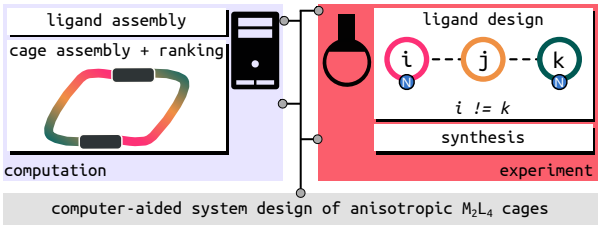
High-throughput Computational Evaluation of Low Symmetry Pd_2L_4 Cages to Aid in System Design

Andrew Tarzia, James E. M. Lewis*, Kim E. Jelfs*

May 17, 2021



*Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, White City Campus, Wood Lane, London, W12 0BZ, UK; james.lewis@imperial.ac.uk, kim.jelfs@imperial.ac.uk

Abstract

The use of unsymmetrical components in metallo-supramolecular chemistry allows for low-symmetry architectures with anisotropic cavities toward guest-binding with high specificity and affinity. Unsymmetrical ditopic ligands mixed with Pd(II) have the potential to self-assemble into reduced symmetry Pd_2L_4 metallo-architectures. Mixtures of isomers can form, however, resulting in potentially undesirable heterogeneity within a system. Therefore it is paramount to be able to design components that preferentially form a single isomer. Previous data suggested that computational methods could predict with reasonable accuracy whether unsymmetrical ligands would preferentially self-assemble into a single isomer under constraints of geometrical mismatch. We successfully apply a collaborative computational and experimental workflow to mitigate costly trial-and-error synthetic approaches. Our low-cost computational workflow rapidly constructs new unsymmetrical ligands (and Pd_2L_4 cage isomers) and ranks their likelihood for forming *cis*-Pd_2L_4 assemblies. From this narrowed search space, we successfully synthesised four new low-symmetry, *cis*-Pd_2L_4 cages, with cavities of different shapes and sizes.

Introduction

Nature has evolved spectacular control over self-assembly processes to produce biological machinery for which high-fidelity of composition and structure is essential for effective functionality. The exploitation of non-covalent interactions allows complex architectures, such as enzymes, to exhibit high substrate specificity through precise control of binding-site size, shape and positioning of functional groups. Over the last few decades, chemists have made great strides in developing approaches to utilise these principles for artificial systems. Metallosupramolecular chemistry has become a prevalent method for assembling ever-more-complex architectures using the predictable coordination geometry of transition metal ions.[1–3]

Since first being reported over twenty years ago,[4] lantern-type Pd_2L_4 cages,[5–7] assembled from "naked" Pd(II) ions and ditopic ligands (L), have become an extensively studied class of metal-organic polyhedra (MOPs).[8–12] Wide-ranging applications for these cages have been investigated, including in drug delivery,[13–15] biomedicine,[16–21] catalysis,[22–25] and guest encapsulation/recognition.[26–32] To simplify the self-assembly process, most previous reports have focussed on high-symmetry systems derived from single, symmetrical ligands. However, it is expected that through the controlled introduction of asymmetry, cages could be designed with more intricate, anisotropic binding sites with specific shapes and functionalities.[33]

Pore asymmetry in M_2L_4 systems has been introduced through the controlled assembly of heteroleptic[34–36] and heteronuclear architectures.[37,38] Mixed-ligand $[Pd_2L^a_2L^b_2]$ assemblies have been realised through both steric[39,40] and geometric control.[41,42] Clever and co-workers have demonstrated the effectiveness of pore asymmetry for improved binding of bent guests over linear counterparts.[43] Crowley and co-workers recently reported a $[PdPtL_4]$ cage in which the different labilities of the two metal ions allowed selective sequestration of the Pd(II) ions to open the cage without complete dissociation of the ligands.[44] We[45,46] and others[47–50] have recently begun to explore an alternative approach that uses unsymmetrical ligands to access lower symmetry structures.[51] The lack of bilateral symmetry introduced into the ligand structure means four possible isomers of the resultant dipalladium cage can form (Figure 1). As with heteroleptic structures, high-fidelity self-sorting into a single isomer can be achieved using steric and/or geometric constraints. However, the inherent difficulty

in designing such ligands that will reliably self-sort helps explain the paucity of examples in the literature.

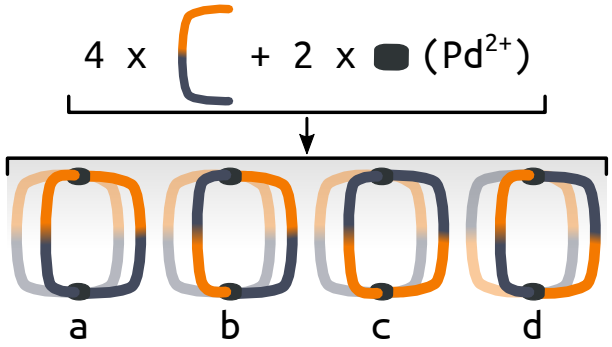


Figure 1: Schematic representation of the self-assembly of an unsymmetrical ditopic ligand and palladium(II) into four possible isomers of the homoleptic Pd_2L_4 cage: (a) "all-up", (b) "three-up-one-down", (c) *cis* and (d) *trans*. Orange and navy colours indicate inequivalent ligand fragments.

Density functional theory (DFT) calculations have been previously used to rationalise experimentally observed self-sorting in low symmetry metallo-supramolecular systems by exploring the relative energies of the potential configurational isomers.[41,43,45,47,48] Indeed, in recent work, we found that the formation of a single Pd_2L_4 cage isomer from the self-assembly of unsymmetrical ditopic ligands with Pd(II) ions only occurred when there was a significant difference in the calculated energies (on the order of at least $5\,kJ\,mol^{-1}$) of the possible isomers.[45] As such, we envisaged that a high-throughput computational workflow could be used to rapidly explore the chemical space of low symmetry MOPs and aid in their design and minimise trial-and-error experimental efforts.

Computational screening has been successfully applied to aid in the rationalisation and/or prediction of self-sorting outcomes of porous organic cages using the relative energetics of possible products.[52–55] Until recently, however, it has not been possible to develop equivalent screening workflows for MOPs because no open-source structure generation software was available. Some of us previously developed the supramolecular toolkit (*stk*),[56,57] an open-source Python framework that handles the structure prediction of supramolecular architectures. Here, we highlight the first use of *stk* to screen candidate MOPs. Young and co-workers also recently developed the software *cgbind*, which performs structure prediction

of M_2L_4 cages.[58] The generalisability of *stk*, however, makes it ideal for this work, where we aim to explore a diverse set of ligand and cage structures.

In this work we present a high-throughput computational workflow that was used to construct 60 unsymmetrical, ditopic ligands and the four possible Pd_2L_4 cage isomers for each *in silico*. Using metrics of geometrical stability and relative cage energies, the ligands were ranked based on their likelihood to form a single Pd_2L_4 isomer. A selection of five ligands with a range of rankings and chemistries were subsequently realised experimentally, and their self-assembly examined. This computer-aided approach facilitates an experimental design with a high success rate, leading to several new unsymmetrical *cis*-Pd_2L_4 cages being prepared. In this manner, the discovery of interesting candidates can be accelerated by providing a likelihood of success, expediting the synthesis of low symmetry MOPs with desirable structural properties. Indeed, this work highlights that a computer-aided approach allows a more efficient exploration of a larger chemical space of potential candidates than a purely experimental approach.

