Host-guest chemistry of polyhedral coordination cages

William Michael Cullen

A thesis submitted to the University of Sheffield in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

December 2015

Department of Chemistry, University of Sheffield, Sheffield, S3 7HF

Author's Declaration

Except where specific references have been made to other sources, the work within this thesis is the original work of the author. It has not been submitted, in whole or part, for any other degree.

William Michael Cullen

December 2015

Acknowledgements

Abstract

Chapter 1 – The background behind the work begins with a general introduction to supramolecular chemistry, then more specifically what cages are and their uses. The basics of host-guest chemistry are also explained, with detail of the hydrophobic effect. Finally previous work conducted by the Ward group that leads on to the work done in this thesis is discussed.

Chapter 2 – With the aim of assembling more water soluble cages, the synthesis of a new ligand ($L^{1,8\text{-naphOH}}$) has been completed and used to prepare an equilibrium mixture of 3 assemblies: a $M_{12}L_{18}$ cage, a M_4L_6 cage and a M_2L_3 mesocate, each of which was fully characterised by X-ray crystallography, NMR and mass spectroscopy. The equilibrium has been investigated by changing the temperature, concentration and solvent of the solution, and its speciation behaviour calculated. The larger assemblies were found to prevail at higher concentrations and lower temperatures; with the smaller assembles prevailing at lower concentrations and higher temperatures, in accordance with Le Chatelier's principle.

Chapter 3 – In order to undertake a more extensive look at guest binding, a high throughput fluorescence displacement assay has been developed to allow us to use a fluorescence plate reader to do simultaneous determination of binding constants. Using this, a more quantitative investigation began using a series of cyclic ketones to investigate the binding limits of the M_8L_{12} cube in water. There was a linear relationship between the hydrophobic surface area and the free energy of binding, equating to a favourable 5 kJ mol^{-1} of binding energy per additional CH_2 in the series. Crystallographic evidence of binding a guest in the cavity was also provided.

Chapter 4 – A computational prediction of guest binding was developed using the molecular docking software GOLD with the creation of our own scoring function with added terms such as the addition of a flexibility term. This method displays a good correlation with both strongly and weakly binding guests and was used to predict the binding of guests from a screen of 3000 molecules. We identified previously unknown guests that have *K* of up to 10^8 M^{-1}.

Chapter 5 – The effect of pH on guest binding in water was investigated. It was found that the neutral form of the guests have binding constants many orders of magnitude larger than the ionised (cationic or anionic) form, since the charged form would rather be solvated in the polar solvent, rather than inside the hydrophobic cage cavity. Using this a multicomponent system was designed, where one specific guest from a mixture (adamantane dicarboxylic acid, cyclononanone, amino adamantane) could be selectively picked to bind in the cage, depending solely on the pH of the solution.

Chapter 6 – Catalysis using the cage was attempted with the decomposition of 1,2-benzosoxazole which forms a charged product (2-cyanophenolate) that does not bind, thus forming the basis of catalytic turnover. The catalytic pathway was found to be essentially independent of pH (from pH 9-11) indicating the mechanism was due to a high local concentration of hydroxide surrounding the cage, and had a rate enhancement (measured) of up to 2 x 10^5, one of the best so far observed.

List of Publications

1) A. J. Metherell, **<u>W. Cullen</u>**, A. Stephenson, C. A. Hunter, M. D. Ward, *Fac and mer isomers of Ru(II) tris(pyrazolyl-pyridine) complexes as models for the vertices of coordination cages: structural characterisation and hydrogen-bonding characteristics*, *Dalton Trans.* **2014**, *43*, 71-84 (Cover article)

2) S. Turega, **<u>W. Cullen</u>**, M. Whitehead, C. A. Hunter, M. D. Ward, *Mapping the Internal Recognition Surface of an Octanuclear Coordination Cage Using Guest Libraries*, *J. Am. Chem. Soc.,* **2014**, *136*, 8475-8483

3) **<u>W. Cullen</u>**, S. Turega, C. A. Hunter, M. D. Ward, *pH-dependent binding of guests in the cavity of a polyhedral coordination cage: reversible uptake and release of drug molecules*, *Chem. Sci.,* **2015**, *6*, 625-631

4) **<u>W. Cullen</u>**, C. A. Hunter, M. D. Ward, *An Interconverting Family of Coordination Cages and a meso-Helicate; Effects of Temperature, Concentration, and Solvent on the Product Distribution of a Self-Assembly Process, Inorg. Chem.,* **2015**, *54*, 2626-2637

5) **<u>W. Cullen</u>**, S. Turega, C. A. Hunter, M. D. Ward, *Virtual Screening for High Affinity Guests for Synthetic Supramolecular Receptors*, *Chem. Sci.,* **2015**, *6*, 2790-2794

6) **<u>W. Cullen</u>**, K. A. Thomas, C. A. Hunter, M. D. Ward, *pH-Controlled selection between one of three guests from a mixture using a coordination cage host*, *Chem. Sci.,* **2015**, *6*, 4025-4028

7) **<u>W. Cullen</u>**, M. C. Misuraca, C. A. Hunter, N. H. Williams, M. D. Ward, *Highly efficient catalysis of the Kemp elimination in the cavity of a cubic coordination cage*, Under review

Table of Contents

Chapter 1 – Introduction

Chapter 4 – Prediction of guest binding using molecular docking

Chapter 5 – The use of pH changes to control guest binding in water

Chapter 6 – Highly efficient catalysis of the Kemp Elimination in the cage cavity

Appendices

Abbreviations

DNA	Deoxyribonucleic acid
TMV	Tobacco Mosaic Virus
2D	2-dimentions
3D	3-dimentions
A	Acceptor
acac	Acetylacetone
AIBN	Azobisisobutyronitrile
anth	Anthracene
ATP	Adenosine triphosphate
biph	Biphenyl
Bu	Butyl
calc	Calculated
CAS	Chemical abstracts service
cat	Catalysed
CCD	Charge-coupled device
CCDC	Cambridge Crystallgraphic Data Centre
CIF	Crystallographic information file
CN	Cyanide
cod	1,5-cycloocatdiene
D	Donor
DCM	Dichloromethane
DFT	Density functional theory
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulphoxide
dppz	Dipyridophenazine
e	Electron
E	Energy
EDTA	Ethylenediaminetetraacetic acid
EPR	Enhance permeability and retention
EPR	Electron paramagnetic resonance
Eq	Equation
ESMS	Electrospray mass spectroscopy
Et	Ethyl
EXAFS	Extended X-ray absorption fine structure
expt	Experimental

fac	Facial
G	Guest
GOLD	Genetic Optimisation of Ligand Docking
H	Host
H-bond	Hydrogen bond
hr	Hour
^{i}pr	Isopropyl
K	Equilibrium constant
k	Rate constant
L	Ligand
M	Metal
M	mol dm^{-3}
m-	Meta-
MCMM	Monte Carlo molecular modelling
mCPBA	Meta-chloroperoxybenzoic acid
Me	Methyl
MeOH	Methanol
mer	Meridional
min	Minute
MMFFs	Molecular mechanics force fields
MOF	Metal Organic Framework
MS	Mass spectroscopy
N,N-DMF-DMA	N,N-Dimethylformamide dimethyl acetal
naph	Napthalene
NBS	N-bromosuccinimide
Nc	Number of carbon atoms
NMR	Nucelar Magnetic Resonance
o-	Ortho-
^{o}C	Degrees Celsius
P	Product
p-	Para-
PCA	Principal component analysis
ph	Phenyl
phen	Phenanthroline
PLP	Piecewise linear potential
Py	Pyridine/pyridile
PyPz	Pyridine pyrazole
Pz	Pyrazole/pyrazolyle

qtpy	Quarter pyridine
R	Idea gas constant
res	Resolution
RMSD	Root mean squared deviation
RT	Room temperature
S	Solvent
s	Second
SA	Surface area
SM	Starting material
SN	Substitution nucleophilic
T	Temperature
TBAF	Tetrabutylammonium fluoride
TBDMS	^{t}butyldimethylsilyl
^{t}Bu	Tertiary butyl
TEMPO	2,2,6,6-Tetramethylpiperidinyloxyl
THF	Tetrahydrofuran
THT	Tail-to-head terpene
TMS	Trimethylsilyl
Tppy	Trispyrazolylborate
TsOH	Toluenesulphonic acid
uncat	Uncatalysed
UV	Ultraviolet
V	Volume
vis	Visible
XED	Extended electron distribution
α	H-bond donor strength
β	H-bond acceptor strength
ΔG	Change in Gibbs free energy
ΔH	Change in entropy
ΔS	Change in enthalpy

Chapter 1

Introduction

1.1 Supramolecular Chemistry

1.1.1 What is supramolecular chemistry?

Supramolecular chemistry is an area of chemistry related to large molecules and assemblies of several smaller molecules based on weak non-covalent interactions such as H-bonding, π-stacking, electrostatics and van der Waals interactions. The word supramolecular comes from Latin where its literal meaning is "above or beyond the molecule". It spans from cycles, cages and knots, to proteins, DNA and molecular machines.[1] Assemblies based on metal coordination complexes are often based on labile metal-ligand bonds. Some examples of supramolecular assemblies are given in **figure 1.1**.[2-4]
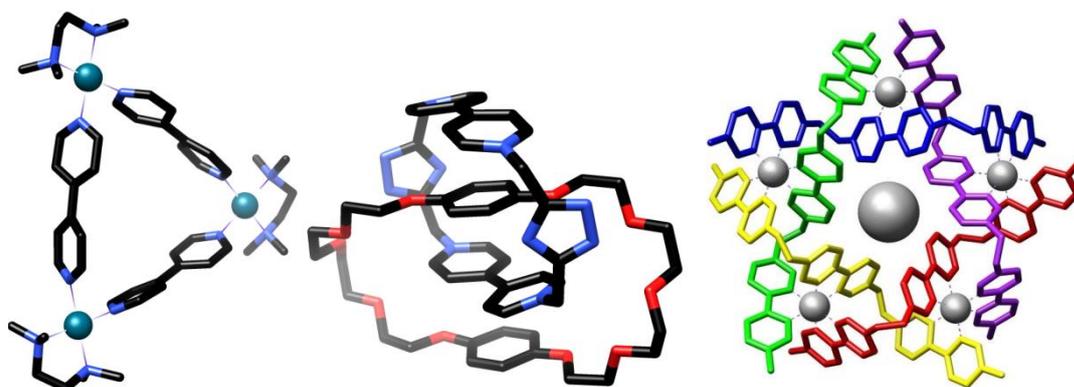
Figure 1.1: A metallomacrocycle, a catenane, and a circular double helicate.[2-4]

1.1.2 Self-assembly

Supramolecular assemblies are usually formed via self-assembly. This is where several components are combined in a specific ratio and spontaneously assemble into higher-ordered structures. This is a thermodynamic process with predominating entropy-enthalpy compensation effects. There are multiple favourable interactions between components, so a favourable enthalpy, but a more ordered structure is formed, so the entropy of the process is unfavourable. For self-assembly to occur, weak or reversible interactions are needed which allow the system to move towards the true thermodynamic

minimum by allowing 'incorrect' assemblies to break and re-form.[1] A classic example of self-assembly is the tobacco mosaic virus (TMV) (**figure 1.2**).[5]
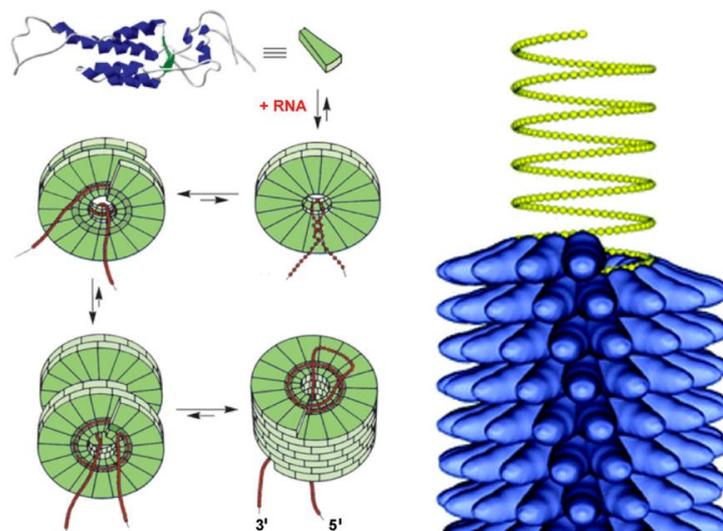
Figure 1.2: Self-assembly of the TMV.[5]

The TMV was the first natural system to show in vitro self-assembly. By changing pH and temperature, the TMV dissociates into its component parts, and then upon return to its original conditions, the components re-assemble back into the fully functional virus.[5]

1.1.3 Supramolecular assemblies based on coordination complexes

The use of metals in supramolecular chemistry has many advantages over assemblies formed via purely organic components using interactions such as hydrogen bonding or π-stacking. By incorporating metal ions into the assembly, a much wider range of predicable geometries and bond angles are available, thus allowing higher complexities of assembly to be created; the metal-ligand bond lability allows for swift assembly (most organic based macrocycles require a template, multi-step synthesis or high dilution methods, all of which are time consuming and can often result in low yields); the metal-ligand bond can be just as strong as a covalent bond, except that coordination bond formation can be reversible which allow robust assemblies to form from bonds that are strong yet labile; and any useful properties such as any photo-physical or electrical-chemical properties associated with the metal ions can provide built-in functionality to the assembly.[1]

A good example of a metallo-supramolecular assembly based on coordination to metal ions is Fujita's Pd(II) square (**figure 1.3**) which can be formed at room temperature in just 10 minutes with high yields.[6,2]
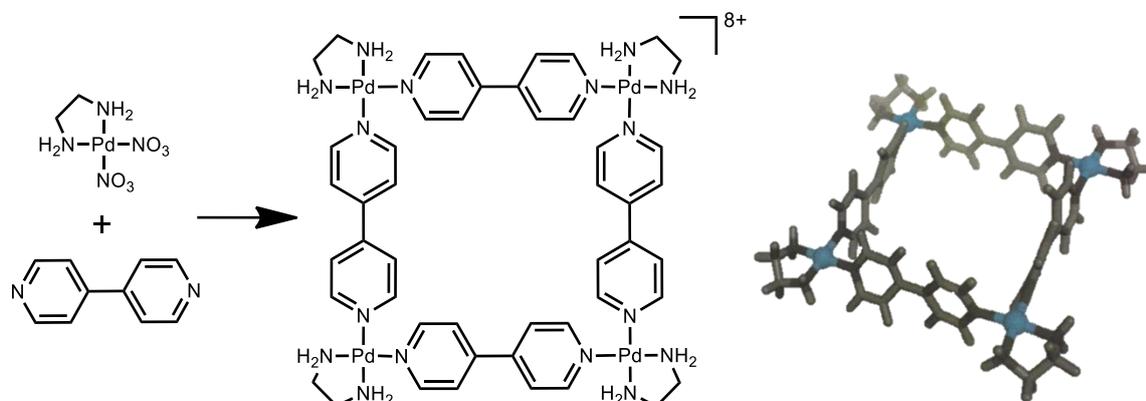
Figure 1.3: Fujita's Pd(II) square.

The lability of Pd(II) works extremely well in allowing the formation of these macrocycles, as it allows both low temperature (often room temperature) and fast synthesis (minutes). However the major disadvantage of Pd(II)'s lability is that the macrocycle is not stable enough for any real applications, and also makes it hard to study. Because of this, the Fujita group repeated the synthesis but using Pt(II) as the corner ion. Pt(II) being more inert, did require some heat to make the reaction proceed, but the final product was much more stable.[6,2]

Another similar example but using a different metal is Hupp's homometallic Re$_4$ square (**figure 1.4**).[7,8]
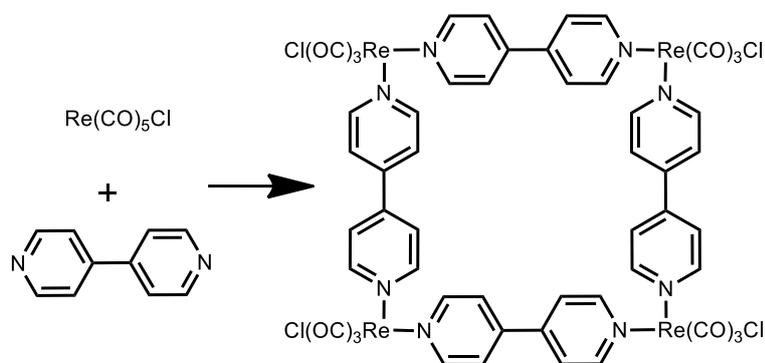
Figure 1.4: Hupp's homometallic Re$_4$ square.

This Re macrocycle formed in near quantitative yields over a period of 2 days in a 1:1 stoichiometric ratio of Re(CO)$_5$Cl and an appropriate bridging ligand such as 4,4-bipyridine. The precipitation of the squares shifting the reaction equilibrium is the most likely driving force for the formation of these macrocyclic compounds.[7]

Metallomacrocycles can also contain two metal types, such as the Ru$_2$Re$_2$ macrocycle from the Thomas group. This macrocycle is assembled by combining the Ru-qtpy (quarter pyridine) mononuclear complex in a 1:1 ratio with ReCl(CO)$_5$. They have used this macrocycle to selectively bind the nucleotide ATP (when dppz is used as the N-N ligand on the Ru centre), which turns on the emission of the host, thus acting as a selective ATP sensor (**figure 1.5**).[9,10]
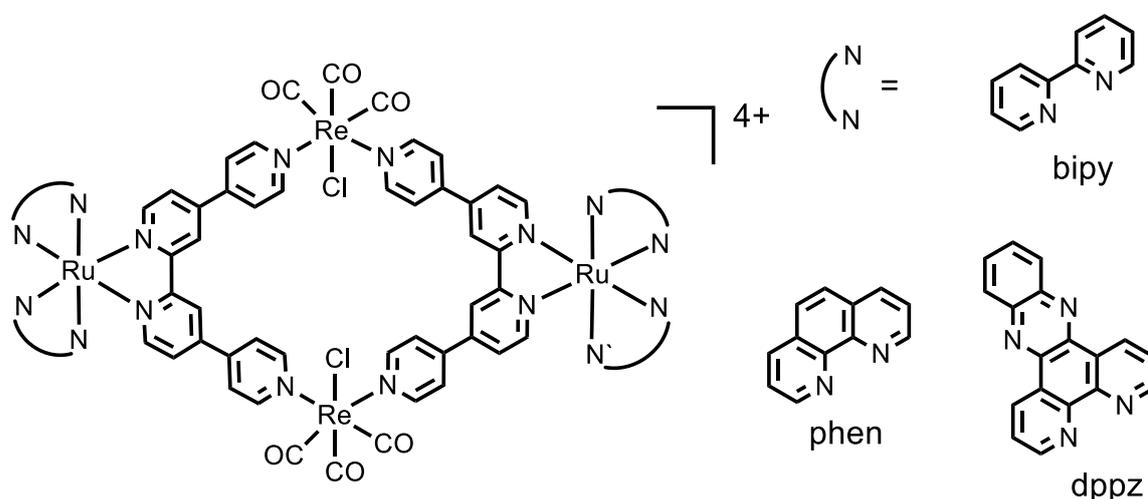
Figure 1.5: Thomas group Ru/Re macrocycle.

Much more complicated assemblies have been made such as helicates (**figure 1.6**).
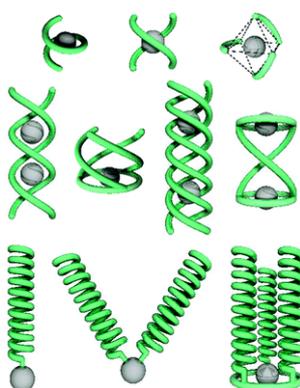
Figure 1.6: Helix architecture with metal complexes at molecular and supramolecular levels.

Helicate assembly displays similar features to biological processes, such as positive cooperativity.[11,12] So once the first metal has coordinated, the coordination of second one becomes easier, and so on.

The Lehn group synthesised a range of oligobipyridine ligands and mixed them with Cu(I) to form helicates of various lengths with each Cu(I) ion binding to 2 bipy units, one from each strand **figure 1.7**). They then investigated what would happen if Cu(I) was added to a mixture of all the ligands. As expected a mixture of helicates formed however this mixture was well defined with each helicate containing identical ligands. This demonstrated another characteristic of helicates, which was self-recognition with each ligand only pairing up with another of the same length, so as to fulfil the maximum site occupancy of the ligand and the Cu(I) ions.[12,13]

n	R	
	H	COOR or CONR2
0	4	8
1	5	9
2	6	10
3	7	11
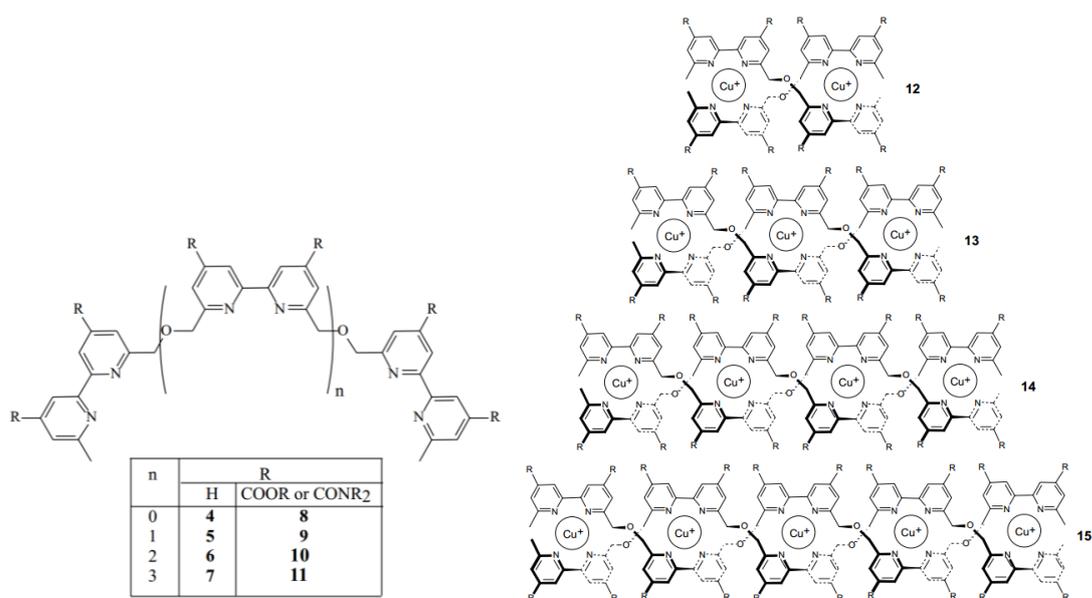
Figure 1.7: Various helicates formed from oligobipyridine ligands with Cu[+] ions by varying n.[12,13]

Other examples of assemblies based on metal coordination complexes include interlocked systems such as borromean rings,[14] knots,[15] catenanes and rotaxanes[16] **figure 1.8**). In all cases the metal ions provide geometric information essential for the assembly by orientation of the component parts in a specific way. E.g. The formation of a trefoil knot.
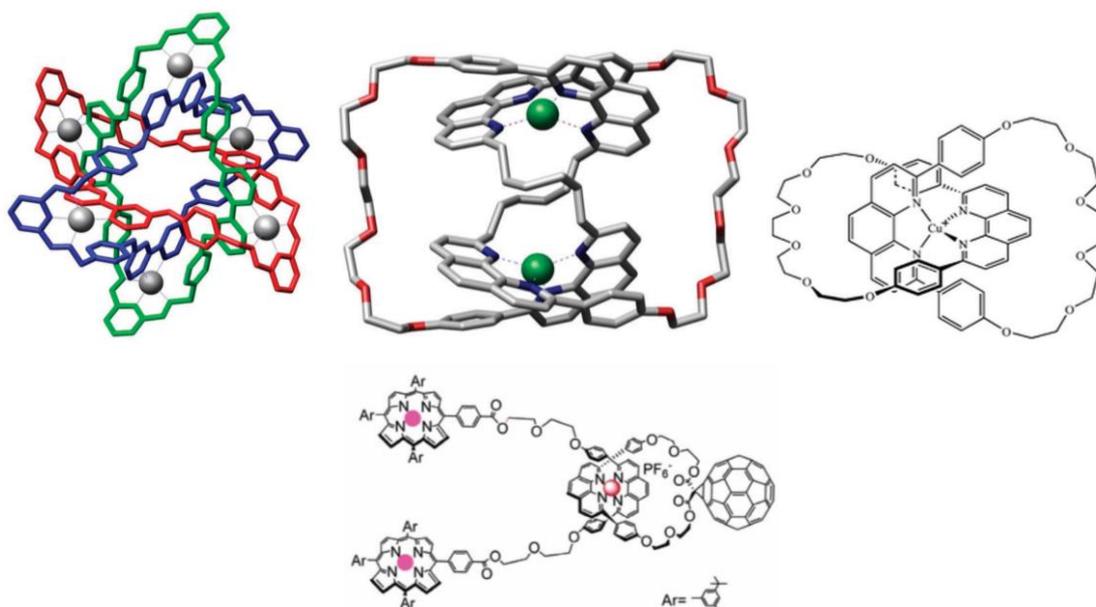
Figure 1.8: Examples of borromean rings,[14] knots,[15] catenanes and rotaxanes.[16]

Another type of coordination assembly that is popular in supramolecular chemistry is the family of coordination cages. Cages are three-dimensional supramolecular assemblies that have a large cavity inside. The first example of this was from Saalfrank with a M_4L_6 tetrahedron,[17] and perhaps one of the most famous is the metal complex cage reported by Fujita, where by just changing the ligand in their Pd(II) square complex from a linear molecule to a triangular one, a three-dimensional octahedral assembly with a central cavity was formed, with a $[Pd(en)]^{2+}$ unit at each of the six vertices, and a triangular face-capping bridging ligand on 4 of the 8 faces, each linking three metal ions (**figure 1.9**).[18,19] This cage assembles with >90% yield with high purity, and can be bulk produced to 100 g, and it can now be bought commercially.
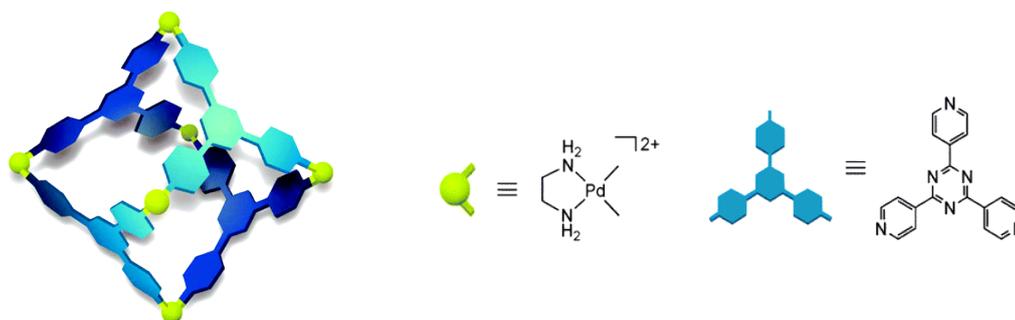
Figure 1.9: Fujita's M_6L_4 Cage.[19]

Since then, a wide range of polyhedral cages have been reported, with the largest currently being the $M_{24}L_{48}$ nanosphere (**figure 1.10**). The ligand bite angle (θ) is used to control which cage forms, with a more linear angle giving less curvature to the surface and hence affording a larger spherical assembly. For the largest $M_{24}L_{48}$ assembly to form, an angle between 134° and 149° is required.[20]
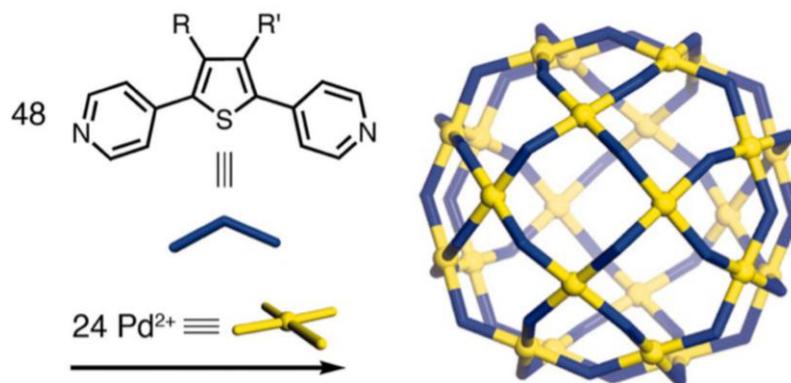
Figure 1.10: $M_{24}L_{48}$ Cage.[20]

1.2 Host-Guest Chemistry

1.2.1 Introduction

An important area of supramolecular chemistry is the study of a molecule (host) binding to another molecule (guest) to produce a host-guest complex, sometimes called a molecular complex. A molecular complex is defined as a non-covalently bound species of definite host:guest (H:G) stoichiometry that is formed in a facile equilibrium process. There is much room for interpretation of what classifies as a "molecular complex".[21]

In supramolecular chemistry the host is typically a large molecule with some form of central cavity that possesses convergent binding sites such as a hydrogen bond donor atom or a Lewis base donor atom. The guest is usually a simple/small molecule that contains divergent binding sites that are complementary to the binding sites of the host.[22]

A good example of a host guest complex is between 18-crown-6 (Host) and K^+ ions (Guest) (**figure 1.11**).[1,22]
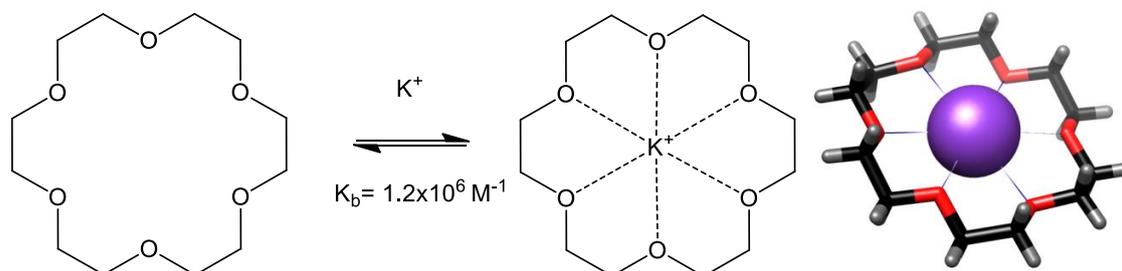
Figure 1.11: Host-guest complex of 18-crown-6 with K^+ ions in methanol.[1,22]

The high binding constant of this host-guest complex is due to the K^+ ion and the crown ether having a good size match allowing strong dative $O\bullet\bullet\bullet K^+$ bonds; the fact that the crown ether requires little rearrangement, so is 'preorganised'; and the presence of multiple interactions between the host and guest resulting in a strong chelate effect.[1,22]

Other examples of host guest complexes are enzyme/substrate complexes. Enzymes have binding sites that are highly selective for guests of specific shape and size. Usually only a small number of substrate molecule(s) can bind in the highly specific binding site. One of the theories to help explain/demonstrate this is one proposed in 1894 by Emile

Fischer called the "lock and key" model. In this model the substrate/guest has geometric size and shape complementarity to the receptor/host and only combinations which match can bind and therefore react (**figure 1.12**).[1,23]
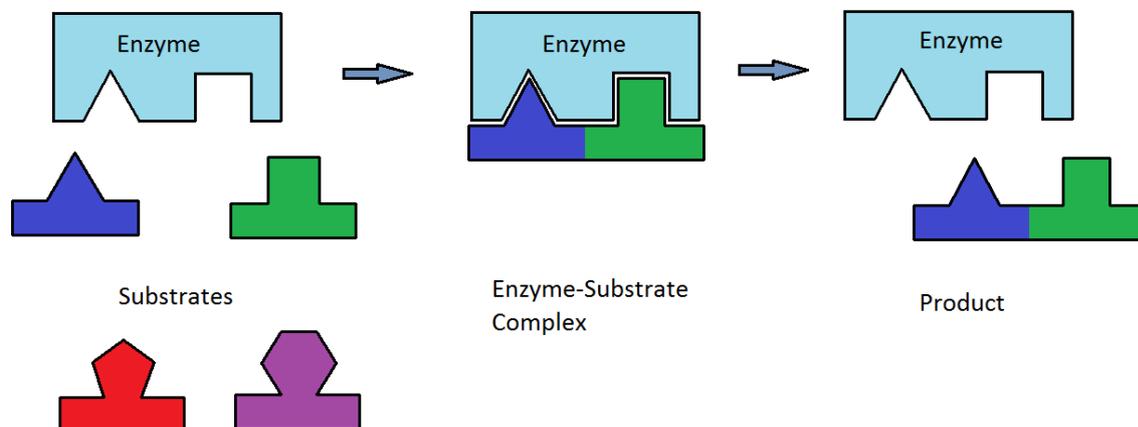
Figure 1.12: The "lock and key" model for enzyme host-guest chemistry.

The "lock and key" model is a good first approximation to explain enzyme host-guest chemistry, but it does not explain all aspects of it fully, so Daniel Koshland proposed a new model called the "induced fit" model (**figure 1.13**).[23] Precise orientation of catalytic groups in the active site is required for enzyme actions to occur. Since enzymes are rather flexible, interactions with the host when the substrate binds reshapes the positions of the amino acids in the active site, which brings the catalytic groups into the required position for catalysis to take place (in the "lock and key" model the host is very rigid). A non-substrate molecule, even if it could bind, would not bring about the required change in the enzyme shape.[1,23]
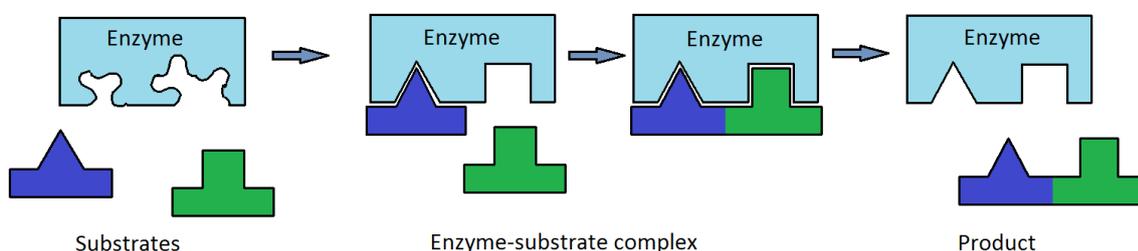
Figure 1.13: The "induced fit" model for enzyme host-guest chemistry.

1.2.2 Binding constants

The binding constant, *K*, is a measure of thermodynamic stability at a given temperature of a host-guest system compared to the separate species, and in most cases it can simply be thought of as merely the ratio of concentrations of each species in an equilibrium, and is therefore an equilibrium constant. The binding constant for a 1:1 host-guest system is given in **figure 1.14**.[1,21,22]

$$Host + Guest \rightleftarrows Host \cdot Guest \quad K = \frac{[H \cdot G]}{[H][G]}$$
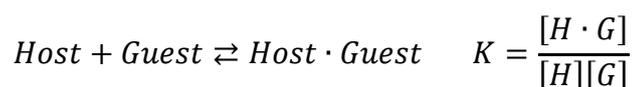
Figure 1.14: Binding constant, K (units of M^{-1}), for a 1:1 host-guest system.

A large binding constant corresponds to a high concentration of HG complex compared to unbound components and therefore corresponds to a more stable complex. The binding constant for the crown ether with K^+ ions shown in **figure 1.1**, is very large (1.2×10^6 M^{-1}), which means that the formation of the complex is highly favoured.

Although many equilibrium constants smaller than 1 M^{-1} have been reported, the interpretation of such small effects is difficult. There is no unambiguous upper limit to the stability of a molecular complex, since this is related to the problem in distinguishing between covalent and non-covalent bonding. For instance an equilibrium constant for the complex Cu(II) with EDTA is 6.3×10^{18} M^{-1}, however it is not clear whether such a complex could be classed as a host-guest complex.[21]

Sometimes host-guest systems do not form in a simple 1:1 ratio (i.e. they have more than one guest molecule per host), so there may be multiple K values, ones for each guest that binds (**figure 1.15**).[1,21,22]

$$H + G \rightleftarrows H \cdot G \quad K_1 = \frac{[H \cdot G]}{[H][G]}$$

$$HG + G \rightleftarrows H \cdot G_2 \quad K_2 = \frac{[H \cdot G_2]}{[HG][G]}$$

$$HG_2 + G \rightleftarrows H \cdot G_3 \quad K_3 = \frac{[H \cdot G_3]}{[HG_2][G]}$$
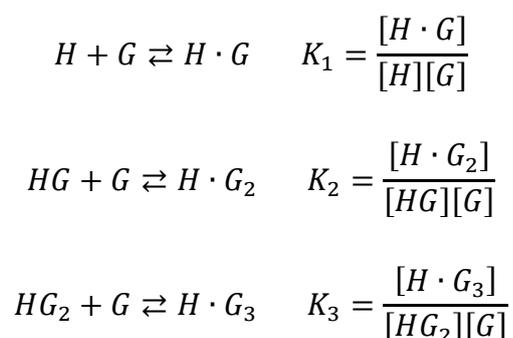
Figure 1.15: Stepwise binding constants in a 1 H: multiple G system

The overall binding constant, β is the product of all stepwise K values (**figure 1.16**).

$$H + G \rightleftarrows H \cdot G \qquad \beta_1 = \frac{[H \cdot G]}{[H][G]} = K_1$$

$$H + 2G \rightleftarrows H \cdot G_2 \qquad \beta_2 = \frac{[H \cdot G_2]}{[H][G]^2} = K_1 K_2$$

$$H + 3G \rightleftarrows H \cdot G_3 \qquad \beta_3 = \frac{[H \cdot G_3]}{[H][G]^3} = K_1 K_2 K_3$$
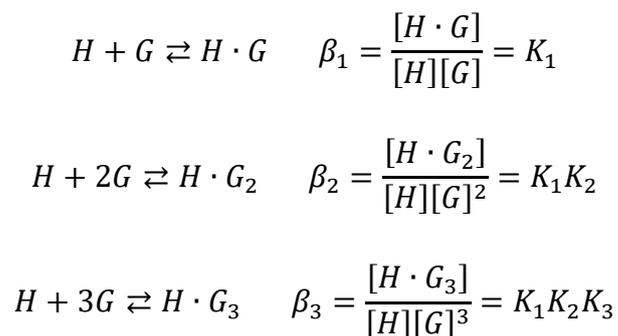
Figure. 1.16: Overall binding constant for a 1:3 H-G system

Binding constants are related to the Gibbs free energy ΔG for complex formation by the equation in **figure 1.17**.[1,21,22]

$$\Delta G = -RTlnK$$

Figure 1.17: Equation relating the Gibbs free energy, ΔG, and binding constant, K, where R is the ideal gas constant and T is the temperature

1.2.3 Measuring binding constants

The measurement of a binding constant is usually done via a titration where one component (guest) is gradually added to the system (host) while monitoring changes in physical properties via NMR, UV-vis, or Fluorescence spectroscopy. Changes in these spectra can be directly related to the concentrations of host, guest and host-guest species present. The resulting changes are then fitted to binding models to obtain information such as the binding constant K.[21,22] UV-vis and fluorescence spectroscopy are perhaps the most widely used techniques for determining binding constants. They are fast to measure, require very little sample due to high sensitivity, which also allows the determination of stronger binding affinities. Fluorescence does require a fluorescent active molecule (either host or guest) to be used however.

NMR spectroscopy is also useful for measuring K values, though it has two limitations: (i) individual spectra take longer to measure than UV/vis or fluorescence spectra, and (ii) it requires higher concentrations which limit the range of binding constants that can be measured. It also can be used to find out much more information from the single titration.

It can identify which parts of the molecule are involved more directly with the binding event (seen by bigger shifts of the corresponding protons), it also can give an indication to the type of interaction (i.e. whether it is hydrogen bonding) by the direction of the movement. It can also be used to measure exchange rates since the timescale of the NMR measurement can be similar to that of the binding event.[21,22]

NMR titrations to obtain a value for *K*

Depending on the timescale of the host-guest equilibrium, NMR titrations can be classified into two categories by relating it to the NMR timescale. These limiting cases are (i) fast exchange (for free/bound guest exchange that occurs faster than the NMR timescale (<ms), so a single averaged peak with a steady change in chemical shift is observed during the titration) and (ii) Slow exchange (for exchange that occurs slower than the NMR timescale (>ms), with free host and host-guest peaks observed separately with one increasing in intensity as the other decreases (**figure 1.18**).[21,22,24]
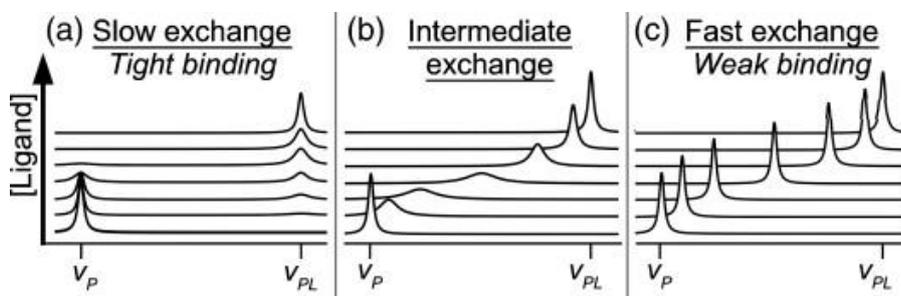
Figure 1.18: NMR titration data showing a) slow exchange, b) intermediate exchange, c) fast exchange.[24]

For slow exchange systems the relative integrals of each peak are taken, and directly related to the concentration of free host, and host-guest species (the guest concentration is known from how much was added). For fast exchange systems the chemical shift values are taken for each peak during the titration and the data are fitted, along with the known concentrations of guest and host added, to the appropriate binding model, such as for a 1:1 or 1:2 host-guest system.[21,22,24]

1.2.4 Binding forces

The forces responsible for the formation of a host-guest complex are the same as those responsible for self-assembly. These are:

<u>Electrostatic interactions</u> – interactions between multipole moments of polar molecules, these moments are charges (C), dipole moments (μ), and quadrupole moments (Q). Apart from charge-charge interactions, the electrostatic potential energies depend on the mutual orientation of the interacting moments. However, the average potential energy, is dependant only on the intermolecular distance. For neutral polar molecules the most important contributor to this is the dipole-dipole interaction that has r^{-6} dependence.[21,25]

<u>Induction/polarisation forces</u> – The effect of a moment in a polar molecule inducing a charge separation in an adjacent molecule.[21,25]

<u>Dispersion (London) forces</u> – This is a quantum mechanical phenomenon, where at any instant the electronic distribution in a molecule may result in an instant dipole moment, even if the molecule in question is a spherical non-polar molecule. This instantaneous dipole induces a moment in an adjacent molecule which interacts with the moment in the original molecule. Simply, these are instantaneous dipole-induced dipoles; often referred to as van der Waals' forces.[21,25]

Other (chemical) interactions are charge transfer and hydrogen bonding, where an electron deficient hydrogen atom (a H-bond donor) forms an interaction with an electron rich atom (a H-bond acceptor).[21,25]

The Solvent (S) also plays an important role. A (solution state) host guest equilibrium can be thought to be not just simply the interaction between H and G, but the H-S, G-S, HG-S and S-S interactions (**figure 1.19**).[21,25]
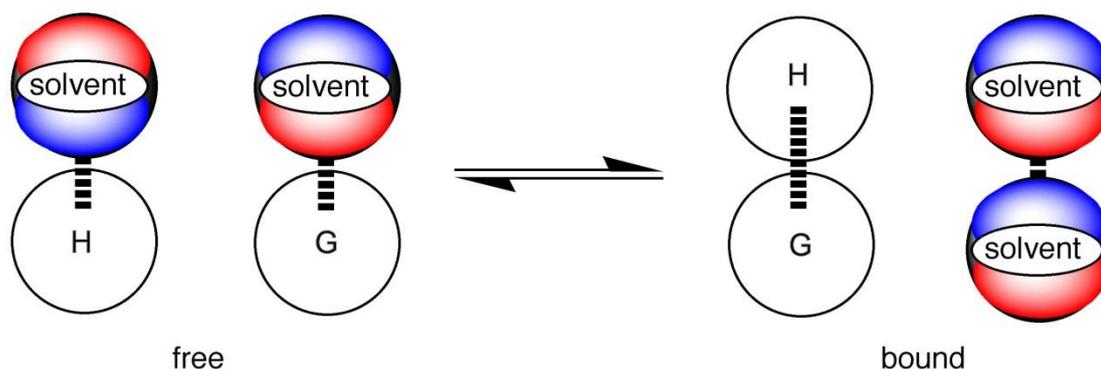
Figure 1.19: pictorial representation of the solvent based host-guest equilibrium; the red and blue parts of the solvent molecules represent acceptor and donor parts.[25]

If the H·S and G·S association constants are low, and the S·S association is high, then the dominant species is the H·G complex. If however the H·S and G·S interactions are high then the equilibrium will lie towards the free (solvated) H and G Species. This becomes a simple representation of solvophobic effects (in water this is referred to as the hydrophobic effect) where the association of the free H and G species with the solvent is disfavoured and the S·S interaction is highly favoured. Thus the H·G species becomes extremely favourable. [21,25]

Most of the content described in this thesis were performed in water, so the hydrophobic effect will play a big part in the chapters to come.

1.2.5 The hydrophobic effect

The hydrophobic effect forms the basis of the separation of oil and water, and the formation of lipid bilayers, and first appeared in the literature in letters between Benjamin Franklin, William Brownrigg and Reverend Mr. Farish in 1773, titled "Of the Stilling of Waves by means of Oil".[26]

Despite being one of the most studied solution effects, it is perhaps the one that is least understood.[27] Over the years many theories have been proposed for its origins however, even today there is an abundance of controversy as to what the actual origin is.

