3. A conformer search is performed starting from the UFF optimized cage structure using high-temperature molecular dynamics (MD). Two sequential MD runs in the NVT ensemble, using the leapfrog verlet integrator, are performed using UFF and GULP at 1000 K. The first run is a short equilibration with a time step of 0.25 fs for 1.0 ps. The production run is performed for 200.5 ps with a time step of 0.75 fs. From the production run, 40 conformers are extracted at 5.0 ps intervals.

4. Each extracted conformer is optimized using UFF in Gulp.

5. The lowest energy cage conformer is optimized using GFN2-xTB[7,8] in the xtb software[8] with the extreme convergence criteria in the gas phase.

6. Finally, using Gaussian16 code[16], density functional theory calculations (DFT) were also performed to optimize the cage structures. For these calculations, hybrid (PBE0D3BJ[17]) and screened-hybrid (HSE06[18,19]) functionals were considered.

5.2 Density functional theory calculations

We performed DFT geometry optimizations to validate the relative energetics of **def**, **A₄-1**, **A₄-2**, **A₄-3**, and **C₄** cage structures. Hybrid and hybrid-screened methods on the models mentioned above were performed using the Gaussian16 code.[16] The calculations were all-electron (*i.e.*, no pseudopotentials were employed) and the Ahlrichs basis set Def2-SVP[20] was used to expand the wavefunctions, which has polarization in all atoms. Each palladium has a charge of 2+ (giving the system a charge of 2*n*, where *n* is the number of Pd atoms) and was assumed to be in the low-spin state.

As shown in Debefve and Pollock's work,[21] the Perdew-Burke-Ernzerhof hybrid functional (PBE0)[17] is suitable for modelling structures of complexes containing third-row transition metals. PBE0 contains calibrated parameters only in the added dispersion term, which promises low-cost computational geometry optimizations for large systems.[22] Grimme's D3BJ dispersion correction[23] was used in PBE0 calculations together with an accurate pruned grid ('Superfinegrid' option in Gaussian16), as recommended for optimizations of large systems.[16] Relative energies obtained from the optimized cages suggest that the three **A₄** isomers are more stable than isomers **C₄** and **def**.

To take into account the effects of nonlocal exchange, we applied the screened hybrid functional developed by Heyd, Scuseria and Ernzerhof (HSE06).[18,19] This functional is similar to PBE0 but uses a screened Coulomb interaction for the exact exchange part with an empirically determined screening parameter.[24] It is well known that the HSE06 functional predicts accurate electronic structures and band gap values[25–29] and exhibits significant improvement with respect to PBE0 as shown in Cui *et al.*.[24] However, this functional allowed us to describe the geometries and relative energies of our models, taking **A₄-3** as the most stable cage which agrees with previous GFN2-xTB calculations and experimental data.

It is imperative to understand that all the results obtained through these approximations (GFN2-xTB and DFT calculations) demonstrate the difficulty in discerning which is the most stable cage, especially between **A₄** isomers (~4 kJ mol⁻¹).

Input files and output files for DFT calculations can be accessed at:

https://github.com/andrewtarzia/citable_data/tree/master/li_2022

5.3 Cage analysis
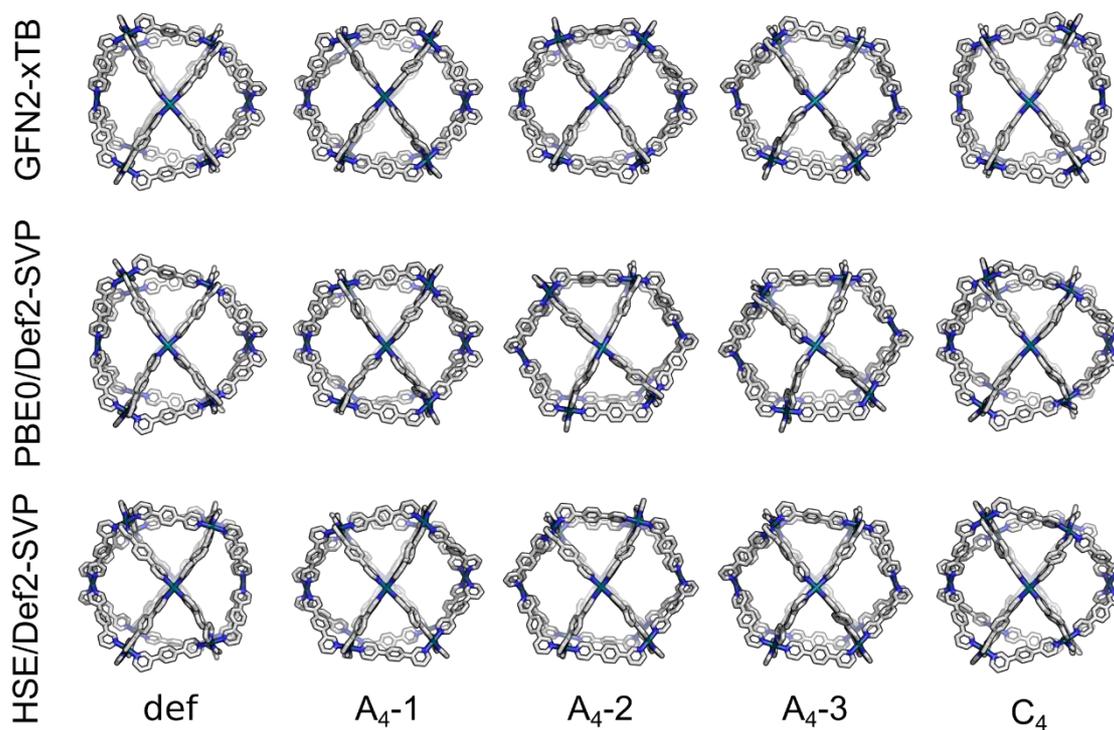
Figure S32. [Pd$_{12}$L$_{24}$]$^{24+}$ structures of five cage isomers after (top) GFN2-xTB, (middle) PBE0/Def2-SVP and (bottom) HSE/Dev2-SVP optimization. Hydrogen atoms are omitted for clarify, carbon atoms are grey, nitrogen is blue, and palladium is cyan.