Results

Semi-empirical methods for cage relative energy evaluation

In previous work it was found that the energy separations of the isomers of $[Pd_2L_4]^{4+}$ cages assembled from unsymmetrical ligands, calculated using DFT methods, correlated well with experimental observations of single isomer formation.[45] For larger data sets it would be desirable to use efficient, semi-empirical methods for geometry optimisations and energy calculations to reduce computational cost and increase throughput. Therefore, we tested whether the xTB family of semi-empirical methods[59] could capture the same relative energy differences between Pd_2L_4 cage isomers as DFT methods. The xTB methods are tight-binding quantum chemical methods for the geometry optimisation of systems containing elements up to $Z = 82$, and represent a robust and significantly cheaper alternative to DFT for metal-containing species.[60] A comparison of the xTB (specifically GFN2-xTB) and DFT-calculated energies of previously reported systems was undertaken (structures were taken directly from

the computational workflow described below). DFT energies were obtained from single-point energy evaluations of xTB geometries using similar methods to those recently applied to related systems[25] (PBE0[61] level of theory using the Ahlrichs basis set def2-SVP,[62,63] Grimme's D3BJ dispersion correction[64] and the polarizable continuum model (PCM)[65] implicit solvation representing DMSO; more details are available in Supporting Information Section S3). The same trends in relative energies were found from both methods (Figure 2), suggesting that GFN2-xTB can reasonably represent the relative energetics of the studied cage systems. Therefore, xTB methods were applied throughout this work for geometry optimisations and for energy evaluations toward a high-throughput workflow.

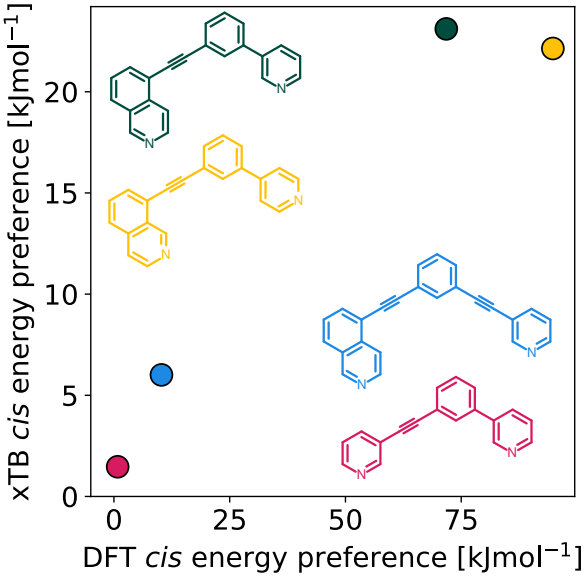


Figure 2: Comparison of GFN2-xTB (DMSO) and DFT (PBE0/def2-SVP/D3BJ/CPCM(DMSO)) energy difference between the *cis* and next most stable isomer of cages formed in ref. 45 from the ligands **3D1** (crimson), **4D2** (yellow), **5D1** (dark green) and **5D3** (blue).

Combined experimental and computational workflow

In this work, a joint computational and experimental workflow (Figure 3(a)) was implemented to facilitate the search for new unsymmetrical ligands with sufficient geometrical constraints to drive the exclusive formation of single Pd_2L_4 cage isomers. This approach

started with an initial experimental choice of building blocks. These were combined *in silico* to form the candidate ligands and their possible Pd_2L_4 cage isomers. The cage systems were then analysed using computationally cheap structural parameters to assess the likelihood of successful self-assembly into a single isomer, thus assisting in the synthetic decision-making process. A focus was placed on using relatively low-cost computational approaches before any experimental investment to minimise wasted efforts. Such an approach opens up the ability to search for new unsymmetrical cages with desirable properties, increasing ligand design efficiency (which is currently based on very few experimental examples).

A major goal of this computational workflow is generalisability, rendering it applicable to a large chemical space that can be adapted for future iterations of the process. To this end, *stk*,[56,57] UFF4MOF (Universal Force Field for MOFs[66,67]) and the xTB family of semi-empirical methods[59,68] were used for the assembly and geometry optimisation, respectively, of ligand and cage structures. *stk* assembles building blocks onto topology graphs; this process includes the placement, alignment and reaction of the building blocks. Before optimisation with xTB, conformer searches were performed on the Pd_2L_4 systems using a broadly applicable force field that can handle common metal complex geometries (UFF4MOF implemented in the General Utility Lattice Program (GULP)[69,70]). Coupling the above methods affords low-cost structure generation, conformer searching and geometry optimisation (Figure 3(e)).

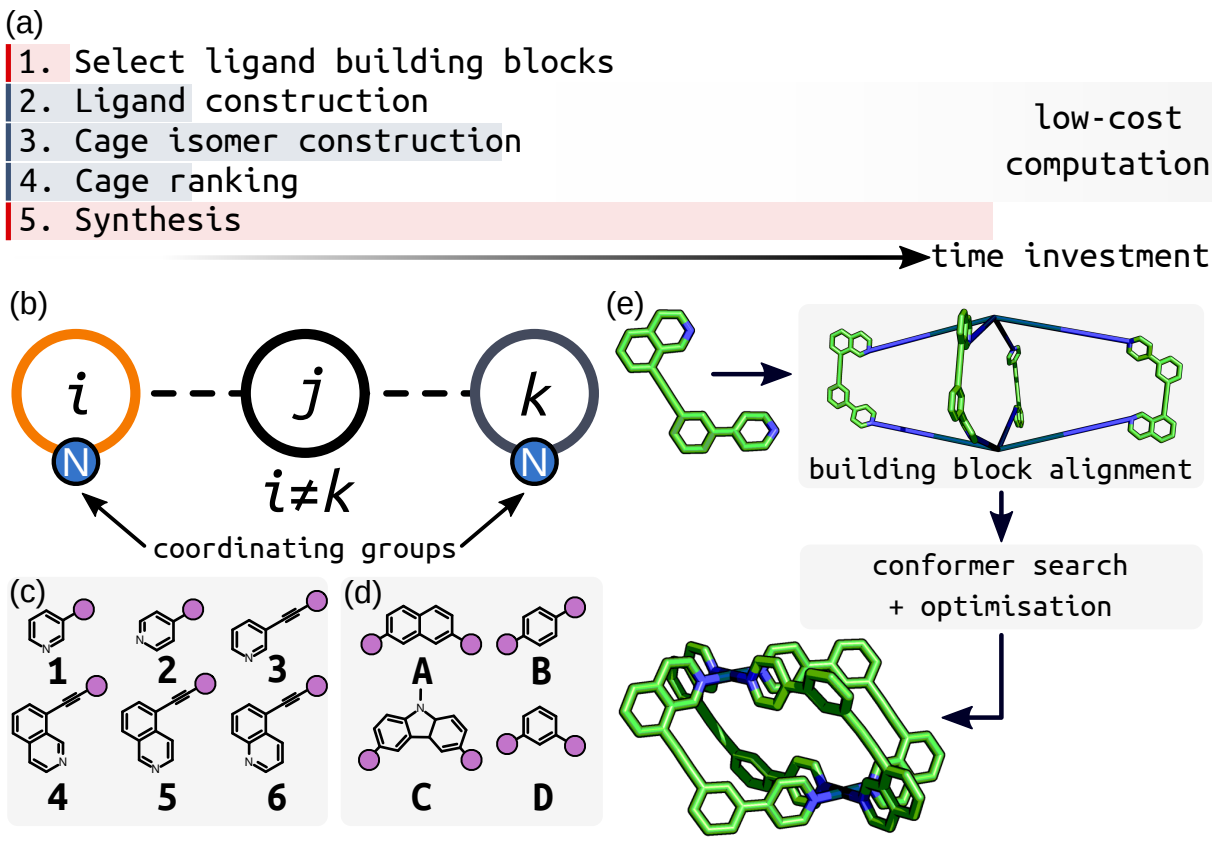


Figure 3: (a) Schematic for the joint experimental and computational workflow showing relative time frames of each step. (b) Schematic of unsymmetrical cage ligands *ijk* formed from three building blocks with the nitrogen coordinating groups highlighted. The library of (c) coordinating building blocks and (d) core building blocks used to construct the cage ligands in this study. Purple circles are the connection points between coordinating and core building blocks. (e) Assembly of a *cis*-Pd_2L_4 structure from an unsymmetrical ligand. Firstly, the Pd(II) ions and ligands are placed and aligned on the M_2L_4 topology vertices by *stk*, generating an expanded structure. Geometry optimisation, coupled with a conformer search, was then performed on this structure to give the optimised geometry.