The simplest and most often referred to model of the hydrophobic effect is the "Iceberg model" (**figure 1.20**), proposed by Frank and Evans in 1945 to explain the large positive change in entropy observed when vaporising a hydrophobic molecule from water.[28] They suggested that when a non-polar molecule is dissolved in water, the structure of water tends towards a greater crystallinity, thus 'freezing' or 'formation of an iceberg' could be thought to occur around the molecule in question.
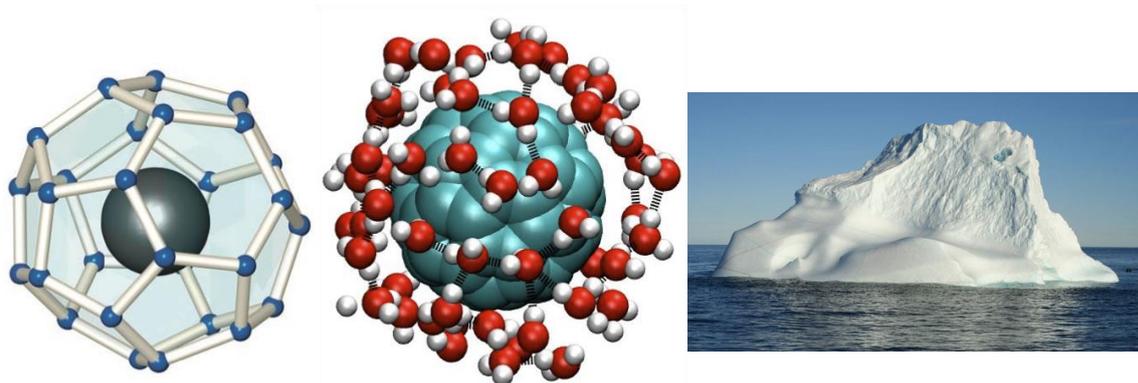
Figure 1.20: The Iceberg model

They also noted that the size of the "iceberg" will be greater the larger the foreign atom. Thus once the molecule is removed from water the "iceberg" melts and there is a large increase in entropy. This Idea explains other observations such as, as the temperature increases, the "iceberg" melts and thus approaches behaviour seen in non-aqueous solvents.[28]

This model also very nicely explains hydrophobic aggregation (**figure 1.21**); Each molecule in solution is surrounded by these "icebergs" (ordered layers of water molecules) and when the molecules come together, some of the ice in-between the two molecules is released back into the bulk solution resulting in an increase in entropy. This is also the main reason why the hydrophobic effect is such a strong driving force in host-guest chemistry since there is a large increase in entropy when a non-polar molecule in water binds inside the hydrophobic cavity of the host.[29]
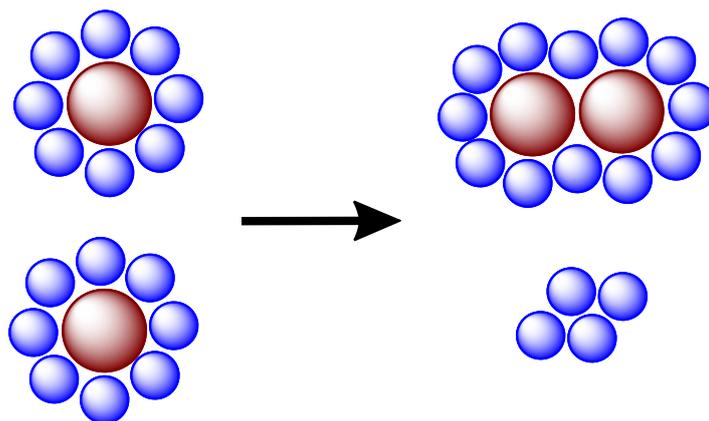
Figure 1.21: Hydrophobic aggregation as described by the iceberg model

Despite its simple explanation for the hydrophobic effect there is very little evidence proving the existence of the ordered water around the molecule, and a lot of evidence that actually disproves this model (experimental studies such as Neutron diffraction, Raman and EXAFS; as well as theoretical and computational studies).[27,30-36] Some studies have even shown a *decrease* in the order of water around the hydrophobic solute molecules. The data however is known to have some degree of interpretation, and could, depending on how you look at it, re-inforce the "iceberg" model, or counter-act it.[37-39]

Because of this, many hypotheses have been proposed however there is always some contradiction in every case.[40-44] One of the more common interpretations of the hydrophobic effect, is an *Enthalpically (H)* driven rather than *entropically (S)* driven hydrophobic effect.[45-49]
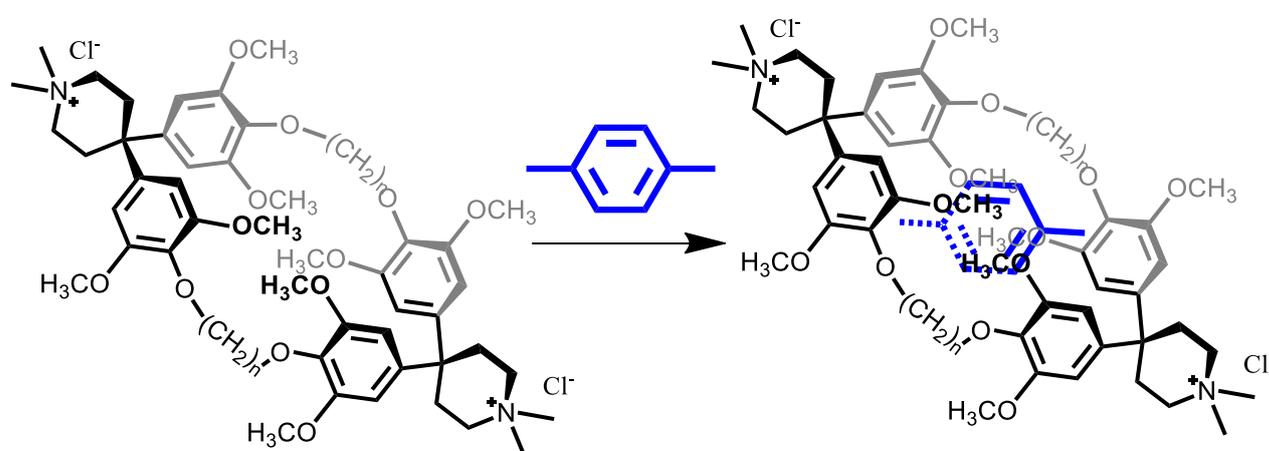
Figure 2.22: Diederich's cyclophane host that has a favourable ΔH and an unfavourable ΔS upon guest binding in water.[45]

This enthalpic hydrophobic effect is seen in some simple host guest systems, such as Diederich's cyclophane host (**figure 1.22**), which for guest binding when bound to aromatic based guests shows, a favourable ΔH and an unfavourable ΔS.[45] This was also observed in various protein-based host guest complexes, and with guest binding to curcurbiturils. This approach is also referred to as being based on "high energy water". It is thought that inside the host cavity the water molecules cannot make an optimum number of hydrogen bonds, thus are in a "high energy" or "frustrated" state. When a guest molecules binds the water is released back into the bulk solution, where the optimum number of H-bonds can be formed, thus producing a favourable enthalpy change (**figure 1.23**).[49] Such "high energy water" effects are often accompanied with extremely high K values of $>10^8$ M^{-1} for guest binding.[48,49]
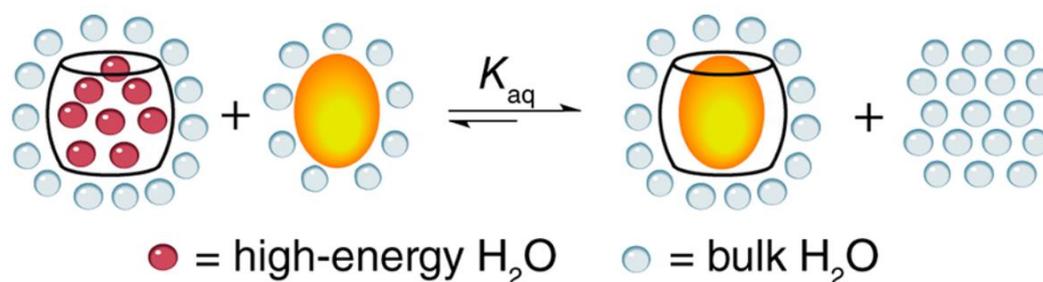
Figure 1.23: Curcubiturils with high energy water binding guest molecules.[49]

The sample set for the "high energy water" driven binding is small. One theory behind this is based on cavity size (**figure 1.24**). A very small cavity, in which no solvent can enter, is essentially void space (vacuum) and binding of extremely small hydrophobic guests is expected to be entropically favourable. As the size of the cavity increases, some water molecules sit in the cavity so as to avoid a vacuum, these molecules cannot form stable hydrogen bonds and so their energetic frustration will be substantial. Once the size of the cavity gets larger the water molecules can form stable clusters of water (the molecular ice). Thus the relative contributions of $\Delta S/\Delta H$ for guest binding in this model will be strongly dependant on cavity size.[48]
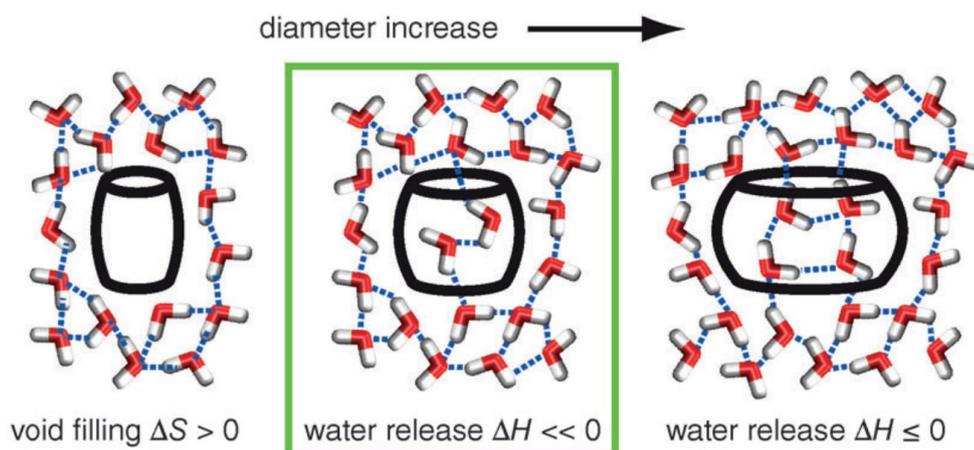
Figure 1.24: pictorial representation of size of cavity and the classical vs non classical hydrophobic effect.[48]

This "high energy water" effect is not really seen in coordination cages, since the size of the cavity is usually large enough for water to not be 'frustrated'. There is even crystallographic evidence of ice-like structure of water molecules inside a cage cavity (**figure 1.25**). Thus the type of hydrophobic effect we expect to see in our cage systems will be most likely the type closer to the "iceberg" interpretation.[50]
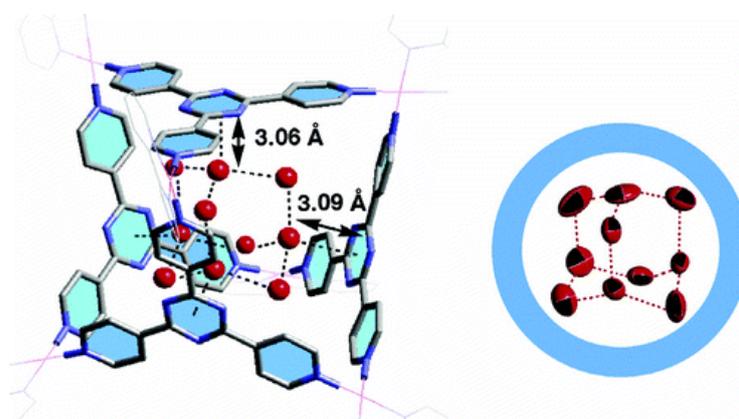
Figure 1.25: X-ray crystal structure of molecular I_c type Ice inside a cage cavity.[50]

1.3 Applications of Cages

1.3.1 Cages as molecular flasks

One way in which cages can be used is as "molecular flasks". The cavity inside the framework can be thought of as a container, or a flask. In a traditional glass flask, the size of the flask is very large in comparison to the molecules reacting in it, so it can be thought that the size and shape of the container does not influence the intermolecular interactions or reactions that occur. In a molecular flask however, the size of the flask is of comparable size to the molecule, so the size and shape of such a flask will change the reactivity and properties of the encompassed molecules.[51]

Biological examples of molecular flasks are enzymes. They have specific size and shaped pockets to bind substrates and then catalyse reactions with them. Reaction pathways in such flasks are most commonly influenced by a variety of non-covalent interactions by either stabilising the transition state, or by bringing 2 molecules together in close proximity (hence producing an increase in effective molarity).
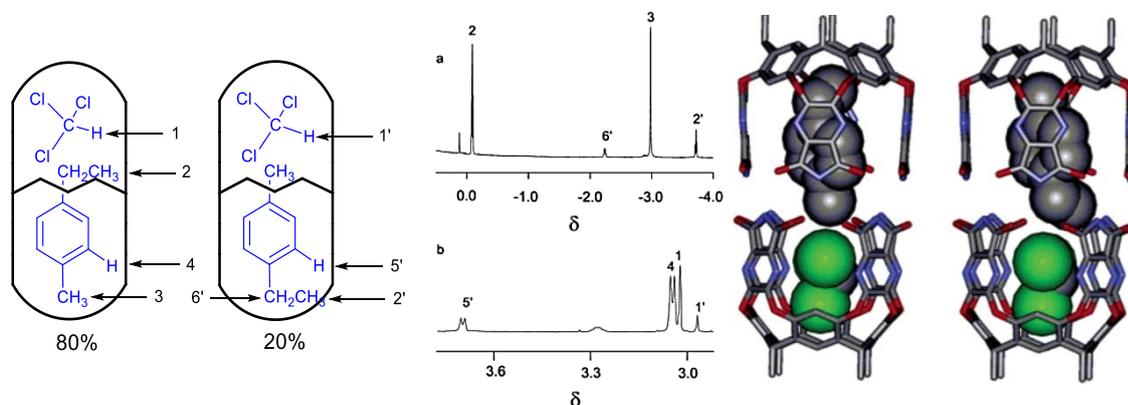
Figure 1.26: Social isomers of 4-ethyl toluene and $CHCl_3$ in a supramolecular capsule.[52]

Social isomers are a type of stereoisomer arising from the relative special positioning and conformation with the host of two or more non-covalently bound guest molecules. This type of isomerism arises because the shape and dimensions of the host prevent the guests from exchanging positions or tumbling on the NMR timescale.[52] An example is 4-ethyl toluene and $CHCl_3$ in a small capsule (**figure 1.26**). Two social isomers arise due to relative positions of the ethyl and methyl substituents of the 4-ethyl-toluene. i.e. whether

the ethyl or the methyl group is at the centre of the cavity. The isomer with the ethyl group in the centre forms as the dominant isomer, possibly due to improved interactions between host and guest.[52]

The water soluble cage in **figure 1.9** was used by the Fujita group to accelerate the room-temperature Diels-Alder reaction of 2-methyl-1,3-butadiene and 1,4-naphthoquinone. In aqueous solution the reactants are driven into the hydrophobic cavity of the cage and react with a 113-fold increase in rate. Quantitative yields were obtained, and the products could be easily extracted with organic solvents leaving the cage in the aqueous layer (**figure 1.27**).[53]
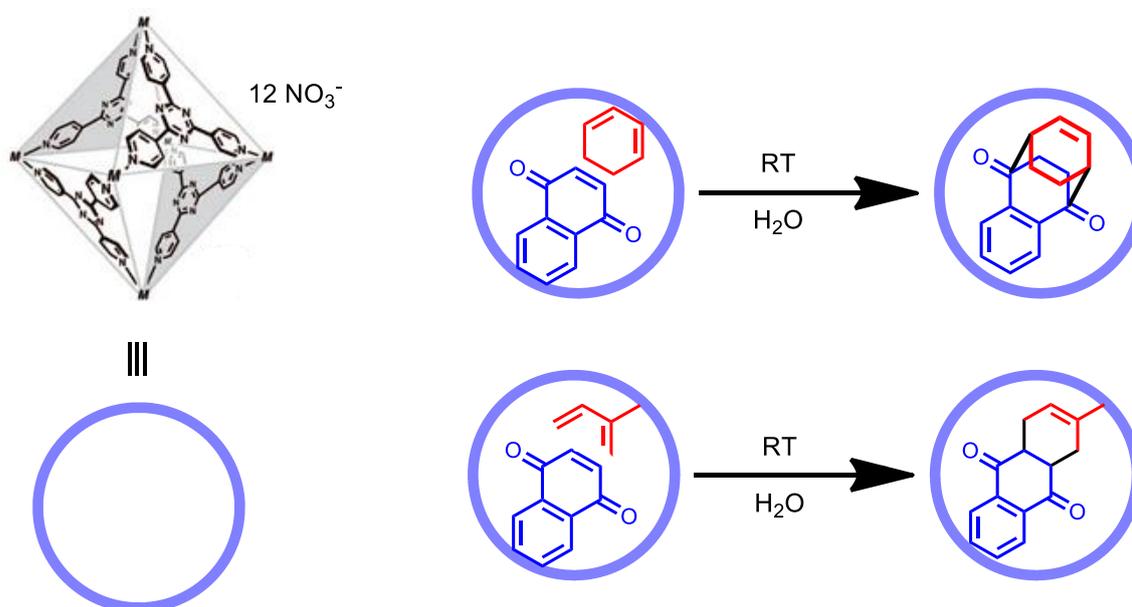
Figure 1.27: Diels-Alder reactions accelerated by the Fujita M_6L_4 octahedral cage.[53]

The steric constriction of the molecular flasks can be utilised to form unusual regio- and stereo-selectivity in Diels-Alder reactions. By encapsulating an appropriate dienophile such as N-cyclohexylmaleimide, along with an anthracene derivative in the cage in the same cage above, the syn 1,4 adduct is formed in high yields (R = CH_2OH, 98% yield; R = CO_2H, 92% yield). The usual product formed is the 9,10-adduct however, due to a restriction in the spatial orientation of the dienophile, inside the cage only the 1,4 adduct can be formed (**figure 1.28**).[53]
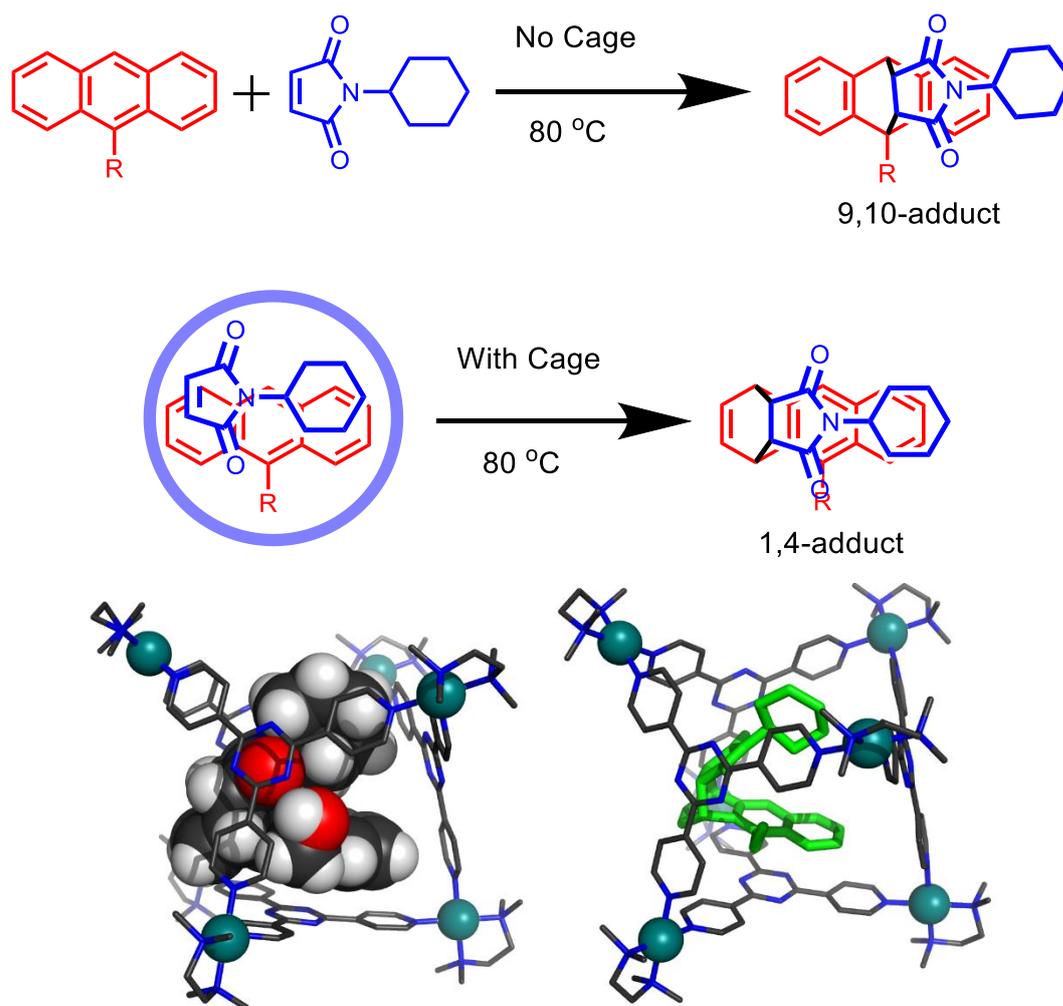
Figure 1.28: Diels-Alder reactions in a molecular flask, yielding the unusual syn 1,4-adduct in the presence of cage; and the 1,9-adduct in the absence of cage.[53]

There are many other examples of the use of cages in this way to produce unusual products and increasing reaction rates. However, despite accelerating the reaction rate, most of these systems do not make good catalysts. This is mainly due to product inhibition which prevents turnover and limits the number of catalytic cycles.[51,54-56]

1.3.2 Cages as drug delivery systems

Cages have the potential to act as nano-scale drug delivery systems. Not much work has been done in this area, but it has the potential to be a rather valuable tool in selective drug delivery. Cages have already been known to bind all sorts of guests inside their cavity, and cages can be assembled and disassembled and therefore release the bound guest. This may involve the use of a stimulus such as a competing ligand to displace the drug guest or even protonation, i.e. a change in pH. By designing the cage's structure so that it

disassembles in the region where the drug is to be targeted, it should be therefore possible to deliver drugs selectively to the desired location.

Tumour cells can be easily targeted by supramolecular assemblies, due to the enhanced permeability and retention (EPR) effect.[57,58] This effect means that tumour cells absorb and retain macromolecules much more than regular cells, so an accumulation of macromolecules within the tumour cells occurs.[57,58] This means that cages carrying particular anti-cancer drugs, or cages which are cytotoxic in their own right, can target and kill specifically cancer cells. Also because cages can bind hydrophobic guests easily, drugs which cannot normally be administered (due to water insolubility) could be delivered straight to the target cell. When inside the cage, or other host system, the anticancer drug is mostly inactive, and so this could be a huge step forward in limiting general cytotoxicity and side effects from drugs, as the effects only become active once released.

This type of drug delivery system has been referred to as a 'Trojan Horse'.[59] The Dyson group has demonstrated this using a ruthenium cage with cytotoxic [M(acac)$_2$] complexes (M = Pd(II), Pt(II)) bound inside, which are released into cancer cells (**figure 1.29**).[59]
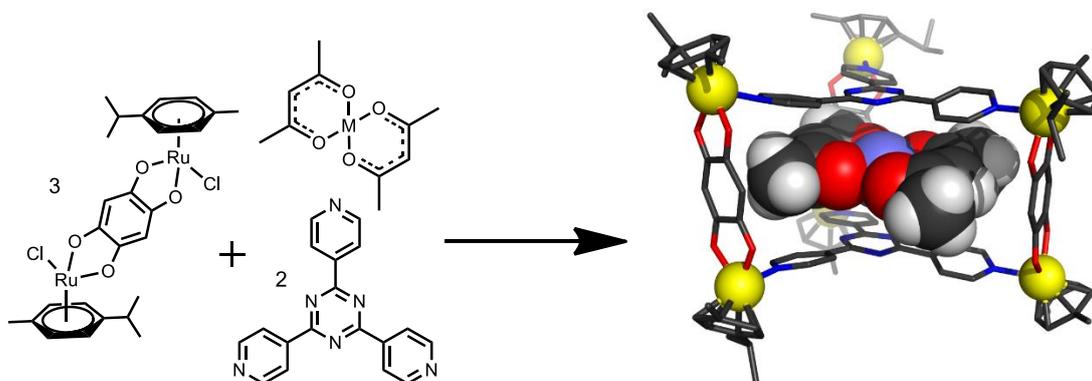
Figure 1.29: Synthesis of the [Ru$_6$(p-^{i}PrC$_6$H$_4$Me)$_6$(tpt)$_2$-(dhbq)$_3$]$^{6+}$ cage, the cage with M=Pd/Pt acac complexes encapsulated and its x-ray crystallographic structure.[59]

Each host-guest system [(acac)$_2$Pt⊂Rucage]$^{6+}$ and [(acac)$_2$Pd⊂Rucage]$^{6+}$ were tested along with free guest [Pd(acac)$_2$] and [Pt(acac)$_2$] and free cage against A2780 human ovarian cancer cells. The results are summarised in **figure 1.30**.

Complex	IC_{50}[a] [μm]
[Pt(acac)$_2$]	Insoluble
[Pd(acac)$_2$]	Insoluble
Ru Cage	23
[(acac)$_2$Pt⊂Rucage]$^{6+}$	12
[(acac)$_2$Pd⊂Rucage]$^{6+}$	1

[a] IC_{50}: drug concentration necessary for 50% inhibition of cell viability.

Figure 1.30: Cytotoxicity tests with A2780 human ovarian cancer cells.[59]

The free [M(acac)$_2$] complexes are insoluble in water on their own and show no signs of cytotoxicity. The free cage is reasonably cytotoxic. The host-guest assemblies are quite a bit more active with the Pt-containing cage being twice as cytotoxic, and the Pd-containing cage being one order of magnitude more cytotoxic than the free cage alone. Once inside the cell the drug is released, and the higher cytotoxicity of Pd(acac)$_2$ compared to Pt(acac)$_2$ could imply that Pd is more easily released.[59]

Another example of this is with cisplatin as a guest. Cisplatin is used to treat a variety of cancers, including ovarian, head and neck, bladder and cervical, and melanomas and lymphomas. It also cures over 90% of testicular cancer cases. The main problem with cisplatin is that it causes many side effects such as kidney damage, damage to the nervous system and bone marrow suppression. It is not very specific and so is cytotoxic to not just the cancer cells but a majority of the cells around the body. Crowley and his group are developing cages to encapsulate cisplatin and deliver it exclusively to cancer cells by taking advantage of the EPR effect. They attempted this with a [Pd$_2$L$_4$](X)$_4$ cage that encapsulates two molecules of cisplatin into its cavity and can be disassembled in the presence of a competing ligand such a 4-dimethylaminopyridine (DMAP) or Cl$^-$ (Bu$_4$NCl) thus releasing the bound cisplatin (**figure 1.31**).[60]
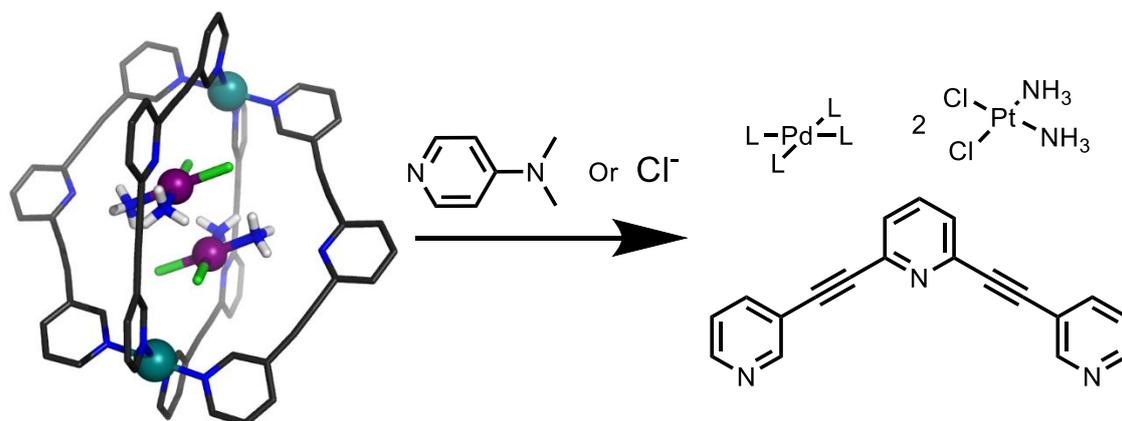
Figure 1.31: X-ray crystallographic structure and the disassembly of the cage-cisplatin host-guest complex where (i) DMAP (8 eq.) or Bu$_4$NCl (8 eq.).[60]

The encapsulation and release of cisplatin can be easily observed by ^{1}H NMR spectroscopy (**figure 1.32**).
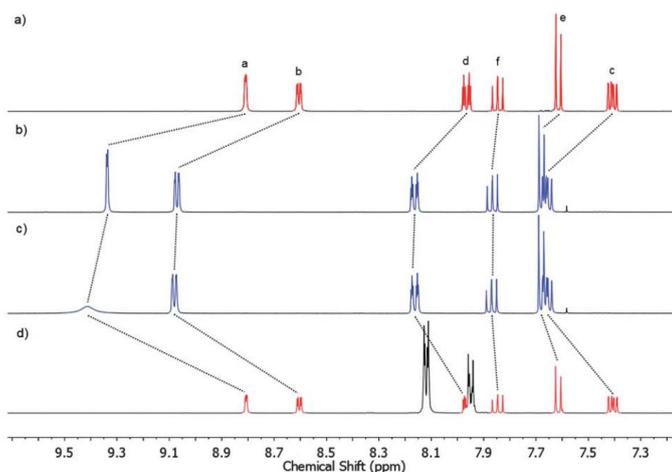
Figure 1.32: ^{1}H NMR spectra in CD$_3$CN of (a) the ligand; (b) the empty cage (X = BF$_4^-$); (c) the cisplatin host–guest adduct; and (d) the cisplatin host–guest adduct after the addition of DMAP(8 eq.).[60]

The cage can be reassembled in quantitative yields by addition of TsOH (8 eq.) or AgSbF$_6$ (excess) as these sequester the competing ligands which frees up coordination sites on Pd^{2+} for reassembly.[60]

1.3.3 Encapsulation of unstable molecules and intermediates

There are many chemical substances that are extremely reactive upon contact with oxygen and/or water. Cages can encapsulate such substances due to the highly hydrophobic binding pocket, thus rendering the once reactive substance effectively inert.

A good example to demonstrate this is the encapsulation of white phosphorus, P_4 in a tetrahedral cage (**figure 1.33**),[61] by Nitschke and co-workers.
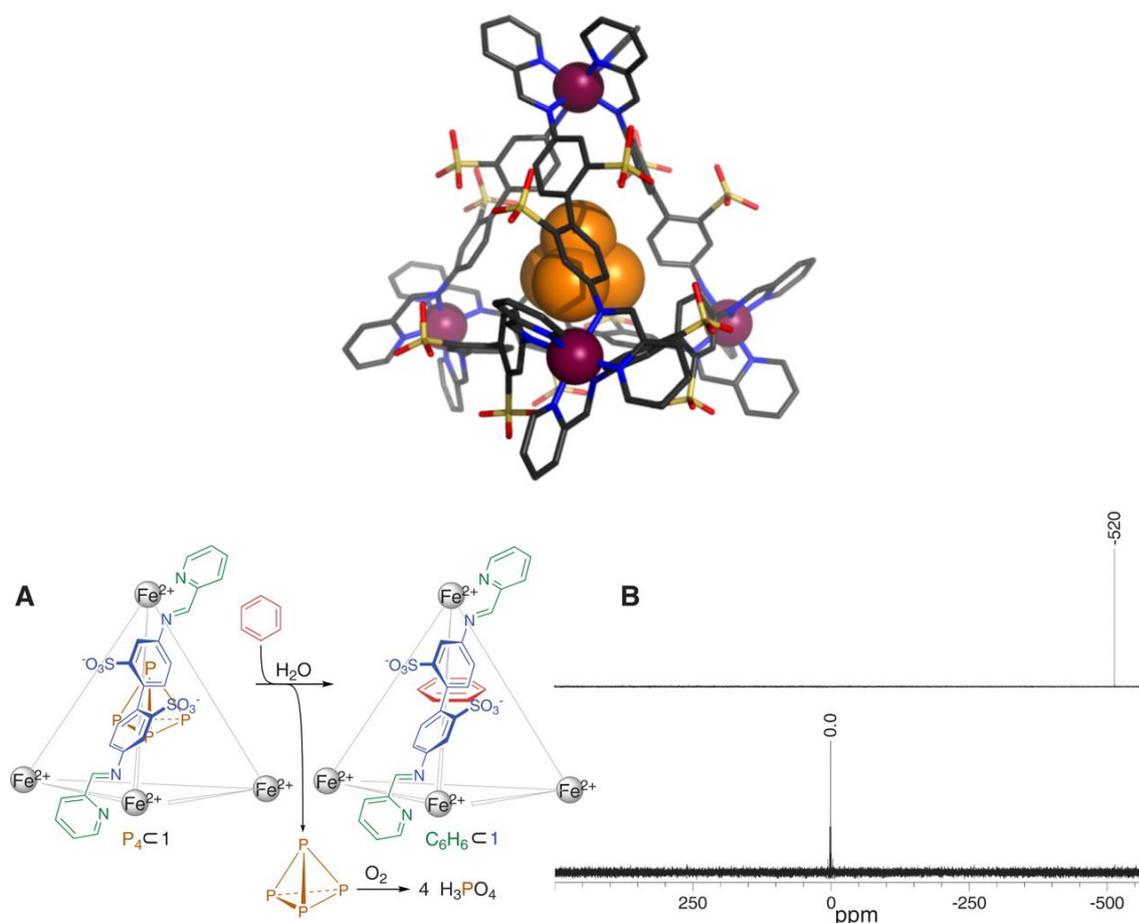
Figure 1.33: X-ray crystal structure showing P_4 encapsulated in the cage cavity (top); a scheme of P_4 encapsulation, displacement (with benzene) and oxidation (A); a ^{31}P NMR showing P_4 in benzene layer after displacement from cage (B, top) and of H_3PO_4 after oxidation with air (B, bottom).[61]

The P_4, when encapsulated, is completely stable in aqueous solution since its decomposition occurs via an intermediate that is too large for the cavity, and so is prevented. It can then be displaced by a competing guest (benzene) when it is quickly oxidised to phosphoric acid.

There have also been reports of cages stabilising reactive intermediates such as some generated by the ruthenium catalyst [CpRuCl(cod)] (cod=1,5-cyclooctadiene), which is used in many C-C bond forming reactions.[62]

In the reaction in **figure 1.34**, when inside the metal–ligand assembly, the reactive intermediates (1 and 3) are stable for several weeks at room temperature in aqueous solution. They are protected from the outside chemical environment and sheltered from possible reaction channels that lead to decomposition.[62] Not only this, the cage allows this catalytic reaction to be performed in aqueous media.
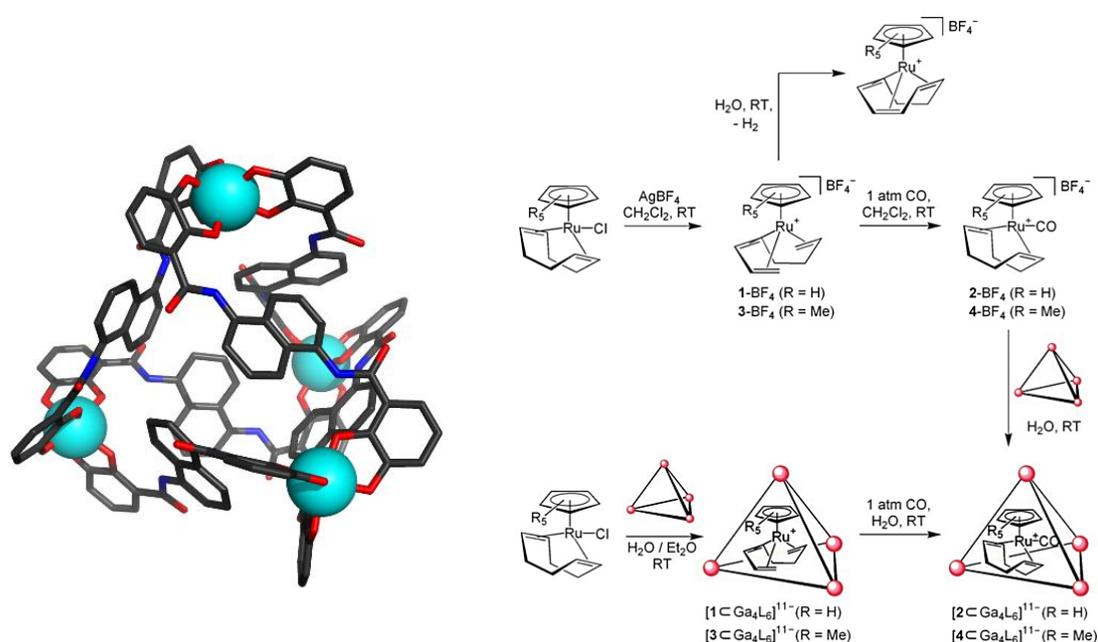
Figure 1.34: Summary of the reactivity of the [CpRuCl(cod)] catalyst showing that the reactive intermediates 1 and 3 are stabilized by the presence of the supramolecular host.[62]

1.4 The Ward Group Cages

1.4.1 Introduction

The Ward group cages first came into being by combining M^{2+} ions with a hexadentate tris(pyrazolyl)borate ligand, $[Tp^{py}]^-$, in a 1:1 stoichiometry. In this ratio the ligand can either act such as to form a mononuclear complex,[63] or alternatively each PyPz arm can act as a bidentate site with the ligand spanning three separate metal ions.[64] When this occurs a tetrahedral cage, $[M_4(Tp^{py})_4]^{4+}$, (M = Mn, Zn)$^{2+}$, is formed with each $[Tp^{py}]^-$ ligand capping each face of the tetrahedron (**figure 1.35**).
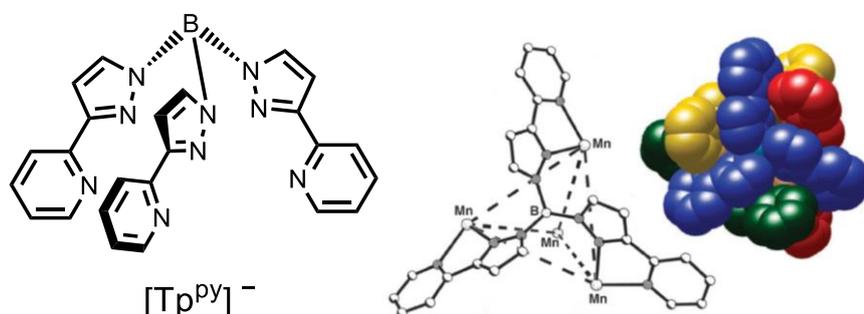
Figure 1.35: $[Tp^{py}]^-$ ligand and its M_4L_4 tetrahedral cage with Mn(II).[64]

From there by using a six-coordinate metal ion such a Co(II) or Zn(II) with the ligand L^{o-Ph}, combined in the correct ratio, another tetrahedral cage was produced (**figure 1.36**).[65]
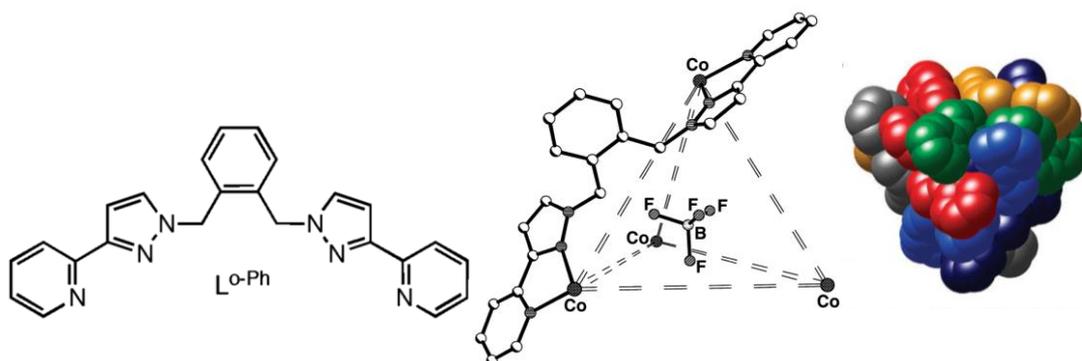
Figure 1.36: L^{o-Ph} ligand and its M_4L_6 tetrahedral cage with Co(II).[65]

The type of ligand used in **figure 1.36** is much better than the $[Tp^{py}]^-$ one (**figure 1.35**), as not only are the N-B bonds fragile and thus $[Tp^{py}]^-$ type ligands are prone to hydrolytic

decomposition, but also the ligand type in **figure 1.36** can offer greater synthetic control by allowing the variation of the phenyl spacer.

The cages assembled by the Ward group are generally prepared by mixing a ligand composed of two PyPz units connected via some spacer unit, such as naphthalene, with the metal ions (Co(II), Cd(II), Zn(II)) in a 3:2 L:M ratio. By varying the spacer unit and the positions from which the PyPz units join the spacer unit, it is possible to assemble cages with wide variety of shapes and sizes. Examples of ligands used in cage preparation are given in **figure 1.37**.
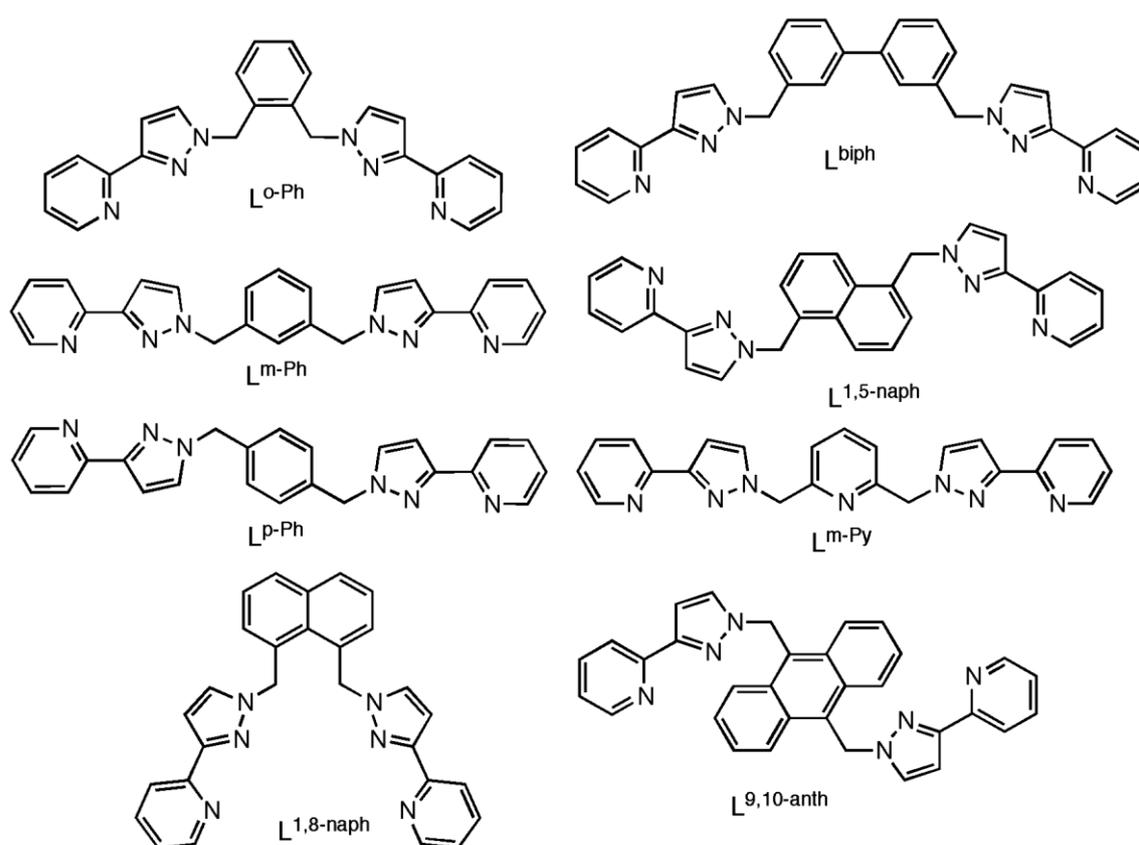
Figure 1.37: Examples of some ligands used by the Ward group in their polynuclear cage assemblies.[66]

The 2M:3L ratio is used to satisfy the 'maximum site occupancy' principle,[13] which suggests that highest stability occurs when the metal and ligand coordination numbers are 'matched'. i.e. 1.5 equivalents of a tetradentate ligand, and a 6-coordinate metal. This results in polyhedral that also have a 2:3 ratio of verticies : edges.[66]

1.4.2 The cubic cage and its host-guest chemistry

The majority of work conducted in this Thesis will be based on the M_8L_{12} cubic cage shown in **figure 1.38** based on Co(II) ions at the vertices and $L^{1,5\text{-naph}}$ ligand along all 12 edges.[67]
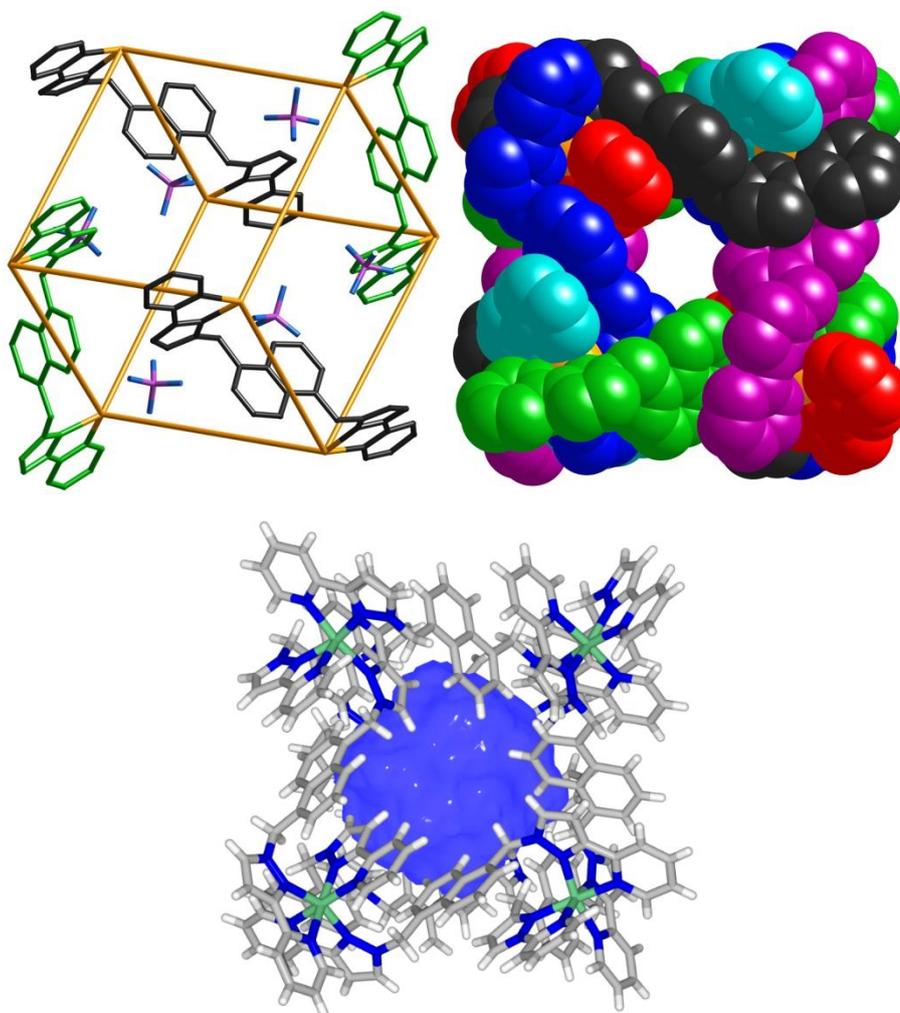
Figure 1.38: X-ray crystal structure of the $[M_8L_{12}]^{16+}$ cube with Co(II) and $L^{1,5\text{-naph}}$ (top), and the internal van der Waals surfaces and cavity volume (bottom, blue) .[67,68]

This cage is not only one of the more stable cages that have been assembled, but it has shown some interesting host-guest chemistry.