Table S3. Relative energies of the five Pd$_{12}$L$_{24}$ cage isomers (in kJ mol^{-1}) for different methods

Method	def	A$_4$-1	A$_4$-2	A$_4$-3	C$_4$
GFN2-xTB	117.73	44.47	7.64	0.00	104.45
PBE0-D3BJ/Def2-SVP	155.05	0.00	7.09	1.93	152.82
HSE/Def2-SVP	158.53	7.33	4.21	0.00	156.51

5.4 Analysis of $Pd_{24}L_{48}$ systems

Models of six $Pd_{24}L_{48}$ isomers were constructed manually as xyz files. Each structure was geometry optimized using GFN2-xTB to the 'normal' convergence criteria (energy converged within 5E-6 E_h and gradient norm converged within 1E-3 $E_h$bohr^{-1}) and their final energies shown in Table S4. Figure S31 shows their structures. All files are available at:

https://github.com/andrewtarzia/citable_data/tree/master/li_2022

Table S4. Relative energies of the six $Pd_{24}L_{48}$ isomers (in kJ mol^{-1}) after GFN2-xTB optimization

Method	H_6-1	H_6-2	H_6-3	H_6-4	H_6-5	H_6-6
GFN2-xTB	0.0	7.9	7.3	5.1	8.4	15.4
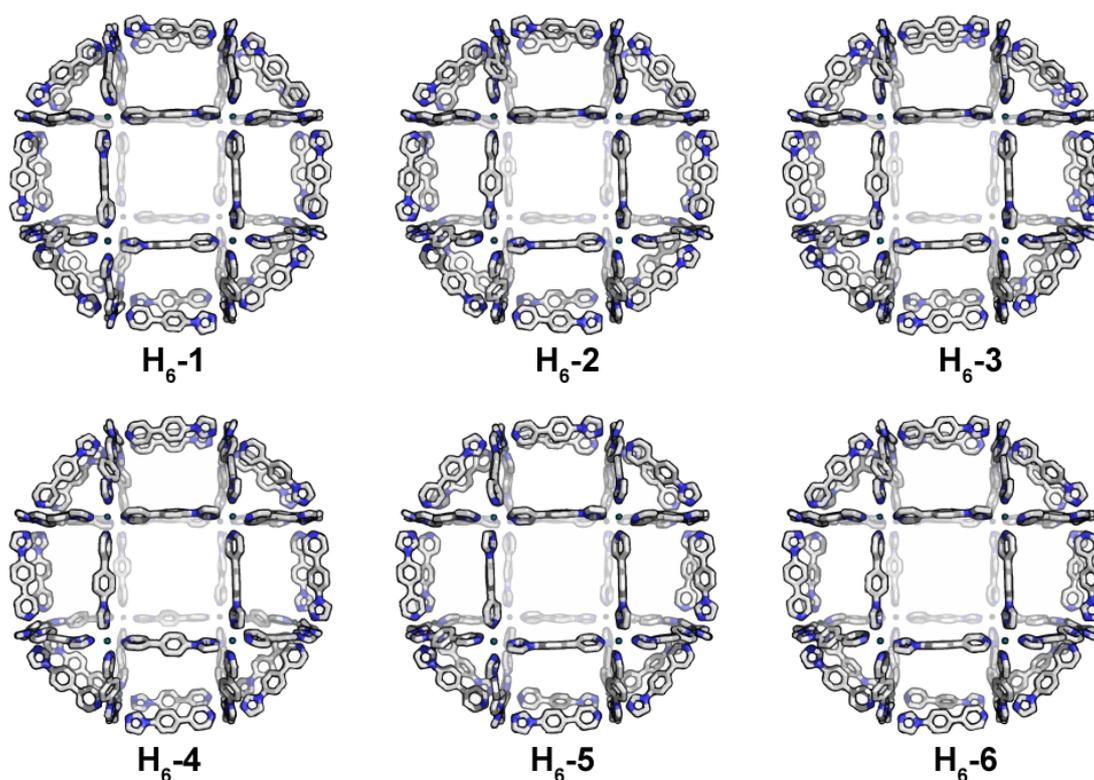
Figure S33. $[Pd_{24}L_{48}]^{48+}$ structures of six cage isomers after GFN2-xTB optimization. Hydrogen atoms are omitted for clarify, carbon atoms are grey, nitrogen is blue, and palladium is cyan.