Automated ligand and cage structure generation

To construct a range of unsymmetrical ditopic cage ligands using *stk*, the ligand structure was partitioned into three building blocks (Figure 3(b)): a core (**A**–**D** in Figure 3(d)) separating two inequivalent coordinating building blocks (**1**–**6** in Figure 3(c)). The building blocks

were selected from structures commonly used in metallo-supramolecular systems. In this initial study, a relatively small selection of building blocks was used, giving a combinatorial library of 60 unsymmetrical ditopic ligands. With four possible Pd_2L_4 cage isomers for each ligand, a library of 240 cages was generated. Through enumeration of this set of common building blocks, a set of unconventional unsymmetrical cage ligands were generated and tested. The step-wise computational assembly and optimisation of ligands and cages (see Methods) through *stk* is automated from the point of input of the ligand building blocks as text-based SMILES strings (Table S1). Therefore, this work is straightforward to extend to a larger chemical space using our hierarchical approach.

Ranking ligands by their cage isomer properties

In the final step of the computational workflow, the assembled cages associated with each ligand were analysed to determine if a single Pd_2L_4 isomer (targeting the *cis* isomer) would be expected to form to the exclusion of others. Our validation of GFN2-xTB (*vide supra*) for providing relative energies of different Pd_2L_4 structures suggested that a GFN2-xTB energy separation of $\approx 6\,\mathrm{kJ\,mol^{-1}}$ between the two lowest energy isomers (ΔE) appeared to be sufficient to drive exclusive formation of the lowest energy cage isomer. Of the 60 ligands in the library, 34 (57 %), including three previously reported examples, had ΔE values of at least $6\,\mathrm{kJ\,mol^{-1}}$, indicating that they might be promising candidates for experimental synthesis.

While the relative xTB energies provide information about the likely self-sorting behaviour of Pd_2L_4 cage isomers, they do not indicate if the desired Pd_2L_4 topology will be the favoured product of self-assembly, rather than a larger species. In fact, the prediction of the preferred topology of palladium cages requires costly computational methods.[71] To bypass these methods, the assumption was made that if the Pd_2L_4 cage is geometrically stable, then, as the smallest possible Pd_nL_{2n} assembly, it is likely to form as the entropically favoured product. As a simple means to probe this, two structural metrics common to every cage in the library were employed (Figure 4): the maximum sum of the deviation of the four nitrogen atoms from each calculated average PdN_4 plane (D_{max}; 0.0 Å indicating no deviation) and the minimum square planar order parameter ($q_{\mathrm{sqp,min}}$;[72] 1.0 indicating a

perfect square planar geometry) of the Pd(II) ions. Both measures quantify the degree of square-planar-likeness in the most strained palladium ion of the cage. It is assumed that if the strain in the cage is significant, Pd_2L_4 formation will be enthalpically unfavourable. By applying ΔE (specifically ΔE_{cis}, which is the relative stability of the *cis* cage isomer), D_{max} and $q_{sqp,min}$ within the workflow as computationally cheap heuristics, cage ligands with desirable properties can be selected for synthesis from the generated rankings favouring those that appear most likely to give successful self-assembly outcomes (where success is defined as exclusive formation of a single cage isomer). Figure 5 shows the relationship between the three heuristics used for ranking candidates.

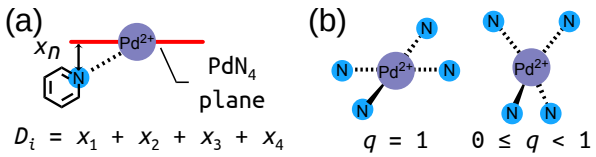


Figure 4: (a) Schematic showing D_i as the sum of the distance of four nitrogen atoms (one is shown) from the plane defined by a PdN_4 unit. D_{max} is the maximum D_i of the two in a Pd_2L_4 cage. (b) Schematic representation of the square planar order parameter q_{sqp}.[72]

Of the 34 ligands with ΔE values $\geq 6\,\mathrm{kJ\,mol^{-1}}$, 12 had lowest energy cage isomers exhibiting very favourable D_{max} and $q_{sqp,min}$ values < 0.1 Å and > 0.95, respectively. Pleasingly, three had previously been synthesised and shown to exclusively form *cis*-Pd_2L_4 isomers.[45] Perhaps unsurprisingly, in each of these 12 instances, the *cis*-Pd_2L_4 isomer was predicted to be the favoured structure, again in agreement with previous work.[45,47,48] Indeed, for most of the ligands (51 of 60; 85 %), the *cis* cage isomer was found to be the most stable (Figure S4(a)).

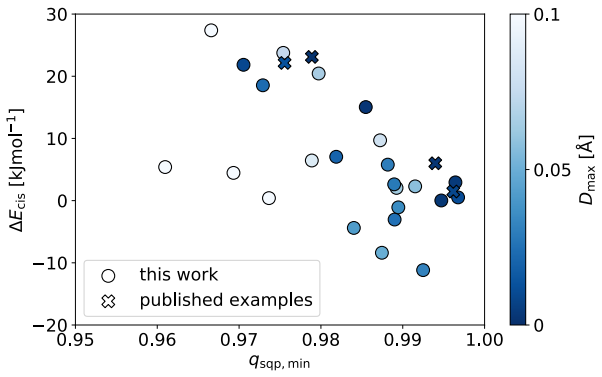


Figure 5: The relative stability of the *cis* cage isomer (ΔE_{cis}) for each ligand as a function of $q_{\text{sqp,min}}$ of the calculated *cis*-Pd_2L_4 structure. Negative values of ΔE_{cis} indicate that the *cis* isomer is not the most stable. Only structures with $q_{\text{sqp,min}} > 0.95$ are shown here; the full data set shown in Figure S3(a).

Experimental realisation of new unsymmetrical cages

Five previously unreported ligands were selected for synthesis to investigate their self-assembly with Pd(II) (Figure 6). These included four ligands from the 12 that adhered to the chosen parameter thresholds (Table 1), including with naphthalene (**5A1**) and para-phenylene (**4B1**, **4B3**, **5B4**) core building blocks, with combinations of pyridyl/isoquinolyl coordinating building blocks. A fifth ligand, **5A3**, was also selected that displayed good structural parameters ($D_{\text{max}} = 0.0\,\text{Å}$; $q_{\text{sqp,min}} = 1.0$) but a low energy separation ($\Delta E_{\text{cis}} = 2.9\,\text{kJ mol}^{-1}$) to probe the fidelity of the ΔE value as a quantitative metric in predicting isomer equilibria, given the necessary simplicity of the workflow's modelling parameters. The ligands were prepared using standard synthetic techniques, and their identities confirmed by NMR spectroscopy and mass spectrometry (MS). In each instance, the ligand self-assembly with Pd(II) was examined by combining the ligand and $[Pd(CH_3CN)_4](BF_4)_2$ in a 2:1 ratio in d_6-DMSO (followed by standing at room or elevated temperature for a period of time, as necessary to reach equilibrium).