It has a large spherical cavity with volume of 407 $Å^3$ (**figure 1.38, bottom**) and six windows, one on each face of the cube, provide access to this cavity. Space-filling models indicate that the cross-section of each window is 4 $Å^3$, which provides sufficient space for a molecular guest to enter the cage. Also the crystal structure of the cube shows that the

central cavity is occupied only by solvent molecules (MeOH), not counter-ions; so the neutral guest molecule will not have to compete with anions for occupation of the cavity of the cationic cage.[68]

Rebek showed that when a guest fills 55% of the total available volume, host–guest interactions in molecular capsules are optimised.[69] So the ideal guest volume for this cage is 224 Å^3. Initially, various guests were screened for binding in MeCN-d_3 via NMR titrations, and the guests which seem to bind the best are coumarin (K = 78 M^{-1}) and its analogues.[68] Although now, several stronger binding guests have been found (**figure 1.39**).[70]

K = 78 M^{-1} K = 600 M^{-1} K = 880 M^{-1} K = 2100 M^{-1} K = 2700 M^{-1}
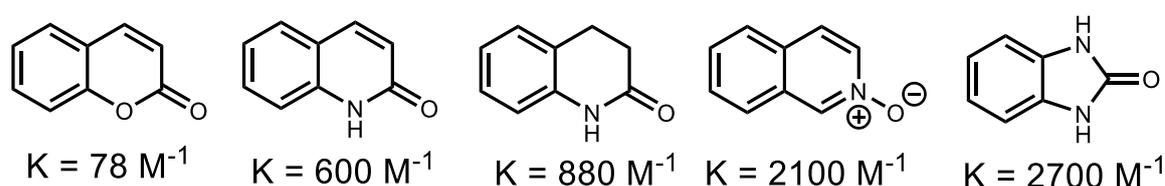
Figure 1.39: Various guests and their binding constants, K, with the $L^{1,5\text{-naph}}$ M_8L_{12} cube in MeCN-d_3.70

The binding studies were all done with the metal as Co(II). Co(II) is paramagnetic and so the ^{1}H NMR peaks spread out over a chemical shift range of +100 to -100 ppm (**figure 1.40**).[67]
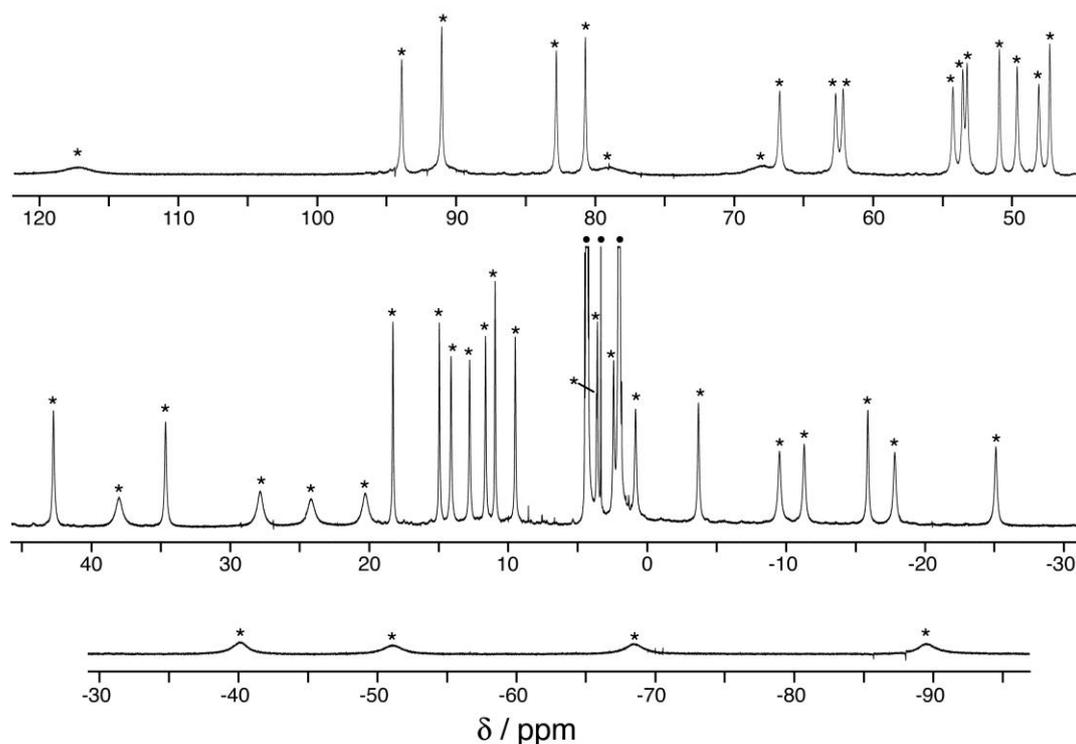
Figure 1.40: Paramagnetic ^{1}H NMR spectrum of the Co(II) $L^{1,5\text{-naph}}$ M_8L_{12} cube.[67]

Not only this, but when a guest binds, the observed peak shifts are large and so it is very easy to see if that particular guest is binding or not. Also it allows for easier calculation and data analysis as peaks very rarely overlap, so there is no need to deconvolute them (**figure 1.41**).[67]
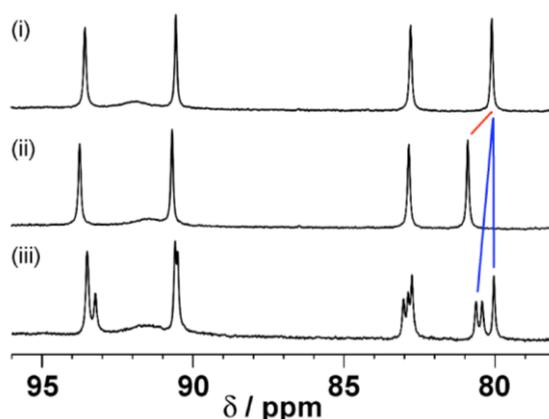
Figure 1.41: ^{1}H NMR titration in CD$_3$CN of the Co(II) L^{1,5nap} M$_8$L$_{12}$ cube. (i) Free host; (ii) addition of a guest that binds in fast exchange and (iii) addition of a guest that binds in slow exchange.[67]

The reason that this cage binds the guests is predominantly thought to be hydrogen bonding with a H-bond donor pocket inside the cage cavity. The ligands coordinate in a meridional tris-chelate geometry around all but two of the metal sites, where they are in a facial arrangement. This has a significant impact on the host guest chemistry, as around the fac tris-chelate sites, the CH$_2$ groups on the ligand converge to a point inside the cage (**figure 1.42**), which allows hydrogen bonding to occur with guest molecules.[67,68] The electrostatic surface potential also backs this hypothesis up, showing a highly positive (blue) potential at this site.
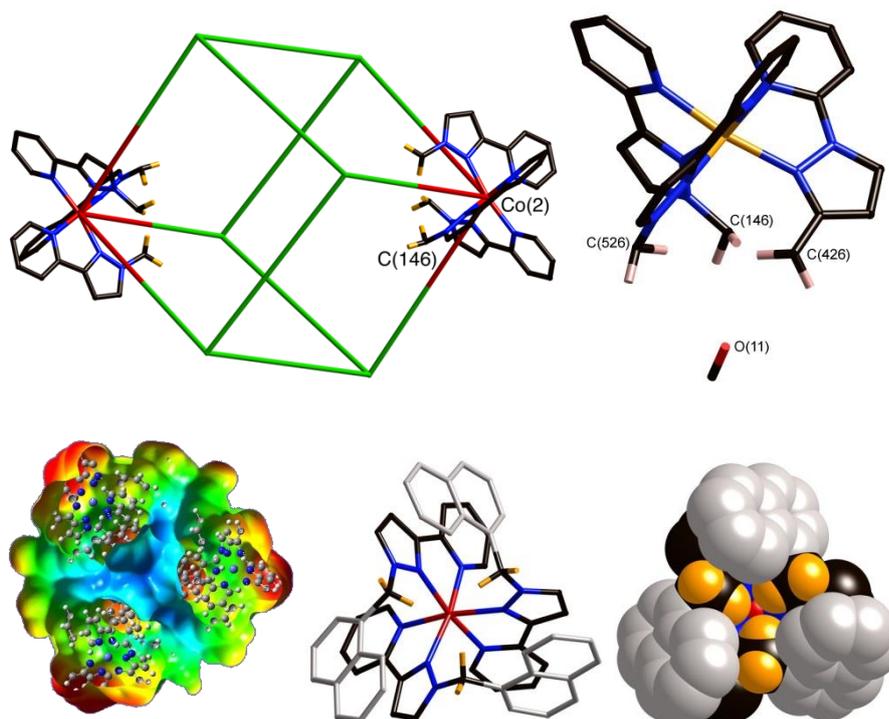
Figure 1.42: X-ray crystal structure of the $L^{1,5-naph}$ M_8L_{12} cube showing the CH_2 groups pointing inside the cage to create a hydrogen bonding binding pocket; a MeOH molecule in the pocket; the electrostatic surface potential map of the fac-site; a stick and space-filled representation of the fac-site.[67,68]

When this cage was assembled using Cd(II), ^{113}Cd NMR spectroscopy clearly showed that 2 of the metal ions were in a different environment to the other 6 with a 3:1 ratio of Cd(II) peaks being observed (**figure 1.43**).[67]
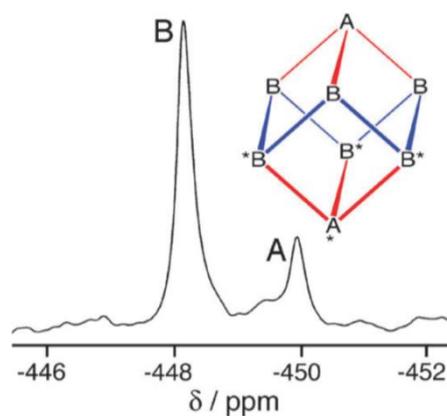
Figure 1.43: ^{113}Cd NMR of the $L^{1,5-naph}$ M_8L_{12} cube, showing a 3:1 ratio of B:A (mer : fac).[67]

This hydrogen bonding site was tested using mononuclear fac and mer isomers synthesised with kinetically inert Ru(II). It was found that the Fac isomer bound guests whereas the mer isomer didn't. The X-ray crystal structure the fac complex included an acetone molecule H-bonding to the complex (**figure 1.44**).[71]
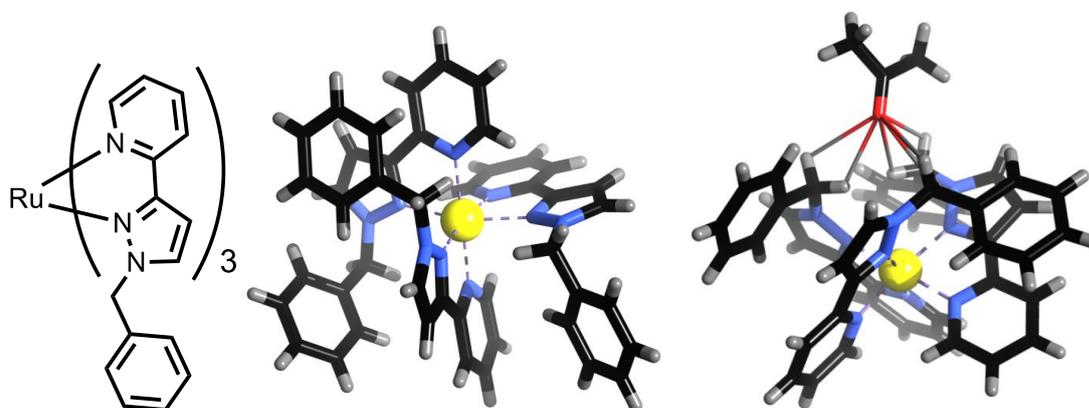
Figure 1.44: Chemdraw of the RuPyPzbenzyl complex and the mer and fac X-ray crystal structures[71]

The donor strength of this hydrogen bonding pocket (α) was determined to be 4.1 (similar to a phenol) using a systematic study in which the hydrogen bond acceptor strength (β) of the guest was steadily increased, resulting in stronger binding (**figure 1.45**).[72]
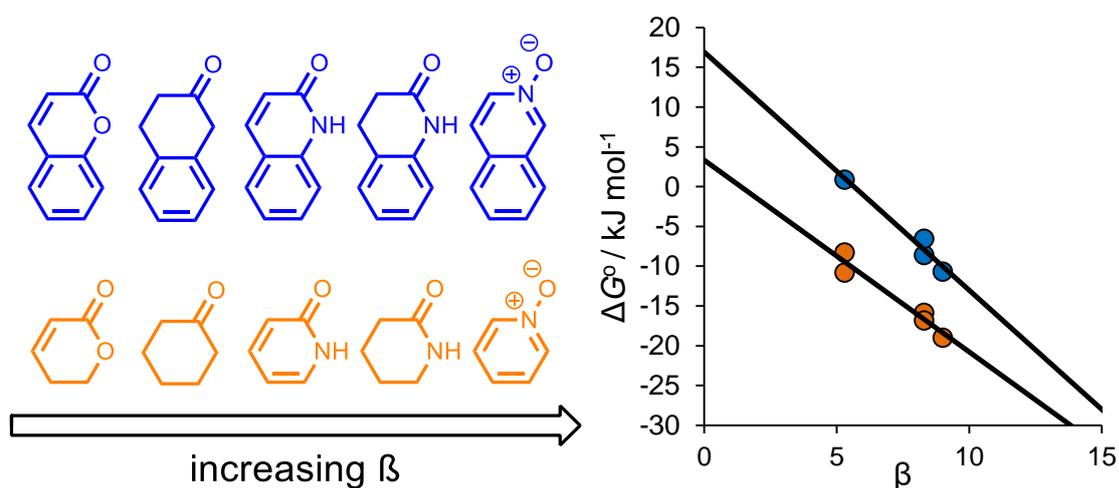
Figure 1.45: ΔG as a function of increasing H-bond acceptor strength (β) of the guests[72]

The binding strength of the guests in MeCN is relatively weak, so an isostructural cage was synthesised that was water soluble (**figure 1.46**). This would allow the guests to bind much stronger by taking advantage of the hydrophobic effect. The new ligand is essentially the same, but with CH_2OH groups on the exterior surface, which when assembled into the cage would cover the exterior surface rendering it water soluble whilst retaining the hydrophobic interior of the cavity.[72]
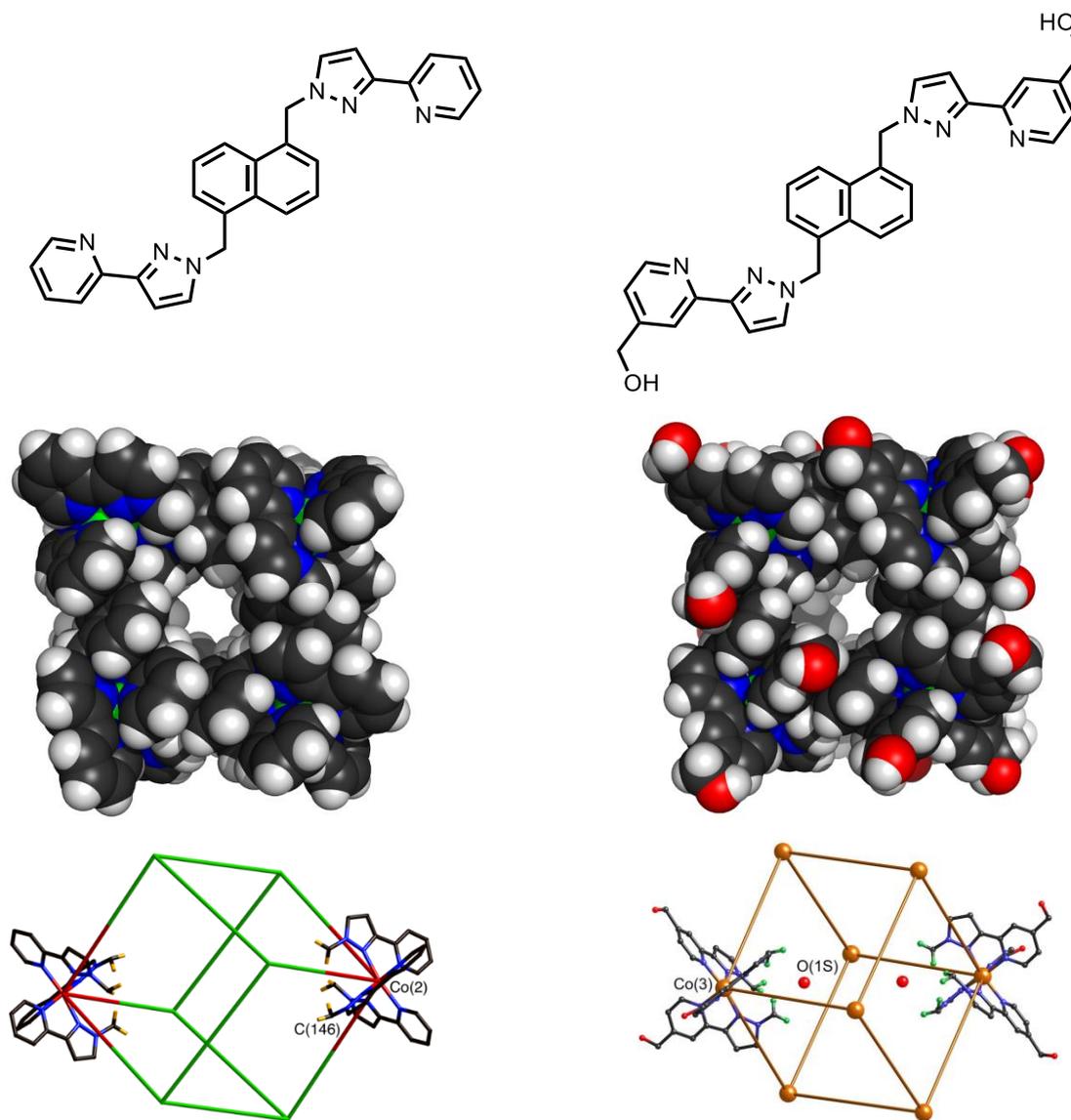
Figure 1.46: Structures of (top) parent and OH-functionalised bridging ligand; space filling model of the M_8L_{12} cubic cages that they form (middle); and their showing the two interior H-bonding sites (bottom).[67,72]

The isostructural water-soluble cage also contains the hydrogen bonding pockets on the interior surface, so initially the same set of guests used for the H-bond strength investigation had their binding constants measured in water to see how well the major driving force in MeCN held up in polar solvents. Due to the strong competing hydrogen bonds that water can make, the strength of the H-bond acceptor of the guest had little effect on binding strength. If anything the stronger the H-bond acceptor property of the guest, the weaker the binding in the cage, presumably because it can make stronger H-bonds with the bulk water (**figure 1.47**).[72]

Figure 1.47: ΔG as a function of increasing H-bond acceptor strength (β) of the guests in MeCN (blue and orange) and water (green and red).[72]

From here studies into the effect of substituents on binding in different solvents was investigated using thermodynamic cycles. From the cycle in **figure 1.48**, it was found that adding an aromatic ring to the guest increased the free energy of binding in water by around 9 kJ Mol^{-1}. I.e. the hydrophobic contribution to guest binding for the aromatic group was 9 kJ Mol^{-1} consistent with expectations based on its surface area.[72]
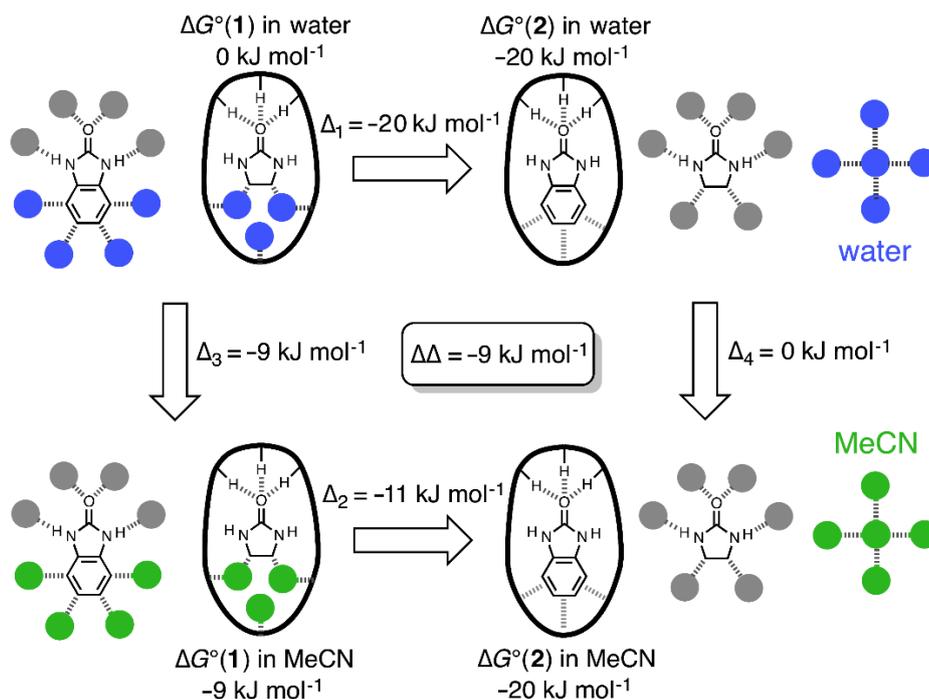
Figure 1.48: Thermodynamic cycle extracting the hydrophobic contribution of an aromatic ring when going from MeCN (green) to H_2O (blue).[72]

1.4.3 Other cage polyhedra

By varying the spacer in the bridging ligand, a multitude of various different polyhedral cages have been assembled, from the small M_4L_6 tetrahedron to the very large (with an equally large name) $M_{16}L_{24}$ tetra-capped-truncated-tetrahedron (**figure 1.49**).[66]
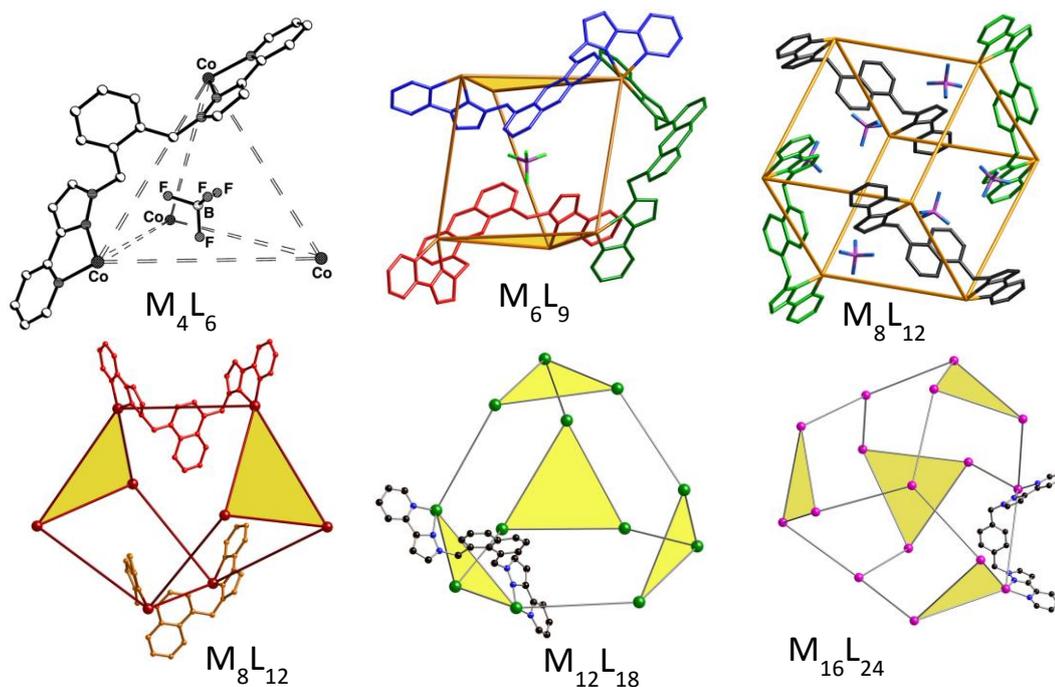
Figure 1.49: Examples of the polyhedral cages assembled in the Ward group.[66]

Some recent cages have been assembled using 2 different metal ions to add extra functionality to the cages. One such example used inert Ru(II) metal ions to form a mononuclear complex of the the fac or mer isomers, which when combined in the correct ratio with a labile metal (Co(II) or Cd(II) for example) resulted in complete assembly of the mixed-metal cage (**figure 1.50**).[73]
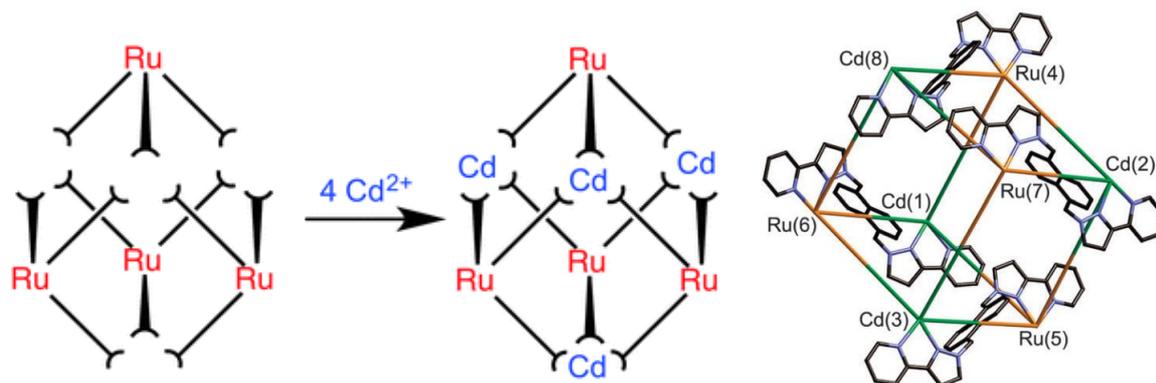
Figure 1.50: Ru$_4$Cd$_4$ mixed metal cube.[73]

1.5 Project Aims

The aim of this project is to synthesise new cages that display host-guest chemistry, and investigate further the host-guest interactions and properties of cages, particularly in water. Then attempt to utilise what we have learned to predict binding and tailor the cages for applicational uses such as catalysis, and reversible uptake and release mechanisms.

1.6 References

1) J. W. Steed; J. L. Atwood, *Supramoleculer Chemistry*, Wiley, **2009**

2) M. Fujita; O. Sasaki; T. Mitsuhashi; T. Fujita; J. Yazaki; K. Yamaguchi; K. Ogura, *Chem. Commun.,* **1996**, *20*, 1535-1536

3) P. R. Ashton; C. L. Brown; E. J. T. Chrystal; T. T. Goodnow; A. E. Kaifer; K. P. Parry; D. Philp; A. M. Z. Slawin; N. Spencer; J. F. Stoddart; D. J. Williams, *J. Chem. Soc., Chem. Commun.,* **1991**, 634-639

4) B. Hasenknopf; J-M. Lehn; b. O. Kneisel; G. Baum; D. Fenske, *Angew. Chem. Int. Ed.,* **1996**, *35*, 1838-1840

5) S. Toksöz; M. O. Guler, *Nano Today*, **2009**, *4*, 458–469

6) M. Fujita; J. Yazaki; K. Oguraa, *J. Am. Chem. Soc.*, **1990**, *112*, 5645–5647

7) R. V. Slone; J. T. Hupp; C. L. Stern; T. E. Albrecht-Schmitt, *Inorg. Chem.*, **1996**, *35*, 4096–4097

8) R. V Slone; D. I. Yoon; R. M. Calhoun; J. T. Hupp, *J. Am. Chem. Soc.,* **1995**, *117,* 11813–11814

9) P. de Wolf; P. Waywell; M. Hanson; S. L. Heath; A. J. H. M. Meijer; S. J. Teat; J. A. Thomas, *Chem. A Eur. J.*, **2006**, *12*, 2188–2195

10) H. Ahmad; B. W. Hazel; A. J. H. M. Meijer; J. A. Thomas; K. A. Wilkinson, *Chem. A Eur. J.*, **2013**, *19,* 5081–5087

11) J-M. Lehn; a Rigault; J. Siegel; J. Harrowfield; B. Chevrier; D. Moras, *Proc. Natl. Acad. Sci.*, **1987**, *84*, 2565–2569

12) A. Pfeil; J-M. Lehn, *J. Chem. Soc. Chem. Commun.*, **1992**, 838-840

13) R. Kramer; J-M Lehn; A. Marquis-Rigault, *Proc. Natl. Acad. Sci.*, **1993**, *90*, 5394–5398

14) K. S. Chichak; S. J. Cantrill; A. R. Pease; S.-H. Chiu; G. W. V Cave; J. L. Atwood; J. F. Stoddart, *Science*, **2004**, *304,* 1308–1312

15) L.-E. Perret-Aebi; A. von Zelewsky; C. Dietrich-Buchecker; J.-P. Sauvage, *Angew. Chem. Int. Ed.,* **2004**, *43*, 4482–448

16) K. M. Mullen; P. D. Beer, *Chem. Soc. Rev.*, **2009**, *38*, 1701–1713

17) R. W. Saalfrank; A. Stark; M. Bremer; G-U. Hummel, *Angew. Chem. Int. Ed.,* **1990**, *3*, 311-314

18) M. Fujita; D. Oguro; M. Miyazawa; H. Oka; K. Yamaguchi; K. Ogura, *Nature*, **1995**, *378*, 469-471

19) T. Kusukawa; M. Fujita, *J. Am. Chem. Soc.*, **2002**, *124*, 13576–13582

20) J. Bunzen; J. Iwasa; P. Bonakdarzadeh; E. Numata; K. Rissanen; S. Sato; M. Fujita, *Angew. Chem. Int. Ed.*, **2012**, *51*, 3161–3163

21) K. A. Connors, *Binding Constants: The Measurement of Molecular Complex Stability,* **1987**

22) P. Thordarson, *Chem. Soc. Rev.*, **2011**, *6*, 1305–1323

23) D. E. Koshland, *Angew. Chem. Int. Ed.,* **1994,** *33*, 2375-2378

24) I. R. Kleckner; M. P. Foster, *Biochim. Biophys. Acta*, **2011**, *1814*, 942–68

25) C. A. Hunter, *Angew. Chem. Int. Ed.,* **2004**, *43*, 5310-5324

26) B. Franklin; W. Brownrigg; Farish, *Phil. Trans.,* **1774**, *64*, 445-460

27) D. Chandler, *Nature*, **2005**, *437*, 640-647

28) H. S. Frank; M. W. Evans, *J. Chem. Phys.,* **1945**, *13*, 507-532

29) W. Kauzmann, *Adv. Protein Chem.,* **1959**, *14*, 1-63

30) W. Blokzijl; J. B. F. N. Engberts, *Angew. Chem. Int. Ed.,* **1993**, *32*, 1545-1579

31) P. Buchanan; N. Aldiwan; A. K. Soper; J. L. Creek; C. A. Koh, *Chem. Phys. Lett.,* **2005**, 89-93

32) K. W. Miller; J. H. Hildebrand, *J. Am. Chem. Soc., 1968, 90*, 3001-3004

33) O. W. Howarth, *J. Chem. Soc. Faraday. Trans. 1,* **1975**, *71* 2303-2309

34) B. Lee, *Biophysical. Chem.,* **1994**, *51*, 271-278

35) B. Madan; B. Lee, *Biophys. Chem.,* **1994**, *51*, 279-289

36) I. T. S. Li; G. C. Walker, *Proc. Natl. Acad. Sci.,* **2011**, *108*, 16527-16532

37) N. Galamba, *J. Phys. Chem. B,* **2013**, *117*, 2153-2159

38) G. Graziano, *J. Phys. Chem. B,* **2014**, *118*, 2598-2599

39) N. Galamba, *J. Phys. Chem. B,* **2014**, *118*, 2600-2603

40) L. R. Pratt; A. Pohorille, *Proc. Natl. Acad. Sci.,* **1992**, *89,* 2995-2999

41) N. Muller, *Acc. Chem. Res.* **1990**, *23*, 23-28

42) D. Mirejovsky; E. M. Arnett, *J. Am. Chem. Soc.,* **1983**, *105*, 1112-1117

43) G. Hummer; S. Garde; A. E. Garcia; A. Pohorille; L. R. Pratt, *Proc. Natl. Acad. Sci.,* **1996**, *93*, 8951-8955

44) J. G. Davis; K. P. Gierszal; P. Wang; D. Ben-Armotz, *Nature*, **2012**, *491*, 582-585

45) S. B. Ferguson; E. M. Seward; F. Diederich; E. M. Sanford; A. Chou; P. Inocenico-Szweda; C. D. Knobler, *J. Org. Chem.,* **1988**, *53*, 5593-5595

46) R. J. Bingham; J. B. C. Findlay; S-Y. Hsieh; A. P. Kalverda; A. Kjellberg; C. Perazzolo; S. E. V. Phillips; K. Seshadri; C. H. Trinh; W. B. Turnbull; G. Bodenhausen; S. W. Homans, *J. Am. Chem. Soc.,* **2004**, *126*, 1675-1681

47) S. W. Homans, *Drug Disc. Today,* **2007**, *12*, 534-539

48) F. Biedermann; W. M. Nau; H-J. Schneider, *Angew. Chem. Int. Ed.,* **2014**, *53*, 11158-11171

49) F. Biedermann; V. D. Uzunova; O. A. Scherman; W. M. Nau; A. De Simone, *J. Am. Chem. Soc.,* **2012**, *134*, 15318-15323

50) M. Yoshizawa; T. Kusukawa; M. Kawano; T. Ohhara; I. Tanaka; K. Kurihara; N. Niimura; M. Fujita, *J. Am. Chem. Soc.,* **2005**, *127*, 2798-2799

51) M. Yoshizawa; J. K. Klosterman; M. Fujita, *Angew. Chem. Int. Ed.*, **2009**, *48*, 3418–3438

52) A. Shivanyuk; J. Rebek, *J. Am. Chem. Soc.*, **2002**, *124*, 12074–12075

53) M. Yoshizawa; M. Tamura; M. Fujita, *Science*, **2006**, *312*, 251–254

54) M. Raynal; P. Ballester; A. Vidal-Ferran; P. W. N. M. van Leeuwen, *Chem. Soc. Rev.,* **2014**, *43*, 1734-1787

55) H. Vardhan; F. Verpoort, *Adv. Synth. Catal.,* **2015**, *357*, 1351-1368

56) C. J. Brown; F. D. Toste; R. G. Bergman; K. N. Raymond, *Chem. Rev.,* **2015**, *115*, 3012-3035

57) Y. Matsumura; H. Maeda, *Cancer Res.,* **1986**, *46*, 6387–6392

58) D. Baban; L. Seymour, *Adv. Drug Deliv. Rev.*, **1998**, *34*, 109–119

59) B. Therrien; G. Süss-Fink; P. Govindaswamy; A. K. Renfrew; P. J. Dyson; *Angew. Chem. Int. Ed*, **2008**, *47*, 3773–3776

60) J. E. M. Lewis; E. L. Gavey; S. A. Cameron; J. D. Crowley, *Chem. Sci.,* **2012**, *3*, 778-784

61) P. Mal; B. Breiner; K. Rissanen; J. R. Nitschke, *Science*, **2009**, *324*, 1697–1699

62) D. Fiedler; R. G. Bergman; K. N. Raymond, *Angew. Chem. Int. Ed.*, **2006**, *118*, 759–762

63) G. M. Davies; J. C. Jeffery; M. D. Ward, *New J. Chem.*, **2003**, *27*, 1550-1553

64) A. J. Amoroso; J. C. Jeffery; P. L. Jones; J. A. McCleverty; P. Thornton; M. D. Ward, *Angew. Chem. Int. Ed.,* **1995**, *34*, 1443-1446

65) J. S. Fleming; K. L. V Mann; C-A. Carraz; E. Psillakis; J. C. Jeffery; J. A. McCleverty; M. D. Ward, *Angew. Chem. Int. Ed.,* **1998**, *37*, 1279–1281

66) M. D. Ward, *Chem. Commum.,* **2009**, 4487-4499

67) I. S. Tidmarsh; T. B. Faust; H. Adams; L. P. Harding; L. Russo; W. Clegg; M. D. Ward, *J. Am. Chem. Soc.*, **2008**, *130*, 15167–15175

68) S. Turega; M. Whitehead; B. R. Hall; M. F. Haddow; C. A. Hunter; M. D. Ward, *Chem. Commun.,* **2012**, *48*, 2752-2754

69) S. Mecozzi; J. Rebek, *Chem. Eur. J.,* **1998**, *4*, 1016–1022

70) S. Turega; M. Whitehead; B. R. Hall; A. J. H. M. Meijer; C. A. Hunter; M. D. Ward, *Inorg. Chem.,* **2013**, *52*, 1122-1132

71) A. J. Metherell; W. Cullen; A. Stephenson; C. A. Hunter; M. D. Ward, *Dalton Trans.,* **2014**, *43,* 71-84

72) M. Whitehead; S. Turega; A. Stephenson; C. A. Hunter; M. D. Ward, *Chem. Sci.,* **2013**, *4*, 2744-2751

73) A. J. Metherell; M. D. Ward, *Chem. Commun.,* **2014**, *50*, 6330-6332

Chapter 2

Synthesis, characterisation and manipulation

of a three component self-assembled system

2.1 Introduction

As mentioned in chapter 1, it is possible to form different shapes of cages by changing the spacer part of the ligand. In the Ward group, currently only the cubic M_8L_{12} cage (made with the 1,5-naphthalene spacer) has demonstrated the ability to bind guests, due mainly to its hydrogen bonding pocket, which helps guests bind in MeCN, and its hydrophobicity which helps guests to bind in water.[1,2]

There are other cages such as the M_4L_6 tetrahedron and the $M_{16}L_{24}$ tetracapped-truncated tetrahedron that also have these H-bond binding sites, but none of them showed guest encapsulation.[3] In the M_4L_6 case, there is a very strongly bound (BF_4^-, or ClO_4^-) anion inside the cavity, which not only templates its formation, but also cannot be removed.[4] In the $M_{16}L_{24}$ case, despite its large cavity, it too is full of anions, which most likely impede binding, particularly given that binding in MeCN, even with the cube, is generally weak.[3]

The $M_{16}L_{24}$ cage also suffers from interconversion between another cage species (M_6L_9). The M_6 species was found to prevail in solution, but it crystallised out as the M_{16} species (**figure 2.1**). Since it rearranges in solution it is not an ideal host despite the presence of a large cavity.[5]
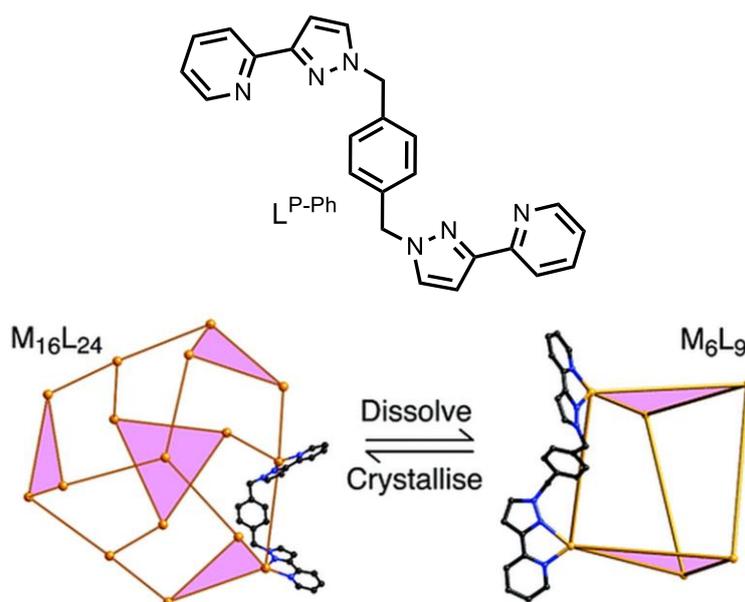
Figure 2.1: Schematic showing the ligand $L^{p\text{-}ph}$ (left) and interconversion between the $M_{16}L_{24}$ and M_6L_9 cages with Cd(II) (right).[5]

$M_{12}L_{18}$ truncated tetrahedron

The $M_{12}L_{18}$ truncated tetrahedron is another possible host worth investigating. The framework is a tetrahedron with its vertices cut off and is assembled with the 1,8-napthalene spaced ligand. It is a polyhedron with hexagonal and triangular faces **(figure 2.2)**. The cage forms in a racemic mixture however; each metal in the same cage has the same optical configuration, which is necessary for a closed pseudo-spherical surface to form. This is quite remarkable as it means 72 metal-ligand bonds need to be formed in the correct optical configuration during the self-assembly if the cage is to be formed.[6]
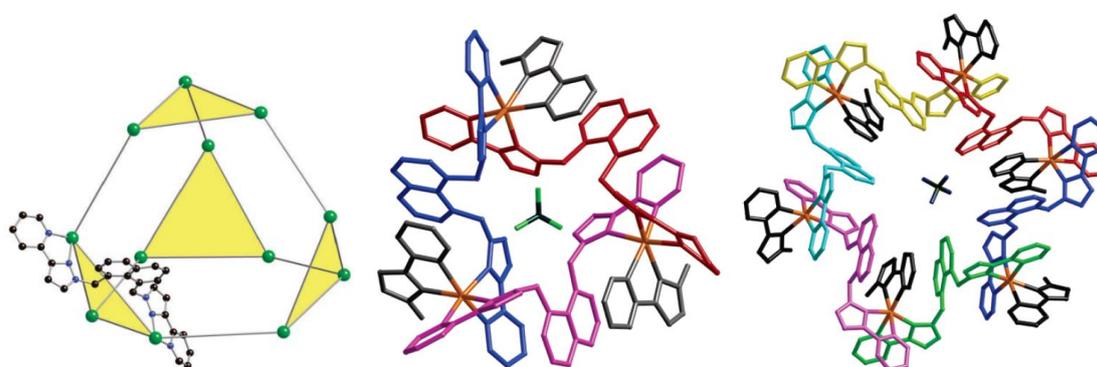
Figure 2.2: X-ray crystallographic structure the Co(II) $L^{1,8nap}$ $M_{12}L_{18}$ cage showing the whole structure, the triangular face and the hexagonal face.[6]

The aim of this work was to synthesise a novel water soluble cage based on the ligand $L^{1,8-naph}$ by incorporating alcohol groups into it, to form the $[M_{12}(L^{1,8-naph})_{18}]^{24+}$ cage and investigate its host guest chemistry in water (**figure 2.3**). We hoped that by taking advantage of the hydrophobic effect, it should be possible to bind guests in water in the same way as we had observed with the M_8L_{12} cube in water.
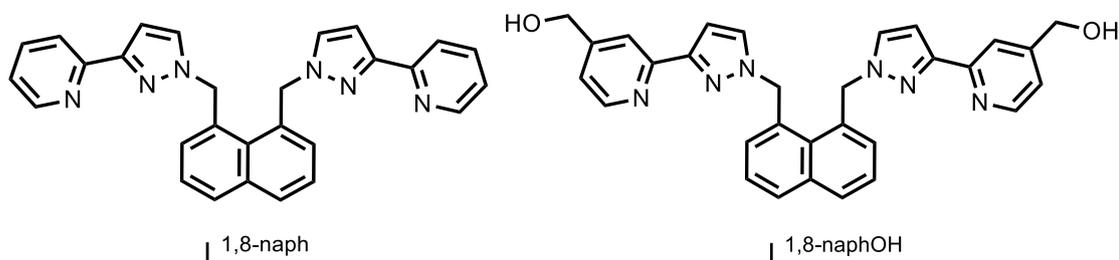
$L^{1,8-naph}$

$L^{1,8-naphOH}$

Figure 2.3: Ligand $L^{1,8-naph}$ (top left) previously used to assemble the $[M_{12}(L^{1,8-naph})_{18}]^{24+}$ cage (top right) for use in organic solvents and Ligand $L^{1,8-naphOH}$ to be synthesised to assemble an isostructural cage for use in water (bottom)

2.2 Results and Discussion

2.2.1 Synthesis of the L$^{1,8\text{-naphOH}}$ ligand

The target ligand is composed of an aromatic spacer unit (naphthalene) separating two pyridine-pyrazole (PyPz) units (in the 1 and 8 positions).

Synthesis of the protected PyPz unit, 7
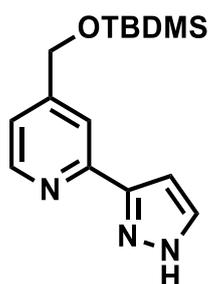
Figure 2.4: Protected PyPz unit, **7**

The target ligand will be water soluble and will contain CH$_2$OH groups attached to the pyridine group on the PyPz unit in the 4 position. The OH groups attached to the pyridine group need to be protected until the synthesis of the full ligand is completed, and so the protecting group ^{t}butyldimethylsilyl (TBDMS) was used.