6 Crystallographic analysis

A single colorless block crystal of $[Pd_{12}(L1)_{24}]Cl_{24}$ was obtained by slow vapor diffusion of THF/Et2O (v/v = 1 : 1) into a DMSO solution of $[Pd_{12}(L1)_{24}](NO_3)_{24}$ (0.375 mM) for several months. The origin of the chloride anions is not known. Maybe there were traces of chloride in the Pd(NO₃) starting material (for a discussion of this problem, see: https://pubs.acs.org/doi/10.1021/acs.inorgchem.1c01714?ref=PDF), and the cage with chloride anions crystallized preferentially, or there was some anion exchange with the glassware during the very long crystallization period.

Bragg-intensities of $[Pd_{12}(L1)_{24}]Cl_{24}$ were collected at 140 K using Cu$K\alpha$ radiation. A suitable crystal with dimensions 0.23 × 0.16 × 0.08 mm3 was selected and mounted on an XtaLAB Synergy R, DW system, and HyPix-Arc 150 diffractometer. The dataset was reduced and corrected for absorption, with the help of a set of faces enclosing the crystals as snugly as possible, with the latest available version of *CrysAlis*^{Pro}.[30]

The structure solution and refinement were performed by the latest available version of *ShelXT*[31] and *ShelXL*[32] using *Olex2*[33] as the graphical interface. All non-hydrogen atoms were refined anisotropically using full-matrix least-squares based on $|F|^2$. The hydrogen atoms were placed at calculated positions employing the "riding" model, where each H-atom was assigned a fixed isotropic displacement parameter with a value equal to 1.2 U_{eq} of its parent C-atom. The RIGU and SIMU restraints were applied to the displacement parameters of all non-hydrogen atoms. Additional counter-ions and solvent molecules, too disordered to be located in the electron density map, were taken into account using the *Olex2*[33] solvent-mask procedure.

Crystallographic and refinement data of $[Pd_{12}(L1)_{24}]Cl_{24}$ are summarized in Table S5. The CCDC number 2161290 contains the crystallographic data for the compound $[Pd_{12}(L1)_{24}]Cl_{24}$, in this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif.
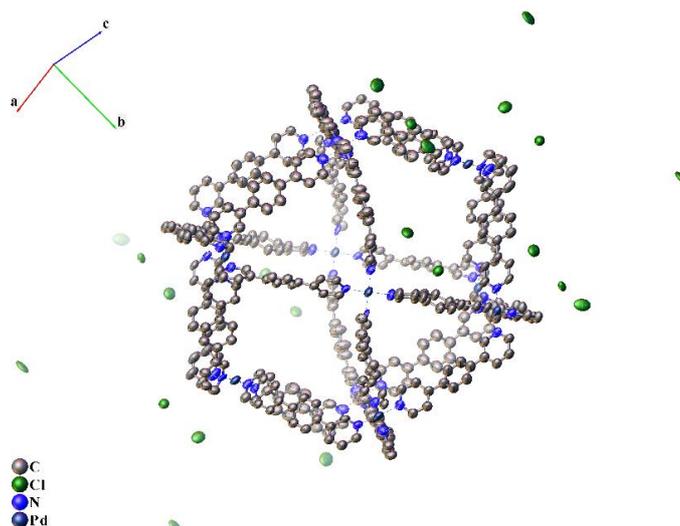
Figure S32. Molecular structures of $[Pd_{12}(L1)_{24}]Cl_{24}$ as determined by X-ray crystallography. Displacements ellipsoids are drawn at the 50% probability level. Hydrogen atoms have been omitted for clarity.