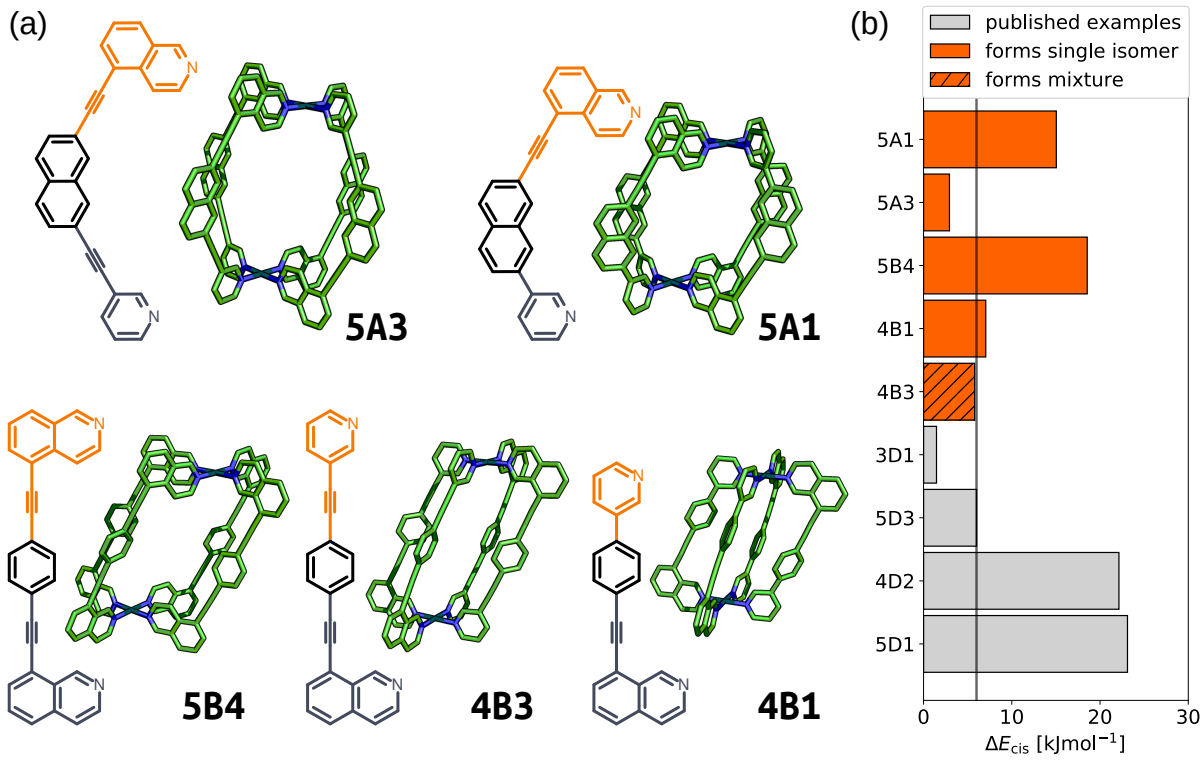


Figure 6: (a) GFN2-xTB optimised structure of the *cis* isomer of selected cage ligands (hydrogen atoms omitted; carbon: green, nitrogen: blue, palladium: cyan). Cage ligands are shown next to each structure with orange and navy colours indicating inequivalent ligand fragments. (b) *cis* isomer GFN2-xTB(DMSO) stability (ΔE_{cis}) for all published[45] and newly selected ligands (patterns distinguish their self-assembly outcomes).

For the three ligands with calculated energy differences in excess of 6.0 kJ mol^{-1} (**5B4**, **5A1** and **4B1**), quantitative conversion to a single species was observed by ^{1}H NMR (Figure 7(a)–(c), respectively) and diffusion-ordered spectroscopy (DOSY). Calculated solvodynamic radii (R_{S}) from the latter (10.2 Å, 8.4 Å and 9.7 Å, respectively) indicated formation of assemblies of similar size to the calculated Pd_2L_4 cage structures. Additionally, isotopic patterns consistent with MOPs of these formulas were found by MS. Through-space interactions between the inequivalent coordinating moieties of the ligands were observed by NOESY which, alongside the symmetry of the ^{1}H NMR spectra, dictated that either the *cis* or *trans* isomers had been formed. Disappointingly, despite multiple attempts, no single crystals suitable for study by X-ray diffraction were obtained. Based on the calculated structures and

extrapolating from previous work,[45] however, we are confident that the structures obtained were the anticipated *cis*-Pd_2L_4 cages.

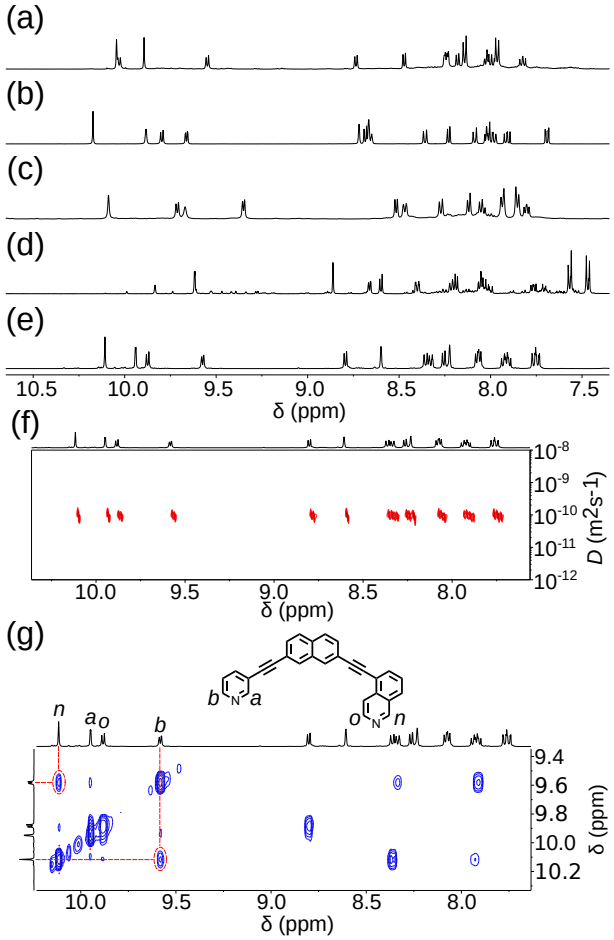


Figure 7: Partial ^{1}H NMR spectra (500 MHz, d_6-DMSO, 298 K) of equilibrated mixtures of 1:2 $[Pd(CH_3CN)_4](BF_4)_2$ and a) **5B4**, b) **5A1**, c) **4B1**, d) **4B3**, and e) **5A3**. f) DOSY, and g) NOESY (500 MHz, d_6-DMSO, 298 K) spectra of $[Pd_2(\mathbf{5A3})_4](BF_4)_4$.

In the case of **4B3** a major product formed but, even after prolonged heating, multiple species could still be observed by ^{1}H NMR (Figure 7(d)). Although it could not be determined absolutely, using similar reasoning to that outlined above, it was concluded that the *cis*-Pd_2L_4 cage was the major species present in solution. It was clear that, under the conditions examined, the difference in energy between this major species and other potential products was not sufficient to drive the exclusive formation of a single assembly.

Intriguingly, despite there being less than 3.6 kJ mol^{-1} difference in energy between the

calculated structures of the *cis*, *trans* and "three-up-one-down" isomers of $[Pd_2(\mathbf{5A3})_4]^{4+}$, a single species was found to form upon the self-assembly of **5A3** with Pd(II) (Figure 7(e)). Once again, DOSY (Figure 7(f); $R_S = 10.4$ Å), MS, the symmetry of the ^{1}H NMR spectrum and cross-peaks observed by NOESY (Figure 7(g)) led to the conclusion that either the *cis* or *trans* assembly had formed. Disappointingly, without single-crystal X-ray diffraction, whether the *cis* or *trans* isomer of the $[Pd_2(\mathbf{5A3})_4]^{4+}$ assembly had formed could not be determined with absolute certainty with the spectroscopic data available.

From the five ligands examined experimentally, the calculated *cis*-Pd_2L_4 structures of three (**5B4**, **5A1** and **4B1**) adhered strictly to our estimated parameter values necessary for quantitative self-assembly of these assemblies. In each instance, exclusive formation of a single Pd_2L_4 isomer was observed and concluded to be the anticipated *cis*-Pd_2L_4 by spectroscopic and computational data. For two of the ligands (**4B3** and **5A3**), the xTB values of ΔE fell at or below the predicted threshold. Interestingly, the ligand associated with the lower value of ΔE (**5A3**) successfully self-assembled into a single Pd_2L_4 cage isomer, whilst the other formed an isomeric mixture. It can be concluded that smaller values of ΔE make predictions of self-assembly outcomes more precarious. This is likely due to effects not taken into consideration within the current computational workflow, such as template effects from anions and/or solvent molecules. Such computationally expensive factors were purposefully omitted to streamline the process and increase throughput. Higher values of ΔE, however, appear to be associated with increased experimental success rates and highlight the efficacy of the workflow for indicating systems with the greatest chance of forming single isomers of the desired cage topology.