The synthesis of this protected PyPz unit followed a route (**figure 2.5**) which included the protection of the OH group (**1** → **2**), formation of an N-oxide (**2** → **3**), followed by the addition of a nitrile group in the 2 position (**3** → **4**) and then a Grignard reaction with MeMgBr to form an acetyl group (**4** → **5**). From here the acetyl group can be readily converted to a pyrazole group using the reagent N,N-DMF-DMA (**5** → **6**), followed by the addition of hydrazine monohydrate (**6** → **7**). The total formation of the PyPz unit follows published synthetic procedures.[2]
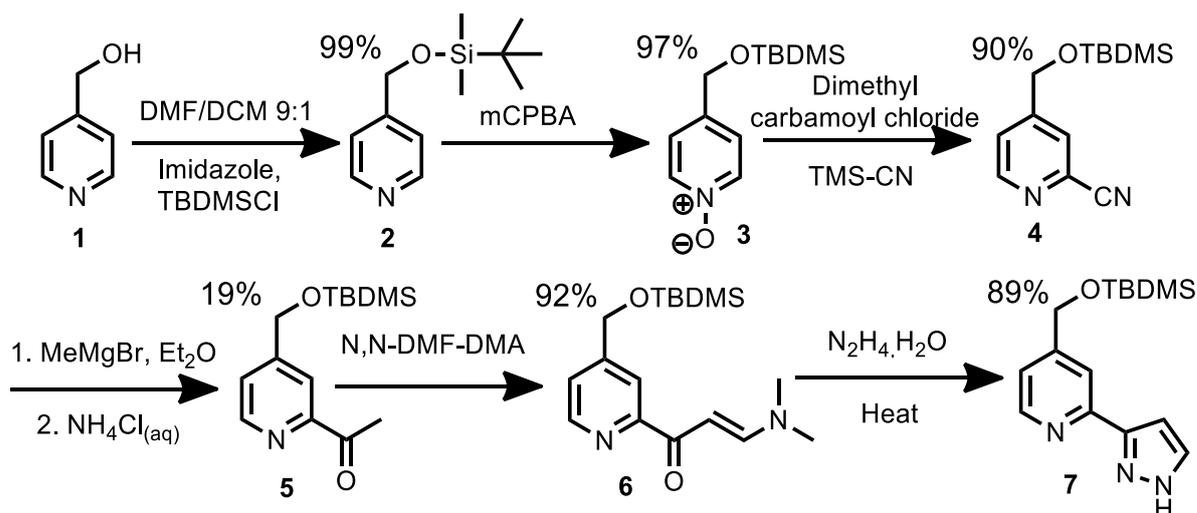
Figure 2.5: Overview of synthetic steps to form protected PyPz, **7**

Synthesis of the 1,8-bis(bromomethyl)naphthalene spacer unit, 9
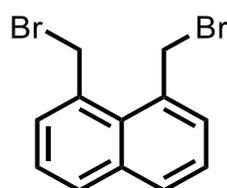
Figure 2.6: 1,8-Bis(bromomethyl)naphthalene spacer unit, **9**

The target ligand consists of an aromatic spacer unit based on naphthalene with the two PyPz units attached to it in the 1 and 8 positions via a methylene bridge. To allow this, 1,8-bis(bromomethyl)naphthalene, **9**, was first synthesised. This was done by mono brominating each of the methyl groups of 1,8-dimethylnaphthalene via a radical reaction using N-bromosuccinimide (NBS), AIBN, CCl$_4$, tungsten lamp and under reflux (**figure 2.7**). A reaction time of 1 hour was optimum with any more allowing over-brominated side products to dominate. The product was purified by crystallisation from hot toluene.
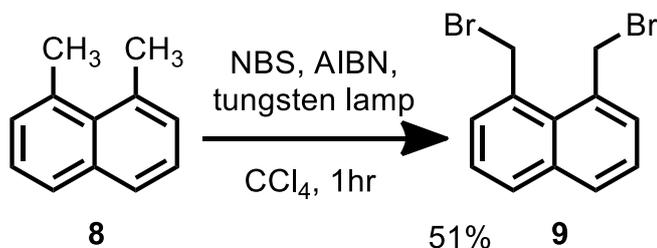
Figure 2.7: Synthesis of 1,8-bis(bromomethyl)naphthalene, **9**

Synthesis of the L$^{1,8\text{-naphOH}}$ ligand
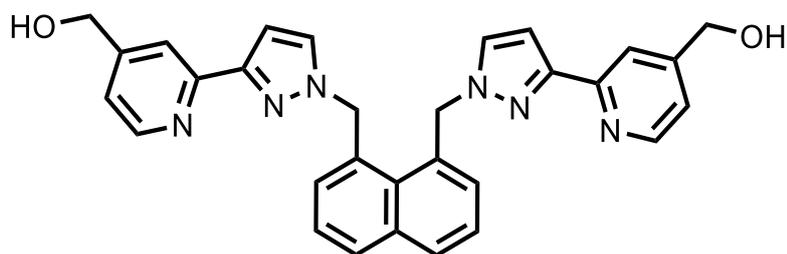
Figure 2.8: L$^{1,8\text{-naphOH}}$ ligand

The synthesis of the target ligand was done by an S$_N$2 reaction with the protected PyPz unit, **7**, and the 1,8-bis(bromomethyl)naphthalene unit, **9** in a 2:1 ratio under reflux using NaH in THF as a base to remove the pyrazole proton. The protected product **10** was purified via column chromatography yielding the protected ligand.

The deprotection (removal of the TBDMS groups) was done at room temperature using tetrabutylammonium fluoride (TBAF) and column chromatography was used to purify the ligand with a 96 % yield (**figure 2.9**).
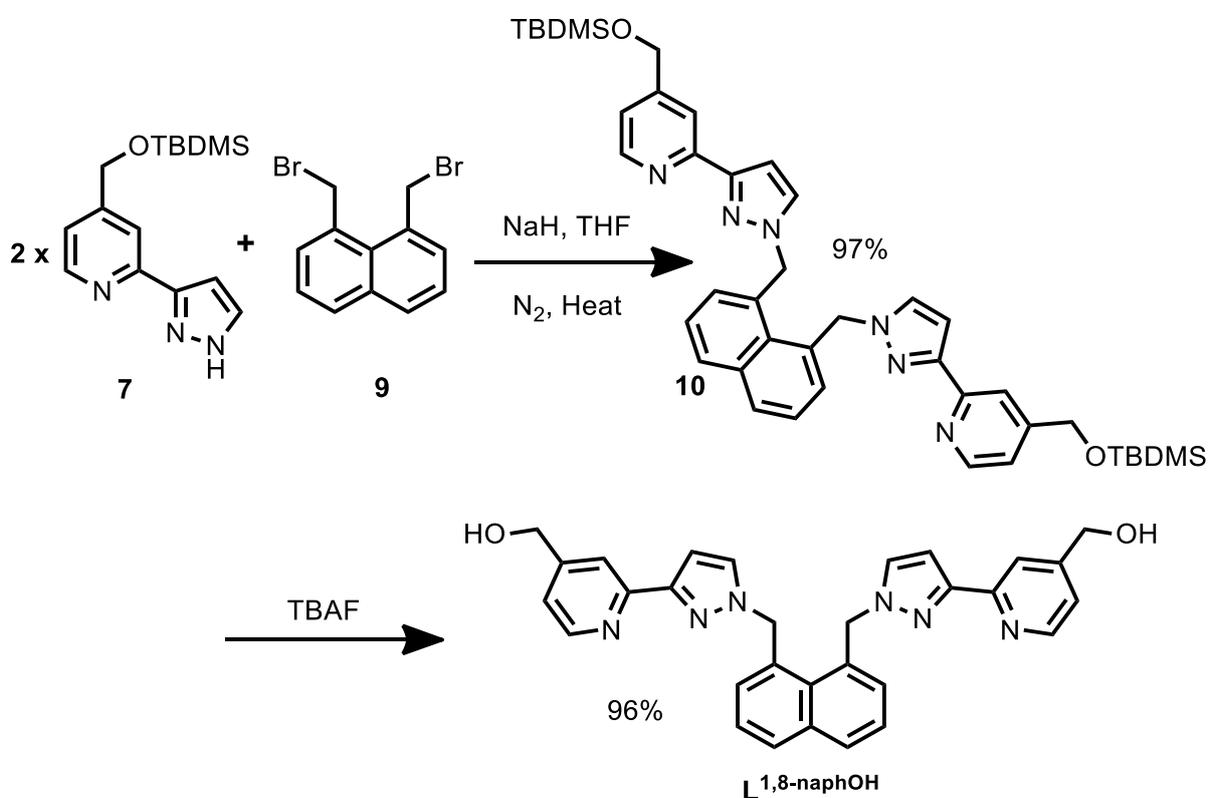
Figure 2.9. Synthesis of L$^{1,8\text{-naphOH}}$

Successful preparation of the ligand was confirmed by mass spectrometry and NMR spectroscopy. The ^{1}H NMR spectrum (**figure 2.10**) shows that the CH_2O-$\underline{H}$ proton couples with the methylene $C\underline{H_2}$ protons to form a triplet (O-H); likewise the CH_2 couples to the hydroxyl proton to give a doublet. This is proved by addition of a drop of D_2O, upon which the O-H signal disappears, and the CH_2 protons become a singlet peak.
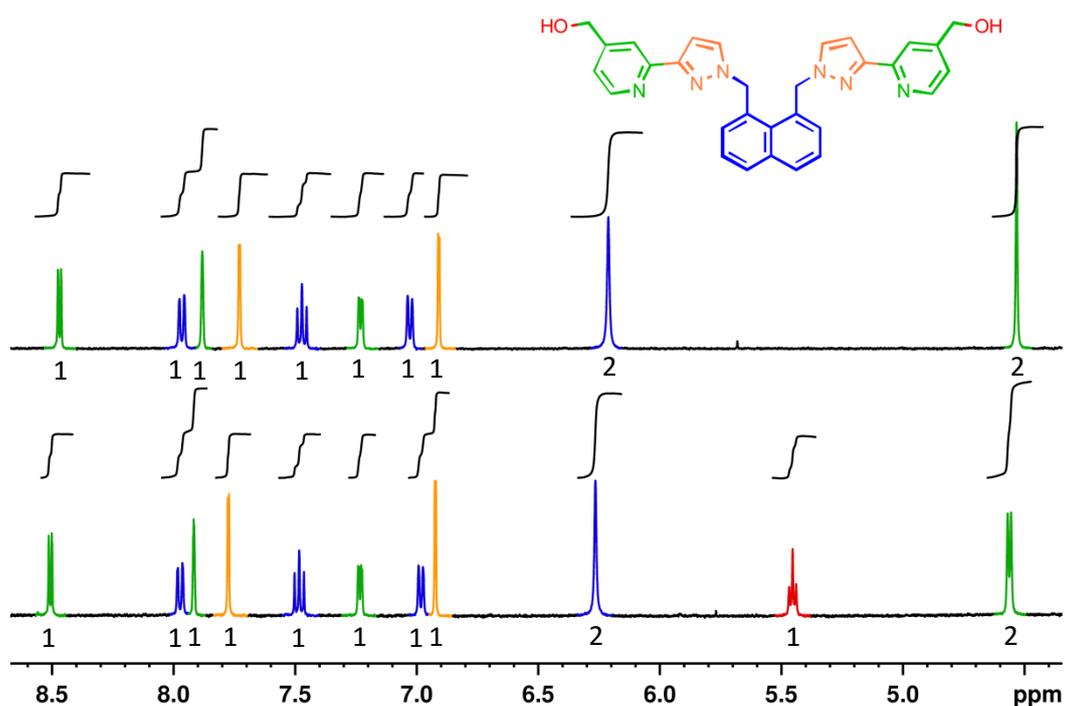
Figure 2.10: ^{1}H NMR spectrum of $L^{1,8\text{-napOH}}$ (in DMSO-d_6) showing the OH proton coupling to the CH_2 protons (bottom) and after addition of D_2O (top)

An X-ray crystal structure was also obtained (**figure 2.11**) by the slow evaporation of a failed cage formation attempt in chloroform. One of the pyridine nitrogen atoms is protonated in the crystal structure. This is most likely due to the slightly acidic conditions of chloroform (the solvent the crystals were grown from) and the presence of a BF_4^- counter ion balances this charge to give a formula in the crystal of $[H(L^{1,8napOH})](BF_4)$.
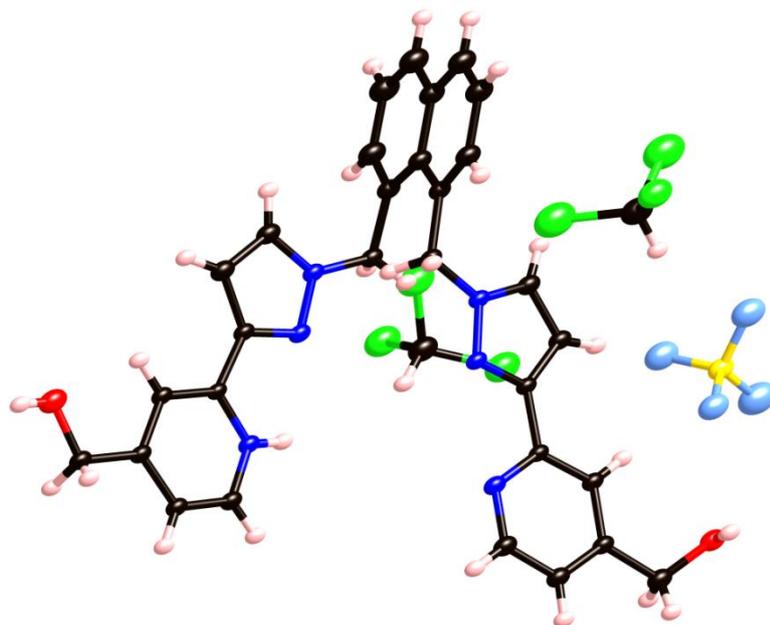
Figure 2.11: X-ray crystallographic structure of $L^{1,8\text{-}napOH}H.BF_4(CHCl_3)_2$ (thermal ellipsoids shown at the 50% level)

The packing observed in the crystal structure reveals π-stacks running through the unit cell (**figure 2.12**).
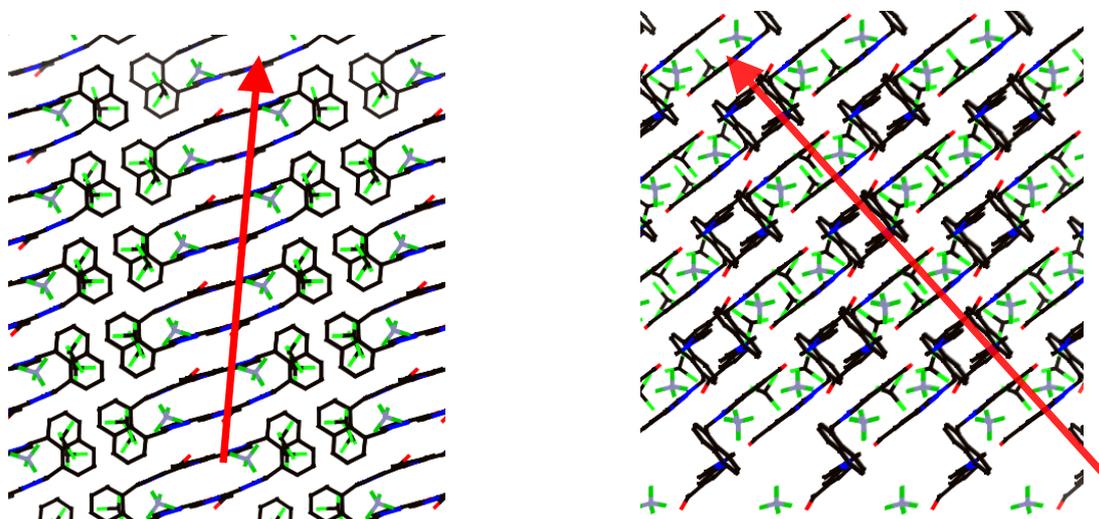
Figure 2.12: X-ray crystal structure $L^{1,8\text{-}naphOH}.HBF_4$ showing the aromatic stacking present in the unit cell

2.2.2 Assembly of coordination cages using $L^{1,8-naphOH}$

The assembly of the cages was done using solvothermal synthesis in a Teflon-lined autoclave. The ligand and metal (Co(II) or Cd(II) as either the BF_4^- or ClO_4^- salt) were mixed in a 3L:2M ratio to fully satisfy all the coordination sites of both the ligand and the metal. Along with some solvent (methanol, acetonitrile, nitromethane, water), the sealed autoclave was heated to 100 °C for 12 hours and then cooled slowly (0.1 °C min^{-1}) to try and help promote crystal formation. Unfortunately crystals did not form, however the orange solution (with Co(II)) in each case was evaporated to dryness and the solid residue was washed with DCM, chloroform and diethyl ether, to remove any unreacted starting materials.

NMR and mass spectrometry studies indicated that, unexpectedly, a mixture of three products had formed ($M_{12}L_{18}$, M_4L_6, M_2L_3) instead of the hoped-for single self-assembled structure of the $M_{12}L_{18}$ cage.

2.2.3 Separation and characterisation of the assemblies

Structure 1 – a Co$_2$L$_3$ dinucelar triple mesocate

Various chromatography-based separation attempts were made but crystallisation of the crude mixture in nitromethane was performed by vapour diffusion of diethyl-ether vapour into the MeNO$_2$ solution, to yield X-ray quality crystals of one of the components. This was the dinuclear triple mesocate [Co$_2$(L^{18napOH})$_3$](BF$_4$)$_4$ (**figure 2.13**), in which all three ligands span both metal centres.
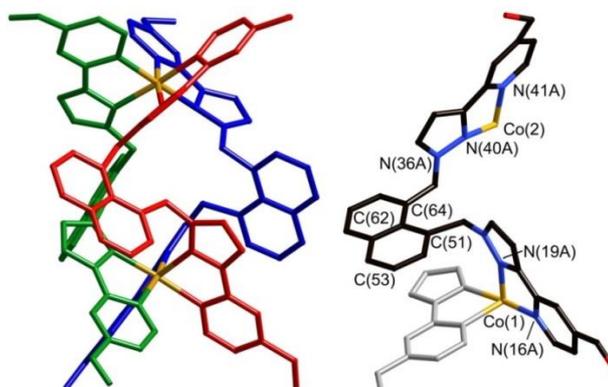
Figure 2.13: X-ray crystallographic structure of the [Co$_2$(L^{18napOH})$_3$](BF$_4$)$_4$ dinuclear triple mesocate

Superficially the structure looks like a triple helicate but in fact the two metal centres within each molecule have opposed chirality so this is an example of a 'meso-helicate' or mesocate, lacking helical chirality. The conformation of the three ligands is clearly not the continuous spiral strand as seen in helicates; instead there is a sharp bend in each ligand allowing one pyrazolyl-pyridine terminus to be angled in the opposite sense to the other. This allows the naphthyl group of each ligand to form a π-π stacking interaction (separation 3.3 – 3.4 Å) with the coordinated pyrazolyl-pyridine group from another ligand – a typical interaction between parallel and overlapping electron rich and electron deficient aromatic fragments. All three such interactions are at the same end of the complex, with the three pyrazolyl-pyridine units around Co(1) all forming stacking interactions with adjacent naphthyl groups; this cannot happen around the other terminus Co(2) without a substantial change in the ligand conformations. Thus the stacking interactions appear to 'lock' the ligands in an asymmetric conformation with two inequivalent termini.

This leads to one unique ligand environment with 24 inequivalent protons present in the NMR spectrum (**figure 2.14**), since the two Co(II) ions are different, and there is a C_3 axis down the centre of the complex making all ligands equivalent.
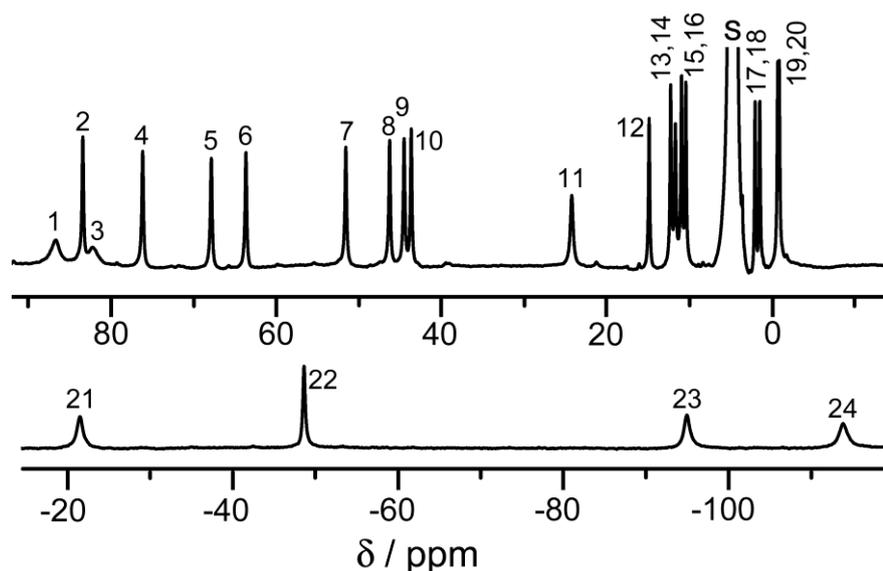
Figure 2.14: ^{1}H NMR spectrum in D_2O at 298 K of $[Co_2(L^{18naphOH})_3](ClO_4)_4$ showing a single ligand environment with no internal symmetry

The paramagnetism of the high-spin co(II) ions disperses the signals over a wide range ($-100 \rightarrow +100$ ppm) making it easy to see the 24 separate signals. The absence of coupling information makes them difficult to assign individually, but the broader signals are from H atoms closest to the Co(II) centres.[4]

The mass spectrum also proved the formation of the M_2L_3 mesocate showing a series of peaks corresponding to $[Co_2(L^{18napOH})_3(ClO_4)_{4-n}]^{n+}$ from successive loss of ClO_4^- anions. High-resolution mass spectrum (**figure 2.15**) shows the isotope pattern for the fragment $[Co_2(L^{18naphOH})_3(ClO_4)_2]^{2+}$.
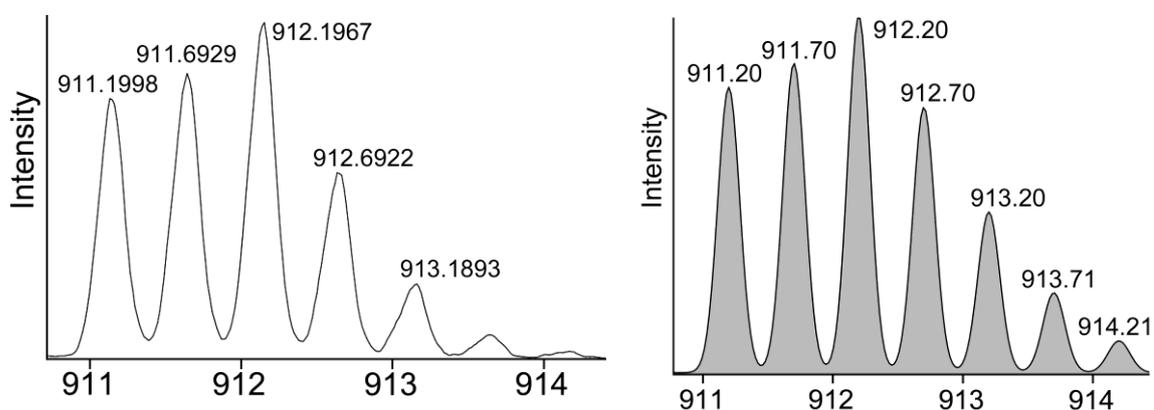
Figure 2.15: Isotope pattern measured (left) and calculated (right) as seen in mass spectrum for the fragment $[Co_2(L^{18naphOH})_3](ClO_4)_2^{2+}$ using high-res ESMS

In the solid-state the cylindrical complex forms columns of packed species which can be seen clearly in **figure 2.16**.
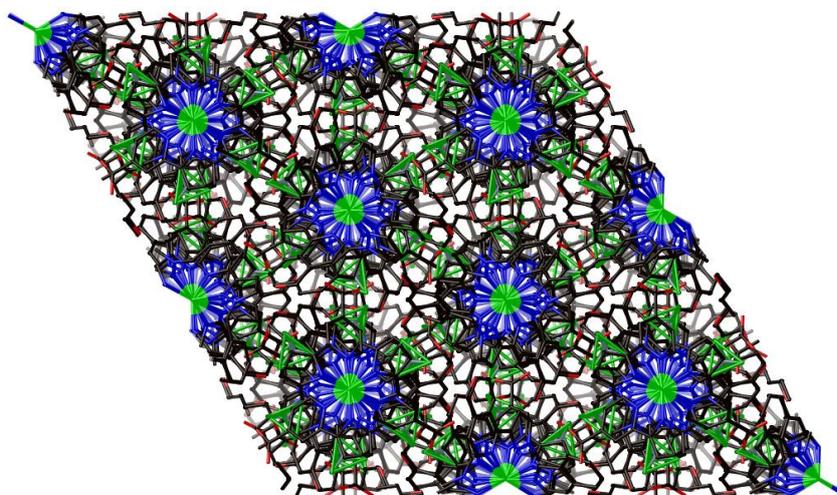
Figure 2.16: Packing diagram for $[Co_2(L^{18naphOH})_3](ClO_4)_4$ showing the unit cell looking from above

There are numerous H-bond interactions between the CH_2OH groups on adjacent molecules and some between a naphthalene or a pyridine C-H and a CH_2OH group (**figure 2.17**).
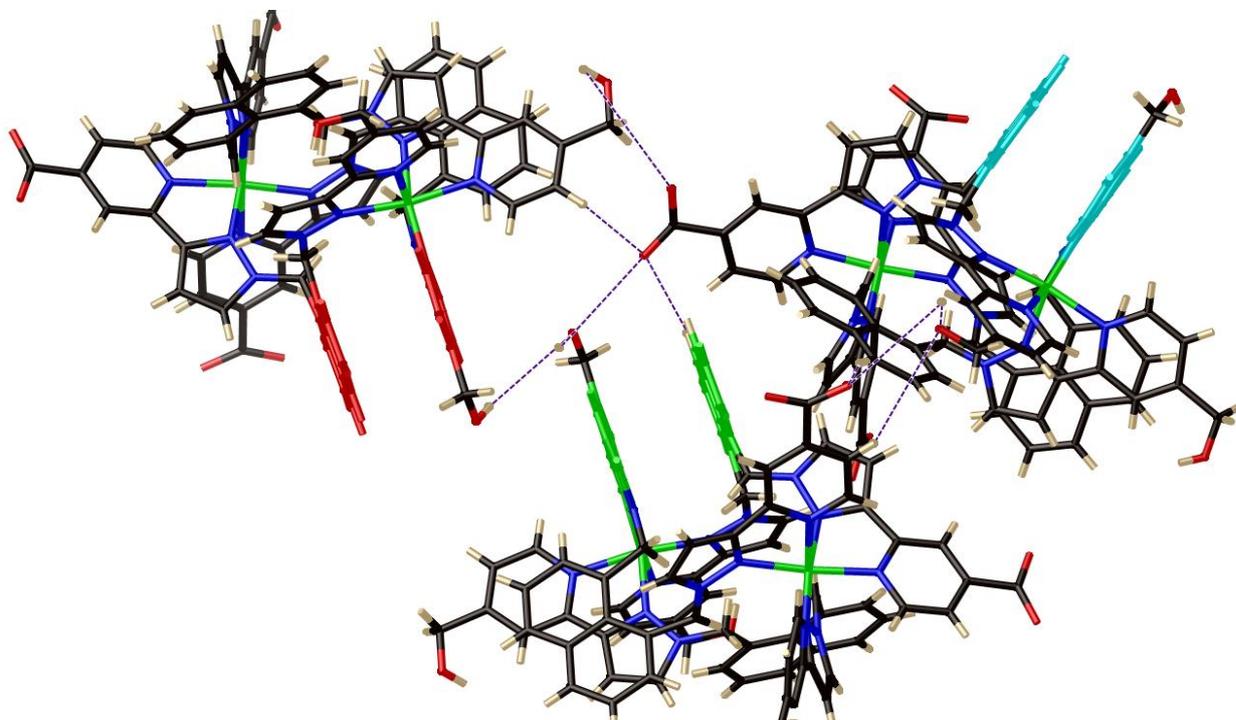
Figure 2.17. X-ray crystalstructure of the $[Co_2(L^{18naphOH})_3](ClO_4)_4$ species showing the pi-stacking (red, green, light blue) and H-bonding contacts (Purple dotted lines) between molecules

Structure 2 – a Co_4L_6 tetrahedral cage

An X-ray quality crystal of the second structure was obtained by slow cooling of a solution of the crude mixture of complexes in D_2O. It is composed of 4 Co(II) centres, all with a fac tris-chelate geometry, arranged in a tetrahedron with a bridging ligand along every edge to give a typical M_4L_6 tetrahedral cage assembly (**figure 2.18**).
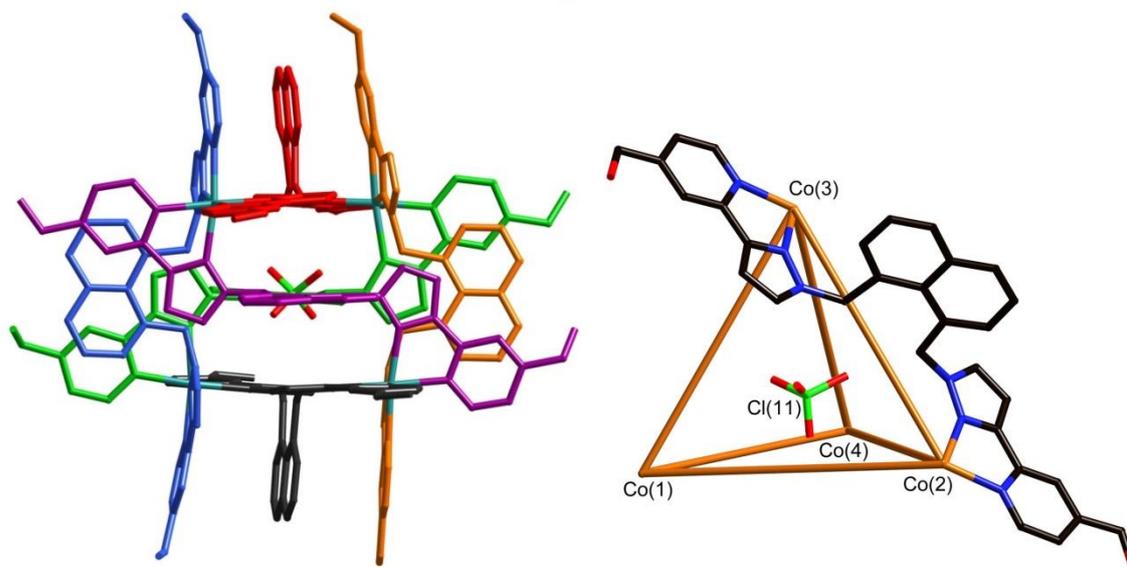
Figure 2.18: X-ray crystal structure of $[Co_4(L^{18naphOH})_6](ClO_4)_8$ showing the whole structure (left) and a skeletal representation with 1 ligand (right)

The naphthyl group of each bridging ligand forms π-stacking interactions with the coordinated pyrazolyl-pyridine termini of two other ligands, forming a three-component A/D/A sandwich (A = electron-deficient pyrazolyl-pyridine acceptor unit, D = electron-rich naphthyl donor unit) along every edge of the tetrahedron. As is usually the case in these tetrahedral cages, an anion occupies the central cavity. This guest anion is inverted with respect to the cage tetrahedron such that each O atom of the perchlorate guest is oriented towards the space in the centre of one of the triangular faces of the Co_4 tetrahedral array. The anion is involved in CH•••O hydrogen-bonding interactions with the interior surface of the cage.

Due to the high symmetry present in the tetrahedron (T symmetry, with 4 C_3 axes through the vertices of the cage and 6 C_2 axes which run through the centres of the ligands of the cage), every ligand has two-fold symmetry such that half a ligand environment (12 independent protons) is observed in the ^{1}NMR spectrum (**figure 2.19**). Again individual signals are not assigned but the higher symmetry compared to the previous example (**figure 2.14**) is obvious.
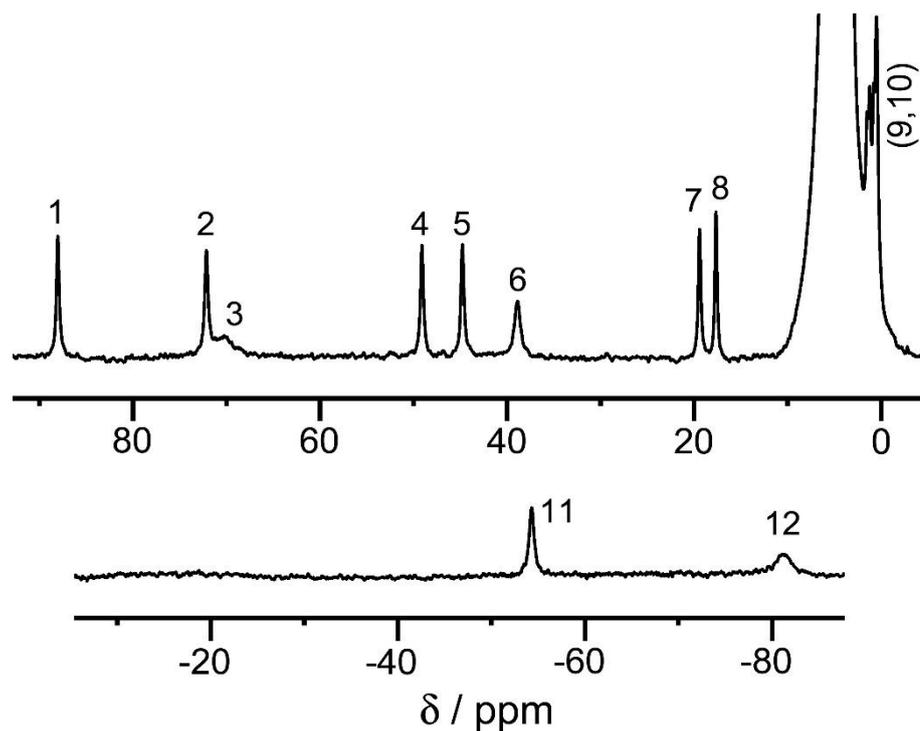
Figure 2.19: ^{1}H NMR spectrum in D_2O at 298 K of $[Co_4(L^{18naphOH})_6](ClO_4)_8$ showing ½ ligand environment

The mass spectrum also proved the formation of the M_4L_6 tetrahedron, with a series of peaks corresponding to $[Co_4(L^{18napOH})_6(ClO_4)_{8-n}]^{n+}$ from successive loss of ClO_4^- anions. High-resolution mass spectrometry (**figure 2.20**) shows the isotope pattern for the fragment $[Co_4(L^{18naphOH})_6(BF_4)_5]^{3+}$.
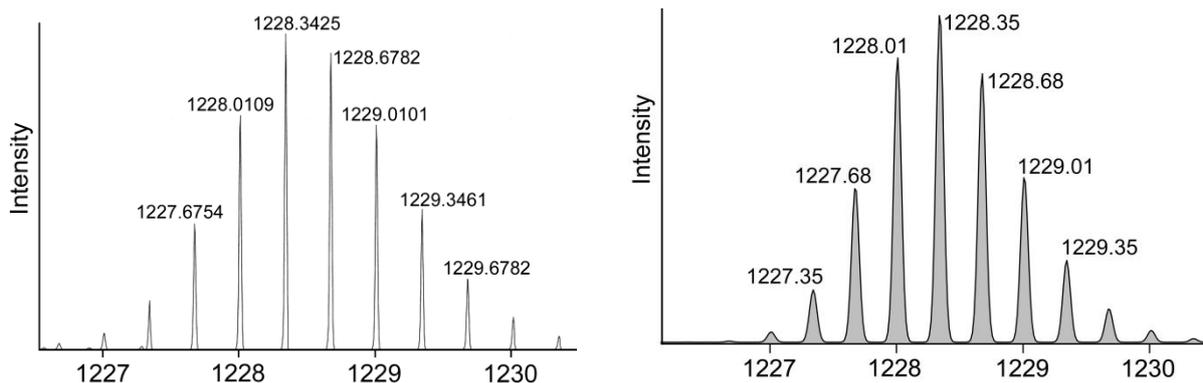
Figure 2.20: Isotope pattern measured (left) and calculated (right) as seen in mass spectrum for the fragment $[Co_4(L^{18naphOH})_6(BF_4)_5]^{3+}$

Also, as with the **Co₂** species, the principal interaction responsible for packing in the unit cell is hydrogen bonding between the methylene alcohol groups on adjacent molecules, and others between a naphthalene or a pyridine C-H and a methylene alcohol group (**figure 2.21**).
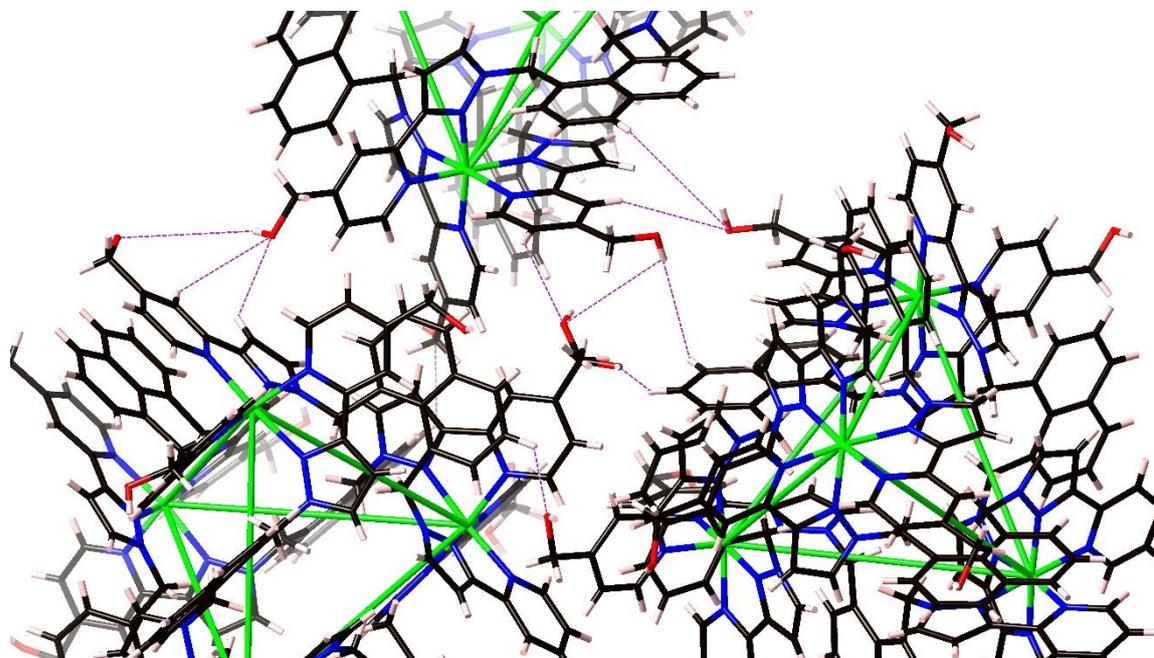
Figure 2.21: X-ray crystal structure of $[Co_4(L^{18naphOH})_6](ClO_4)_8$ showing the H-bonding present within the molecule; anions are removed for clarity

Structure 3 – a Co₁₂L₁₈ truncated tetrahedral cage

It was possible to separate out the third self-assembled component of the crude mixture from reaction of Co(II) salts with $L^{18nahOH}$ using size exclusion chromatography (SEPHADEX G50, eluted with water). The column developed as a diffuse orange band (a mixture of all three species) and a thick, well defined band (the pure **Co₁₂** cage). However, X-ray quality crystals of the **Co₁₂** cage were grown from slow cooling of a solution of the crude mixture in D_2O. The structure indicated it is a $M_{12}L_{18}$ truncated tetrahedron (**figure 2.22**), which is isostructural to the cage formed with the parent unsubstituted ligand ($L^{1,8naph}$), and was the initial intended structure. This cage is different from the other two structures in that all of the metal centres have a meridional tris-chelate geometry.
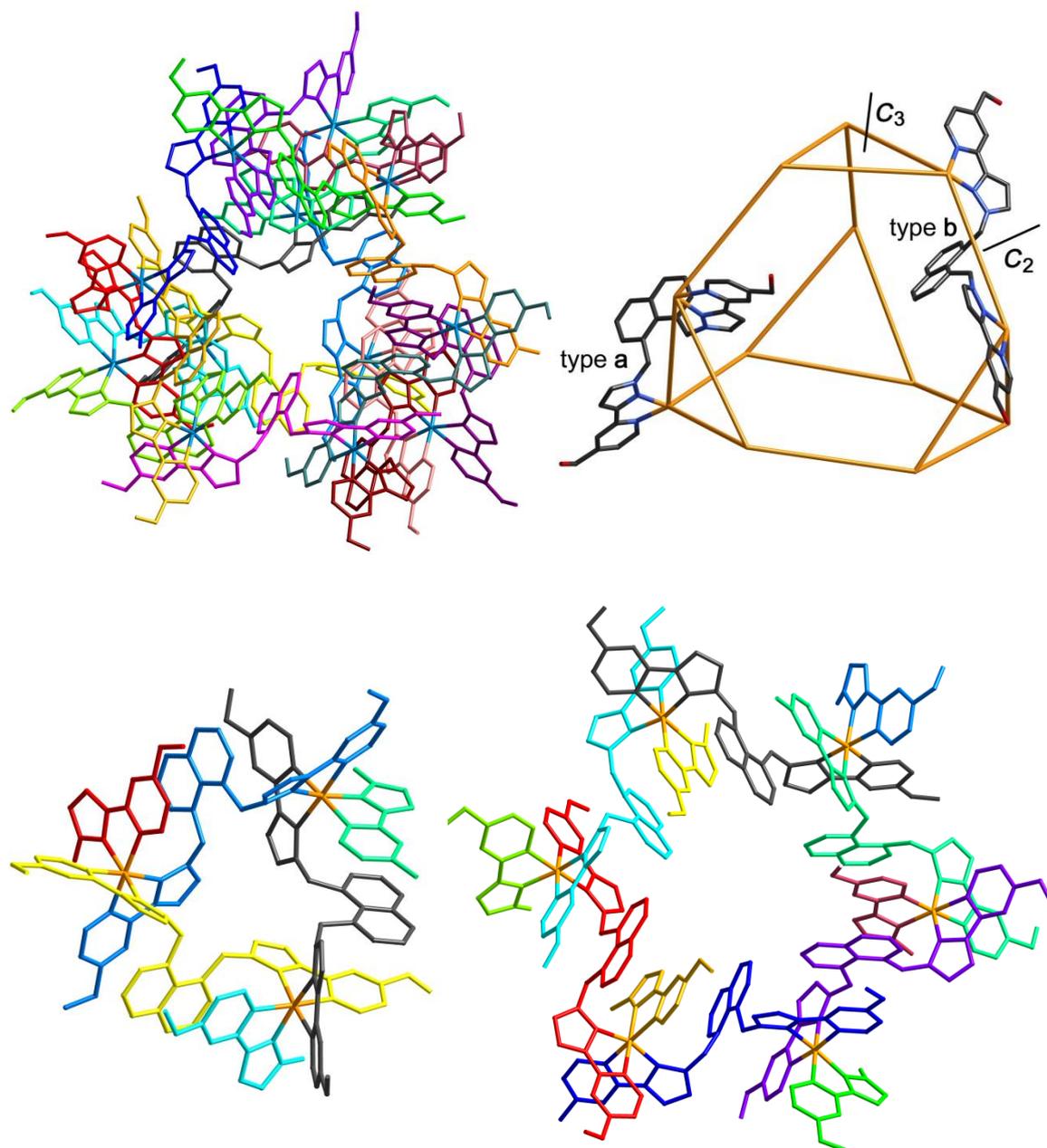
Figure 2.22: X-ray crystal structure of the $[Co_{12}(L^{1,8\text{-naphOH}})_{18}](BF_4)_{24}$ truncated tetrahedral cage showing (a) the whole structure (left) and a skeletal representation with 2 ligands showing (right); and (b) The triangular face (left) and the hexagonal face (right); all counter ions are removed for clarity

A truncated tetrahedron is an Archimedean solid with all vertices equivalent, but two types of face – triangular and hexagonal – and two types of edge. The two types of edge may be described as type 'a', which are the 12 edges associated with the four triangular faces, and type 'b', which are the six edges connecting these triangles – these are the edges of the parent tetrahedron before it was truncated. The ligands spanning these

edges may therefore be abbreviated as L^a, of which there are 12 (with no internal symmetry), and L^b, of which there are 6 (all lying on a twofold axis).

Each $M_3(L^a)_3$ triangular face is a cyclic helicate, with four of these linked in a tetrahedral array by additional bridging ligands L^b. All metal centres in this structure are meridional tris-chelates – in contrast to the first two structures – and all metal centres are homochiral. The arrangement of ligands, and in particular the flexibility associated with the methylene groups that link the pyrazolyl-pyridine termini to the central aromatic core, permits extensive aromatic stacking (**figure 2.23**) between ligands with six 7-membered A-D-A-D-A-D-A stacks around the periphery of the complex.
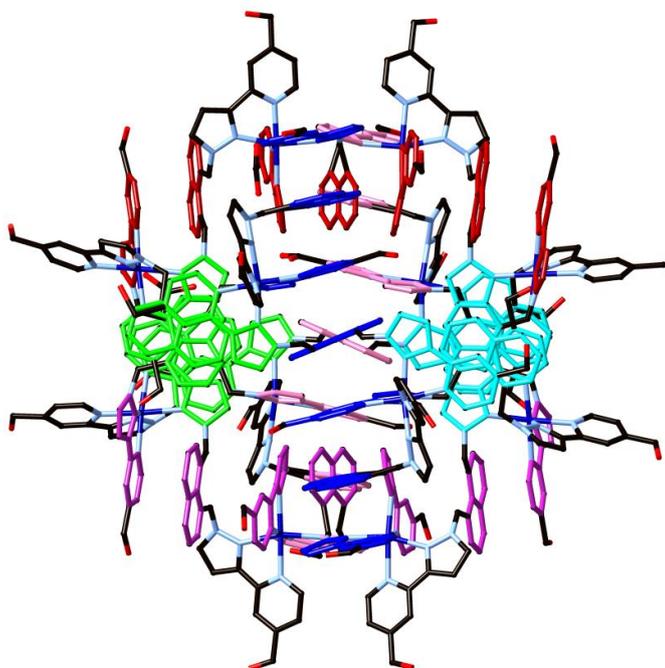
Figure 2.23: X-ray crystal structure of $[Co_{12}(L^{18naphOH})_{18}](BF_4)_{24}$ showing the π-stacking within the molecule, each colour represents one stack.

The spaces in the centre of each triangular and hexagonal face provide pockets which each accommodate a tetrafluoroborate anion (**figure 2.24**) that forms CH•••F interactions with the surrounding ligand; thus, eight anions are associated with the surface of the cage. The rest are sat around the cage exterior however many could not be located as they were severely disordered and could not be modelled successfully in the crystallographic refinement.
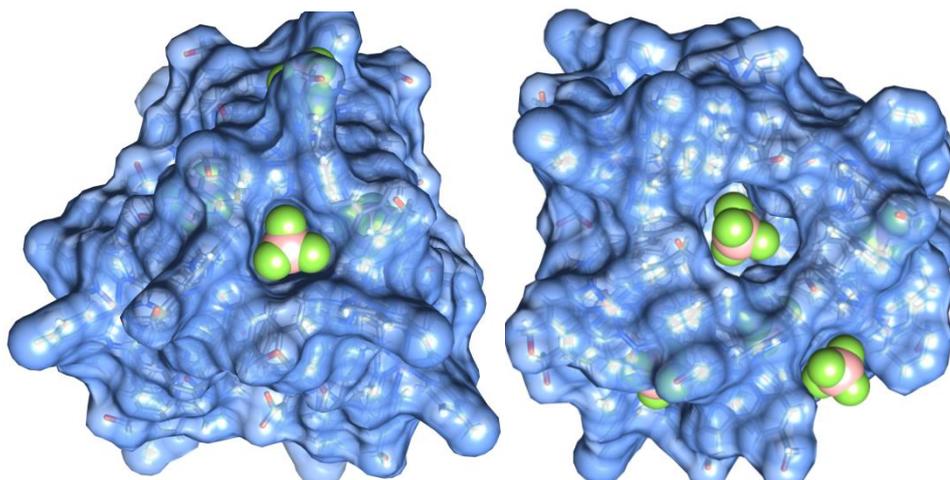
Figure 2.24: BF_4^- counter ions nestled in one of the triangular faces (left) and one of the hexagonal faces (right) of $[Co_{12}(L^{18naphOH})_{18}](BF_4)_{24}$

There is a C_3 axis through the centre of each $M_3(L^a)_3$ triangular face (and also through the opposite hexagonal face), as well as three C_2 axes, each of which bisects an opposite pair of L^b ligands along the type 'b' edges. The result of this is that there must be 1.5 magnetically independent ligand environments.