Table S5. Crystal data and structure refinement for $[Pd_{12}(L1)_{24}]Cl_{24}$

Compound	$[Pd_{12}(L1)_{24}](Cl)_{24}$
Formula	$C_{384}H_{288}Cl_{24}N_{48}Pd_{12}$
$D_{calc.}$/ g cm^{-3}	0.267
μ/mm^{-1}	1.286
Formula Weight	7702.21
Color	colourless
Shape	irregular-shaped
Size/mm^3	0.22×0.13×0.07
T/K	140.00(10)
Crystal System	orthorhombic
Space Group	*Cmme*
a/Å	54.9549(13)
b/Å	61.3516(13)
c/Å	56.7355(12)
α/°	90
β/°	90
γ/°	90
V/Å^3	191288(7)
Z	4
Z'	0.25
Wavelength/Å	1.54184
Radiation type	Cu*Kα*
Θ_{min}/°	2.121
Θ_{max}/°	44.573
Measured Refl's.	243338
Ind't Refl's	38617
Refl's with I > 2σ(I)	14506
R_{int}	0.1401
Parameters	871
Restraints	1639
Largest Peak/e Å^{-3}	1.366
Deepest Hole/e Å^{-3}	-0.948
GooF	1.107
wR_2 (all data)	0.4058
wR_2	0.3536
R_1 (all data)	0.1868
R_1	0.1224
CCDC number	2161290

7 References

1   R.-J. Li, A. Marcus, F. Fadaei-Tirani and K. Severin, *Chem. Commun.*, 2021, **57**, 10023–10026.
2   B. Kandasamy, E. Lee, D.-L. Long, N. Bell and L. Cronin, *Inorg. Chem.*, 2021, **60**, 14772–14778.
3   D. S. Dummit and R. M. Foote, *Abstract Algebra*, 3rd *Ed.*, Wiley-VCH, 2003.
4   L. Turcani, A. Tarzia, F. T. Szczypiński and K. E. Jelfs, *J. Chem. Phys.*, 2021, **154**, 214102.
5   https://github.com/lukasturcani/stk, version 2022.1.26.0.
6   P. Pracht, F. Bohle and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 7169–7192.
7   C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *WIREs Comput. Mol. Sci.*, 2021, **11**, e1493.
8   C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
9   A. Tarzia, J. E. M. Lewis and K. E. Jelfs, *Angew. Chem. Int. Ed.*, 2021, **60**, 20879–20887.
10 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 2002, **114**, 10024–10035.
11 M. A. Addicoat, N. Vankova, I. Farjana Akter and T. Heine, *J. Chem. Theory. Comput.*, 2014, **10**, 880–891.
12 D. E. Coupry, M. A. Addicoat and T. Heine, *J. Chem. Theory Comput.*, 2016, **12**, 5215–5225.
13 D. J. Gale and L. A. Rohl, *Molecular Simulation*, 2003, **29**, 291–341.
14 D. J. Gale, *J. Chem. Soc., Faraday Trans.*, 1997, **93**, 629–637.
15 G. A. Landrum, RDKit: Open-Source Cheminformatics., http://www.rdkit.org/, (accessed September 5, 2021).
16 M. J. et al. Frisch, *Gaussian 16 Rev. C.01.*
17 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
18 J. Heyd, E. G. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.
19 J. Heyd, E. G. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **124**, 219906.
20 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297.
21 L. M. Debefve and C. J. Pollock, *Phys. Chem. Chem. Phys.*, 2021, **23**, 24780–24788.
22 D. Bakowies and O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2021, **17**, 4872–4890.
23 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
24 Z.-H. Cui, Y.-C. Wang, M.-Y. Zhang, X. Xu and H. Jiang, *J. Phys. Chem. Lett.*, 2018, **9**, 2338–2345.
25 G. Zhu, Q. Sun, Y. Kawazoe and P. Jena, *Int. J. Hydrog. Energy*, 2015, **40**, 3689–3696.
26 V. Posligua, A. Aziz, R. Haver, M. D. Peeks, H. L. Anderson and R. Grau-Crespo, *J. Chem. Phys. C*, 2018, **122**, 23790–23798.
27 J.-W. Song, G. Giorgi, K. Yamashita and K. Hirao, *J. Chem. Phys.*, 2013, **138**, 241101.
28 V. Posligua, D. Pandya, A. Aziz, M. Rivera, R. Crespo-Otero, S. Hamad and R. Grau-Crespo, *J. Phys: Energy*, 2021, **3**, 034005.
29 J. Zhao, Q. Du, S. Zhou and V. Kumar, *Chem. Rev.*, 2020, **120**, 9021–9163.
30 *CrysAlis^{Pro}* software system, Rigaku Oxford Diffraction, (2021).
31 G. M. Sheldrick, *Acta Cryst.*, 2015, **A71**, 3–8.
32 G. M. Sheldrick, *Acta Cryst.*, 2015, **C71**, 3–8.
33 O. V. Dolomanov, J. L. Bourhis, R. J. Gildea, J. A. K. Howard and H. Puschmann, *J. Appl. Cryst.*, 2009, **42**, 339–341.