Table 1: Calculated properties of the *cis* isomer of the selected ligands. Energy separations (ΔE) are the difference in energy from the *cis* isomer to the next most stable isomer at both levels of theory. $D_{\max}$ is in Å.

	cis isomer		ΔE [kJ mol^{-1}]		
ligand	$q_{\text{sqp,min}}$	$D_{\max}$	xTB	DFT	experimental outcome
5B4	0.973	0.023	18.5	71.0	*cis*-Pd_2L_4
5A1	0.985	0.003	15.0	24.6	*cis*-Pd_2L_4
4B1	0.982	0.020	7.0	21.1	*cis*-Pd_2L_4
4B3	0.988	0.030	5.8	16.0	isomeric mixture
5A3	0.996	0.005	2.9	7.0	*cis*-Pd_2L_4

For the five synthesised ligands, we compared the xTB-calculated energy separations of their isomers with DFT single point calculations and found that the relative energy relationships were similar. The relative stabilities of the *cis* isomers do change among these candidates. This suggests that ranking the ability of each ligand to self-sort using xTB, whilst qualitatively equivalent to DFT, is not quantitative. Additionally, the value of the relative energy threshold changes from $\approx 6\,\text{kJ mol}^{-1}$ to $\approx 10\,\text{kJ mol}^{-1}$ for the methods applied here based on the energy separation of **5D3**. These energy differences are well within DFT error for such complex systems and GFN2-xTB[59] is not parameterised to produce energies accurately. However, this validation supports that the suggestions made by the computational rankings would be equivalent if more costly DFT methods were used.

Low-cost exploration of cage property space

Although five high-ranking candidates were selected for synthesis, models of many potential structures were generated in the computational workflow. This allows for an exploration of the calculated structural properties of these systems. Of particular interest to this work is the controlled introduction of anisotropy toward high affinity and specificity binding of guests. For these Pd_2L_4 systems, anisotropy can be simply defined as the displacement of the Pd(II) centres from alignment perpendicular to the PdN_4 planes (Δ_{Pd}, shown inset in

Figure 8(a), for Pd_2L_4 systems assembled from symmetrical ligands Δ_{Pd} should be 0).

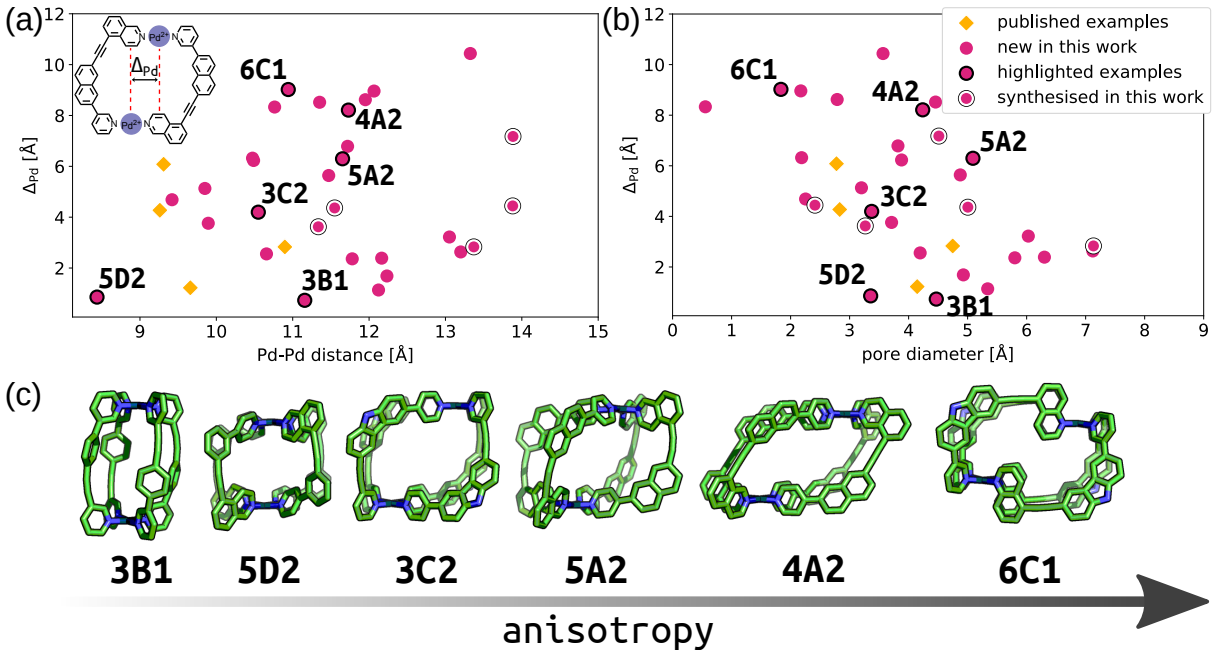


Figure 8: Cage anisotropy (Δ_{Pd}) as a function of (a) Pd· · ·Pd distance and (b) pore diameter. Inset of (a) shows Δ_{Pd} schematically. (c) A set of example cages with increasing anisotropy from left to right; none of these examples are synthesised and do not necessarily pass the synthesisability criteria we apply above (only cages with high geometrical stability, i.e. $q_{sqp,min} > 0.9$, are shown). These cages are highlighted in (a) and (b).

Even though many of the calculated *cis*-Pd_2L_4 structures do not appear to be synthetically accessible systems, an exploration of the predicted property space within this library would serve to highlight the principle of screening for high-value target assemblies. To this end, an analysis of Δ_{Pd} values compared to Pd· · ·Pd distance (Figure 8(a)) and pore size[73] (Figure 8(b)) was undertaken. Example cage structures with increasing anisotropy are shown in Figure 8(c). Hypothetical cages from ligands **4A2** and **6C1**, for example, are of interest as they represent structures with high anisotropy and large or small pore diameters, respectively. Indeed, the structure from **6C1** affords two virtually isolated binding pockets within the cavity, in contrast to the large, single pore of **4A2**.

The combined family of *cis*-Pd_2L_4 cages previously reported[45] and realised in this work are represented within this analysis and demonstrate a diverse anisotropy-property space.

For example, the previously reported cages assembled from ligands **3D1**, **5D3**, **5D1** and **4D2** have Δ_{Pd} values ranging from 1.2 to 6.1 Å and Pd$\cdots$Pd distance ranging from 9.3 to 10.9 Å. The cages successfully synthesised in this work have Δ_{Pd} values ranging from 2.8 to 7.2 Å and Pd$\cdots$Pd distances all larger than 11.3 Å. Therefore, we have not only doubled the number of Pd_2L_4 cages formed from unsymmetrical ligands here, but also increased the explored cage size and anisotropy. The successfully synthesised cages also show interesting relationships between their size (Pd$\cdots$Pd) and porosity, which we expect to be a crucial property to control for the application of reduced-symmetry cages. For example, $[Pd_2(\mathbf{4B1})_4]^{4+}$ has a calculated Pd$\cdots$Pd distance of 11.3 Å, pore diameter of 3.3 Å and Δ_{Pd} of 3.6 Å. As such the utility of the approach presented in this work for the rapid identification of accessible architectures with highly anisotropic structures is demonstrated.

Discussion

The synthesis and investigation of unsymmetrical Pd_2L_4 assemblies with asymmetric pores, with potential utility in high specificity and affinity guest binding properties, is a growing field. The design of such systems to ensure high-fidelity self-assembly, however, remains non-trivial. Here we have shown that a simple and low-cost computational workflow can be used to inform decisions in experimental work, resulting in a "high hit-rate" synthesis of targeted unsymmetrical *cis*-Pd_2L_4 cages. The open-source and generalisable computational procedure provided efficient, and sufficiently accurate, predictions of cage structures starting from a combinatorially constructed library of 60 unsymmetrical ligands. Using a small number of cheap and calculable metrics, based entirely on experimental results, we have realised four previously unreported low-symmetry, *cis*-Pd_2L_4 cages from this investigation, greatly expanding the existing repertoire of these systems and validating our workflow. Additionally, this work is a platform for further exploring the chemical space of unsymmetrical Pd_2L_4 assemblies. In this manner, the synthetic chemist can choose ligands and/or cages with desirable properties with confidence in the reliability of their self-assembly profile.

It was shown that a hierarchical and combinatorial computational screening approach, facilitated by open-source software, allowed the construction of large precursor and cage libraries for high-throughput screening. While focussed initially on a limited number of common building blocks, expanding the initial precursor library is trivial because the only required inputs to the automated workflow are SMILES strings. Additionally, the use of common building blocks did not limit the generation of unconventional and novel unsymmetrical cage ligands in this work. The applied computational workflow can be generalised to future problems to explore a much larger chemical space of metal-organic cages and other materials classes.