In the twelve L^a ligands, all protons are inequivalent due to the helical chirality of the $M_3(L^a)_3$ triangular array which means that the ligands have distinct 'head' and 'tail' ends. The six L^b ligands are all bisected by C_2 axes, generating 12 equivalent halves of the L^b-type ligands. The result is 36 magnetically inequivalent protons with the same abundance (excluding exchangeable OH protons), and the ^1H NMR spectrum is consistent with this (**figure 2.25**).
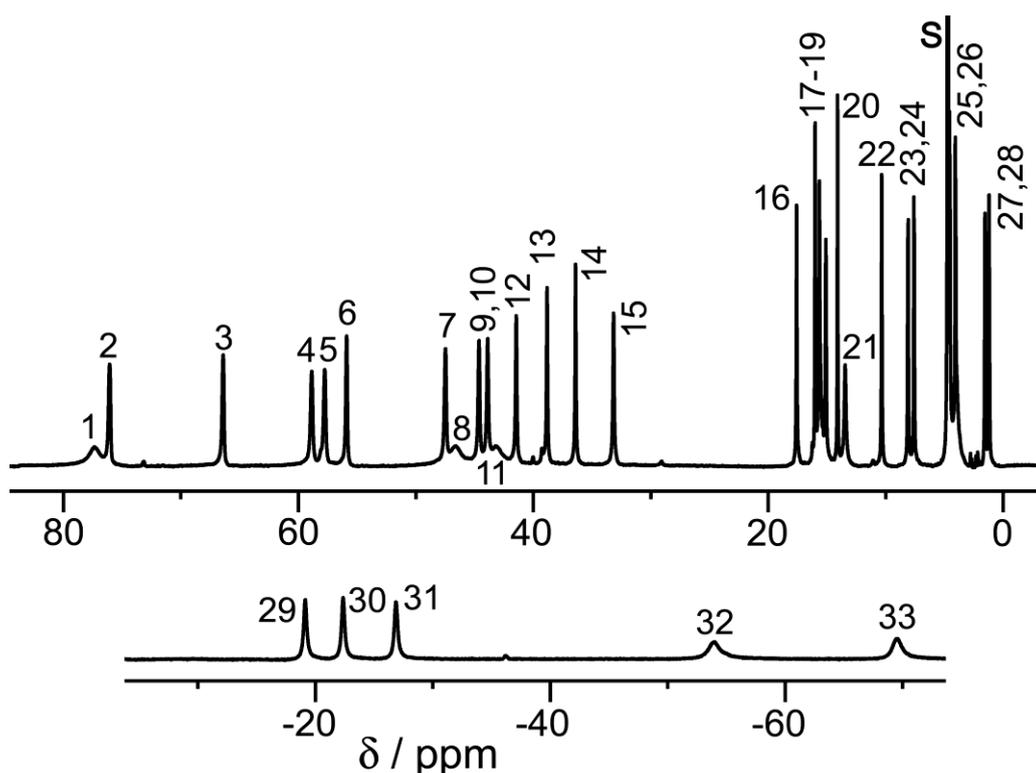
Figure 2.25: ^{1}H NMR spectrum in D_2O at 90 °C of $[Co_{12}(L^{1,8\text{-naphOH}})_{18}](BF_4)_{24}$ showing 1.5 ligand environments

We can easily identify 33 of the expected 36 signals; the missing ones may be obscured under the HOD peak (there are many closely-spaced signals in this region) or may still be too broad to detect. However this spectrum is clearly in agreement with the symmetry of the solid-state structure, and in particular we can see how some signals occur in sets of three corresponding to the three independent ligand halves [*e.g.* signals 1, 8 and 11 in **figure 2.25** are the three pyridyl H^6 environments, and signals 29 – 31 arise from one of the protons on each of the three independent methylene groups].

The mass spectrum also proved the formation of the $M_{12}L_{18}$ species with a series of peaks corresponding to $[Co_{12}(L^{18napOH})_{18}(BF_4)_{8-n}]^{n+}$ from successive loss of BF_4^- anions. High-res MS (**figure 2.26**) shows the isotope pattern for the fragment $[Co_4(L^{18naphOH})_6(BF_4)_{16}]^{8+}$.
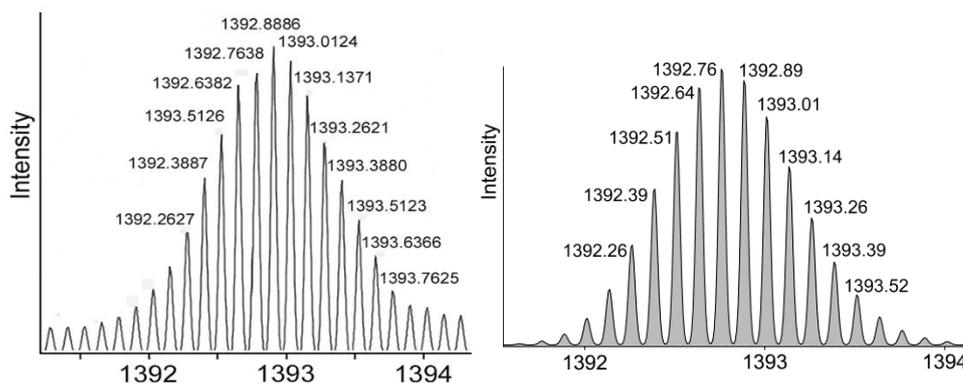
Figure 2.26: Isotope pattern measured (left) and calculated (right) as seen in mass spectrum for the fragment $[Co_{12}(L^{18naphOH})_{18}(BF_4)_{16}]^{8+}$

2.2.4 Interconversion between the assemblies in aqueous solution

Since multiple attempts at changing the experimental conditions for the synthesis results in a similar mixture of components in every case, the possibility suggests itself that the different species isolated as crystals could be in slow equilibrium in solution. The fact that we could obtain clean ^{1}H NMR spectra of each species independently, using redissolved crystals of each component, means that any equilibrium must be on a timescale of hours or longer at room temperature, which made separation and individual identification of the components possible (*e.g.* we could isolate pure **Co$_{12}$** by size-exclusion column chromatography). ^{1}H NMR spectroscopy provides a convenient tool to study any equilibration between components, in particular because (i) the paramagnetism disperses the signals over such a wide chemical shift range, such that most individual signals are clearly resolved; and (ii) the different symmetries results in different numbers of independent signals for each complex (12, 24 or 36) which allows each set of signals to be identified easily. Accordingly we investigated the equilibria between the three complexes in aquoues solution using ^{1}H NMR spectroscopy.

Effects of temperature

The first experiment involved dissolving some of the crude mixture of the **Co$_2$**, **Co$_4$** and **Co$_{12}$** complexes (6.0 mg in 0.6 ml D$_2$O) and its ^{1}H NMR spectrum was measured. The

equilibrium at this point contained mainly the **Co$_{12}$** and **Co$_4$** species (bottom spectrum in

figure 2.27)
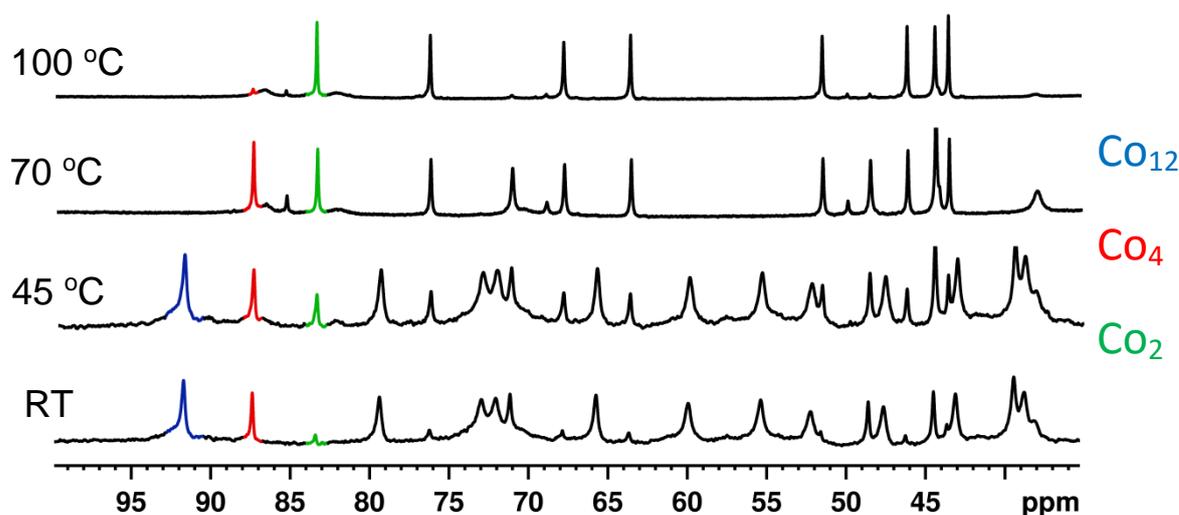
Figure 2.27: ^{1}H NMR spectra measured at 298 K, after equilibration of the sample at
25 °C, 45 °C, 70 °C, 100 °C, with 6 mg in 0.6 ml D$_2$O

The temperature was increased (25 → 45 °C) and the ^{1}H NMR spectrum was again

obtained. Some of the peaks in the NMR spectrum seemed to have changed in intensity

so it was further monitored at this temperature until a new equilibrium (4 days) had been

reached. A slight increase in the **Co$_2$** species was apparent. From there, the temperature

was increased again (70 °C) and again monitored by NMR until the new equilibrium had

been reached (1 day). At this point a large shift in equilibrium had occurred with the major

species being **Co$_4$** and **Co$_2$** with no **Co$_{12}$** being present. Finally the temperature was raised

once more (100 °C) and monitored by NMR until a new equilibrium had been reached (12

hours). At this point there remained no **Co$_{12}$** species and now hardly any M$_4$ species; the

M$_2$ species now dominates (**figure 2.27**).

Reversibility

The second experiment was to dissolve crystals of the pure **Co$_2$** complex in D$_2$O at the

same concentration as in the first experiment and follow the equilibration process at 25

°C. A ^{1}H NMR spectrum recorded immediately looked like pure **Co$_2$**, however over time,

the intensity of the **Co$_2$** peaks diminished and the increase of **Co$_4$** and **Co$_{12}$** peaks occurred

until equilibrium was reached (36 hours). The composition was very similar to the equilibrium mixture of the first experiment (mainly **Co₄** and **Co₁₂** with almost no **Co₂**) (**figure 2.28**).
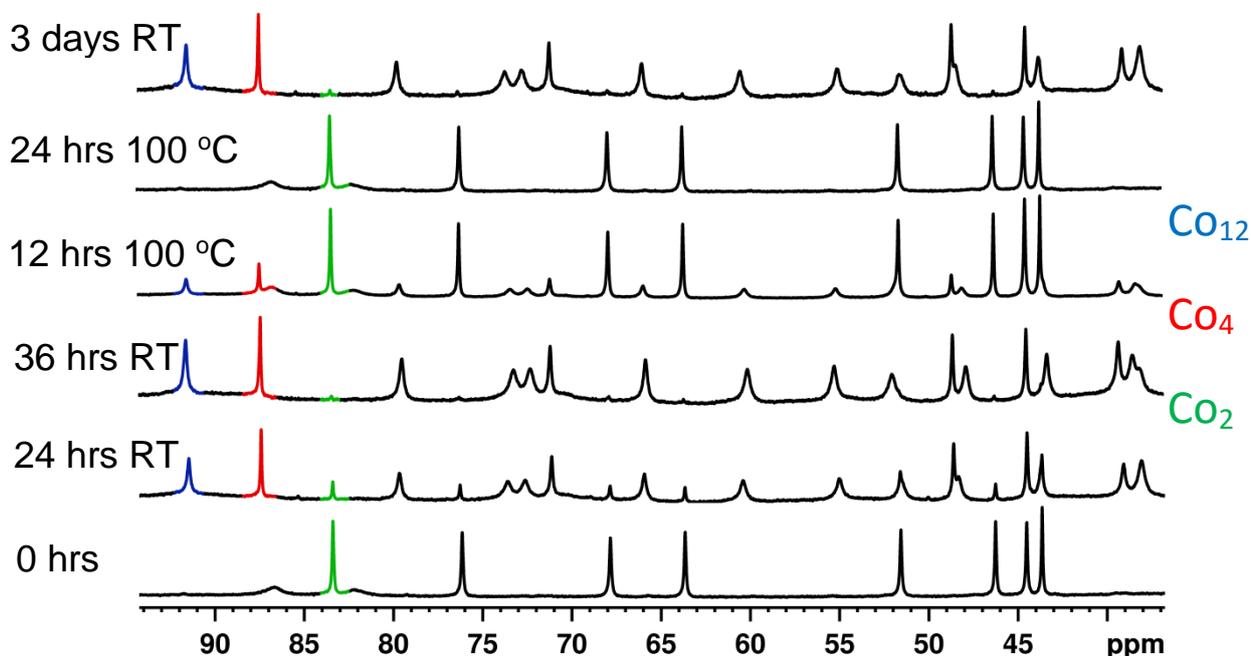
Figure 2.28: ^{1}H NMR spectra measured at 298 K, of a solution (6 mg **Co₂** in 0.6 ml D₂O) at 25 °C for 0, 24 and 36 hrs, then 100 °C for 12 and 24 hrs and back to 25 °C for 3 days

The reversibility of this interconversion was also investigated by taking the same sample and heating it to 100 °C. The equilibrium fully reverted back to its starting point (all M₂) and the temperature-change cycle was then repeated, each time fully reverting back to all **Co₂** at the high temperature limit (**figure 2.28**).

Similar experiments using redissolved crystals of pure **Co₄** or **Co₁₂** [again, with the same total concentration of Co(II) ions] gave the same results: the initially-obtained spectrum of pure complex in each case evolved slowly to show the same equilibrium mixture of **Co₂**, **Co₄** and **Co₁₂** appropriate to the temperature. Thus identical behaviour in solution is seen whether started from the as-isolated mixture of complex components or from redissolved crystals of *any one component*, proving that the **Co₂/Co₄/Co₁₂** system exists in aqueous solution as a fully reversible equilibrium. The effect of changing the temperature can be simply explained in terms of entropy. As the temperature increases, the value of T$\Delta S°$ increases, so entropy effects become more important, and the equilibrium shifts towards smaller assemblies.

In addition, this fragmentation disperses the positive charges over more particles, which is electrostatically favourable (a $\Delta H°$ effect). This raises the question of why the larger assemblies form at all: there must be an additional factor specifically favouring formation of **Co₁₂** in preference to six molecules of **Co₂** or three molecules of **Co₄**, given that fragmentation has ΔH and ΔS contributions which are both favourable.

Many of the specific interactions that contribute to $\Delta H°$ for formation of an individual complex scale linearly with complex size so do not provide a driving force for formation of larger assemblies. Thus the total number of metal-ligand bonds is independent of the size of the assembly: one **Co₁₂** complex contains the same number of Co–N bonds (of similar length, according to the crystal structures, and therefore similar strength) as three **Co₄** complexes or six **Co₂** complexes. Similarly, there are more pairwise π-π stacking interactions in larger assemblies, but the crystal structures show 6 such interactions in **Co₂**, 12 in **Co₄** and 36 in **Co₁₂**, so the number of π-π stacking interactions is two per ligand in each case.

The main thing that would favour formation of larger assemblies is a decrease in the surface area (SA) to volume (V) ratio: the larger the assembly, the smaller is the proportion of hydrophobic ligand backbone that is exposed to water at the surface, and the greater is the proportion that is buried in the interior and protected from solvent. Thus, the hydrophobic effect would be expected to favour larger assemblies in which a higher proportion of the hydrophobic ligand surface is shielded from the solvent.

The surface areas of the complexes can be estimated by using the X-ray crystal structures. Using a water molecule as the probe, the solvent-accessible surface areas of the complex cations of **Co₄** and **Co₁₂** are 2076 and 4885 Å^2 respectively. Thus, three complex cations of **Co₄** have an external surface area of *ca.* 6200 Å^2 in contact with the aqueous solvent, and reorganising them into a single **Co₁₂** complex cation reduces the hydrophobic surface area by *ca.* 1300 Å^2, providing a strong driving force for formation of the larger assembly in water. This competition between an increased hydrophobic effect which promotes larger assemblies, with other entropic / electrostatic factors promoting fragmentation into smaller assemblies, qualitatively explains the concentration and

temperature dependence of the equilibrium composition of the **Co₂/Co₄/Co₁₂** mixture in aqueous solution.

This may be the reason the **Co₂** (the smallest of the assemblies) could be crystallised only from nitromethane where the hydrophobic induced aggregation would be absent. Thus the effect of solvent upon this equilibrium was investigated.

Effects of solvent

The solubility of these assemblies is poor in non-aqueous solvents, so a saturated nitromethane solution was prepared, filtered, allowed to equilibrate for several days, and has its ^{1}NMR spectrum taken (**figure 2.29**).
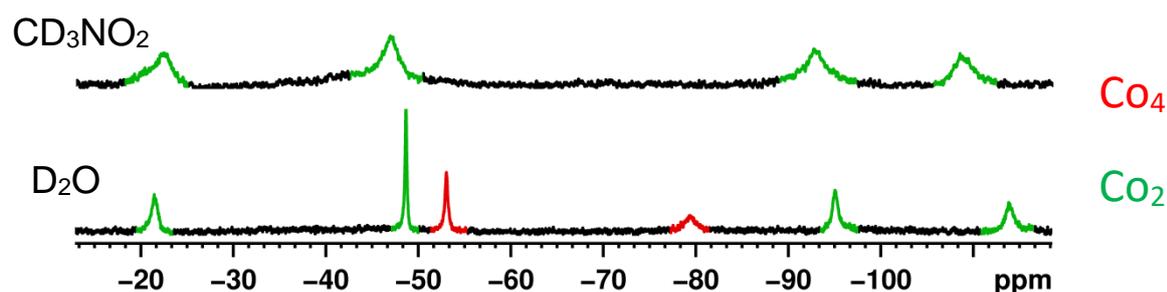
Figure 2.29: ^{1}H NMR spectra at 25˚C of the crude reaction product in CD_3NO_2 (top) and in D_2O (bottom)

Only the presence of **Co₂** can be seen in the nitromethane solution. The solution was evaporated to dryness and then re-dissolved in D_2O (The concentration will be the same in this case), and left to equilibrate for several days. The presence of the **Co₄** assembly indicates that the formation of larger assemblies is favoured in water. There was no presence of the **Co₁₂** in this sample, since the sample was too dilute for it to be present.

Effects of concentration

The concentration of the solution also plays a vital role in determining which assemblies are favoured in the equilibrium. In order to investigate its effect, an experiment analogous to the first temperature-dependence experiment, but at a much

higher concentration (60 mg of the **Co₂/Co₄/Co₁₂** mixture in 0.6 ml D₂O; 10 times higher), was performed (**figure 2.30**).
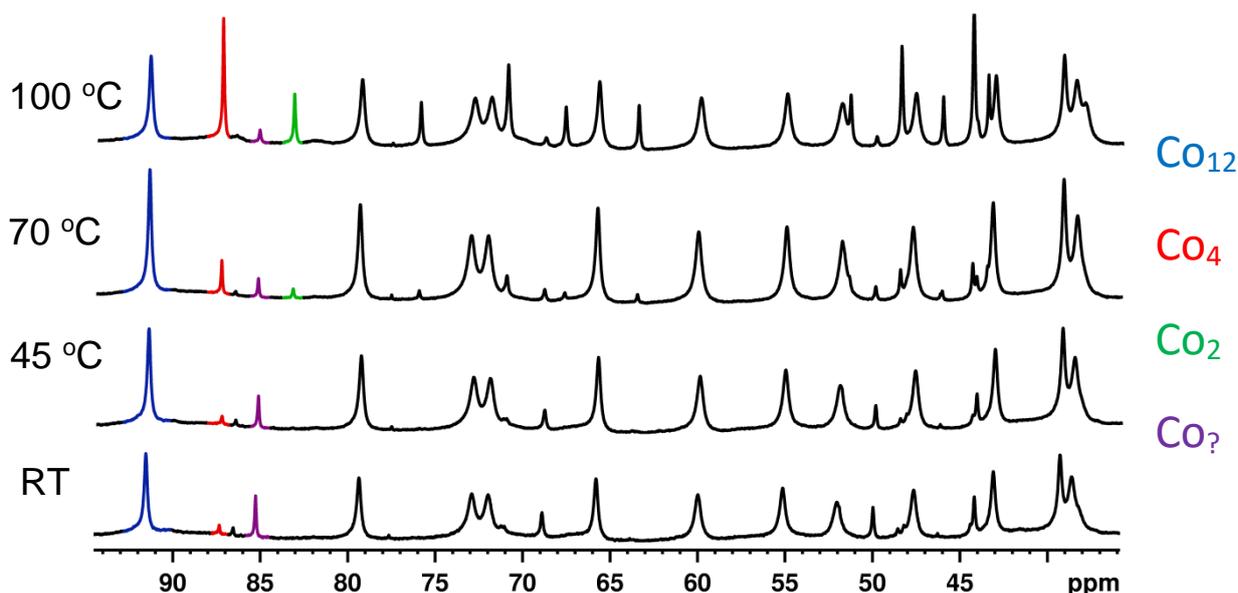
Figure 2.30: ^{1}H NMR spectra measured at 298 K, after equilibration of the sample at 25 °C, 45 °C, 70 °C, 100 °C, with 60 mg in 0.6 ml D₂O

The difference between the speciation behaviour of this and the sample 10 times more dilute (**figure 2.27**) is very apparent. Firstly the **Co₁₂** is now present at all temperatures. The **Co₄** grows in as the temperature increases, followed by the **Co₂**. In the 100 °C sample, all 3 species are in high abundancy, in contrast to the dilute solution, in which the **Co₂** clearly dominates.

Essentially, at higher concentrations, the larger assemblies dominate, whereas at lower concentrations, the smaller assemblies dominate. This can be accounted for simply by applying the Le Chatelier principle. At high concentrations the equilibrium shifts to decrease the number of molecules, generating a smaller number of big assemblies. At low concentrations, the equilibrium shifts to increase the number of molecules, so in this case, generating a larger number of smaller assembles.

Consider the equilibrium between three molecules of **Co₄** and one of **Co₁₂** (**Eq. 1**).

$$3Co_4 \rightleftharpoons Co_{12} \qquad \textbf{(Eq. 1)}$$

$$[Co_{12}] = K \bullet [Co_4]^3 \qquad \textbf{(Eq. 2)}$$

The associated equilibrium constant (**Eq. 2**) shows that as the total concentration increases the balance will quickly shift towards the larger complex. A factor of 10 increase in the equilibrium concentration of **Co₄** requires a factor of 1000 increase in the concentration of **Co₁₂** to maintain the equilibrium constant, *i.e.* the [**Co₁₂**]/[**Co₄**] ratio will increase by a factor of 100. Similarly, in the 2**Co₂** ⇌ **Co₄** equilibrium, increasing the concentration of **Co₂** by a factor of 10 requires [**Co₄**] to increase by a factor of 100, *i.e.* a factor of 10 increase in the [**Co₄**]/[**Co₂**] ratio. A shift in concentration domain by a factor of 10 – as per the difference between spectra in **figures 2.27 and 2.30** – therefore strongly increases the proportions of the larger complexes present in the equilibrium mixture.

It is also worth looking at the peaks in **figure 2.30** at 85 ppm labelled in purple. These belong to none of the three species identified earlier and so must correspond to an unknown species also present in the equilibrium. Since it is mainly present at very high concentrations and low temperatures (diminishes as temperature increases) it can be assumed to be larger than the **Co₁₂** species. However it does not show up in mass spectrometry studies and this concentration is at the limit before saturation occurs so a more concentrated solution cannot be obtained to study this species in more detail (at the current concentration only a small percentage of this species is present compared to the **Co₁₂**). It may well be the **Co₁₆** species, since it is currently the largest assembly we have observed.

Calculation of speciation behaviour in water

From integration of signals associated with different species in equilibrium their relative concentrations were determined. This requires careful consideration of the symmetry of the complexes, as a single signal corresponds to a different number of protons in each case. In **Co₂** one signal corresponds to 3H as there are three equivalent ligands with no internal symmetry; in **Co₄**, with six equivalent ligands all having twofold symmetry, each signal corresponds to 12H; and in **Co₁₂**, with 18 ligands split into 12 equivalent sets (each of 1.5 magnetically equivalent ligands), each signal again corresponds to 12H. Taking this into account, and knowing the total amount of complex used, the concentration of each species could be calculated.

From the set of spectra at 25 ˚C, for the 2**Co₂** ⇌ **Co₄** equilibrium, an equilibrium constant of 8.4 x 10³ M⁻¹ (Δ$G°$ = −2 kJ mol⁻¹); and for the 3**Co₄** ⇌ **Co₁₂** equilibrium, an equilibrium constant of 1.5 x 10⁷ M⁻² (Δ$G°$ = −41 kJ mol⁻¹) were obtained. These values are averaged from several NMR measurements at different concentrations.   From these equilibrium constants, for the 6**Co₂** ⇌ **Co₁₂** equilibrium, the equilibrium constant is 9.1 x 10¹⁸ M⁻⁵ [ΔG = 108 kJ mol⁻¹, *i.e.* 3 x (−22) + (−41) kJ mol⁻¹ within rounding errors]. **Table 1** lists the equilibrium constants at four different temperatures (25, 45, 70 and 100 ˚C).

Temperature / ˚C	K_{2-4} / M⁻¹	K_{4-12} / M⁻²	K_{2-12} / M⁻⁵
25	8.4 x 10³	1.5 x 10⁷	9.1 x 10¹⁸
45	2.8 x 10³	4.8 x 10⁷	1.1 x 10¹⁸
70	7.4 x 10²	2.1 x 10⁷	8.7 x 10¹⁵
100	3.6 x 10²	1.0 x 10⁷	4.6 x 10¹⁴

Table 1: Equilibrium constants for interconversions between **Co₂, Co₄, Co₁₂** at different temperatures based on integration of signals in ¹H NMR spectra.

From these equilibrium constants, the speciation behaviour for the whole three-component system at a range of temperatures were determined, as shown in **figure 2.31**. The accuracy of the speciation diagrams is limited by uncertainty in measurements of integral values of weak signals in paramagnetic complexes – in the 6 **Co₂** ⇌ **Co₁₂** equilibrium constant, for example, the equilibrium constant includes an intensity measurement with an estimated uncertainty of ±20% raised to the sixth power – but the general behaviour is clear.
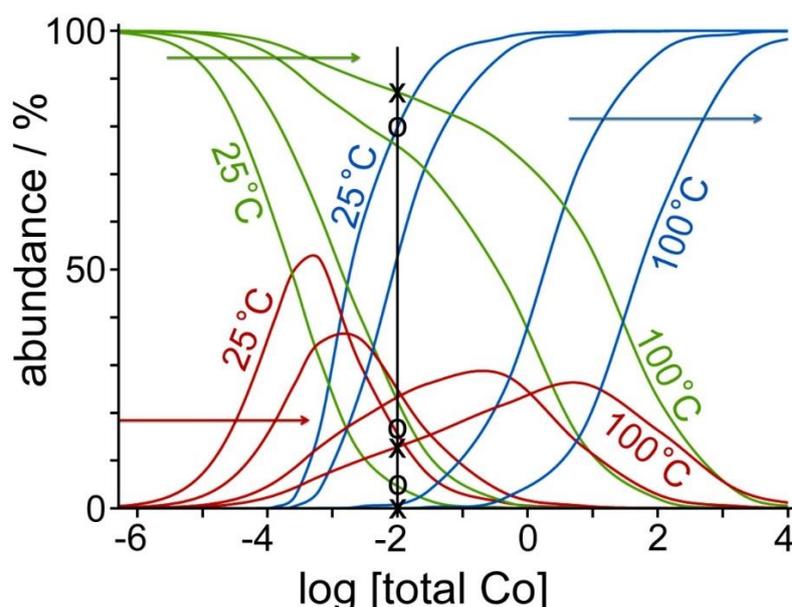
Figure 2.31: Speciation behaviour of the **Co₂** (green)/ **Co₄** (red)/ **Co₁₂** (blue) system in aqueous solution at 25, 45, 70, 100 ˚C based on the stepwise *K* values (Table 1).

As the temperature increases the curves that describe the proportions of each species at that temperature shift to the right such that the concentration at which **Co$_{12}$** disappears, and the smaller complexes appear, increases. Thus at higher temperatures, a given concentration results in more fragmentation.

The black line in **figure 2.31** drawn at [Co] = 10^{-2} M corresponds to the concentration used for the ^{1}H NMR spectra in **figure 2.27**. From the intersections of this line with the 25 ˚C curves (marked by **circles on figure 2.31**) the equilibrium solution contains **Co$_{12}$** as the major component, **Co$_4$** as a significant minor component and almost no **Co$_2$**, which agrees with the RT spectrum in **figure 2.27**. Conversely the intersections of the black line with the 100 ˚C speciation curves (marked by crosses) shows that the equilibrium solution is dominated by **Co$_2$** with a small amount of **Co$_4$** and virtually no **Co$_{12}$**, which again agrees well with the 100 ˚C spectrum (**top of figure 2.27**). The match between the observed ^{1}H NMR spectra in figure 2.30, recorded at the higher concentration of [Co] = 0.1 M, and the calculated speciation behaviour in **figure 2.31**, is less quantitatively convincing – presumably because our model does not take into account the formation of the additional fourth species, larger than **Co$_{12}$**, which starts to appear at high concentrations (purple peaks in **figure 2.30**).

We note also that **figure 2.31** shows how fortunate it was to be able to isolate crystals of **Co$_4$** from cold aqueous solution: the high concentrations (molar) in developing crystals should give almost exclusively **Co$_{12}$** under those conditions, except that the interconversion from **Co$_4$** to **Co$_{12}$** was clearly very slow at that temperature compared to the timescale of crystal growth (hours).

The effect of a guest molecule

The original aim of this chapter was to see if we could take advantage of the hydrophobic effect to observe guest binding in cages other than the cube. The addition of a guest molecule into the equilibrium may drive it to the formation of one sole species if the guest binds perfectly in the cavity of one particular size of cage.
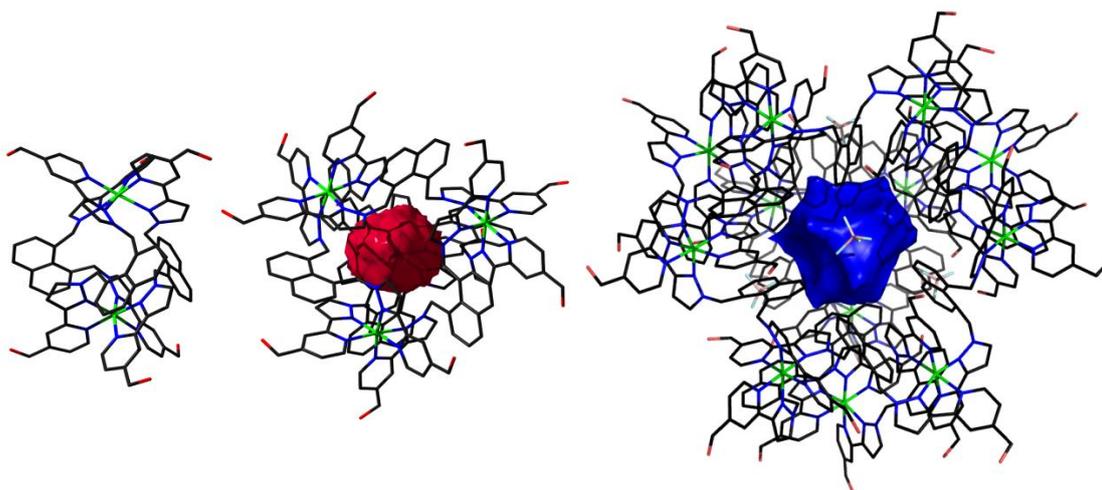
Figure 2.32: X-ray crystallographic structures of Co_2L_3, Co_4L_6 and $Co_{12}L_{18}$. The blue and red blobs represent the cavity of each species from its solvent accessible surface

The smallest component (**Co₂**) has no cavity for a guest to bind. The **Co₄** species does have a cavity however no guest is expected to bind since the cavity is very small (91 Å^3) and is already occupied by a tightly bound BF_4^- counter ion, which effectively prevents guest binding. Therefore the only structure of this series which could accommodate a guest is the largest assembly (**Co₁₂**), which has a cavity volume of 320 Å^3 (**figure 2.32**). According to Rebek, the ideal guest volume would be around 55% of the cavity volume (180 Å^3).

Various guest molecules were screened with different shapes and sizes however no clear evidence was observed for binding in any of the three components. There were a couple of guests (adamantane carboxylic acid, and adamantanone) that showed some slow exchange binding, however not enough to quantify, nor enough to perturb the equilibrium.

Possibly the cavity contains anions that can impede the guest binding (they were present in the parent **Co₁₂** cage), this was not seen in this crystal structure, however the data quality was not good enough to locate most of the anions. Another factor could be the cavity shape (**figure 2.33**), the naphthalene parts of the ligand protrude deep inside the cages cavity, which is the reason despite being technically a larger assembly than the **Co₈** cube, the **Co₁₂** has a smaller cavity. Also due to its awkward shape, the usable cavity volume for a guest molecule to fit, may be even smaller.
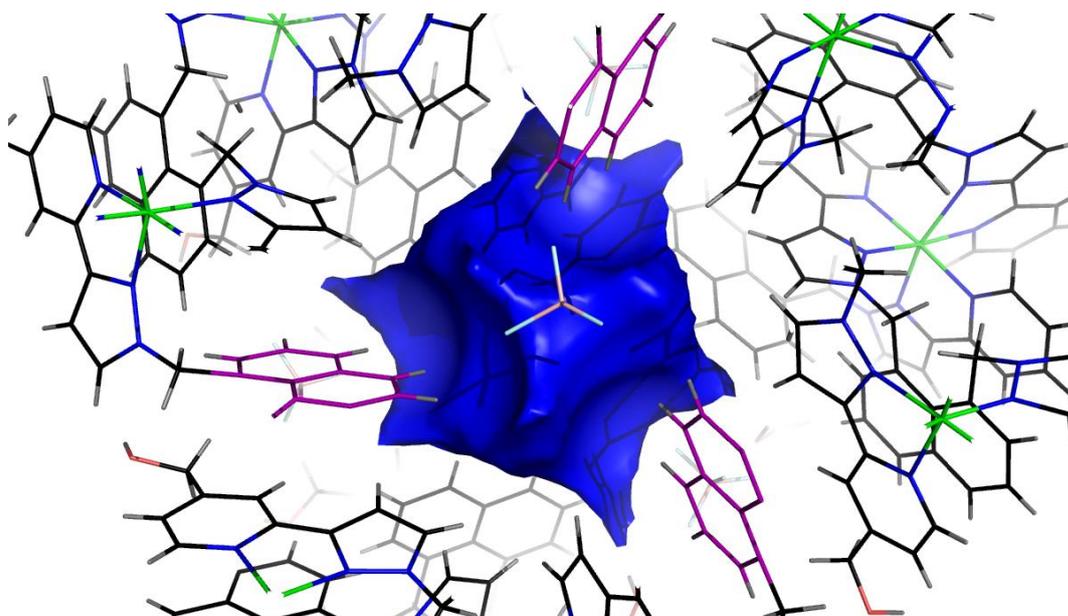
Figure 2.33: Cavity shape with naphthalenes (purple) digging in the cavity

2.2.5 Structure of an unexpected Cd network

When L$^{1,8\text{-naphOH}}$ was mixed in a 1:1 ratio with Cd(ClO$_4$)$_2$ some crystalline material formed after solvothermal synthesis. From analysis by X-ray crystallography and mass spectrometry, the assembled structure was found to be a two-dimensional coordination network.

The repeating unit (**figure 2.34**) is composed of 4 Cd(II) ions connected in a square by four L$^{1,8\text{-naphOH}}$ ligands as the base. Each repeating unit is joined together by two bridging CH$_2$OH from the PyPz of opposing repeating units to form a double helix-like structure.
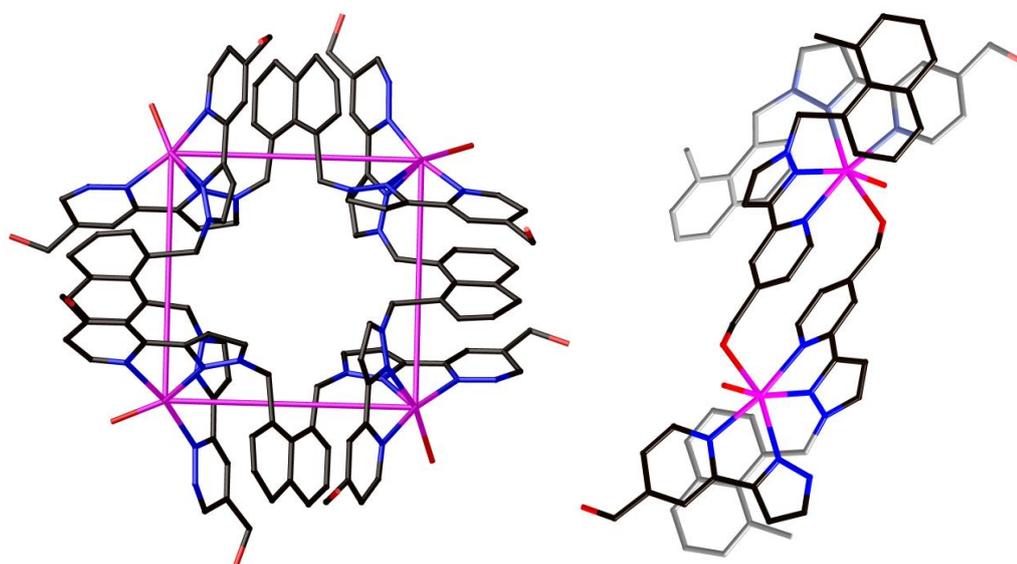
Figure 2.34: M$_4$L$_4$ square repeating unit (left) and the bridging unit (right)

It is worth pointing out that there are only two PyPz units per metal ion (rather than 3). One of the remaining coordinating sites at each metal is taken up from a PyPz methylene alcohol oxygen atom from the neighbouring Cd atom. This is an important consequence associated with putting CH$_2$OH groups on the ligands, as this O-bridged structure could not form with the unsubstituted PyPz unit. The final coordination site is occupied by an oxygen atom from the aqueous solvent, hence Cd(II)N$_4$O$_2$ environments (**figure 2.35**).
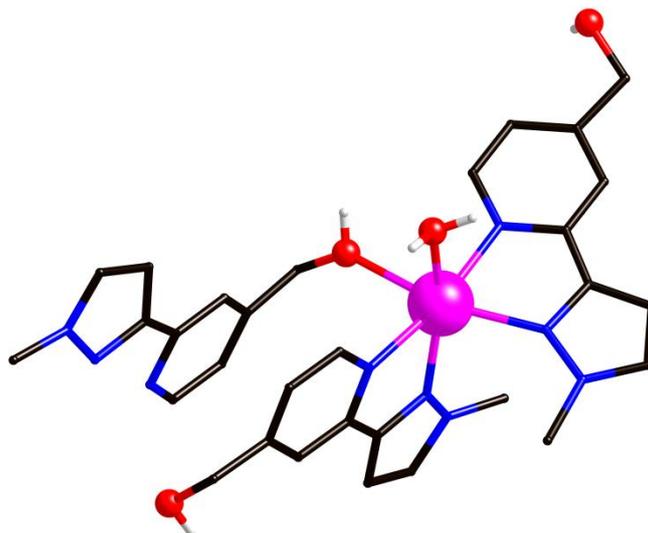
Figure 2.35: The environment around each Cd(II) ion

The arrangement of the ligands in this way forms a dinuclear double helix-like structure at each Cd_2 pair. The square base with the bridging sides creates a pseudo-bowl structure, and there is always a ClO_4^- counter ion present in the square (**figure 2.36**).
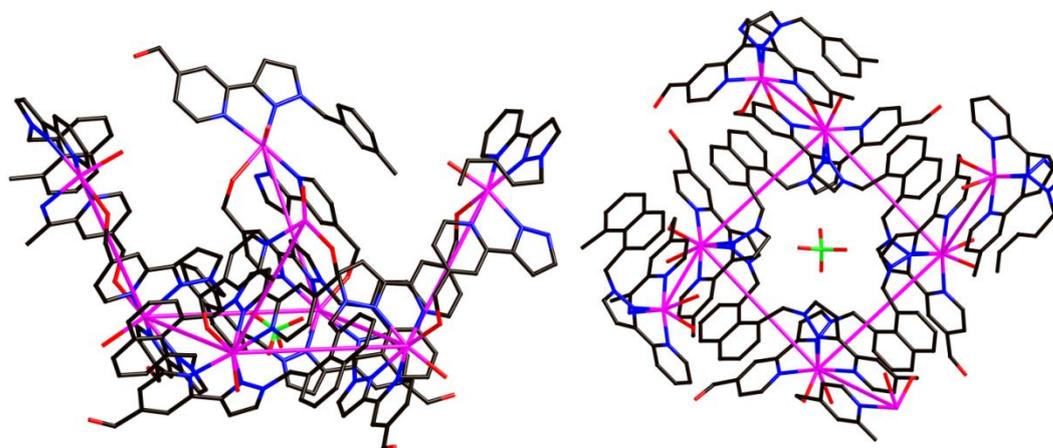
Figure 2.36: the bowl shape viewed from the side (top) and viewed from the top to show the ClO_4^- counter ion in the square (bottom)

In the fully grown structure, when viewed from the top, it possible to see each sheet and the packing between the units and the intricacy and shape of this sheet can been beautifully seen when all but the Cd atoms are removed (**figure 2.37**).
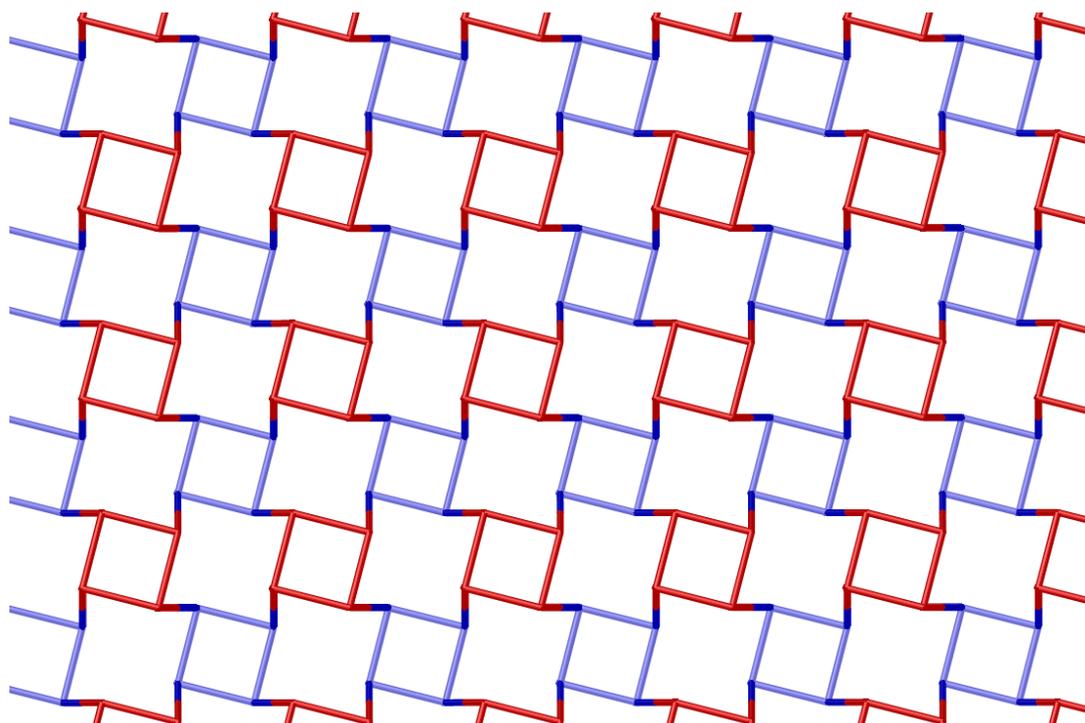
Figure 2.37: X-ray crystal structure showing the 2D array of the
$[Cd_4(L^{1,8\text{-naphOH}})_4]n(ClO_4)n.4nH_2O$ network viewed from the top to show the intricate
design of each sheet; all ligands and counter ions removed for clarity

When the 2D sheet is viewed from its side, it is clear that has a corrugated structure
(**figure 2.38**).
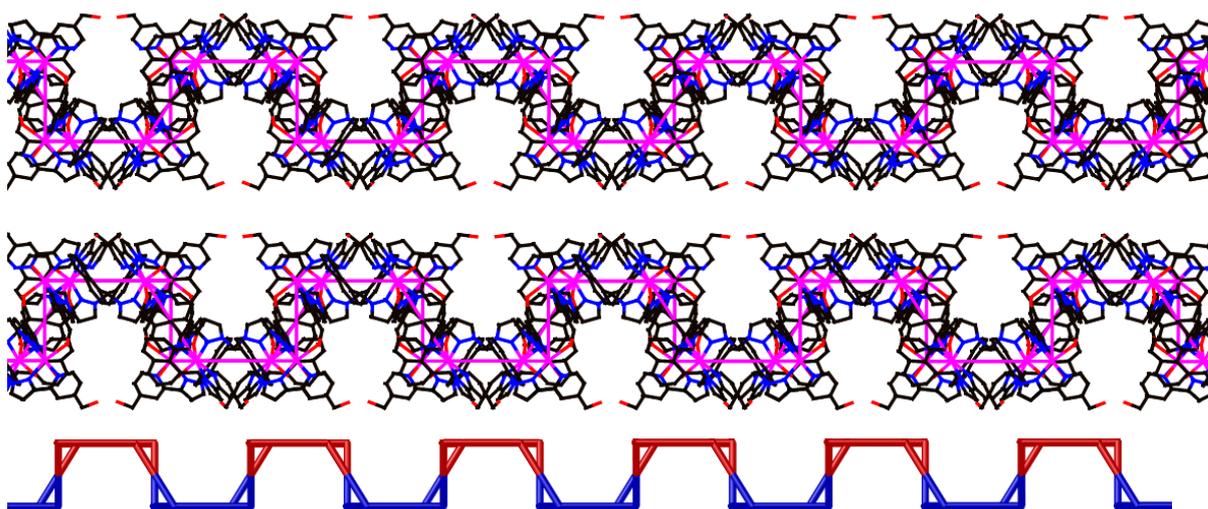
Figure 2.38: X-ray crystal structure showing the $[Cd_4(L^{18naphOH})_4]n(ClO_4)n.4nH_2O$ network
viewed from the side to show the corrugated shape; the pink/red/blue lines represent
each ligand demonstrating the geometry of the unit

There is pi-stacking that can be seen running through the whole network. These pi-stacks consist of many short range stacks and also longer ranger pi-stacks which run through the whole network in almost a spiral-like fashion (**figure 2.39**).
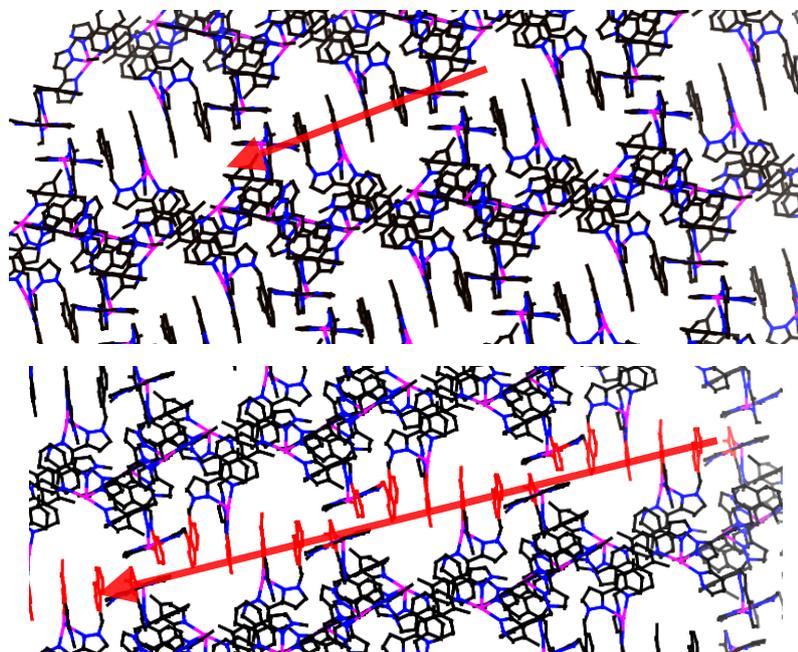
Figure 2.39: X-ray crystal structure of $[Cd_4(L^{1,8-naphOH})_4]n(ClO_4)n.4nH_2O$ showing the short range π-stacking.

2.3 Conclusions

The target L$^{1,8\text{-naphOH}}$ ligand has been successfully synthesised and has been used to assemble the desired water-soluble Co$_{12}$L$_{18}$ cage. However this was not the only assembled structure. As it turns out there are at least two other assemblies which are a Co$_4$L$_6$ tetrahedron and a Co$_2$L$_3$ dinuclear triple mesocate.

Isolation of each of the three major structures has been completed successfully, and have been characterised by NMR spectroscopy, mass spectrometry and X-ray crystallography.

This is a very rare and thoroughly characterised example of the effects of temperature/concentration/solvent on the course of a self-assembly process, showing how the different components can each be targeted by controlling the conditions.

Guest binding tests in the Co$_{12}$L$_{18}$ cavity were not conclusive, mainly due to strange shaped cavity and possible anions inside the cavity.

A novel M1:L1 Cd network has also been isolated and characterised by X-ray crystallography and mass spectrometry.

2.4 Experimental

2.4.1 Synthetic procedures
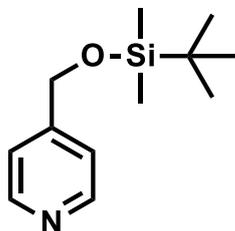
4-ᵗbutyldimethylsilyl-O-methyl pyridine, **2**

Under a N_2 atmosphere, imidazole (14.0 g, 234 mmol) was added to a 2-neck round bottomed flask (250 ml), followed by dimethylformamide (DMF)/dichloromethane (DCM) (90:10, 100 ml, dry). ᵗButyldimethylsilyl (TBDMS) chloride (24.9 g, 165 mmol) was added slowly. Once added, the reaction mixture was left to stir at room temperature for 10 minutes and then 4-methanol pyridine (15.0 g, 136 mmol) was added slowly. Once added, the reaction mixture was stirred for 18 hours at room temperature. After this time, the solvent was removed on a rotary evaporator, followed by the addition of water (150 ml) and extracted with ethyl acetate/hexane (1:1, 4 x 100 ml). The organic layers were combined, dried over magnesium sulphate, and the solvent was removed on a rotary evaporator to give a creamy solid product.

Yield 30.5 g, 99 %;

ES-MS m/z (%) 224.1 [M + H]⁺;

¹H NMR (250 MHz, CDCl₃) δ 0.14 (s, 6H), 0.97 (s, 9H), 4.78 (s, 2H), 7.31 (d, J = 6.0 Hz, 2H), 8.58 (d, J = 6.0 Hz, 2H).
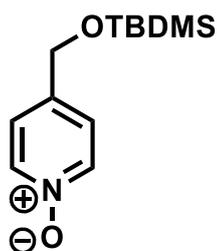
4-TBDMSO-methyl pyridine N-oxide, **3**

2 (30.5 g, 137 mmol) was placed in a 1-neck round bottomed flask (500 ml) and dissolved in DCM (200ml). Metachloroperbenzoic acid (mCPBA) (70 %, 41.4 g, 234 mmol) was then added slowly, and left to stir for 18 hours at room temperature. Following this, sodium hydroxide (200 ml, 1 M) was added, washed with water (100 ml), and the aqueous layers were combined and extracted with DCM (3 x 100 ml). The organic layers were combined, dried over magnesium sulphate, and solvent was removed on a rotary evaporator to give a very pale yellow oil product.

Yield 31.8 g, 97 %;

ES-MS m/z (%) 240.1 [M + H]$^+$;

^{1}H NMR (400 MHz, CDCl$_3$) δ 0.12 (s, 6H), 0.95 (s, 9H), 4.71 (s, 2H), 7.26 (d, J = 7.0 Hz, 2H), 8.20 (d, J = 7.0 Hz, 2H).
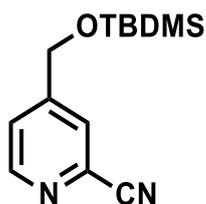
4-TBDMSO-methyl cyano pyridine, **4**

3 (31.8 g, 133 mmol) was dissolved in a 1-neck round bottomed flask (1 L) using DCM (300 ml, dry). To this trimethylsilyl cyanide was added dropwise and stirred for a further 10 minutes at room temperature. Dimethylcarbamoyl chloride was then added dropwise and then stirred for 18 hours at room temperature. Aqueous potassium carbonate (350 ml, saturated) was added, stirred for a further 10 minutes and extracted with DCM (3 x 150 ml). The organic layers were combined, dried over magnesium sulphate and the solvent was removed on a rotary evaporator. The red oil was subsequently stirred with water (150 ml) for 1 hour to remove any remaining water-soluble impurities. The mixture was again extracted with DCM (3 x 150 ml), dried over magnesium sulphate and the solvent was removed on a rotary evaporator to give a red oil. This oil was purified by column chromatography (silica, 1 % MeOH in DCM) to give a pale yellow oil.

Yield 29.8 g, 90 %;

ES-MS m/z (%) 249.1 [M + H]$^+$;

^{1}H NMR (400 MHz, CDCl$_3$) δ 0.15 (s, 6H), 0.98 (s, 9H), 4.8 (s, 2H), 7.48 (d, J = 5.0 Hz, 1H), 7.70 (s, 1H), 8.67 (d, J = 5.0 Hz, 1H).
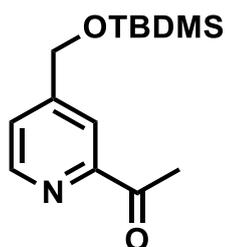
4-TBDMSO-methyl-2-acetyl pyridine, **5**

4 (3.00 g, 12.1 mmol), was placed into a flame-dried 2-neck round bottomed flask (100 ml) under a N_2 atmosphere. After diethyl ether (40.0 ml, dry) was added, the solution was cooled to 0 °C, and MeMgBr (3 M solution, 5.00 g, 14.5 mmol) was added dropwise. After 10 minutes, the reaction mixture was brought to room temperature, where it was left to stir for 3 hours. Aqueous ammonium chloride (50.0 ml, saturated) was added to quench the reaction followed by vigorous stirring for 2 minutes. The organic layer was separated, and the aqueous layer was extracted with DCM (3 x 50 ml). The organic layers were combined, dried over magnesium sulphate and the solvent was removed on a rotary evaporator to give an orange oil. This oil was purified by column chromatography (silica, 2 % MeOH in DCM) to give a yellow oil.

Yield 0.69 g, 19 %;

ES-MS m/z (%) 266.2 $[M + H]^+$;

^{1}H NMR (400 MHz, CDCl$_3$) δ 0.12 (s, 6H), 0.95 (s, 9H), 2.73 (s, 3H), 4.79 (s, 2H), 7.49 (d, J = 5.0 Hz, 1H), 7.95 (s, 1H), 8.64 (d, J = 5.0 Hz, 1H).
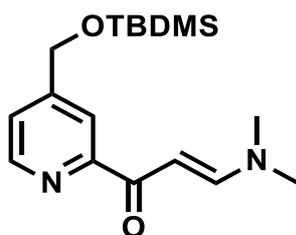
4-TBDMSO-methyl-2-acetyl dimethyl enamine pyridine, **6**

5 (2.52 g, 7.68 mmol) was placed in a 1-neck round bottomed flask (25 ml) and N,N-Dimethylformamide dimethyl acetal (N,N-DMF-DMA) (97 %, 2.00 g, 16.3 mmol) was added as both the solvent and the reagent. The reaction mixture was stirred at 110 °C for 18 hours, and the solvent was then removed on a rotary evaporator to give a brown oil. This oil was purified by column chromatography (silica, 10 % MeOH in DCM) to give a brown solid.

Yield 2.81 g, 92 %;

ES-MS m/z (%) 321.2 [*M* + H]⁺;

¹H NMR (400 MHz, CDCl₃) δ 0.13 (s, 6H), 0.98 (s, 9H), 3.03 (s, 3H), 3.20 (s, 3H), 4.80 (s, 2H), 6.47 (d, *J* = 12.5 Hz, 1H), 7.45 (d, *J* = 5.0 Hz, 1H), 7.94 (d, *J* = 12.5, 1H), 8.07 (s, 1H), 8.61 (d, *J* = 5.0 Hz, 1H).
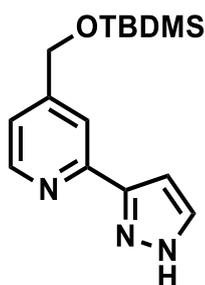
4-TBDMSO-methyl-2-pyrazole pyridine, **7**

6 (2.81 g, 8.74 mmol) was placed into a 1-neck round bottomed flask (50 ml) and dissolved in ethanol (20 ml). Hydrazine monohydrate was added (5.60 g, 175 mmol), and the reaction mixture was stirred at 60 °C for 30 minutes. The reaction was then cooled to room temperature, and the solvent was removed on a rotary evaporator. The residue was dissolved in DCM (50 ml) and washed with water (3 x 50 ml). The aqueous layers were combined and extracted with DCM (1 x 50 ml). The organic layers were combined, dried over magnesium sulphate and solvent removed on a rotary evaporator to give a dark brown oil. This oil was purified by column chromatography (silica, 5 % MeOH in DCM) to give a light brown solid.

Yield 2.25 g, 89 %;

ES-MS m/z (%) 290.2 [M + H]$^+$;

Accurate mass calculated for $C_{15}H_{24}N_3OSi$ [MH$^+$]: 290.1689, observed: 290.1687

^{1}H NMR (400 MHz, CDCl$_3$) δ 0.15 (s, 6H), 0.99 (s, 9H), 4.82 (s, 2H), 6.82 (d, J = 2.0 Hz, 1H), 7.25 (d, J = 5.0 Hz, 1H), 7.69 (d, J = 2.0, 1H), 7.72 (s, 1H), 8.61 (d, J = 5.0 Hz, 1H);

^{13}C NMR (100.6 MHz, CDCl$_3$) δ -5.32, 18.40, 25.90, 63.51, 103.41, 117.14, 119.85, 137.96, 144.52, 149.16, 149.34, 151.91.
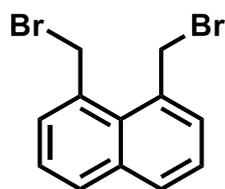
1,8-Bis(bromomethyl)naphthalene, **9**

1,8-Dimethylnaphthalene (0.50 g, 3.20 mmol) was added to a 1-neck round bottomed flask (100 ml) followed by *N*-bromosuccinimide (1.63 g, 9.16 mmol) and azobisisobutyronitrile (AIBN) (30.0 mg, 0.190 mmol). CCl_4 (45 ml) was added and the reaction mixture was refluxed at 83 °C and irradiated with a tungsten lamp. The reaction mixture was cooled to room temperature, and the white precipitate by-product formed was filtered off. The solvent was removed from the filtrate on a rotary evaporator. The residue was dissolved in DCM (50 ml), washed with water (3 x 50 ml), dried over magnesium sulphate and the solvent removed on a rotary evaporator to give a yellow solid. This solid was re-purified by crystallisation from toluene to give yellow crystals.

Yield 0.51 g, 51 %;

ES-MS m/z (%) 313 $[M + H]^+$;

^{1}H NMR (400 MHz, CDCl$_3$) δ 5.33 (s, 4H), 7.48 (dd, *J* = 8.0, 7.0 Hz, 2H), 7.65 (dd, *J* = 7.0, 1.5 Hz, 2H), 7.91 (dd, *J* = 8.0, 1.5 Hz, 2H);

^{13}C NMR (100.6 MHz, CDCl$_3$) δ 37.20, 125.74, 129.6, 131.96, 133.08, 133.46, 136.16.
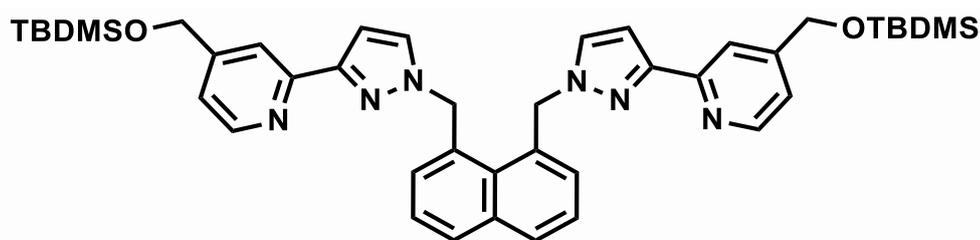
L$^{\text{1,8-naphOTBDMS}}$

7 (0.50 g, 1.73 mmol) was placed in a 2-neck round bottomed flask (100 ml) under a nitrogen atmosphere and sodium hydride (60 % dispersion in mineral oil, 0.0690 g, 1.73 mmol) was added and after 20 minutes tetrahydrofuran (THF) (25 ml, dry) was added. The reaction mixture was stirred for 10 minutes and then **9** (0.270 g, 8.60 mmol) was added and then stirred at 70 °C. After 8 hours more sodium hydride (60 % dispersion in mineral oil, 0.0690 g, 1.73 mmol) was added. The reaction was monitored by thin layer chromatography (TLC) (silica, 5 % MeOH in DCM) until all the starting material had disappeared (24 hours). After this time, the reaction mixture was cooled to room temperature, and quenched by the addition of methanol (10 ml). The solvent was removed on a rotary evaporator and the resulting orangey-brown solid was purified by column chromatography (silica, 5 % MeOH in DCM) to give a yellow oil.

Yield 0.61 g, 97 %;

ES-MS m/z (%) 731.4 [M + H]$^+$;

^{1}H NMR (400 MHz, CDCl$_3$) δ 0.13 (s, 6H), 0.96 (s, 9H), 4.79 (s, 2H), 5.95 (s, 2H), 6.89 (d, J = 2.5 Hz, 1H), 7.17 (d, J = 2.5 Hz, 1H), 7.26 (d, J = 5.0 Hz, 1H), 7.48 (t, J = 8.0 Hz, 1H), 7.88 (s, 1H), 7.94 (d, J = 8.0 Hz, 1H), 8.60 (d, J = 5.0 Hz, 1H);

^{13}C NMR (100.6 MHz, CDCl$_3$) δ -5.32, 18.40, 25.90, 56.89, 63.68, 104.92, 116.99, 119.56, 125.60, 130.42, 130.90, 130.99, 131.02, 131.42, 135.90, 149.37, 151.29, 151.87, 152.24.
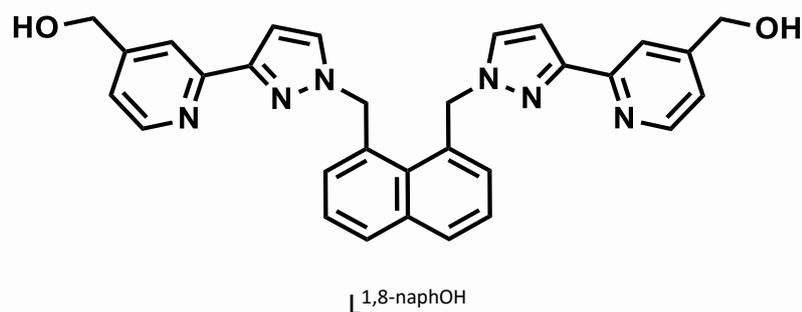
L$^{1,8\text{-naphOH}}$

L$^{1,8\text{-naphOTBDMS}}$ (0.610 g, 0.834 mmol) was dissolved in THF (30 ml) and placed in a 1-neck round bottomed flask (100 ml). tetrabutylammonium fluoride (TBAF) (0.530 g, 1.67 mmol) was added, and the reaction mixture was stirred for 18 hours at room temperature. $CHCl_3$ (40 ml) was added and stirred for a further 5 minutes, then washed with water (3 x 50 ml), dried over magnesium sulphate and solvent removed on a rotary evaporator to give a yellow solid. This was purified by column chromatography (silica, 10 % MeOH in DCM) to afford a white solid.

Yield 0.41 g, 96 %;

ES-MS m/z (%) 503.2 [M + H]$^+$;

Accurate mass calculated for $C_{30}H_{27}N_6O_2$ [MH$^+$]: 503.2195, observed: 503.2173

^{1}**H NMR** (400 MHz, DMSO-d$_6$ with drop of D$_2$O) δ 4.53 (s, 2H), 6.215 (s, 2H), 6.91 (d, J = 2.5 Hz, 1H), 7.02 (d, J = 8.0 Hz, 1H), 7.23 (d, J = 5.0 Hz, 1H), 7.47 (t, J = 8.0 Hz, 1H), 7.73 (d, J = 2.5 Hz, 1H), 7.97 (d, J = 8.0 Hz, 1H), 8.47 (d, J = 5.0 Hz, 1H);

^{13}**C NMR** (100.6 MHz, CDCl$_3$) δ 56.16, 61.91, 105.18, 116.95, 120.54, 125.95, 128.84, 129.93, 130.53, 132.85, 133.84, 135.62, 149.38, 151.87, 151.97, 152.77.

[Co$_2$L$_3$](BF$_4$)$_4$/(ClO$_4$)$_4$, [Co$_4$L$_6$](BF$_4$)$_8$/(ClO$_4$)$_8$ and [Co$_{12}$L$_{18}$](BF$_4$)$_{24}$/(ClO$_4$)$_{24}$;
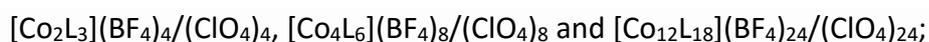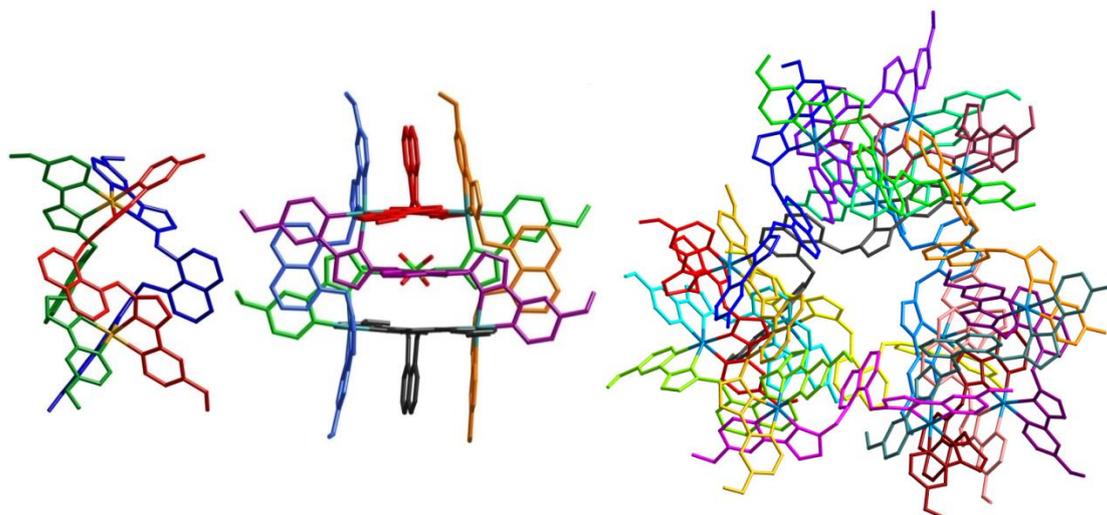
where L = L$^{1,8\text{-naphOH}}$

L$^{1,8\text{-naphOH}}$ (30.0 mg, 0.06 mmol) and either Co(ClO$_4$)$_2$.6H$_2$O (14.6 mg, 0.04 mmol) or Co(BF$_4$)$_2$.6H$_2$O (13.6 mg, 0.04 mmol) and methanol (8 ml) were added to a Teflon lined autoclave. The autoclave was sealed, placed in an oven and heated to 100 °C for 12 hours and cooled slowly to room temperature at a rate of 1 °C min^{-1}. The resulting orange solution was a mixture of the three assembled structures M$_2$L$_3$, M$_4$L$_6$ and M$_{12}$L$_{18}$.

ES-MS m/z (%) M$_{12}$L$_{18}$: 1887 [M − 6BF$_4^-$]$^{6+}$, 1611 [M − 7BF$_4^-$]$^{7+}$, 1394 [M − 8BF$_4^-$]$^{8+}$,

1229 [M − 9BF$_4^-$]$^{9+}$, 1098 [M − 10BF$_4^-$]$^{10+}$, 990 [M − 11BF$_4^-$]$^{11+}$,

900 [M − 12BF$_4^-$]$^{12+}$;

M$_4$L$_6$: 1926 [M − 2ClO$_4^-$]$^{2+}$, 1251 [M − 3ClO$_4^-$]$^{3+}$, 913 [M − 4ClO$_4^-$]$^{4+}$;

M$_2$L$_3$: 1923 [M − 1ClO$_4^-$]$^{1+}$, 912 [M − 2ClO$_4^-$]$^{2+}$, 575 [M − 3ClO$_4^-$]$^{3+}$,

406 [M − 4ClO$_4^-$]$^{4+}$;

Accurate mass M$_{12}$L$_{18}$: calculated for C$_{570}$H$_{494}$N$_{114}$O$_{38}$B$_{16}$F$_{64}$Co$_{12}$ [M − 8BF$_4^-$]$^{8+}$: 1392.7638, observed: 1392.7642;

M$_4$L$_6$: calculated for C$_{180}$H$_{156}$N$_{36}$O$_{12}$B$_5$F$_{20}$Co$_4$ [M − 3BF$_4^-$]$^{3+}$: 1228.3448, observed: 1228.3425;

M$_2$L$_3$: calculated for C$_{90}$H$_{78}$N$_{18}$O$_{14}$Cl$_2$Co$_2$ [M − 2ClO$_4^-$]^2: 911.1993, observed: 911.1998

^{1}H NMR (400 MHz, D$_2$O) For M$_{12}$, M$_4$ and M$_2$ see main text.
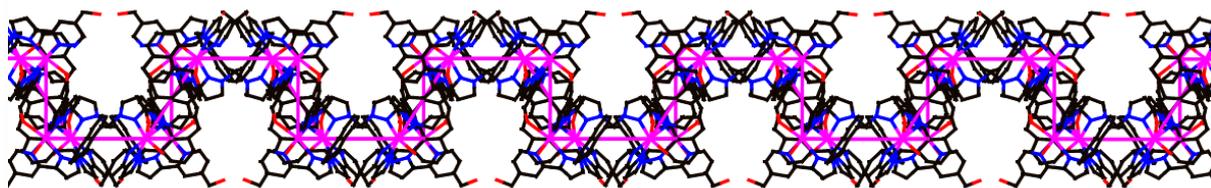
$[Cd_4(L^{1,8-naphOH})_4]n(ClO_4)n.4nH_2O$ network

$L^{1,8-naphOH}$ (30.0 mg, 0.06 mmol) and $Cd(ClO_4)_2.6H_2O$ (25.2 mg, 0.06 mmol) or $Co(BF_4)_2.6H_2O$ (13.6 mg, 0.04 mmol) and methanol (8 ml) were added to a Teflon lined autoclave. The autoclave was sealed, placed in an oven and heated to 100 ^oC for 12 hours and cooled slowly to room temperature at a rate of 1 ^oC min^{-1}. The resulting colourless crystals were pure Cd network.

2.4.2 X-ray crystallography

The crystal structure data collection of the ligand $L^{1,8-napOH} \bullet HBF_4 \bullet 2CHCl_3$ was performed at the University of Sheffield using a Bruker Apex-2 diffractometer with a Mo-Kα sealed tube source; data collection, solution and refinement were routine.

For $[Co_2(L^{1,8-napOH})_3](BF_4)_4$, $[Co_4(L^{1,8-napOH}_6)](ClO_4)_8 \bullet 18H_2O$, $[Co_{12}(L^{1,8-napOH})_{18}](BF_4)_{24} \bullet 1.5H_2O$ and $[Cd_4(L^{1,8-naphOH})_4]n(ClO_4)n \bullet 4nH_2O$, data collections in each case were performed at the EPSRC National Crystallography Service at the University of Southampton, UK, using a Rigaku FR-E+ diffractometer equipped with a Saturn 724+ CCD detector, using high-intensity Mo-Kα radiation from either a rotating anode or a microfocus sealed-tube source.[7] Structure solution and refinement was with the SHELX suite of programmes.[8] In all cases crystals exhibited the usual problems of this type of structure, *viz.* weak scattering due to a combination of poor crystallinity, extensive solvation, and disorder of anions / solvent molecules. In each case the basic structure and connectivity of the complex cation could be unambiguously determined, which is all that is required for the purposes of this work. Extensive use of geometric restraints on aromatic rings and anions, and restraints on aromatic displacement parameters, were required to keep refinements stable. Solvent molecules that could be modelled satisfactorily were included in the final refinements; in all cases large regions of diffuse electron density that could not be modelled (from disordered solvents / counter ions) were removed from the refinement, using either the SQUEEZE function in PLATON (for **Co₄**)[9,10] or the 'Solvent Mask' function in OLEX-2 (for **Co₂**, **Co₁₂** and Cd network).[11] Full details are in the individual CIFs.

The compositions given are approximate; not just because of severe disorder of anions / solvents but because the number of anions may be lower than expected (*i.e.* less than two per Co^{2+} ion) if some of the OH groups on the complex cations are deprotonated in the crystals: the high positive charge on the complex cations renders the OH groups acidic in aqueous solution which makes this plausible. For **Co₂** the total electron count removed by the 'solvent mask' in OLEX was 551 e / unit cell, which amounts to *ca.* 46 electrons per dinuclear complex unit. Only 1.5 of the expected 4 $[BF_4]^-$ anions could be located per dinuclear complex cation. The 46 e / complex unit removed during the refinement by the

'solvent mask' function is equivalent to *ca.* one additional $[BF_4]^-$ anion, implying partial deprotonation of peripheral OH groups on the complex cation and therefore fewer anions than expected.

For **Co₄** the total electron count per unit cell removed by the 'SQUEEZE' function in PLATON was 133 e / unit cell, which amounts to *ca.* 33 electrons per tetranuclear complex unit. We could only locate 7 $[ClO_4]^-$ anions per complex unit rather than the expected 8. The 'SQUEEZED' electron density is insufficient to account for this missing anion (49e) so we suggest that there are only 7 $[ClO_4]^-$ anions per complex cation in the crystal due to loss of one acidic proton from the cation, with the 33 e / complex being equivalent to *ca.* three water molecules per complex.

For **Co₁₂** only four $[BF_4]^-$ anions could be located *per* **Co₁₂** cation. The total electron count per unit cell removed by the 'solvent mask' in OLEX was 3943 e / unit cell, or *ca.* 657 e / complex unit which is consistent with *ca.* 16 $[BF_4]^-$ anions, giving a (maximum) total of 20 anions, or fewer anions plus solvent molecules. This is again consistent with partial deprotonation of OH groups to reduce the high positive charge of the cage.

For the Cd network, all four $[ClO_4]^-$ anions could be located in the asymmetric unit. The total electron count per unit cell removed by the 'solvent mask' in OLEX was 129 e / complex unit which is consistent with *ca.* 7 MeOH molecules, 13 H_2O molecules, or a combination of both.

Crystallography data tables

Complex	[Co$_2$(L$^{1,8\text{-napOH}}$)$_3$](BF$_4$)$_4$	[Co$_4$(L$^{1,8\text{-napOH}}$)$_6$](ClO$_4$)$_8$ •18H$_2$O	[Co$_{12}$(L$^{1,8\text{napOH}}$)$_{18}$](BF$_4$)$_{24}$ •1.5H$_2$O
Formula	C$_{90}$H$_{78}$N$_{18}$B$_4$Co$_2$F$_{16}$O$_6$	C$_{180}$H$_{192}$N$_{36}$Cl$_8$Co$_4$O$_{62}$	C$_{540}$H$_{471}$N$_{108}$B$_{24}$Co$_{12}$ F$_{96}$O$_{37.5}$
Molecular weight	1972.8	4371.0	11863.8
T, K	100(2)	100(2)	100(2)
Crystal system	Trigonal	Triclinic	Trigonal
Space group	*R*–3c	*P*–1	*R*–3
a, Å	18.885(3)	19.2188(13)	44.705(9)
b, Å	18.885(3)	29.070(2)	44.705(9)
c, Å	101.44(2)	35.460(3)	68.621(14)
α,˚	90	90.322(3)	90
β,˚	90	98.971(3)	90
γ,˚	120	98.598(3)	120
V, Å^3	31333(11)	19340(2)	118768(54)
Z	12	4	6
ρ, g cm^{-3}	1.255	1.501	0.995
Crystal size, mm^3	0.22 x 0.08 x 0.03	0.18 x 0.11 x 0.05	0.1 x 0.1 x 0.1
Data, restraints, parameters	4534, 463, 335	88225, 5308, 4968	23528, 2282, 1767
Final *R*1, *wR*2^b	0.188, 0.519	0.128, 0.406	0.199, 0.522

Complex	$L^{1,8-napOH}$ •HBF_4•$2CHCl_3$	$[Cd_2(L^{1,8-napOH})_2(H_2O)_2]_n(ClO_4)_{2n}$ •$2n(MeOH)$
Formula[a]	$C_{32}H_{29}BCl_6F_4N_6O_2$	$C_{62}H_{66}Cd_2Cl_4N_{12}O_{25}$
Molecular weight	829.12	1745.86
T, K	100(2)	100(2)
Crystal system	Triclinic	Tetragonal
Space group	$P-1$	$I4$
a, Å	9.5725(3)	21.8575(16)
b, Å	11.1173(3)	21.8575(16)
c, Å	17.7204(5)	38.353(3)
α,°	83.021(2)	90
β,°	77.285(2)	90
γ,°	84.835(2)	90
V, Å^3	1813.47(9)	18323(3)
Z	2	8
ρ, g cm^{-3}	1.518	1.266
Crystal size, mm^3	0.35 x 0.33 x 0.08	0.26 x 0.11 x 0.04
Data, restraints, parameters	6141, 0, 462	14148, 993, 877
Final $R1$, $wR2$[b]	0.044, 0.112	0.069, 0.1785

a   These formulae (and consequently the crystal densities) are necessarily approximate given that large amounts of diffuse electron density in solvent-accessible voids was removed from the refinements using either the 'SQUEEZE' function in PLATON or the OLEX 'Solvent Mask' function. See CIFs, and comments in experimental section, for details.

b   The value of $R1$ is based on 'observed' data with $I > 2\sigma(I)$; the value of $wR2$ is based on all data.

2.5 References

1) S. Turega; M. Whitehead; B. R. Hall; A. J. H. M. Meijer; C. A. Hunter; M. D. Ward, *Inorg. Chem.,* **2013**, *52*, 1122-1132

2) M. Whitehead; S. Turega; A. Stephenson; C. A. Hunter; M. D. Ward, *Chem. Sci.,* **2013**, *4*, 2744-2751

3) M. D. Ward; *Chem. Commun.,* **2009**, *30*, 4487-4499

4) B. R. Hall; L. E. Manck; I. S. Tidmarsh; A. Stephenson; B. F. Taylor; E. J. Blaikie; D. A. Vander Griend; M. D. Ward, *Dalton Trans.,* **2011**, *40*, 12132-12145

5) A. Stephenson; S. P. Argent; T. Riss-Johannessen; I. S. Tidmarsh; M. D. Ward, *J. Am. Chem. Soc.,* **2011**, 113, 858-870

6) S. P. Argent; H. Adams; T. Riis-Johannessen; J. C. Jeffery; L. P. Harding; O. Mamula; M. D. Ward, *Inorg. Chem.,* **2006**, *45*, 3905-3919

7) S. J. Coles; P. A. Gale, *Chem. Sci.,* **2012**, *3*, 683-689

8) G. M. Sheldrick, *Acta Crystallogr. Sect. A, Found. Crystallogr.,* **2008**, *64*, 112-122

9) A. L. Spek, *J. Appl. Crystallogr.,* **2003**, *36*, 7-13

10) P. van der Sluis; A. L. Spek, *Acta Crystallogr. Sect. A: Found. Crystallogr.,* **1990**, *46*, 194-201

11) O. V. Dolomanov; L. J. Bourhis; R. J. Gildea; J. A. K. Howard; H. Puschmann, *J. Appl. Cryst.,* **2009**, *42*, 339-341

Chapter 3

A fluorescence displacement assay for guest binding: exploring the limits of binding in the cavity using a guest library

3.1 Introduction

Over the years, there have been many published studies on cages and their host-guest capabilities for various applications such as catalysis, drug delivery, stabilisation of reactive species and sensing. However there has been a significant lack of systematic studies that do not rely purely on crystal structures to display guest binding. This is of high importance since a greater understanding of what makes a good guest will bring greater predictability to identification of guests which will allow some design aspects to be brought to the area, which in turn will give the ability to create a cage for their specific purposes and thus enhance their abilities for their proposed applications.

The Rebek group has investigated binding of alkanes in hydrogen bonded capsules which led to the 55% rule for the optimum guest size for binding in a cavity. However, very little of comparable work has been conducted using cages.[1] One of these rare examples was conducted by the Raymond group who encapsulated and stabilised a range of reactive phosphonium/ketone adducts. It was found that the size, shape and pD of the guest cations played an important role in the degree of stabilisation of the phosphonium salt in the cavity (**figure 3.1**).[2]
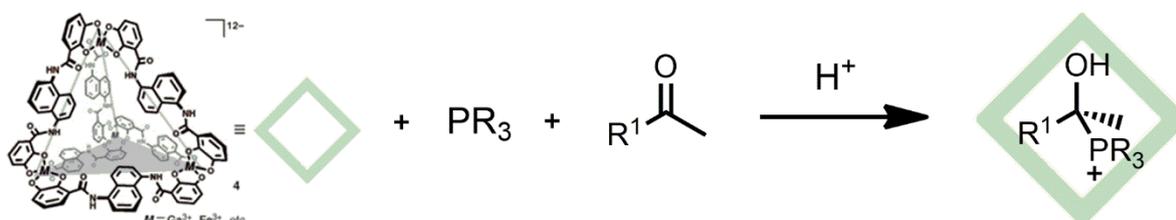
Figure 3.1: Overview of encapsulation and stabilisation of phosphonium salts where R = Me, Et, PhMe$_2$, Ph$_2$Me and R^1 = Me, Et, CFH$_2$, CF$_3$[2]

It was found that with larger groups on the ketone, the stability of the reactive guest was far greater (several weeks) than with a smaller ketone such as acetone (several days). Also aromatic groups on the phosphine could pi-stack to the cavity surface to aid stability, yet the inclusion of fluorinated groups decreases the stability.[2]

The Nitschke group have also done some work in this area. They performed studies to investigate the thermodynamic and kinetic factors that relate to guest uptake, and

performed a principal component analysis (PCA) which considers multiple factors to try to gain a greater understanding of what factors best control guest binding. They concluded that the most influential factor was the size of the guest. They used their results to demonstrate various achievements such as designing a time-dependant sequential uptake and release of specific guests (**figure 3.2**).[3]
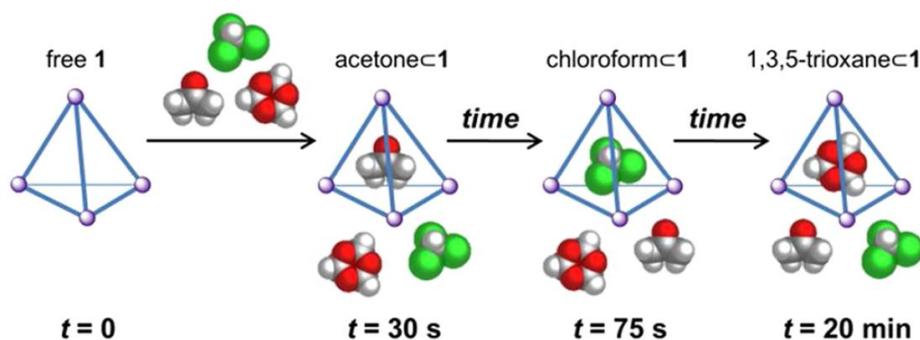
Figure 3.2: Sequential formation of acetone, chloroform and 1,3,5-trioxane host-guest complexes following simultaneous addition of all three guests[3]

In order to look into these interactions, a large number of guests will need to be investigated and so a quicker, more high-throughput, method of guest screening needs to be developed to replace NMR titrations. We need a screening system that will allow the quantification of a large library of guests in the shortest possible time to rapidly gain a grasp of the factors that govern the host-guest interactions responsible for guest binding.

Using fluorescence as a basis for evaluating guest binding, the aim is to conduct a systematic study to investigate the factors responsible for guest binding in water.

The work in the chapter was undertaken together with Simon Turega.

3.2 Results and Discussion

3.2.1 Using fluorescence when the host isn't fluorescent!

The cage currently used for host guest studies is a M_8L_{12} cubic cage assembled using Co(II). This works well for NMR studies since the paramagnetism of Co(II) allows for easy detection of guest binding (magnified shifts and large dispersion of peaks makes it easy to see spectral changes when guests bind). However using NMR spectroscopy to assess guest binding has some limitations. Firstly in order to attain the good signal to noise required for accurate measurements a large number of scans need to be run, which takes many minutes per measurement. This is fine for high concentrations of the monitored species, but in order to measure higher binding constants, lower and lower concentrations are needed and thus the time it would take to perform an NMR titration experiment would increase.

Because of these limitations, fluorescence spectroscopy is a good alternative. The concentrations commonly used are much lower than in NMR due to the ease of measuring fluorescence. This allows higher values of *K* to be measured and, by utilising a fluorescence plate-reader, many parallel measurements can be performed in a single experiment.

The main problem that first needs to be overcome in using fluorescence spectroscopy with this cage, is that due to low lying d-d transitions associated with Co(II), the emission of the normally highly fluorescent naphthalene groups in this case is quenched, and so the cage itself is not fluorescent. This can be overcome by deploying a competition experiment with a fluorescent guest (fluorophore). The emission of this fluorophore will be quenched by the Co(II) centres when binding into the cavity occurs. Then, with the addition of a competing guest, the quenched fluorophore is released back into the free solution and its emission is switched back on. By monitoring this emission change during the competition experiment, and knowing the binding affinity of the fluorophore being displaced, it is possible to calculate the binding affinity of the competing guest.

3.2.2 Choice of fluorophore

Since it is already known that coumarin, along with its simple substituted derivatives, bind within the cage, a coumarin based fluorophore was chosen as the initial guest. The selection criteria for this fluorophore were that it had to have a reasonable binding affinity (> 1000 M^{-1}), and more importantly have an absorption maximum away from that of the cage itself (**figure 3.3**) so that it can be selectively excited without the interfering absorption of the cage. This is important so it is just the guest that is monitored, and so it is possible to use the assumption that all light is absorbed by the fluorophore and not by the cage, as this would affect the intensity of the emission.

The fluorophore also needs to have an emission at high enough energy to allow energy transfer to the Co(II) ions in the cage, so that quenching can occur when it is encapsulated, but not it is when free in solution.
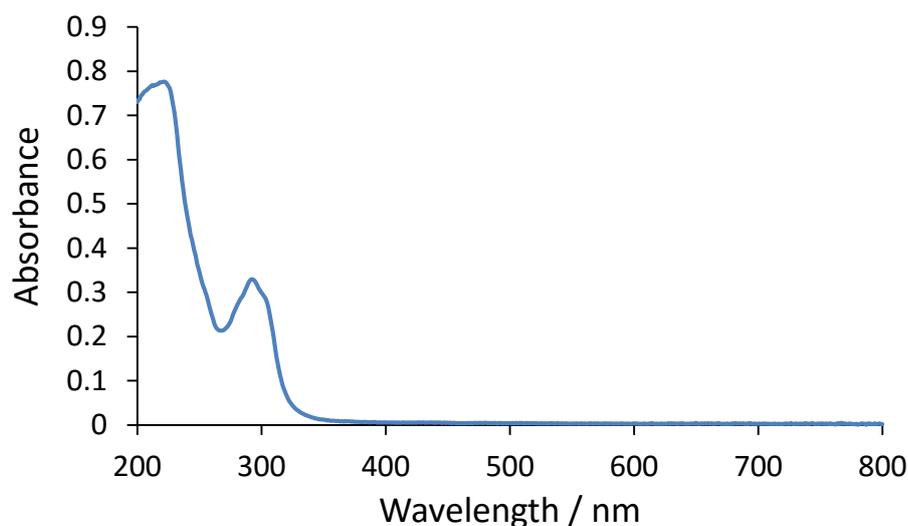
Figure 3.3: Absorption spectrum of $[Co_8(L^{1,5-naphOH})_{12}](BF_4)_{16}$ in water

The cage shows strong absorption at wavelengths below 350 nm, so a fluorophore with an absorption maximum of ≥400 nm would be ideal.

Potential coumarin based fluorophores are outlined in **figure 3.4**.
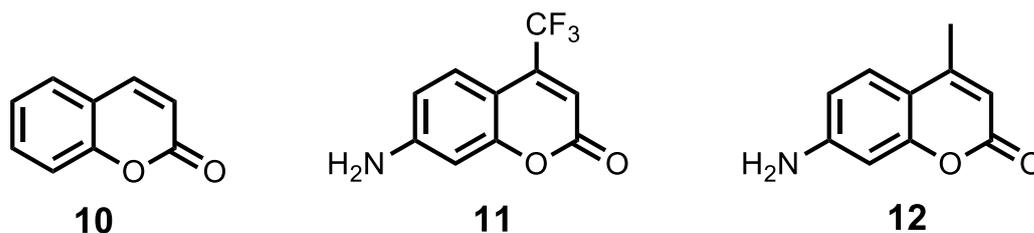
Figure 3.4: Coumarin based fluorophores: coumarin, **10**, 7-amino-4-methylcoumarin, **11** and 7-amino-4-(trifluoromethyl)coumarin, **12**

Coumarin **10**, although already known to bind in the Co_8L_{12} cage (7600 M^{-1} in water), has its longest wavelength absorption maximum at around 300 nm, which is too close to the cage absorption maximum to be of use. To encourage the absorption maximum to shift to longer wavelengths, a coumarin with electron withdrawing CF_3 and electron donating NH_2 groups was investigated (coumarin **11**) to generate a low energy charge-transfer transition. However no binding was observed with the cage. This was thought to be that the CF_3 group made **11** too bulky for the cavity. Finally a coumarin with just the NH_2 group, Coumarin **12** (aka coumarin 120) was then investigated. This would have an absorption maximum at just a long enough wavelength (343 nm) in water (**figure 3.5**) to be free of interference from cage absorption and therefore appropriate for our purposes.
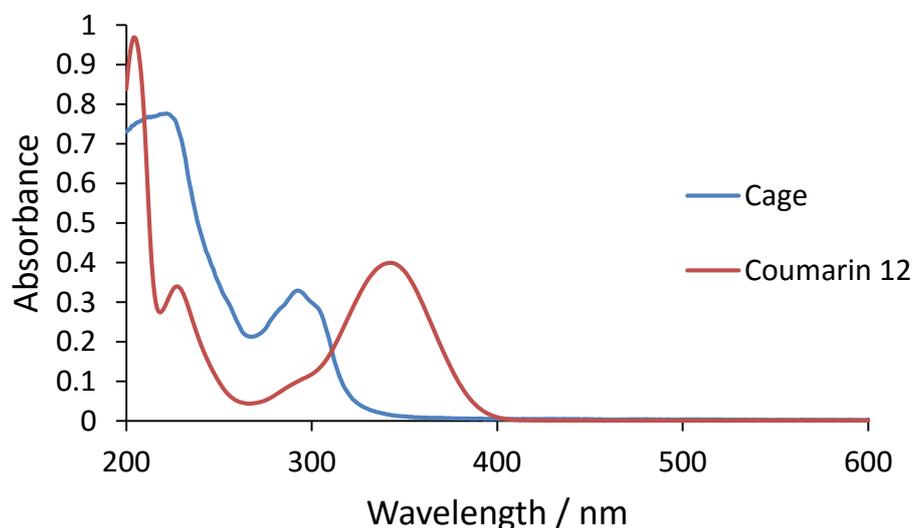
Figure 3.5: Absorption spectrum of $[Co_8(L^{1,5-naphOH})_{12}](BF_4)_{16}$ cubic cage (blue) and coumarin **12** (red) in water

A titration experiment was conducted by monitoring the emission from coumarin **12** following addition of cage solution (**figure 3.6**). The emission of **12** was indeed quenched

by the cage as it is taken up and by fitting this data to a 1:1 binding model, the binding
constant could be determined (20000 ± 2000 M^{-1}).
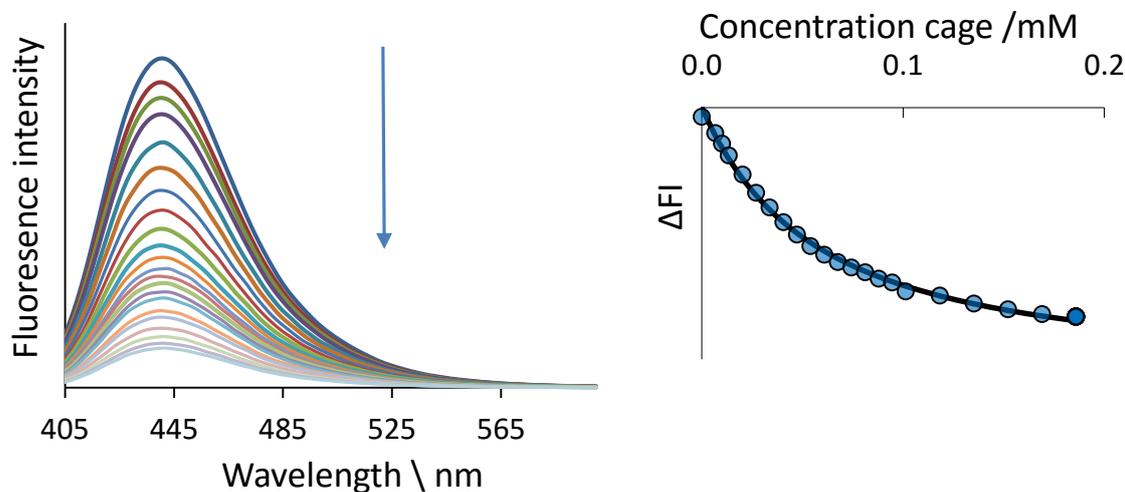
Figure 3.6: Change in fluorescence intensity of coumarin **12**, with increasing
concentration of cage and its binding isotherm in water at 298 K

3.2.3 Displacement titration test

Now that the fluorophore was selected and its binding affinity with the cage
determined, the competition experiment was then conducted to see if (i) the emission of
the fluorophore would increase with addition of a competing guest, and then if (ii) the
binding affinity of the competing guest could be determined that was consistent with the
previously determined value obtained by NMR.