Finally, our experimental efforts led to one ligand out of five for which self-assembly with Pd(II) did not produce a single species. In agreement with the metrics employed, however, a single cage isomer did appear to be predominate. This highlights that the heuristics applied do not capture all of the necessary information to ensure absolute fidelity of self-sorting, and we suggest that the role of ligand flexibility and explicit solvent/counter-ion templation could

be significant. However, given the simplicity and high-throughput nature of this approach, it is remarkably effective for informing experimental decisions. As experimental data in this field is still limited, additional information obtained from this and future studies will help recognise metrics of importance to incorporate into the workflow, leading to a refinement of the process. Ultimately this will lead to improved certainty in future synthetic decisions using a joint computational and experimental discovery workflow.

Methods

Ligand and cage assembly

The cage ligands and cages were assembled with our Python software, *stk*,[56,57] using the linear polymer and M_2L_4 lantern topology graphs, respectively. Cage ligand conformations were selected to have two coordinating nitrogen atoms pointing in the same direction to improve robustness of the cage construction process (Figure S1). All structure optimisation was performed using our software package, *stko* (github.com/JelfsMaterialsGroup/stko). *stko* provides methods to optimise metal-containing systems including the UFF4MOF force-field[66,67] in GULP (version 5.1),[69,70] and the xtb software.[59,60] See Supporting Information Section S2 for further details. After construction with *stk*, the lowest energy cage conformer was found using the following sequence of optimisation steps:

1. *stk* assembles structures based on predefined topology graphs with unphysical, long bonds between building blocks (nodes). The expanded structure is collapsed to a realistic size, while maintaining the shape of the assembled structure, by translating each rigid building block toward the centre of mass of the assembled structure. The algorithm ("Collapser" in *stko*) stops when the inter-building block distance is less than 2 Å to avoid steric clashes.

2. The cage structure is geometry optimised using UFF[74] in GULP. The atom typing is handled by a Python implementation of the "ForceFieldHelpers" module in RDKit,[75] 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, "Pd4+2". The bonding used within GULP matches the bonding in the *stk* molecule.

3. A conformer search is performed starting from the UFF optimised 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 100.5 ps with a time step of 0.75 fs. From the production run, 100 conformers are extracted at 1.0 ps intervals.

4. Each extracted conformer is optimised at the "normal" level using GFN2-xTB with DMSO as the implicit solvent (with the "very tight" solvent grid option).

5. The lowest energy cage conformer is optimised using GFN2-xTB with the "extreme" convergence criteria and DMSO as the implicit solvent (with the "very tight" solvent grid option). Frequency calculations were not performed in xtb.

Cage analysis

Computationally efficient methods were used to analyse the stability and energetic preference of metal-organic cage isomers in a screening workflow. Firstly, isomer preference was determined based on the relative GFN2-xTB[59,68] energy of the lowest energy conformer obtained of each isomer with DMSO as the implicit solvent (with the "very tight" solvent grid option). Secondly, the stability of a cage was screened based on simple geometric descriptors that determine how far from an ideal square planar geometry the palladium metal-centres deviate when in the cage. Two cheap methods were used for calculating the geometric stability of a cage: the maximum plane deviation (D_{max}) and the minimum square planar order parameter ($q_{\mathrm{sqp,min}}$)[72] of both metal centres in a cage. Both D_{max} and $q_{\mathrm{sqp,min}}$ are designed to quantify the degree of distortion in the most strained metal centre in a cage and are reasonably well correlated (Figure S3(b)). The plane deviation of a metal centre was calculated as the sum of the shortest distance of each bound nitrogen atom from the plane of best fit defined by the palladium atom and the four nitrogen atoms (Figure 4). $q_{\mathrm{sqp,min}}$ was calculated using the

default implementation in pymatgen (version 2019.5.8), where the neighbours of the palladium atoms were manually set to the four coordinated nitrogen atoms.[76] Throughout this work, we used the geometrical measures of stability (D_{max} and $q_{\text{sqp,min}}$) mostly as guidelines when making experimental decisions. All pore sizes were calculated using our open-source Python package pyWindow.[73]

Code Availability

All code used in this work is available at github.com/andrewtarzia/unsymm_match.

Data Availability

All structure data and ligand ranking is available at github.com/andrewtarzia/citable_data/tarzia_lewis_2021.

Acknowledgements

KEJ thanks the Royal Society for a University Research Fellowship and a Royal Society Enhancement Award 2018, and the ERC through Agreement Number 758370 (ERC-StG-PE5-CoMMaD). JEML is grateful for an Imperial College Research Fellowship. Dr Lukas Turcani is thanked for assistance with *stk* and *stko* development. Dr Alejandro Santana-Bonilla is thanked for assistance with DFT calculations. Dr Nike Dattani and Andrew Rosen are thanked for online discussions. Professor Matthew J. Fuchter is thanked for useful discussions and access to equipment and resources.

Author Contributions

All authors contributed to the design of the project. AT wrote the Python code and performed computations, supervised by KEJ. JEML carried out the synthesis and characterisation studies. AT wrote the manuscript with input and feedback from JEML and KEJ. All

authors discussed the results and approved the final manuscript.

Competing Interests

The authors declare no competing interests.

References

1. M. Fujita, Metal-Directed Self-Assembly of Two- and Three-Dimensional Synthetic Receptors, *Chem. Soc. Rev.*, 1998, **27**, 417–425.

2. S. Leininger, B. Olenyuk and P. J. Stang, Self-Assembly of Discrete Cyclic Nanostructures Mediated by Transition Metals, *Chem. Rev.*, 2000, **100**, 853–908.

3. B. J. Holliday and C. A. Mirkin, Strategies for the Construction of Supramolecular Compounds through Coordination Chemistry, *Angew. Chem. Int. Ed.*, 2001, **40**, 2022–2043.

4. D. A. McMorran and P. J. Steel, The First Coordinatively Saturated, Quadruply Stranded Helicate and Its Encapsulation of a Hexafluorophosphate Anion., *Angew. Chem. Int. Ed.*, 1998, **37**, 3295–3297.

5. A. Schmidt, A. Casini and F. E. Kühn, Self-Assembled M2L4 Coordination Cages: Synthesis and Potential Applications, *Coord. Chem. Rev.*, 2014, **275**, 19–36.

6. M. Han, D. M. Engelhard and G. H. Clever, Self-Assembled Coordination Cages Based on Banana-Shaped Ligands, *Chem. Soc. Rev.*, 2014, **43**, 1848–1860.

7. S. Saha, I. Regeni and G. H. Clever, Structure Relationships between Bis-Monodentate Ligands and Coordination Driven Self-Assemblies, *Coord. Chem. Rev.*, 2018, **374**, 1–14.

8. K. Harris, D. Fujita and M. Fujita, Giant Hollow MnL2n Spherical Complexes: Structure, Functionalisation and Applications, *Chem. Commun.*, 2013, **49**, 6703–6712.

9. M. D. Ward, Polynuclear Coordination Cages, *Chem. Commun.*, 2009, 4487–4499.

10. M. M. J. Smulders, I. A. Riddell, C. Browne and J. R. Nitschke, Building on Architectural Principles for Three-Dimensional Metallosupramolecular Construction, *Chem. Soc. Rev.*, 2013, **42**, 1728–1754.

11. N. B. Debata, D. Tripathy and H. S. Sahoo, Development of Coordination Driven Self-Assembled Discrete Spherical Ensembles, *Coord. Chem. Rev.*, 2019, **387**, 273–298.

12. T. R. Cook and P. J. Stang, Recent Developments in the Preparation and Chemistry of Metallacycles and Metallacages via Coordination, *Chem. Rev.*, 2015, **115**, 7001–7045.

13. J. E. M. Lewis, E. L. Gavey, S. A. Cameron and J. D. Crowley, Stimuli-Responsive Pd2L4 Metallosupramolecular Cages: Towards Targeted Cisplatin Drug Delivery, *Chem. Sci.*, 2012, **3**, 778–784.