13
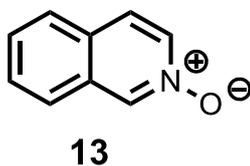
Figure 3.7: Competing guest, isoquinoline N-oxide, **13**

The competing guest isoquinoline N-oxide, **13**, was used first (**figure. 3.7**). The
experiment was conducted using an amino-methyl-coumarin/cage solution in which the
coumarin **12** was around 50% bound. An increase in fluorescence intensity was indeed
observed as the concentration of the competing guest **13** was increased and displaced the

fluorophore (**figure 3.8**). Using this data the binding constant was determined to be 4000 $\pm$ 300 M^{-1}, which is consistent with that determined previously by NMR spectroscopy (3100 $\pm$ 400 M^{-1}).
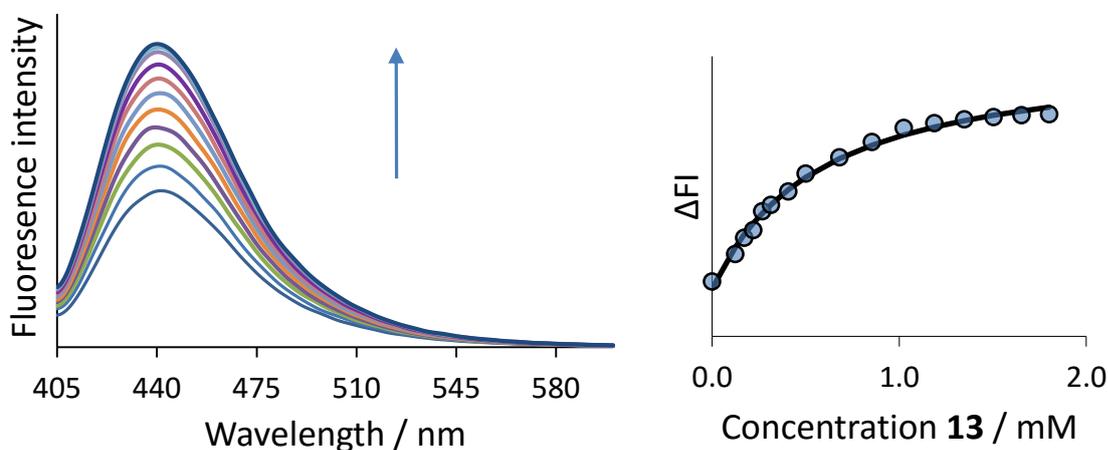
Figure 3.8: Change in fluorescence intensity of coumarin **12** (0.01 mM in a 0.055 mM cage solution), with increasing concentration of competing guest **13** (1 mM) and its binding curve in water at 298 K

The competition experiment is summarised in **figure 3.9**.
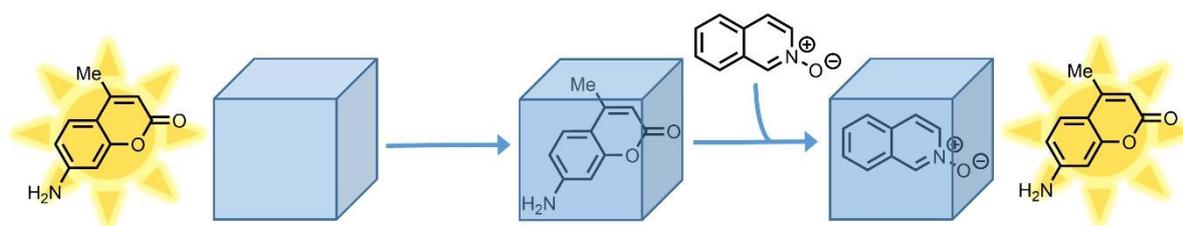
Figure 3.9: Summary of the displacement assay to measure binding affinities using fluorescence spectroscopy

At the point where there is 50% bound coumarin, there is only ever a maximum of 9 % bound cage. This means that there is actually a large excess of unbound cage for the guest molecule to bind in. Because of this, the guest will most likely bind in the free cage since not only will it cost less energy to get inside, but also will be statistically more favourable. Due to this, this experiment is not a typical displacement assay as such; when the guest binds to the free cage, the coumarin/cage equilibrium is shifted to restore the unbound

cage that was lost in complexation with the competing guest, thus incidentally forcing the coumarin out of the cage.
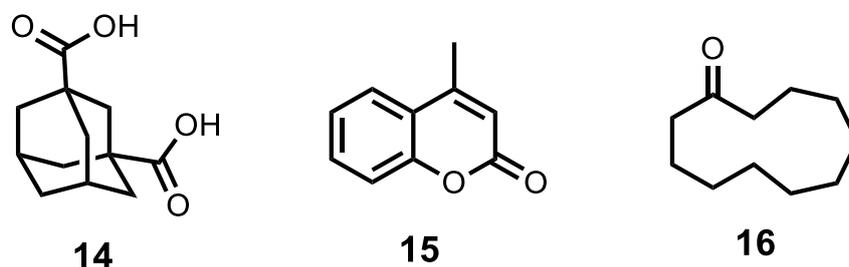
Figure 3.10: 1,3-Adamantane dicarboxylic acid **14**, 4-methylcoumarin **15** and cycloundecanone **16**

Using this assay, it has been possible to obtain binding constants for some guests (**figure 3.10**) that have binding affinities that are too strong to measure by NMR spectroscopy. The results are in **table 3.1**.

Compound	K / M^{-1}
14	2.3(1) x 10^5
15	1.1(4) x 10^5
16	1.20(7) x 10^6

Table 3.1: Binding constants of compounds **14**, **15** and **16** obtained using the fluorescence displacement assay

The displacement assay was then used on a fluorescence plate reader (as opposed to a standard cuvette in a fluorescence spectrometer) to allow for a high throughput titration method to be used (see Chapter 4 for more details on this).

3.2.4 Investigation of guest size

Now that we have fluorescence on our side, we decided to utilise this method for obtaining binding constants to test the limits of guest binding in the M_8L_{12} cubic cage in water. For this a large number of cyclic ketones were used, from a 5 carbon cyclic ketone to a 14 carbon cyclic ketone (**figure 3.11**). These guests were chosen for their large hydrophobic groups to take advantage of the hydrophobic effect in water to aid binding strength, and also for their ketone group to aid solubility and to allow for hydrogen bonding to occur inside the cage cavity with the convergent methylene H-bond donors.
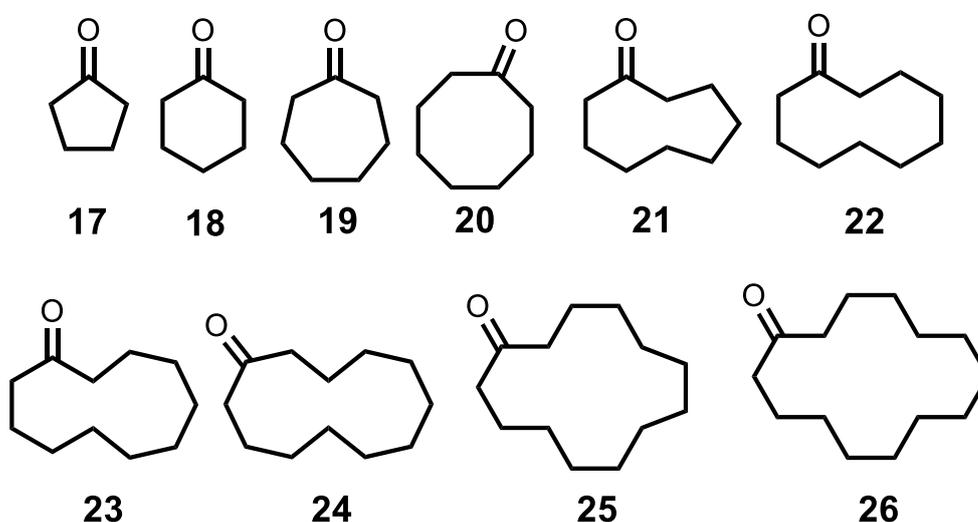
Figure 3.11: Cyclic alkane ketone guests containing 5 – 14 carbon atoms

Each guest was titrated with cage in water using the fluorescence displacement assay to obtain binding affinities. Guest length, surface area and volume were also calculated. The binding constants for **17/18** were too small to measure accurately by fluorescence spectroscopy and so NMR spectroscopy was used instead (**table 3.2**).

Guest	No. of carbon atoms	K / M^{-1}	$-\Delta G^{o}$ / KJ Mol^{-1}	Length / Å	Surface area / Å^2	Volume / Å^3
17	5	13(2)	6.3(4)	4.29	96.16	117.40
18	6	56(7)	9.8(6)	5.08	115.03	133.42
19	7	420(40)	15.0(3)	5.14	132.81	150.30
20	8	$2.1(5) \times 10^3$	19.0(6)	5.58	149.56	164.92
21	9	$1.1(3) \times 10^4$	23.1(7)	5.96	169.13	198.34
22	10	$1.4(6) \times 10^5$	29.4(11)	6.57	186.01	212.26
23	11	$1.2(7) \times 10^6$	34.69(2)	7.36	204.60	237.90
24	12	$1.5(2) \times 10^5$	29.5(3)	7.99	224.14	255.40
25	13	$1.9(5) \times 10^4$	24.4(7)	8.58	242.38	182.91
26	14	insoluble	-	9.13	260.81	277.73

Table 3.2: Binding results of guests **17** to **26** along with the corresponding length, surface area and volume of each guest

From looking at these results a few trends are immediately obvious. The most obvious is the large increase in binding affinity with increased size of guest (**figure 3.12**). This is not so unexpected since the larger the guest, the more hydrophobic surface area it contains and so there is a greater contribution to the binding strength. This increase continues up to cycloundecanone **23** (N_C = 11; K = 1.2 x 10^6 M^{-1}), then drops by an order of magnitude for cyclododecanone **24** (N_C = 12; K = 1.5 x 10^5 M^{-1}) and again by another order of magnitude for the cyclotridecanone **25** (N_C = 13; K = 1.9 x 10^4 M^{-1}). Molecular volume calculations indicate that the strongest-binding guest **23** has a volume of 205 Å^3, equivalent to 50% of the host cavity volume, whereas **24** has a volume of 224 Å^3, equivalent to 55% of the cavity volume. Based simply on volume, therefore, **24** is the guest that best matches Rebek's 55% rule.[1] The lower binding affinity of **24** compared to **23** suggests that the disc-shaped nature of guest **24** is not an ideal match for the more spherical cavity of the cage: we provide structural evidence to support this point later.
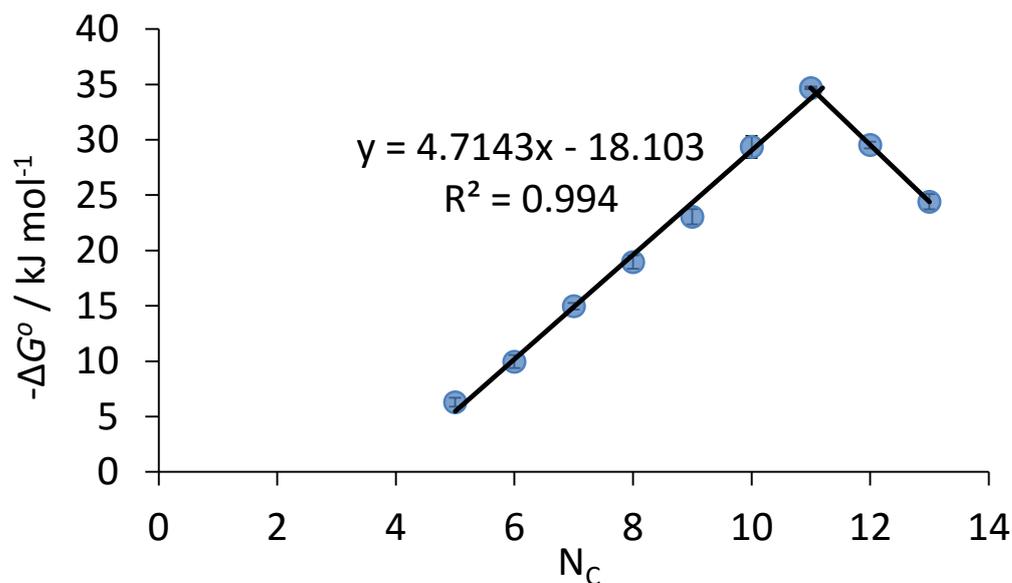
Figure 3.12: Plot of -ΔG for binding against the number of carbon atoms (N_c) for guests **17** to **25**

The relationship between the free energy change on complexation, $\Delta G°$, and the number of carbons atoms, N_C, for the series of ketones is shown in **figure 3.11**. From **17** to **23**, $\Delta G°$ is a linear function of N_C ($R^2 > 0.99$). There is a fairly systematic increase of on average 5 KJ Mol^{-1} per additional CH_2 group. For guests **24 – 25**, we see a reversal of the previous trend with a steady decrease in binding strength as the guests become too large. There are only three points in this series (**26** was too insoluble for the assay), but the steady decrease in $\Delta G°$ with increasing steric bulk is clear once the capacity of the host is exceeded.

According to previous work published on the hydrophobic effect, the addition of an extra CH_2 group to hexadecane in water destabilises it by about 3.5 kJ mol^{-1} (desolvation of the guest when removed from water to bind within the cavity provides an extra 3.5 kJ mol^{-1}).[4] If the CH_2 makes an optimal contact with the cavities interior surface then the equivalent hydrophobic surface area in the host would also be desolvated and so the maximum contribution would be around 7 kJ mol^{-1} (i.e. double). Similarly, Fersht and co-workers showed that burial of methylene chains on protein folding increased stability by 6.5 kJ mol^{-1} per CH_2 group,[5] and in a separate study using a range of alanine to glycine mutations at different positions on a protein they derived a coefficient of 4.1 kJ mol^{-1} for change in stability associated with change in hydrophobic surface area.[6] The increase of 5

kJ mol^{-1} per additional CH_2 group observed in our system is consistent with these results. It may be lower than the 7 kJ mol^{-1} maximum contribution because the cavity of our host is not entirely hydrophobic due to the presence of two H-bond donor sites in the cavity.

We can consider the guests as being composed of a non-polar (hydrocarbon) region and a polar (carbonyl) region. The surface areas of the guests can be described by **Eq. 1**, which is the best-fit straight line to the graph of SA *vs.* $N(CH_2)$, the number of CH_2 groups in the guest (**figure 3.13**).

$$SA / Å^2 = 54 + 16\ N(CH_2) \qquad \textbf{(Eq. 1)}$$

The constant, 54 Å^2, represents the surface area of the carbonyl group, and each CH_2 group adds a surface area of 16 Å^2.
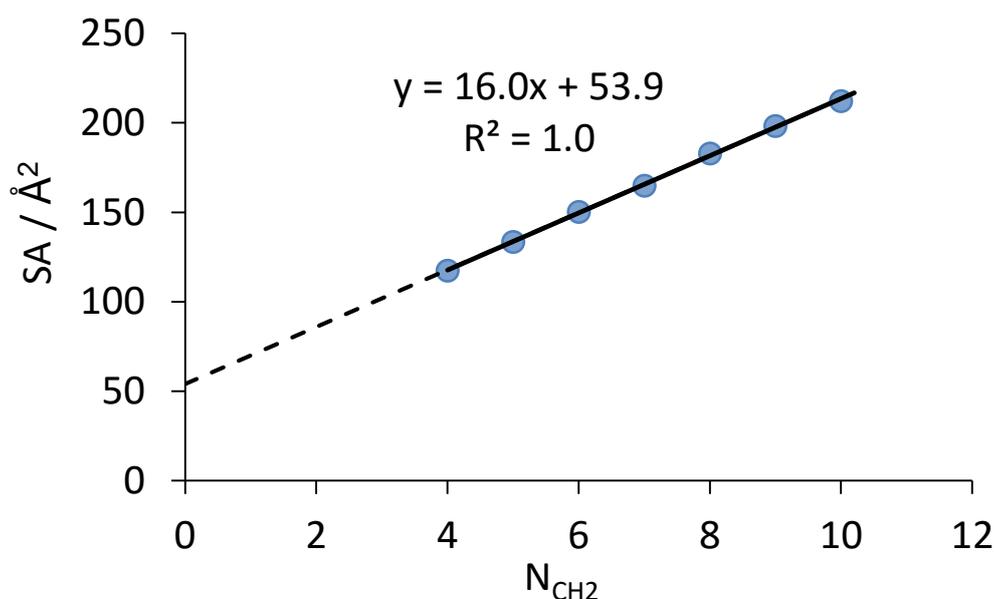
Figure 3.13: Plot of surface area (SA) against number of CH_2 groups for guests **17** to **23**

We can use **Eq. 1** to define the total surface area of the CH_2 groups, SA_{CH2} in **Eq. 2**.

$$SA_{CH2} / Å^2 = SA_{Total} - 54 \qquad \textbf{(Eq. 2)}$$

By plotting this new hydrophobic surface area (SA_{CH2}) from **Eq. 2** with $\Delta G°$ of binding (**figure 3.14**), we get **Eq. 3**.

$$\Delta G° / kJ\ mol^{-1} = +13 - 0.3\ SA(CH_2) / Å^2 \qquad \textbf{(Eq. 3)}$$
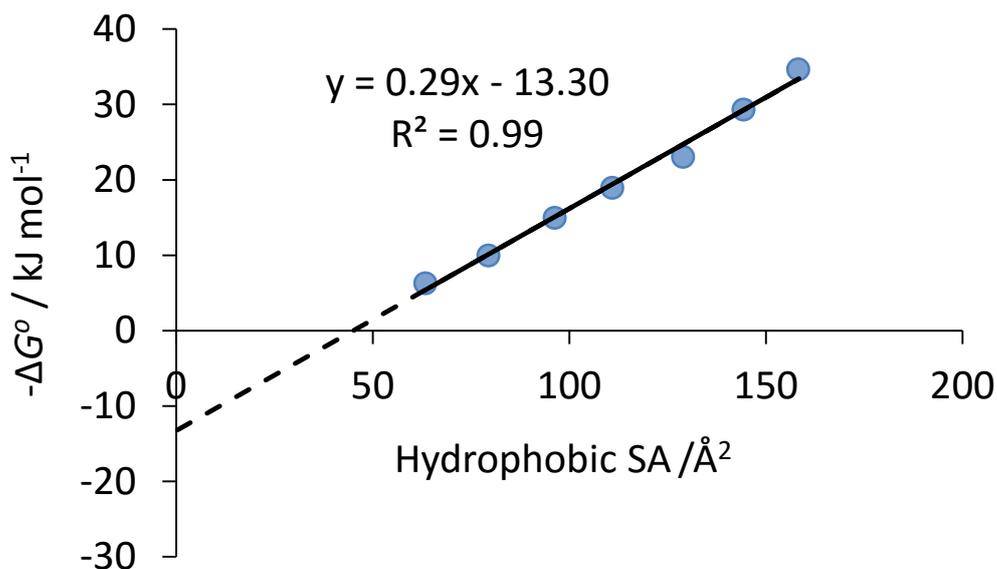
Figure 3.14: Plot of -$\Delta G°$ of binding against the hydrophobic SA for guests **17** to **23**

The first term is a constant unfavourable contribution of 13 kJ mol^{-1} to the change in free energy on complexation due to binding of the polar carbonyl group. The unfavourable free energy change associated with formation of a bimolecular complex in solution is 6 kJ mol^{-1},[7,8] which implies that binding of the carbonyl group in the cage is associated with an unfavourable free energy change of 7 kJ mol^{-1}. This adverse free energy change reflects the thermodynamic cost of desolvation of the carbonyl oxygen on removal from water, which is not fully compensated by formation of weaker interactions with the CH groups in the polar binding site inside the cage.

The second term is the favourable contribution of 0.3 kJ mol^{-1} per Å^2 of hydrophobic surface area, which equates to around 5 kJ mol^{-1} per CH$_2$ group (SA = 16 Å^2).

We can use **Eq. 3** to predict the binding of similar guests very accurately (**figure 3.15**) and the results are outlines in **table 3.3** and **figure 3.16**.

27   **28**
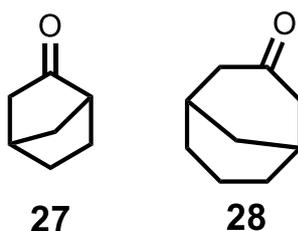
Figure 3.15: Guests used with **Eq. 3** to predict the $\Delta G°$ of binding based of hydrophobic surface area

Guest	No. of carbon atoms	K / M^{-1}	-ΔG^o / KJ Mol^{-1}	Predicted −ΔG^o / KJ Mol^{-1}	Surface area / Å^2	Hydrophobic Surface area / Å^2
27	7	130(30)	12.1(6)	11.7	123.22	84.45
28	9	4.0(16) x 10^3	20.5(1)	20.2	156.59	113.79

Table 3.3: Binding results binding for the two predicted guests **3.22** and **3.23** along with their surface area and hydrophobic surface area
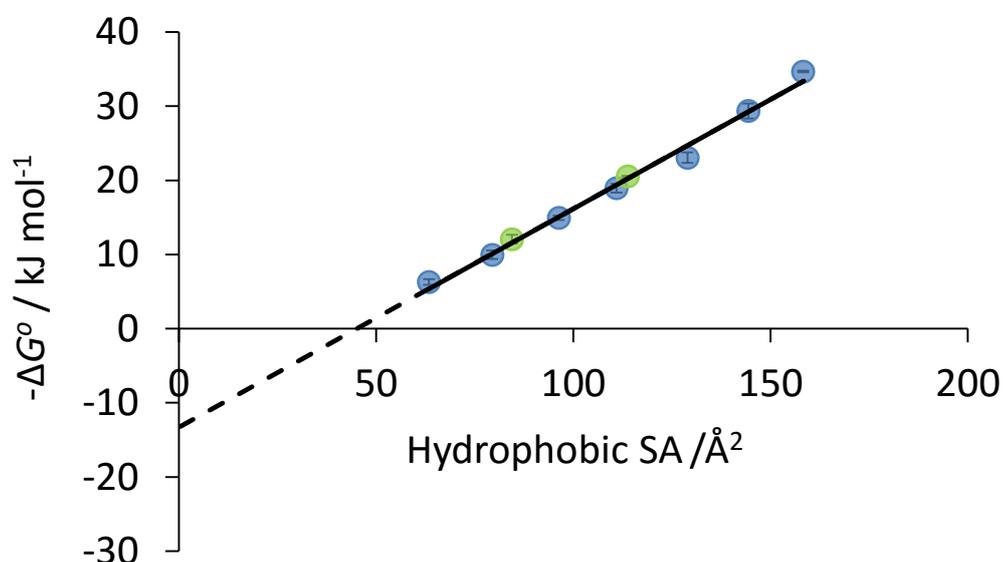
Figure 3.16: Plot of -ΔG° of binding for guests **17** to **23** (blue) and the measured ΔG° of binding for the two predicted guests **27** and **28** (red) against hydrophobic SA

3.2.5 Crystal structure of a cage-guest complex

By soaking the empty cage crystals in pure oil of guest **23** (cycloundecanone), the strongest binding guest in the series, X-ray quality crystals of the host-guest complex were obtained (**figure 3.17**). This is a method similarly used in metal organic frameworks (MOFs) to obtain crystal structures of guests encapsulated in its pores.
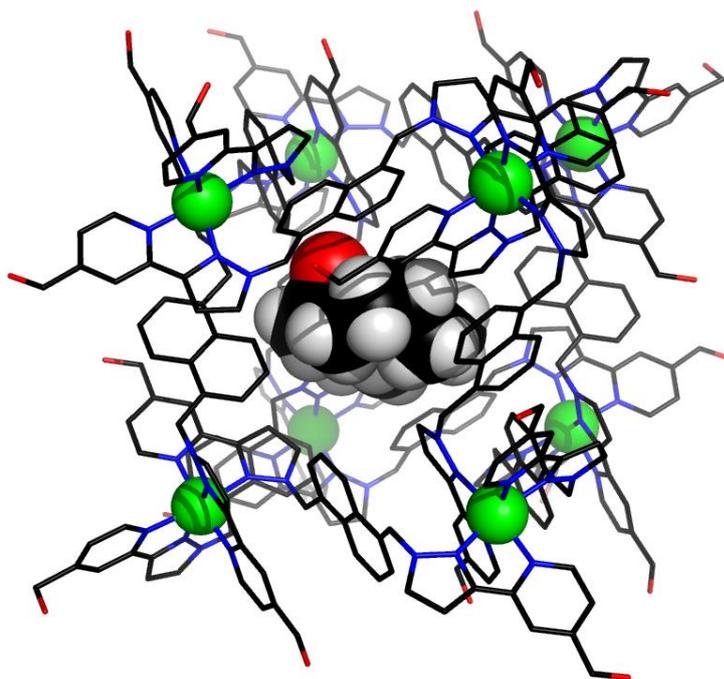
Figure 3.17: The X-ray crystal structure of the cage•**23** complex, [Co$_8$L$_{12}$•(**23**)](BF$_4$)$_{16}$: the [Co$_8$L$_{12}$]$^{16+}$ complex cation is shown in wireframe mode, with the guest shown in space-filling mode.

The cage framework is unaffected by the presence of the guest, which is disordered over two symmetrically equivalent orientations (only one shown). The oxygen atom of the guest carbonyl group was straightforward to identify in the electron density map due to polar interactions with well-defined sites on the interior surface of the cage which hold it in place (**figure 3.18**), but the carbon atoms required geometric restraints to allow a reasonable model to be constructed which gave a stable refinement.

The disc-shaped guest is located centrally in the cavity with the carbonyl group projected towards one of the two regions of high positive electrostatic potential associated with the *fac* tris-chelate sites, which lie at either end of the long diagonal of the cube. The carbonyl oxygen makes short contacts of between 2.54 and 3.06 Å with four CH protons in this pocket (**figure 3.18**). The Co•••O separation in this site is 5.72 Å, which is similar to that observed involving solvent molecules which occupy this position in other crystal structures. These polar interactions, although they may not contribute to the guest binding affinity in water, are important in determining the orientation of the guest.
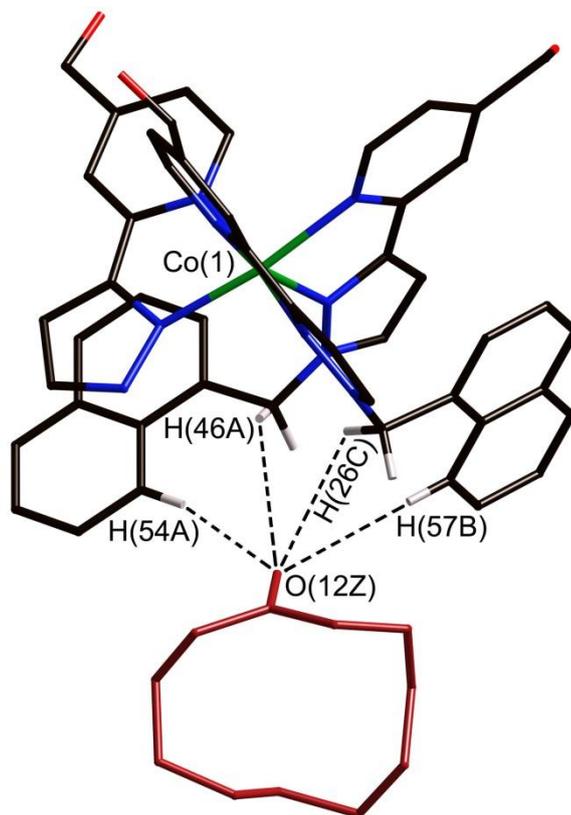
Figure 3.18: A close-up view from the crystal structure of the closest contacts between the carbonyl group of **23** and some of the naphthyl and methylene CH protons of the cage

Looking at a slice through the centre of the structure (**figure 3.19**), in a space-filling mode with the guest shown in green for clarity, it can be seen that the guest does not completely fill the cavity of the cage; it makes contact with the cage surface around its 'equator' but there is space on either side of the guest disc where contact with the cage is poorer. This mismatch between the shape of the pseudo-spherical cavity and the disc-shaped guest explains why the volume of the highest affinity guest in this series is slightly smaller than expected on the basis of Rebek's 55% rule.
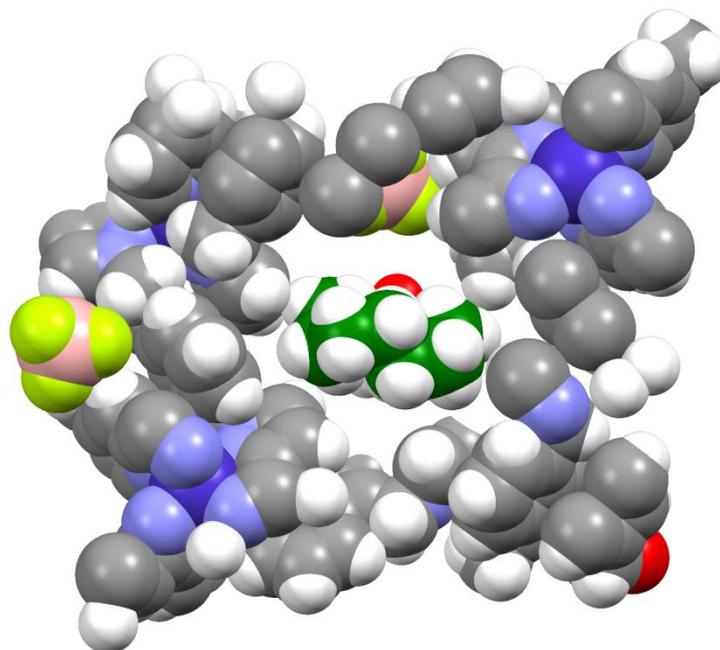
Figure 3.19: A slice through a space-filling model of the crystal structure of
[Co$_8$L$_{12}$•(**23**)](BF$_4$)$_{16}$ illustrating the extent to which guest **23** fills the cavity

The conformation of **23** bound inside the cage is very similar to the minimum-energy conformation calculated for **23** *in vacuo* (**figure 3.20**), indicating that **23** is almost perfectly preorganised for guest binding.
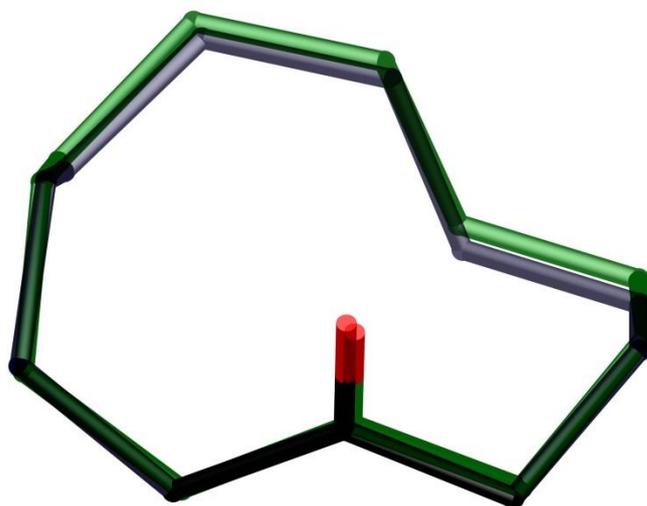
Figure 3.20: Superposition the calculated minimum-energy conformation of **23** *in vacuo* (green bonds) and the observed conformation in the X-ray crystal structure of the cage•**23** complex (dark blue bonds).

The soaking method for guest encapsulation in the crystal works due to large channels caused by packing in the crystal. Each cage makes 4 hydrogen bonds to another in all

directions. This causes the cavities to align creating a channel for easy diffusion of guest, even in the solid state (**figure 3.21**).
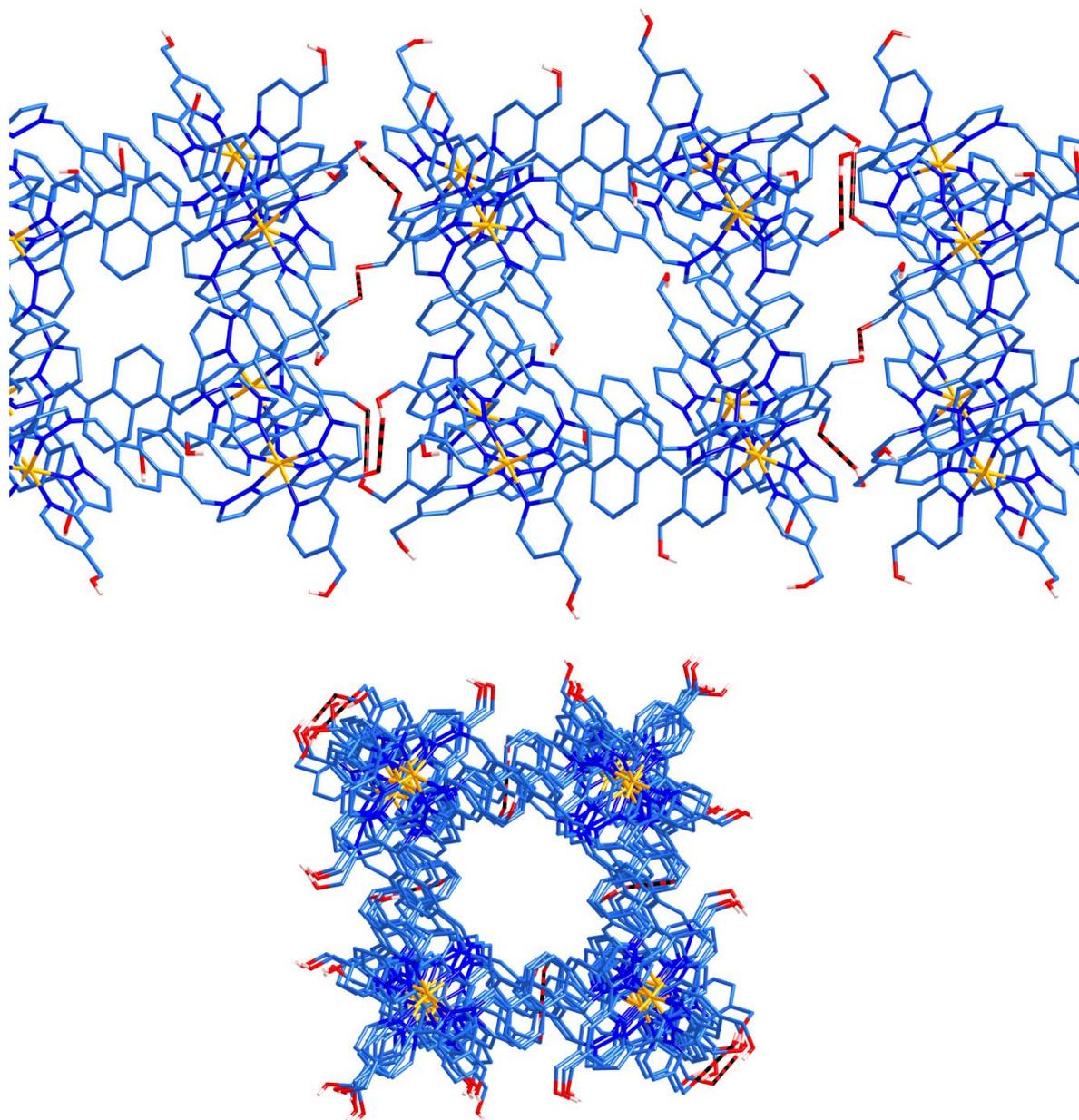
Figure 3.21: views of the packing of the M_8L_{12} water soluble cube showing (i) the four hydrogen bonds (black and red lines) between each cage (top); and (ii) looking down on of the channels (bottom)

3.2.6 Investigation of guest flexibility, rigidity and shape

A second set of guests was used (**figure 3.22**) similar to that just described, i.e. containing a ketone group and a hydrophobic section, however this set of guests have very similar molecular weight, surface area and volume (all contain 10 carbon atoms and one ketone). They differ in length with a range of 12.08 Å to 5.45 Å and therefore flexibility ranging from a very compact ridged tetracyclic adamantanone **32**, to the very long and flexible linear ketones **36** and **37**.
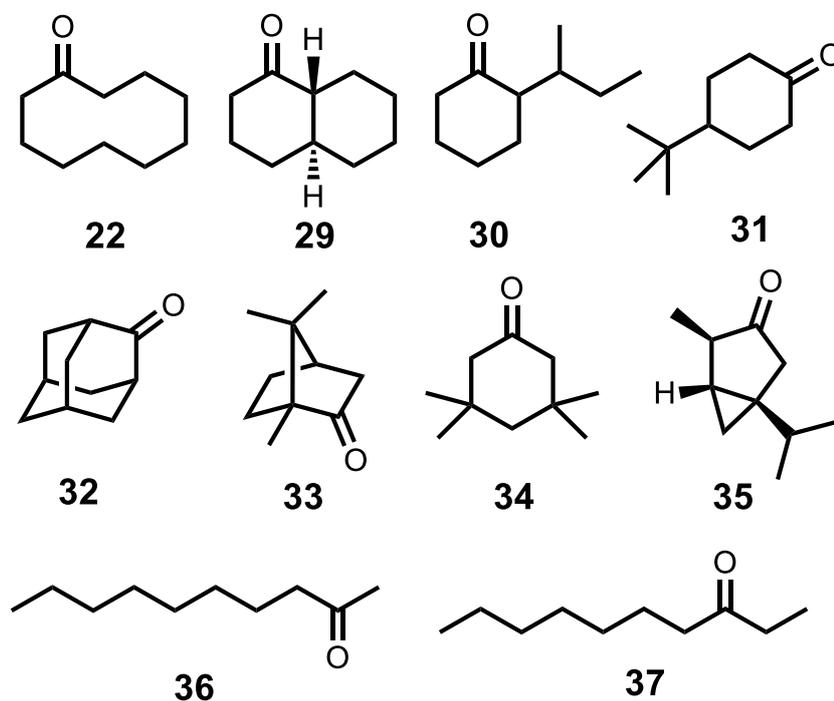
Figure 3.22: 10 carbon ketone guests

Each guest was titrated with cage in water using the fluorescence displacement assay to obtain binding affinities. Guest length, surface area, volume and number of conformations each guest can adopt (in a 50 KJ Mol^{-1} window from the lowest energy conformation) were also calculated. (**Table 3.4, figure 3.23**).

Guest	No. rings	No. confs	K / M^{-1}	-ΔG^o / KJ Mol^{-1}	Length / Å	Surface area / Å^2	Volume / Å^3
22	1	73	1.4(6) x 10^5	29.4(11)	6.570	174	170
29	2	7	9.5(10) x 10^3	22.7(3)	7.226	170	160
30	1	58	1.6(1) x 10^4	24.0(2)	7.913	183	170
31	1	5	8.7(20) x 10^3	22.5(6)	7.459	179	170
32	4	1	1.9(1) x 10^4	24.41(13)	5.445	154	146
33	3	1	1.8(3) x 10^5	30.0(4)	5.673	162	165
34	1	6	7.5(20) x 10^4	27.8(7)	6.830	176	172
35	2	4	2.0(1) x 10^4	24.54(12)	7.308	174	167
36	0	1167	Nb	-	13.08	215	179
37	0	1647	Nb	-	13.08	215	180

Table 3.4: Binding results of guests **22, 29** to **37** along with the corresponding length, surface area, volume and number of conformations
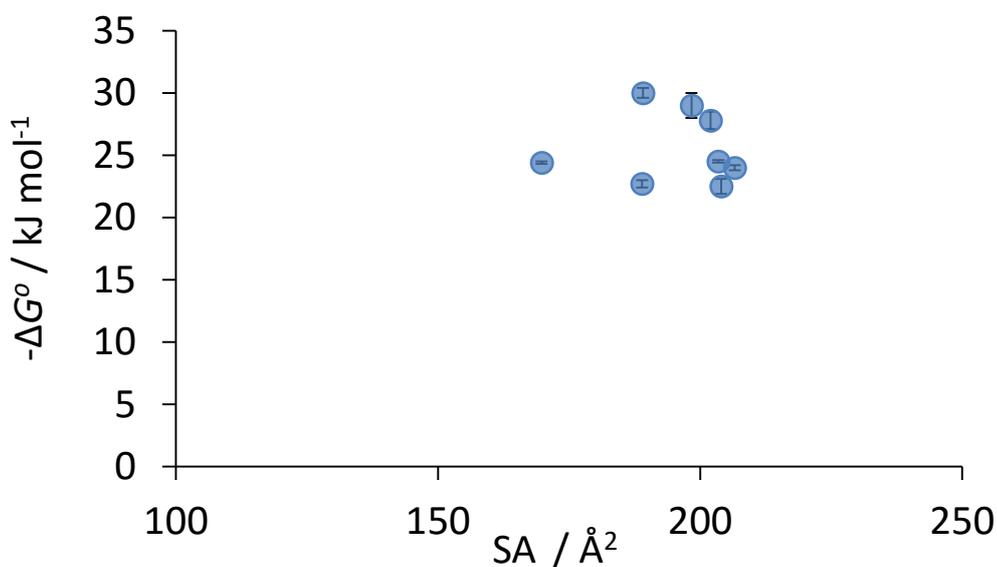
Figure 3.23: Plot of -$\Delta G°$ of binding against surface area for guests **22, 29** to **37**

Unlike the previous set of guests (homologous series of cyclic ketones), there is no clear correlation, so the guests of this set will be looked at in smaller groups.

The first immediate observation that can be easily made is that the long linear ketones **36** and **37** do not bind at all yet all other guests in this series bind reasonably strongly with K values around 10^3 - 10^4 M^{-1}. In order to help describe this, energy minimised

conformational searches were run on each guest to determine the number of conformers that each guest can adopt that is within a 50 kJ Mol⁻¹ energy window of the lowest energy conformer. The linear ketones **36** and **37** have over 1000 conformers. Once the first ring is formed (cyclodecanone **22**) the number of conformers drops significantly to just under 100 which is a reduction of >90 % in number of conformers. The entropy penalty for reorganisation to bind within the cavity will therefore be very large for the linear ketones with its larger number of conformers, however this penalty will largely disappear once the first ring is made (the guest is much more preorganised), hence cyclodecanone **22** has a binding strength of around 29 kJ Mol⁻¹. There is also an enthalpic penalty for these linear ketones too; in their lowest energy conformation they are extended with all of the carbon-carbon bonds in a staggered conformation, and these bonds must adopt high energy *gauche* conformations to fold the molecules up into compact structures that will fit inside the cage.

22   **29**
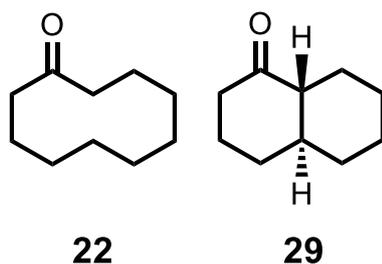
Figure 3.24: Cyclodecanone **22** and trans-1-decalone **29**

If we compare cyclodecanone **22** and trans-1-decalone **29** (**figure 3.24**), the two guests are structurally similar but **29** contains an extra C-C bond which results in rigidification and a small decrease in surface area and volume: the decrease in surface area between **22** and **29** is just 4 Å². However the decrease in binding free energy for the smaller guest **29** is ≈ 7 kJ mol⁻¹ which is much too large to explain based on the hydrophobic effect alone. The conclusion here is that the relative rigidity of **29** is starting to cause steric problems for binding, whereas the more flexible **22** can better adjust to the cavity's steric constraints as it can adopt a disc-shaped conformation similar to that seen for **23**.
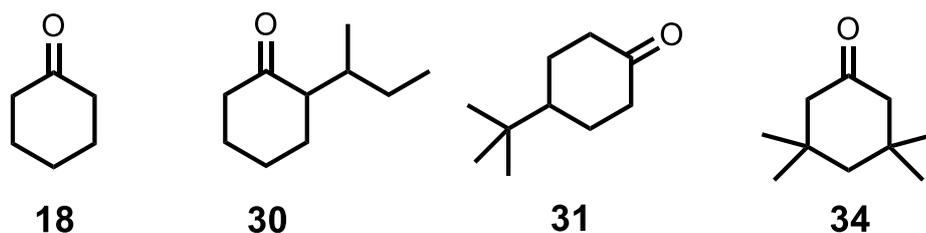
Figure 3.25: Guests cyclohexanone **18**, 2-sec-butylcyclohexanone **30**, 4-^{t}butylcyclohexanone **31** and 3,3,5,5-tetramethylcyclohexanone **34**

Another observation that can be seen is a comparison with the previously discussed guest **18** with the three isomeric analogues **30**, **31** and **34** (**figure 3.25**). Whilst **18** binds rather weakly, **30**, **31** and **34** all bind relatively strongly, due to the increased hydrophobic surface area of around 60 Å^2 for each. Despite this the origin of the differences between the isomers **30**, **31** and **34** is not immediately obvious as the guest with the largest surface area **30** is not the strongest binder and they all have similar volumes (around 170 Å^3). It is possible that since the limits of size for optimum binding is being reached, small steric changes in certain positions have large impacts on binding strengths, i.e. the more symmetric shape of **34** is favoured by the highly symmetric cavity of the cage more than the more asymmetric shapes of **30** and **31**.
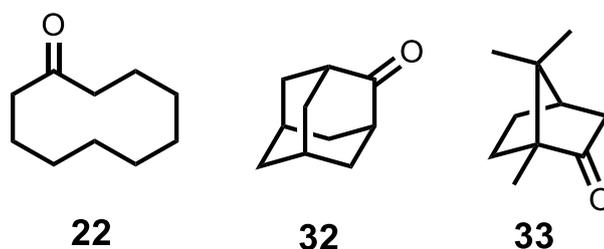
Figure 3.26: Guests cyclodecanone **22**, adamantanone **32** and camphor **33**

The strongest binding guest in this series (**figure 3.26**) is camphor **33**. Despite having a smaller surface area than cyclodecanone **22** by 12 Å^2 it binds slightly more strongly. This is most probably due to **33** being more spherical in shape which is a better match for the spherically shaped cavity. Adamantanone, **32**, binds an order of magnitude weaker than camphor **33** despite it being the most rigid and preorganised structure. It does however have the smallest surface area of the guests and its rigidity means if it is not a perfect fit for the cavity then there is no possibility for it to change to a shape that does.