14. A. Schmidt, V. Molano, M. Hollering, A. Pöthig, A. Casini and F. E. Kühn, Evaluation of New Palladium Cages as Potential Delivery Systems for the Anticancer Drug Cisplatin, *Chem. Eur. J.*, 2016, **22**, 2253–2256.

15. J. Han, A. F. B. Räder, F. Reichart, B. Aikman, M. N. Wenzel, B. Woods, M. Weinmüller, B. S. Ludwig, S. Stürup, G. M. M. Groothuis, H. P. Permentier, R. Bischoff, H. Kessler, P. Horvatovich and A. Casini, Bioconjugation of Supramolecular Metallacages to Integrin Ligands for Targeted Delivery of Cisplatin, *Bioconjugate Chem.*, 2018, **29**, 3856–3865.

16. S. M. McNeill, D. Preston, J. E. M. Lewis, A. Robert, K. Knerr-Rupp, D. O. Graham, J. R. Wright, G. I. Giles and J. D. Crowley, Biologically Active [Pd2L4]4+ Quadruply-Stranded Helicates: Stability and Cytotoxicity, *Dalton Trans.*, 2015, **44**, 11129–11136.

17. D. Preston, S. M. McNeill, J. E. M. Lewis, G. I. Giles and J. D. Crowley, Enhanced Kinetic Stability of [Pd2L4]4+ Cages through Ligand Substitution, *Dalton Trans.*, 2016, **45**, 8050–8060.

18. A. Ahmedova, D. Momekova, M. Yamashina, P. Shestakova, G. Momekov, M. Akita and M. Yoshizawa, Anticancer Potencies of PtII- and PdII-Linked M2L4 Coordination Capsules with Improved Selectivity, *Chem. Asian J.*, 2016, **11**, 474–477.

19. A. Ahmedova, R. Mihaylova, D. Momekova, P. Shestakova, S. Stoykova, J. Zaharieva, M. Yamashina, G. Momekov, M. Akita and M. Yoshizawa, M2L4 Coordination Capsules with Tunable Anticancer Activity upon Guest Encapsulation, *Dalton Trans.*, 2016, **45**, 13214–13221.

20. R. A. S. Vasdev, L. F. Gaudin, D. Preston, J. P. Jogy, G. I. Giles and J. D. Crowley, Anticancer Activity and Cisplatin Binding Ability of Bis-Quinoline and Bis-Isoquinoline Derived [Pd2L4]4+ Metallosupramolecular Cages, *Front. Chem.*, 2018, **6**, 563.

21. S. M. McNeill, N. M. Giles, D. Preston, P. P. Jones, J. D. Crowley and G. I. Giles, Quadruply Stranded Metallo-Supramolecular Helicate [Pd2(Hextrz)4]4+ Acts as a Molecular Mimic of Cytolytic Peptides, *Chem. Res. Toxicol.*, 2020, **33**, 1822–1834.

22. J. Wang, T. A. Young, F. Duarte and P. J. Lusby, Synergistic Noncovalent Catalysis Facilitates Base-Free Michael Addition, *J. Am. Chem. Soc.*, 2020, **142**, 17743–17750.

23. V. Martí-Centelles, A. L. Lawrence and P. J. Lusby, High Activity and Efficient Turnover by a Simple, Self-Assembled "Artificial Diels–Alderase", *J. Am. Chem. Soc.*, 2018, **140**, 2862–2868.

24. R. L. Spicer, A. D. Stergiou, T. A. Young, F. Duarte, M. D. Symes and P. J. Lusby, Host–Guest-Induced Electron Transfer Triggers Radical-Cation Catalysis, *J. Am. Chem. Soc.*, 2020, **142**, 2134–2139.

25. T. A. Young, V. Martí-Centelles, J. Wang, P. J. Lusby and F. Duarte, Rationalizing the Activity of an "Artificial Diels-Alderase": Establishing Efficient and Accurate Protocols for Calculating Supramolecular Catalysis, *J. Am. Chem. Soc.*, 2020, **142**, 1300–1310.

26. N. Kishi, Z. Li, K. Yoza, M. Akita and M. Yoshizawa, An M2L4 Molecular Capsule with an Anthracene Shell: Encapsulation of Large Guests up to 1 Nm, *J. Am. Chem. Soc.*, 2011, **133**, 11438–11441.

27. N. Kishi, Z. Li, Y. Sei, M. Akita, K. Yoza, J. S. Siegel and M. Yoshizawa, Wide-Ranging Host Capability of a PdII-Linked M2L4 Molecular Capsule with an Anthracene Shell, *Chem. Eur. J.*, 2013, **19**, 6313–6320.

28. M. Yamashina, Y. Sei, M. Akita and M. Yoshizawa, Safe Storage of Radical Initiators within a Polyaromatic Nanocapsule, *Nat. Commun.*, 2014, **5**, 4662.

29. D. P. August, G. S. Nichol and P. J. Lusby, Maximizing Coordination Capsule–Guest Polar Interactions in Apolar Solvents Reveals Significant Binding, *Angew. Chem. Int. Ed.*, 2016, **55**, 15022–15026.

30. T. Tsutsui, L. Catti, K. Yoza and M. Yoshizawa, An Atropisomeric M2L4 Cage Mixture Displaying Guest-Induced Convergence and Strong Guest Emission in Water, *Chem. Sci.*, 2020, **11**, 8145–8150.

31. K. Matsumoto, S. Kusaba, Y. Tanaka, Y. Sei, M. Akita, K. Aritani, M.-a. Haga and M. Yoshizawa, A Peanut-Shaped Polyaromatic Capsule: Solvent-Dependent Transformation and Electronic Properties of a Non-Contacted Fullerene Dimer, *Angew. Chem. Int. Ed.*, 2019, **58**, 8463–8467.

32. D. Preston, K. M. Patil, A. T. O'Neil, R. A. S. Vasdev, J. A. Kitchen and P. E. Kruger, Long-Cavity [Pd2L4]4+ Cages and Designer 1,8-Naphthalimide Sulfonate Guests: Rich Variation in Affinity and Differentiated Binding Stoichiometry, *Inorg. Chem. Front.*, 2020, **7**, 2990–3001.

33. S. Pullen, J. Tessarolo and G. H. Clever, Increasing Structural and Functional Complexity in Self-Assembled Coordination Cages, *Chem. Sci.*, 2021.

34. W. M. Bloch and G. H. Clever, Integrative Self-Sorting of Coordination Cages Based on 'Naked' Metal Ions, *Chem. Commun.*, 2017, **53**, 8506–8516.

35. S. Pullen and G. H. Clever, Mixed-Ligand Metal–Organic Frameworks and Heteroleptic Coordination Cages as Multifunctional Scaffolds—A Comparison, *Acc. Chem. Res.*, 2018, **51**, 3052–3064.

36. D. Bardhan and D. K. Chand, Palladium(II) Based Self-Assembled Heteroleptic Coordination Architectures: A Growing Family, *Chem. Eur. J.*, 2019, **25**, 12241–12269.

37. F. Li, L. F. Lindoy, F. Li and L. F. Lindoy, Metalloligand Strategies for Assembling Heteronuclear Nanocages – Recent Developments*, *Aust. J. Chem.*, 2019, **72**, 731–741.

38. H. Li, Z.-J. Yao, D. Liu and G.-X. Jin, Multi-Component Coordination-Driven Self-Assembly toward Heterometallic Macrocycles and Cages, *Coord. Chem. Rev.*, 2015, **293-294**, 139–157.

39. D. Preston, J. E. Barnsley, K. C. Gordon and J. D. Crowley, Controlled Formation of Heteroleptic [Pd2(La)2(Lb)2]4+ Cages, *J. Am. Chem. Soc.*, 2016, **138**, 10578–10585.

40. R. Zhu, W. M. Bloch, J. J. Holstein, S. Mandal, L. V. Schäfer and G. H. Clever, Donor-Site-Directed Rational Assembly of Heteroleptic Cis-[Pd2L2L′2] Coordination Cages from Picolyl Ligands, *Chem. Eur. J.*, 2018, **24**, 12976–12982.

41. W. M. Bloch, J. J. Holstein, W. Hiller and G. H. Clever, Morphological Control of Heteroleptic *Cis*- and *Trans*-$Pd_2L_2L'_2$ Cages, *Angew. Chem. Int. Ed.*, 2017, **56**, 8285–8289.

42. S. Saha, B. Holzapfel, Y.-T. Chen, K. Terlinden, P. Lill, C. Gatsogiannis, H. Rehage and G. H. Clever, Rational Design of an Amphiphilic Coordination Cage-Based Emulsifier, *J. Am. Chem. Soc.*, 2018, **140**, 17384–17388.

43. W. M. Bloch, Y. Abe, J. J. Holstein, C. M. Wandtke, B. Dittrich and G. H. Clever, Geometric Complementarity in Assembly and Guest Recognition of a Bent Heteroleptic *Cis*-$[Pd_2L^A{}_2L^B{}_2]$ Coordination Cage, *J. Am. Chem. Soc.*, 2016, **138**, 13750–13755.

44. L. S. Lisboa, J. A. Findlay, L. J. Wright, C. G. Hartinger and J. D. Crowley, A Reduced-Symmetry Heterobimetallic [PdPtL4]4+ Cage: Assembly, Guest Binding, and Stimulus-Induced Switching, *Angew. Chem. Int. Ed.*, 2020, **59**, 11101–11107.

45. J. E. M. Lewis, A. Tarzia, A. J. P. White and K. E. Jelfs, Conformational Control of Pd2L4 Assemblies with Unsymmetrical Ligands, *Chem. Sci.*, 2020, **11**, 677–683.

46. J. E. M. Lewis, Multi-Functional, Low Symmetry Pd2L4 Nanocage Libraries**, *Chem. Eur. J.*, 2021, **27**, 4454–4460.

47. D. Ogata and J. Yuasa, Dynamic Open Coordination Cage from Nonsymmetrical Imidazole–Pyridine Ditopic Ligands for Turn-On/Off Anion Binding, *Angew. Chem. Int. Ed.*, 2019, **58**, 18424–18428.

48. S. S. Mishra, S. V. K. Kompella, S. Krishnaswamy, S. Balasubramanian and D. K. Chand, Low-Symmetry Self-Assembled Coordination Complexes with Exclusive Diastereoselectivity: Experimental and Computational Studies, *Inorg. Chem.*, 2020, **59**, 12884–12894.

49. S. K. Sen and R. Natarajan, Influence of Conformational Change and Interligand Hydrogen Bonding in a Chiral Metal–Organic Cage, *Inorg. Chem.*, 2019, **58**, 7180–7188.

50. S. Samantray, S. Krishnaswamy and D. K. Chand, Self-Assembled Conjoined - Cages, *Nat. Commun.*, 2020, **11**, 880.

51. J. E. M. Lewis and J. D. Crowley, Metallo-Supramolecular Self-Assembly with Reduced-Symmetry Ligands, *ChemPlusChem*, 2020, **85**, 815–827.

52. E. Berardo, R. L. Greenaway, L. Turcani, B. M. Alston, M. J. Bennison, M. Miklitz, R. Clowes, M. E. Briggs, A. I. Cooper and K. E. Jelfs, Computationally-Inspired Discovery of an Unsymmetrical Porous Organic Cage, *Nanoscale*, 2018, **10**, 22381–22388.

53. V. Abet, F. T. Szczypiński, M. A. Little, V. Santolini, C. D. Jones, R. Evans, C. Wilson, X. Wu, M. F. Thorne, M. J. Bennison, P. Cui, A. I. Cooper, K. E. Jelfs and A. G. Slater, Inducing Social Self-Sorting in Organic Cages To Tune The Shape of The Internal Cavity, *Angew. Chem. Int. Ed.*, 2020, **132**, 16898–16906.

54. R. L. Greenaway, V. Santolini, A. Pulido, M. A. Little, B. M. Alston, M. E. Briggs, G. M. Day, A. I. Cooper and K. E. Jelfs, From Concept to Crystals via Prediction: Multi-Component Organic Cage Pots by Social Self-Sorting, *Angew. Chem. Int. Ed.*, 2019, **131**, 16421–16427.

55. R. L. Greenaway, V. Santolini, M. J. Bennison, B. M. Alston, C. J. Pugh, M. A. Little, M. Miklitz, E. G. B. Eden-Rump, R. Clowes, A. Shakil, H. J. Cuthbertson, H. Armstrong, M. E. Briggs, K. E. Jelfs and A. I. Cooper, High-Throughput Discovery of Organic Cages and Catenanes Using Computational Screening Fused with Robotic Synthesis, *Nat. Commun.*, 2018, **9**, 2849.

56. L. Turcani, E. Berardo and K. E. Jelfs, Stk: A Python Toolkit for Supramolecular Assembly, *J. Comput. Chem.*, 2018, **39**, 1931–1942.

57. L. Turcani, A. Tarzia, F. Szczypiński and K. E. Jelfs, Stk: An Extendable Python Framework for Automated Molecular and Supramolecular Structure Assembly and Discovery, *ChemRxiv*, 2021.

58. T. A. Young, R. Gheorghe and F. Duarte, Cgbind: A Python Module and Web App for Automated Metallocage Construction and Host–Guest Characterization, *J. Chem. Inf. Model.*, 2020, **60**, 3546–3557.

59. C. Bannwarth, S. Ehlert and S. Grimme, GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.

60. M. Bursch, H. Neugebauer and S. Grimme, Structure Optimisation of Large Transition-Metal Complexes with Extended Tight-Binding Methods, *Angew. Chem. Int. Ed.*, 2019, **58**, 11078–11087.

61. C. Adamo and V. Barone, Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model, *J. Chem. Phys.*, 1999, **110**, 6158–6170.

62. F. Weigend and R. Ahlrichs, Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

63. F. Weigend, Accurate Coulomb-Fitting Basis Sets for H to Rn, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.

64. S. Grimme, S. Ehrlich and L. Goerigk, Effect of the Damping Function in Dispersion Corrected Density Functional Theory, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

65. J. Tomasi, B. Mennucci and R. Cammi, Quantum Mechanical Continuum Solvation Models, *Chem. Rev.*, 2005, **105**, 2999–3094.

66. D. E. Coupry, M. A. Addicoat and T. Heine, Extension of the Universal Force Field for Metal–Organic Frameworks, *J. Chem. Theory Comput.*, 2016, **12**, 5215–5225.

67. M. A. Addicoat, N. Vankova, I. F. Akter and T. Heine, Extension of the Universal Force Field to Metal–Organic Frameworks, *J. Chem. Theory Comput.*, 2014, **10**, 880–891.

68. C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, Extended Tight-Binding Quantum Chemistry Methods, *WIRES Comput. Mol. Sci.*, 2021, **11**, e1493.

69. J. D. Gale, GULP: A Computer Program for the Symmetry-Adapted Simulation of Solids, *J. Chem. Soc. Faraday Trans.*, 1997, **93**, 629–637.

70. J. D. Gale and A. L. Rohl, The General Utility Lattice Program (GULP), *Mol. Simul.*, 2003, **29**, 291–341.

71. D. A. Poole, E. O. Bobylev, S. Mathew and J. N. H. Reek, Topological Prediction of Palladium Coordination Cages, *Chem. Sci.*, 2020, **11**, 12350–12357.

72. N. E. R. Zimmermann and A. Jain, Local Structure Order Parameters and Site Fingerprints for Quantification of Coordination Environment and Crystal Structure Similarity, *RSC Adv.*, 2020, **10**, 6063–6081.

73. M. Miklitz and K. E. Jelfs, Pywindow: Automated Structural Analysis of Molecular Pores, *J. Chem. Inf. Model.*, 2018, **58**, 2387–2391.

74. A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

75. G. A. Landrum, *RDKit: Open-Source Cheminformatics.*, http://www.rdkit.org/, (accessed March 1, 2020).

76. S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, Python Materials Genomics (Pymatgen):

A Robust, Open-Source Python Library for Materials Analysis, *Comput. Mater. Sci.*, 2013, **68**, 314–319.