3.3 Conclusions

A fluorescence displacement assay has been developed by using the fluorescent guest 7-amino-4-methyl coumarin **12** as the fluorophore. This works well allowing for fast titrations and quick screening of guests to be run with reasonably high accuracy.

This displacement assay has been the used to probe the limits of the cubic M_8L_{12} cage in water by undertaking titrations with a range of guests with different size, shape and flexibility of guest molecules with a high throughput and fast method using a plate reader.

Using this method our strongest binding guest so far (cycloundecanone; $K \approx 10^6$ M^{-1}) has been found and our understanding of the types of guests that bind stronger and their interactions that govern this has been greatly enhanced. Mainly there is a strong correlation of guest binding with hydrophobic surface area until a point is reached where the guest is too large.

3.4 Experimental techniques and procedures

3.4.1 Measurements and calculations

Chemicals

All chemicals were purchased from commercial sources and used as supplied unless otherwise stated.

The host cage $[Co_8(L^{1,5\text{-}naphOH})_{12}](BF_4)_{16}$ was prepared according the published method.[9]

Nuclear magnetic resonance spectra

All NMR data were collected using a Bruker Avance III 400 MHz NMR spectrometer at 298 K with the parameters: 3072 scans, 284 ppm sweep width, O1p value of 0 ppm, D1 0.5 s, aq 0.5 s. The NMR data was processed using Bruker Topspin 3.1

Fluorescence spectra

Fluorescence data was collect either using a Horbia Jobin Yvon Fluoromax-4 spectrofluorometer and a quartz cuvette of 1 cm path length. Or using a BMG FLUOstar Omega plate reader and either a 300 µl Hellma 96 well quartz microplate or a 100 µl Griener Bio-one µClear black 384 well plate. All spectra were collected at 298 K.

Fluorescence titration with 7-amino-4-methylcoumarin, 12

A solution of **12** was made up in water (1×10^{-5} M), and 2000 µl of this was placed in a quartz cuvette. A solution of cage (1 mM) was made up using the stock solution of **12**. 1 to 100 µl portion of this cage solution was pipetted in to the cuvette containing **12**. The cuvette was mixed 20 times, all bubbles removed and then left for 5 minutes to equilibrate. After equilibration, the fluorescence spectrum was run using slit widths of 2 nm, excitation wavelength of 400 nm and data were collected data between 405 nm and 600 nm. The addition of cage, equilibrating and measuring was repeated until the fluorescence intensity reached a constant value.

Using Microsoft Excel, the concentration of cage was plotted against fluorescence intensity and the resulting curve fitted to produce a value for the binding constant, K.

Fluorescence displacement titration with a guest

The host solution was made up of cage (5.5×10^{-5} M) and coumarin **12** (1×10^{-5} M) with water.

The guest solution was made up of guest (10^{-1} mM) using the host solution.

1 to 100 µl portion of guest solution was pipetted in to the cuvette containing cage (2000 µl). The cuvette was mixed 20 times, all bubbles removed and then left for 5 minutes to equilibrate. After equilibration, the fluorescence spectrum was ran using slit widths of 2 nm, excitation wavelength of 400 nm and collected data between 405 nm and 600 nm. The addition of cage, equilibrating and measuring was repeated until the fluorescence intensity reached a constant value.

Using Microsoft Excel, the concentration of guest was plotted against fluorescent intensity and the resulting curve fitted (using an algorithm to take into account the second equilibrium between cage and fluorophore) to produce a value for the binding constant, K.

Fluorescence titrations using the plate reader

The host solution was made up of cage (5.5×10^{-5} M) and coumarin **12** (1×10^{-5} M) with water.

The guest solution was made up of guest (1 - 10 mM) using the host solution.

To each well, different amounts of host and guest were added to a total volume of 300 µl (96 well plate, for a 384 well plate 100 µl) from all host solution to all guest solution using 12-24 wells per titration. The plate was then heated to 35 °C for 20 minutes to allow for mixing to occur, and then cooled to 25 °C and equilibrated for 20 minutes before the fluorescence emission at 450 nm (using a 400 nm excitation wavelength and the receiver gain was set to a well containing free coumarin **12**) of each well was measured.

Each titration was repeated twice and by using Microsoft Excel, the concentration of guest was plotted against fluorescent intensity and the resulting curve fitted (using an algorithm

to take into account the second equilibrium between cage and fluorophore) to produce a value for the binding constant, *K*.

Molecular modelling

The molecular modelling was done using MacroModel[10]. The crystal structure of the $[Co_8L_{12}][BF_4]_{16}$ cage was imported and the guest was positioned inside its cavity. The cage structure was frozen we assumed that it will undergo no changes when a guest is present in the cavity. The guest was energy minimised using a molecular mechanics energy minimisation (molecular mechanics force fields (MMFFs)), and conformational search (in a 50 kJ mol^{-1} window) using a Monte Carlo molecular modelling (MCMM) conformational search.

Inter-atom distances were also measured using Macromodel, Molecular volumes and surface areas were calculated from the 0.002 Bohr Å^{-3} isodensity surface from B3LYP 6-31G* DFT calculations implemented in Spartan.[11]

3.4.2 X-ray crystallography

The crystal structure data collection of the complex $[Co_8L_{12} \bullet (\textbf{23})](BF_4)_{16}$ was performed at the EPSRC National Crystallography Service at the University of Southampton, UK. The structure was solved and refined using the SHELX suite of programs. The asymmetric unit contains one half of the cage complex which lies astride an inversion centre, as well as one complete guest molecule whose atoms all have site occupancies of 0.5. Thus, the complete complex contains one guest molecule disordered over 2 symmetrically equivalent (and spatially overlapping) orientations with the O atom pointing towards diagonally opposite corners Co(1) and Co(1A). The usual disorder of anions / solvent molecules and solvent loss characteristic of cage complexes of this type resulted in weak scattering, necessitating use of extensive geometric and displacement restraints to keep the refinement stable: these are described in detail in the CIF. We could locate and refine five of the expected eight $[BF_4]^-$ anions in the asymmetric unit. The presence of large regions of diffuse electron density which could not be modelled, accounting for the remaining anions plus solvent molecules, required use of the 'Solvent Mask' function in the OLEX-2 software package. The thermal displacement parameters of the atoms of the guest molecule in the cage cavity are larger than those of the rest of the cage structure. This could arise from unresolved positional disorder, or from the fact that the fraction of cage cavities occupied by guest molecules is less than 100%. We have assumed the former explanation and left the site occupancies at 0.5 for each disordered component for the final refinement, *i.e.* one complete guest molecule per host cage. Overall the final $R1$ value of 18.7% is typical of cage structures of this type and is sufficient to establish the identity and connectivity of the complex.

Crystallography data table

Complex	$[Co_8L_{12}](BF_4)_{16} \bullet$(cycloundecanone)
Formula	$C_{371}H_{330}B_{16}Co_8F_{64}N_{72}O_{25}$
Molecular weight	8057. 5
T, K	100(2)
Crystal system	Monoclinic
Space group	*C2/c*
a, Å	27.5037(19)
b, Å	39.282(3)
c, Å	42.103(3)
$\alpha,°$	90
$\beta,°$	106.2580(10)
$\gamma,°$	90
V, Å^3	43440(5)
Z	4
ρ, g cm^{-3}	1.232
Crystal size, mm^3	0.13 x 0.13 x 0.05
Data, restraints, parameters	38209, 2424, 1855
Final *R*1, *wR2*^{b}	0.187, 0.489

a   These formulae (and consequently the crystal densities) are necessarily approximate given that large amounts of diffuse electron density in solvent-accessible voids was removed from the refinements using the OLEX 'Solvent Mask' function. See CIFs, and comments in experimental section, for details.

b   The value of *R*1 is based on 'observed' data with $I > 2\sigma(I)$; the value of w*R*2 is based on all data.

3.5 References

1) S. Mecozzi; J. Rebek, *Chem. Eur. J.,* **1998**, *4,* 1016-1022

2) J. L. Brumaghim; M. Michels; K. N. Raymond, *Eur. J. of Org. Chem*., **2004**, *22*, 4552-4559

3) M. M. J. Smulders; S. Zarra; J. R. Nitschke, *J. Am. Chem. Soc.,* **2013**, *135*, 7039-7046

4) C. A. Hunter, *Chem. Sci.,* **2013**, *4*, 834-848

5) L. Serrano; J. T. Kellis; P. Cann; A. Matouschek; A. R. Fersht, *J. Mol. Biol.,* **1992**, *224*, 783-804

6) L. Serrano; J Sancho; M. Hirshberg; A. R. Fersht, *J. Mol. Biol.,* **1992**, *227*, 544-599

7) C. A. Hunter, *Angew. Chem. Int. Ed.,* **2004**, *43*, 5310-5324

8) P. R. Andrews; D. J. Craik; J. L. Martin, *J. Med. Chem.,* **1984**, *27*, 1648-1657

9) M. Whitehead; S. Turega; A. Stephenson; C. A. Hunter; M. D. Ward, *Chem. Sci.,* **2013**, *4*, 2744-2751

10) MacroModel, version 9, Schrödinger Inc., New York, NY, 2011

11) Spartan'06, Wavefunction Inc., Irvine, CA, 2006

Chapter 4

Prediction of guest binding using molecular docking

4.1 Introduction

Currently there is a multitude of artificial host molecules ranging from simple 2-dimensional cycles, to 3-dimensional capsules, and even interlocked systems, that have shown the ability to bind guest molecules. Despite this, there has been no attempt to predict how well guest molecules might bind to these artificial hosts. Currently most binding studies on new host molecules are performed using trial and error methods, which is very laborious and takes time; or by using knowledge of a very similar host that has known binding which only works in limited cases.

For natural host molecules such as proteins and enzymes, the prediction of guest molecule binding is of great importance. If guests that can specifically target such hosts can be identified, they may be used as drugs. The drug discovery industry uses host-guest binding predictions to identify leads so the ability to predict which guests will bind and how strongly is potentially commercially important.[1-3]

4.1.1 Molecular docking

The software used for these predictions is molecular docking software. The prediction of guest binding is a complex task, particularly with the demand for more and more complex molecules and architectures to be investigated. Because of this, molecular docking is often a multi-step procedure, but can be considered as having two main parts: Posing and Scoring.[1-4]

Posing

Posing (**figure 4.1**) is the prediction of the orientation and conformation of the guest, made by the use of a docking algorithm. This is perhaps the most complex part of the docking procedure, since even small molecules can have a large degree of conformational freedom. This step has to be done accurately enough to give the structure that best matches the receptor site, but also fast enough to allow screening of huge libraries of potential guest molecules.
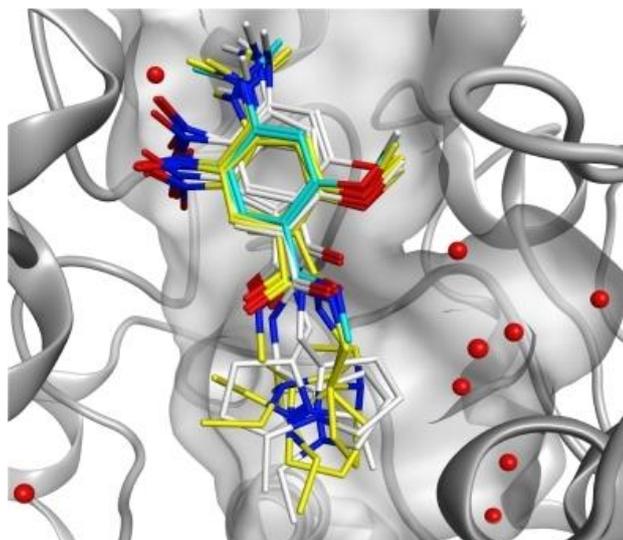
Figure 4.1: posing[5]

The *posing* of a guest is done using one of three methods: systematic, random/stochastic, and simulation methods.[1-3]

Systematic methods try to explore all the degrees of freedom in a molecule, usually by stepwise or incremental searches to avoid combinatorial explosions; or by the use of pre-generated libraries of conformations.

Random/stochastic methods use mainly Monte Carlo or genetic algorithms to make random changes to the guest, which is subsequently evaluated using a pre-defined probability function.

Simulation methods use molecular dynamics simulations. These methods often only find local minima on the energy surface due to time constraints with high-energy barriers. To overcome this issue, simulation methods are often accompanied by other search methods.

Posing of the host-guest complex does not require any calculations as to how strongly the guest may bind, this is dealt with in the second part of the process (scoring).

Scoring
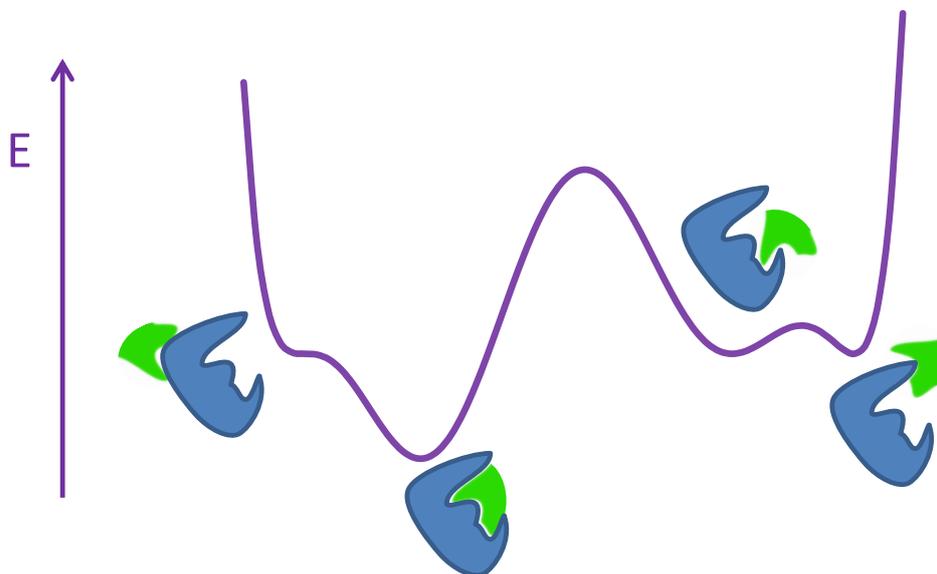
Figure 4.2: Finding the 'real' binding mode based of interaction energies with host and guest generated by posing step

The second part of the process is *scoring*, in which a scoring function is generated to estimate the interaction energy between the host and guest. This is also the step where the software tries to find the experimental (real) binding mode (**figure 4.2**) from those generated by the posing step. Early examples of this step were based upon approximate shape and electrostatic complementarities, however now this has been expanded to a more detailed treatment of electrostatic and van der Waals' interactions, and inclusion of at least some solvation or entropic effects. Despite the fact that guest binding is driven by a combination of enthalpic and entropic effects, often the entropic effects are ignored in the scoring since they are difficult to quantify.[1-3]

Scoring functions

Scoring functions are mathematical approximations for predicting the free energy of binding. They are split into three categories: empirical, force field based and knowledge based.[1-4]

Empirical scoring functions are the weighted sum of several intermolecular interaction terms. The weighting of each factor is 'trained' through regression analysis, in which

theoretical values are fitted to experimental data (aka a 'training set'). The different terms reflect the different types of interaction established between ligand and target, such as hydrogen bonding, ionic and van der Waals' interactions.

Force field based scoring functions use molecular mechanics force fields to quantify the sum of two energies, the host-guest interaction energy and internal ligand energy (such as steric strain).

Knowledge based scoring functions are used to reproduce experimental structures rather than binding energies. They are based on statistical observations of intermolecular contacts identified from large datasets of experimental 3D structures.

Scoring functions are often thought to be the part of modelling the docking process that results in failure to predict the binding. Thus *posing* may work well but numerical estimation of binding strength fails.

4.1.2 GOLD

The docking software chosen to be used as the basis for the work in this chapter is GOLD (Genetic Optimisation of Ligand Docking). It uses random/stochastic methods with a genetic algorithm for the posing step, and mostly empirical based scoring functions for the scoring step. Its current most 'accurate' scoring function is 'CHEMPLP' (**Eq. 1**), which is a combination of a piecewise linear potential (PLP) that takes into account steric complementarity between host and guest, and GOLD's previous scoring function 'Chemscore' that takes into account angle and geometrical terms.[6-9] The individual terms are:

- ligand_clash – Steric clashes with the host and guest
- part_buried – The burial of a polar group in a non-polar environment
- non-polar – Hydrophobic interactions
- ligand_torsion – The torsional strain induced in the ligand (guest) on binding
- H-bond_donor / H-bond_acceptor – Hydrogen bonding terms
- metal_coordination – Interactions of ligands (guests) with metal ions in the receptor

CHEMPLP score = $w_{lc} \bullet f$(ligand_clash) + $w_{pb} \bullet f$(part_buried) + $w_{np} \bullet f$(non-polar) +

$w_{lt} \bullet f$(ligand_torsion) + $w_{mc} \bullet f$(metal_coordination) + $w_{hbd} \bullet f$(H-bond_donor)

$+ w_{hba} \bullet f$(H-bond_acceptor);

Eq. 1

where w_i are the weightings of each function, f.[9]

4.2 Results and Discussion

4.2.1 The training set

At the start of this work we had amassed binding data for 54 guest molecules in water (**figure 4.3; appendix 2**), which provided the basis of how well we can predict the binding of the guests to the cage. The data is from Chapters 3 and 5 in this thesis, as well as from previously published work.[10-12]
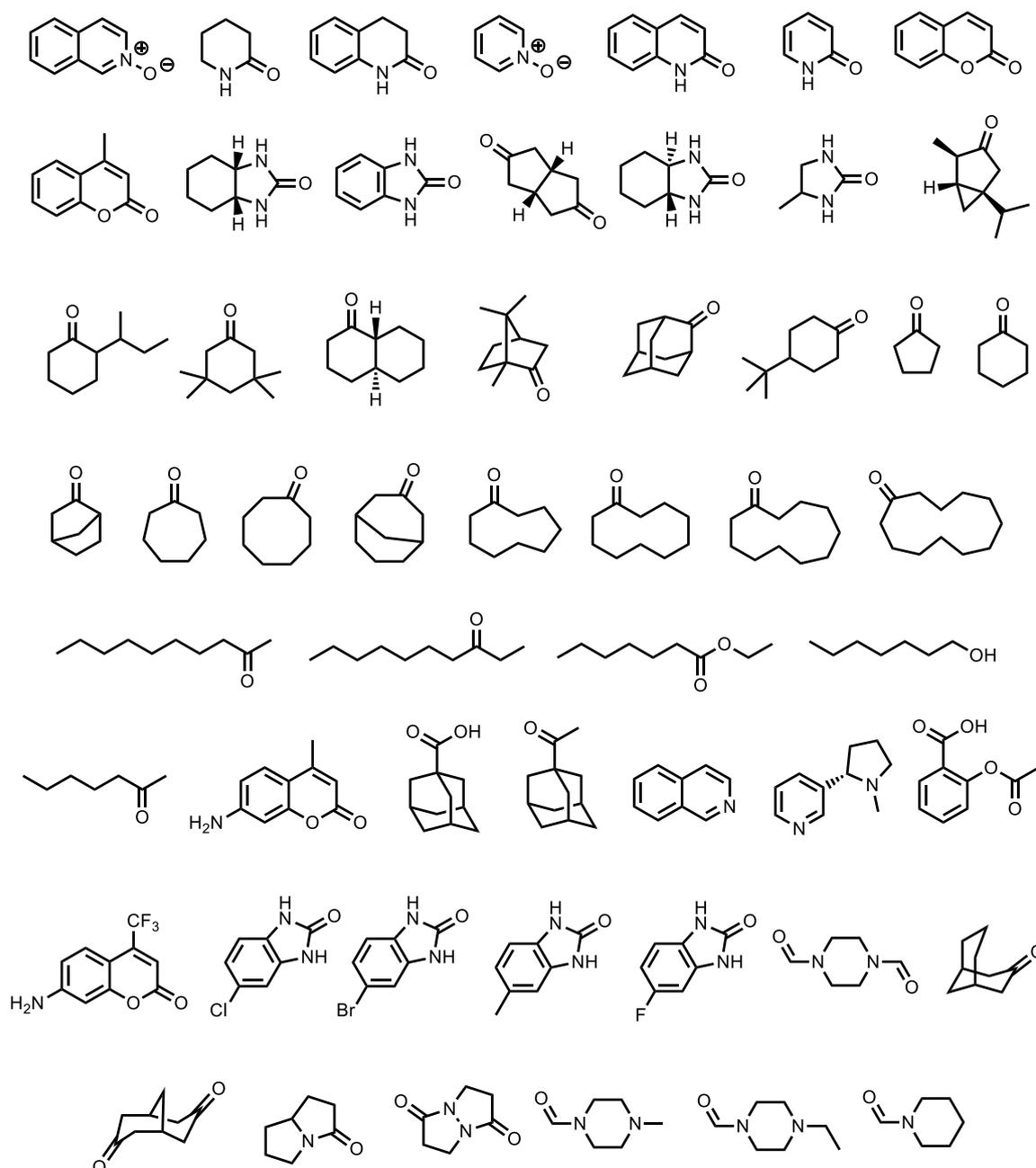
Figure 4.3: The training set of 54 molecules with known binding constants in water

We initially wanted to see how our measured data matched the binding strength predicted by the standard CHEMPLP scoring function (outlined in the introduction to this chapter). The fit is shown in **figure 4.4**. It is immediately obvious there is very little correlation ($r^2 = 0.02$) between our measured and calculated values. In particular, molecules known to bind weakly ($\log K < 0$) were predicted to bind strongly using this scoring function, which actually gave little difference in predicted binding across the whole series. This is not surprising since GOLD uses empirical based scoring functions and so needs to be 'trained' alongside our data to get an accurate fit for our system.
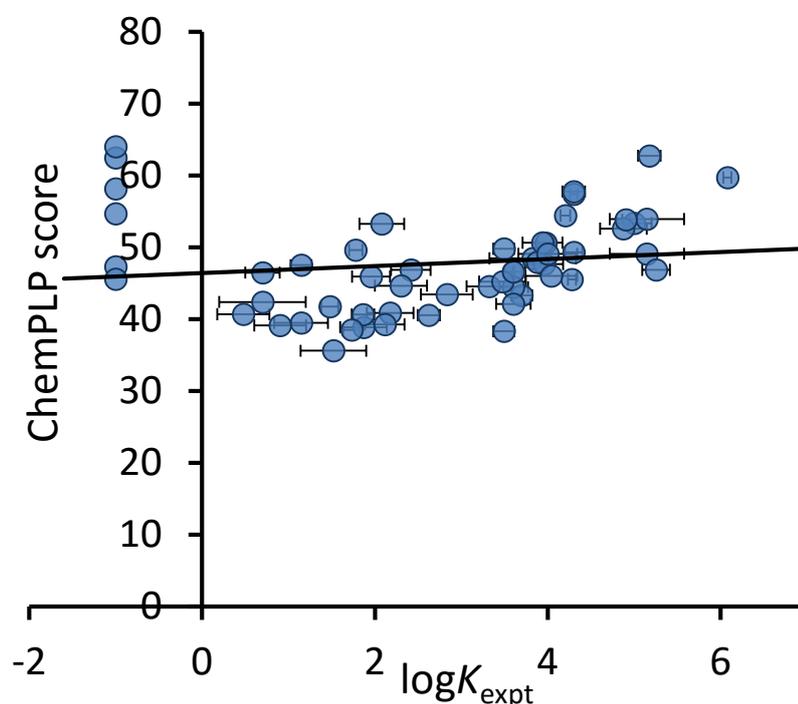
Figure 4.4: Plot of the score from GOLD's CHEMPLP scoring function, with our experimental binding data ($\log K_{expt}$) showing the line of best fit.

4.2.2 Training the scoring function

In order to 'train' the scoring function for our system, we did a non-linear least squares regression between the calculated values ($\log K_{calc}$) and the experimental values ($\log K_{expt}$), letting the weightings of each interaction in the CHEMPLP score (w_i in **Eq. 1**) vary. This allowed us to immediately ignore some of the interactions which had a weighting of 0, leaving us with only 4 major contributions: Ligand_clash, ligand_torsion, part_buried and non-polar, to generate our initial scoring function (**Eq. 2**).

$$\log K_{calc} = -3.83\, f(\text{ligand\_clash}) + 0.12\, f(\text{part\_buried}) - 0.08\, f(\text{non-polar})$$
$$- 2.71\, f(\text{ligand\_torsion})$$

Eq. 2

The correlation between observed $\log K$ values and predicted ones based on this revised scoring function is in **figure 4.5**. Ideally the points should lie on the dashed line ($y = x$).
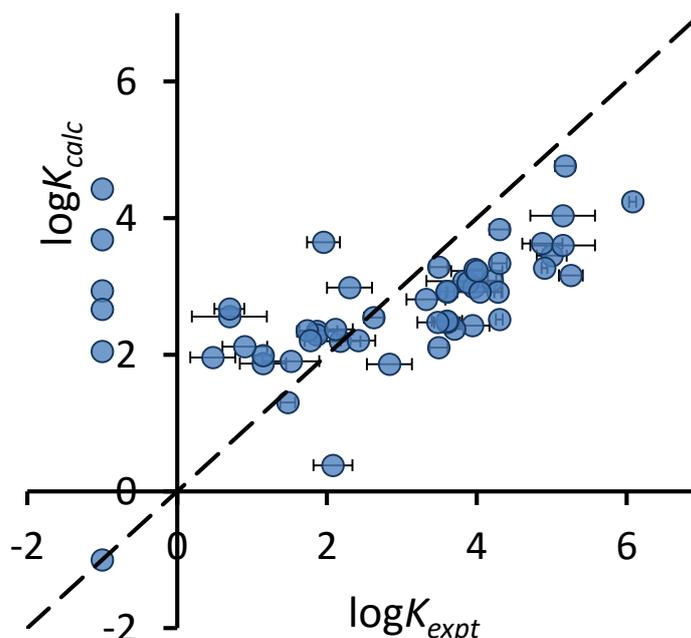
Figure 4.5: Comparison of experimental binding constants for the training set ($\log K_{expt}$) with binding constants calculated using **Eq. 2** ($\log K_{calc}$). The dotted line corresponds to $y = x$ (RMSD = 1.66).

There is a significant improvement of the calculated data with our experimental data ($r^2 = 0.21$), with a particularly good correlation for the high affinity guests. However this fit becomes less good for the weaker binding guests, and an extremely bad fit for 5 of the non-binding guests, which are all still predicted to bind strongly.

These five guests are all open-chain linear molecules that have a high degree of conformational flexibility. This is a manifestation of a point raised in the introduction to this chapter, whereby in molecular docking, entropic effects are often ignored in favour of enthalpic ones, despite the fact they often play a vital role in host-guest interactions. This effect is summarised in **figure 4.6**.

$$K < 1 \text{ M}^{-1} \qquad K = 1.5 \times 10^5 \text{ M}^{-1}$$
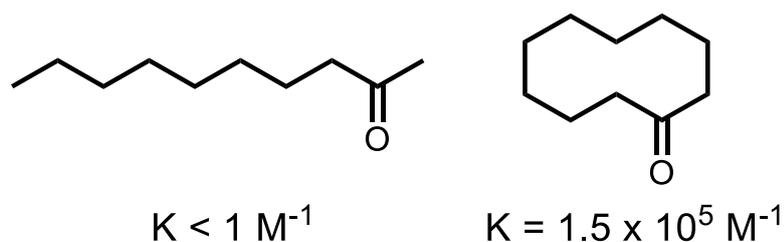
Figure 4.6: linear ketone doesn't bind; cyclic ketone binds strongly, despite similar hydrophobic surface areas and identical functional groups.

It was observed in chapter 3 that linear ketones such as the one in **figure 4.6** above do not display binding to the cage (K < 1 M^{-1}), whereas cyclic ones bind strongly despite containing the same number of carbons (10 in this case). According to hydrophobic surface area and other interactions these two molecules can make with the cavity, they should bind equally strongly (as predicted by GOLD), however due to entropic penalty associated with linear molecules being conformationally restricted in a cage cavity, they do not bind, and their cyclic, more pre-organised, analogues bind strongly.

In the GOLD docking process, a search of different guest conformations is performed, and it is possible to find a conformation of the open-chain ketone that fits as well into the cage as the cyclic ketone. The ligand_torsion term in **Eq. 2** describes the torsional strain, in other words the **enthalpy** penalty associated with putting a guest into a high energy conformation. However, the scoring function does not account for the **entropy** penalty of restricting the freedom of the inherently flexible guest, which we believed to be the problem.

To estimate the loss of conformational mobility when flexible guests bind, we used the program XedeX to calculate the number of rotatable bonds in each guest.[13] This number was used as an additional term, ligand_flexibility, in the scoring function.

With this term added to the scoring function, we re-optimised our scoring function in the same way as before, by allowing the individual weightings of each term to vary to give the best match between predicted and experimental K values. To generated a new improved scoring function shown below (**Eq. 3**).

$$\log K_{calc} = -4.48\, f(\text{ligand\_clash}) + 0.20\, f(\text{part\_buried}) - 0.10\, f(\text{non-polar})$$
$$+ 0.90\, f(\text{ligand\_torsion}) - 0.93 f(\text{ligand\_flexibility}) \qquad \textbf{Eq. 3}$$

This new scoring function gives a much improved fit ($r^2 = 0.82$) over the whole range of binding strength (non-binding, weakly, and strongly binding guests) to give the calculated vs experimental graph shown in **figure 4.7**. This is a substantial improvement with the data lying close to the y = x line as required – note improvement in prediction in weakly binding guests
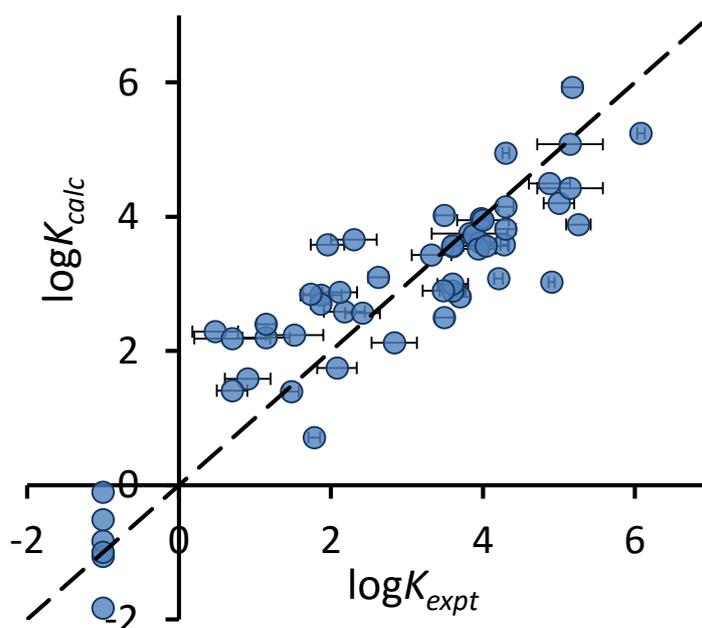
Figure 4.7: Comparison of experimental binding constants for the training set ($\log K_{\text{expt}}$) with binding constants calculated using **Eq. 3** ($\log K_{\text{calc}}$). The dotted line corresponds to *y = x* (RMSD = 0.79).

4.2.3 Constraints and assumptions

The scoring function we used was not purely an empirical scoring function, but a combination of *empirical* and *knowledge-based* scoring functions. We added a small constraint on the docking procedure which tells GOLD to place H-bond acceptor groups (if present) as best they can in the H-bond donor site in the cage (see chapter 1), directed by positions of methanol molecules from the X-ray crystal structure that sit in these sites. Although we do not use this orientating effect as a contribution in our score, this positioning gives the best fit possible. Without this positioning information, the fit is not as good (**figure 4.8**; RMSD = 1.11; $r^2 = 0.71$)
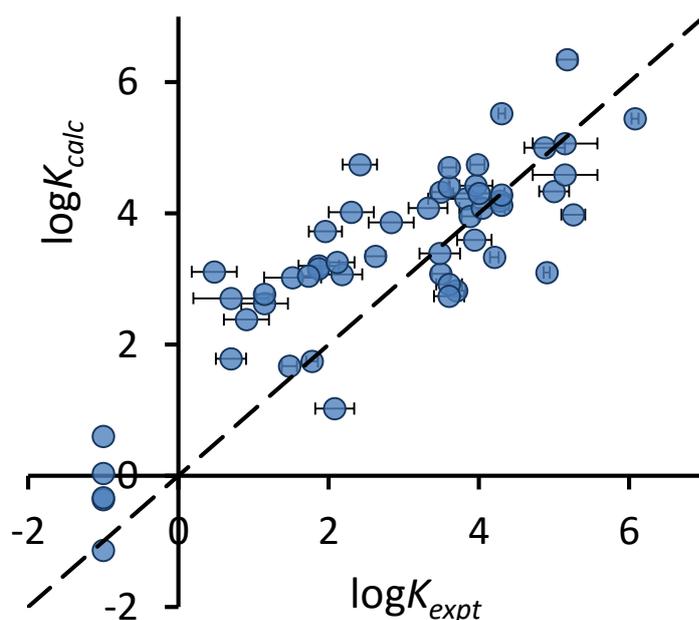
Figure 4.8: Comparison of experimental binding constants for the training set (logK_{expt}) with binding constants calculated using **Eq. 3** (logK_{calc}) without the H-bond acceptor positing constraint. The dotted line corresponds to $y = x$ (RMSD = 0.79).

We have assumed that the cage cavity is rigid and does not change upon guest binding. This is a reasonable assumption since the overlay of the X-ray crystal structures for the empty cage, and cage with bound cycloundecanone, are essentially identical (**figure 4.9, left**). Also shown in **figure 4.9 (right)** is the overlay of the guest cycloundecanone from the X-ray crystal structure with the GOLD predicted structure.
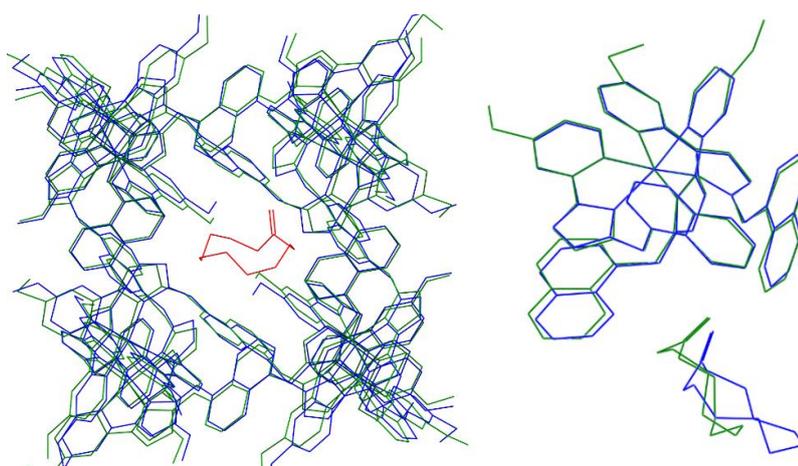
Figure 4.9: Overlay of crystal structures of free cage (blue) and cage with guest (green host, red guest) (left); and crystal structure of the cycloundecanone guest (green) and that generated by GOLD (blue) (right).

4.2.4 Virtual screen of a library of 3000 molecules

To test the predictive ability of this new scoring function (**Eq. 3**), we screened an in-house library of ca. 3000 compounds to identify new guests. The guests were ranked from strongest to weakest, and the top 120 compounds were chosen to do a mass-experimental screen using the fluorescence displacement assay with a plate reader. The top 15 predicted guests are shown with the corresponding $\log K_{calc}$ values in **figure 4.10**.
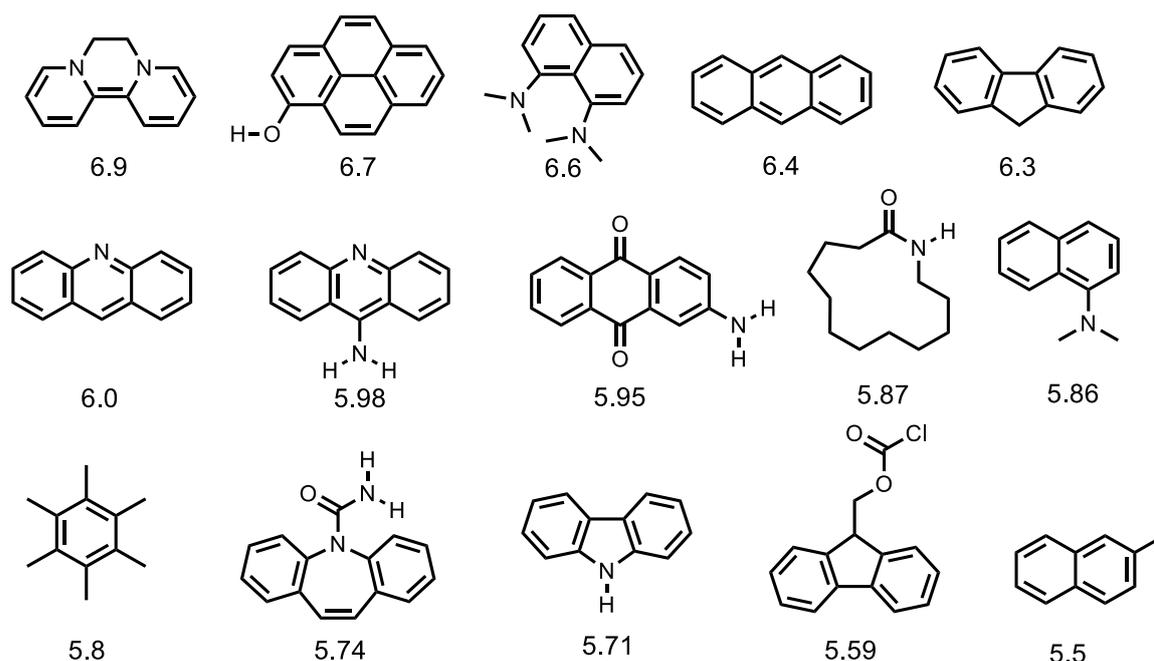
Figure 4.10: The top 15 guests from the virtual screen of 3000 compounds with $\log K_{calc}$ values underneath each compound.

4.2.5 Fluorescence plate-reader for mass guest screening

A fluorescence plate-reader uses a 'plate' of multiple wells (often 96 or 384 wells) (**figure 4.11**) and the fluorescence intensity of each well can be read simultaneously using a fluorescence spectrometer. It should be possible to obtain a binding affinity from a single data point: i.e. the addition of a fixed amount (e.g. 1 mM) of a guest to the test solution (cage + coumarin) and then by seeing how much of the fluorophore is displaced (see chapter 3). Using this method a screen could be developed where the addition of a different guest could be added to the test solution in each of the 96 or 384 wells and get

almost instant read-out fluorescence intensity values that can be directly related to the binding affinities of those 96 or 384 guests in a single experiment.
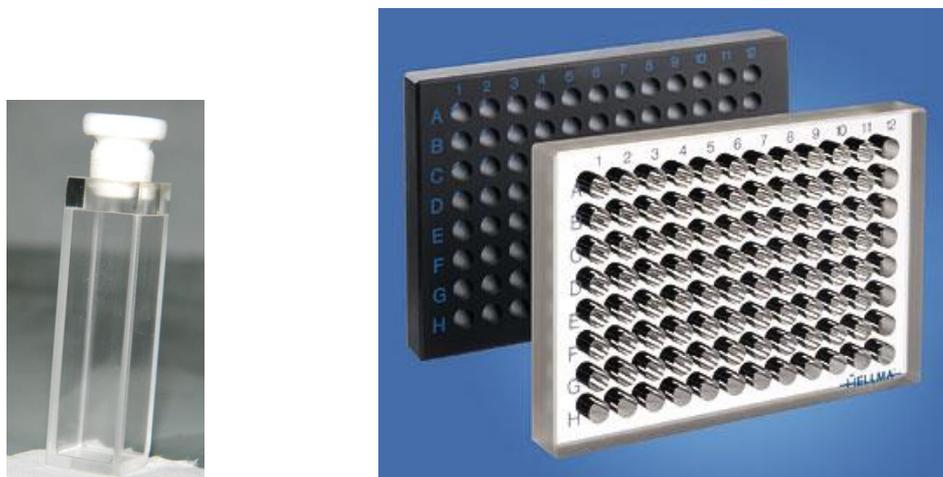
Figure 4.11: Standard cuvette (left), 96 well plate (right)

Knowing the binding constant of the coumarin fluorphore, and the flourescence intensity of the starting cage + coumarin mixture, and of 100 % free coumarin, it is possible to simulate what the readout should be for binding affinity values from log K = 0 to 10 (**figure 4.12**). The results fit a sigmoidal shaped curve with the best distinguishability being between a log K of 2 and to a log K of 6. This is an ideal range since a majority of the guests bind within this range. It is however possible to shift this inflection point if needed by altering the cage concentration. It is also worth noting that the concentration of the coumarin essentially has no effect on the outcome of the result.
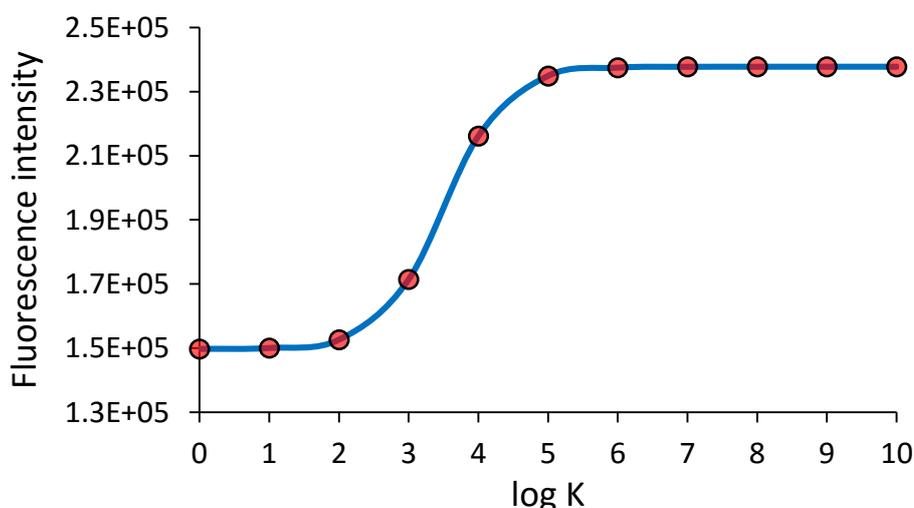
Figure 4.12: Simulation plot of log K against fluorescence intensity, with 0.055 mM cage, 0.01 mM coumarin, 1 mM guest.
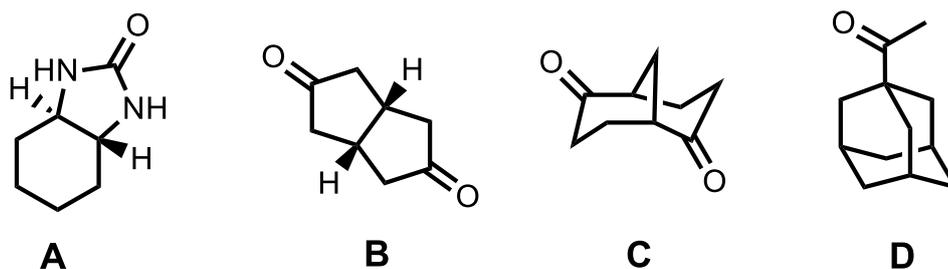
Figure 4.13: Compounds with known binding affinities used for a test screen.

Some compounds (**figure 4.13**) known to already bind with a range of binding constants from K = 150 to 15000 M^{-1} were tested using this plate reader assay to see if this method of using a single data point actually works as a quick test for screening guests. The results are shown in **figure 4.14** and **table 4.1**.
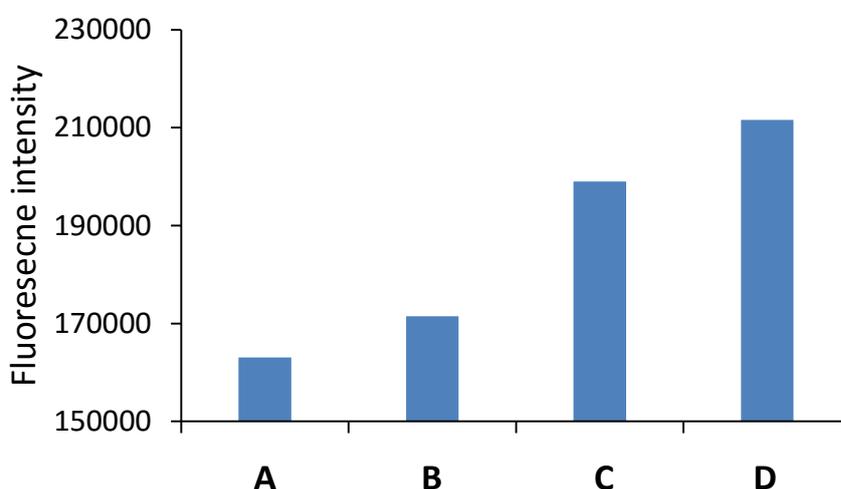
Figure 4.14: Plot of the fluroescence out-put for guests **A**, **B**, **C** and **D**

Guest	Fluorescence intensity	log*K* from fluorescence assay	log*K* from NMR
3.8	163033.8	2.10	2.18
3.9	171467.8	2.80	2.83
3.10	198972.3	3.75	3.70
3.11	211593	4.23	4.18

Table 4.1: Results of fluorescence read-out and the comparison of corresponding log K values, with the actual pre-measured log K values (determined by NMR spectroscopy)

From looking at these results it is clear there is a good correlation between the log*K* values obtained from a one point fluorescence read-out, with the log*K* values measured

from full titrations using NMR spectroscopy. We plan to use this method to screen 384 possible guests in a single experiment, or in this case 120 possible guests with each repeated 3 times for accuracy.

4.2.6 The big screen

Despite the great potential of this screening method, there were some difficulties. Firstly due to the generation of a few thousand 3D compounds, there were some mistakes in which the in-silico 3D representation of a molecule was flawed, not matching the real-world structure. Because of this, each in-silico representation was checked to see if it was feasible. This limited our 120 compounds down to 110. In particular the best binding predicted guest (**figure 4.15**) does not actually exist, and is in fact the compound shown to its right, which will not bind due to its charge (see chapter 5).
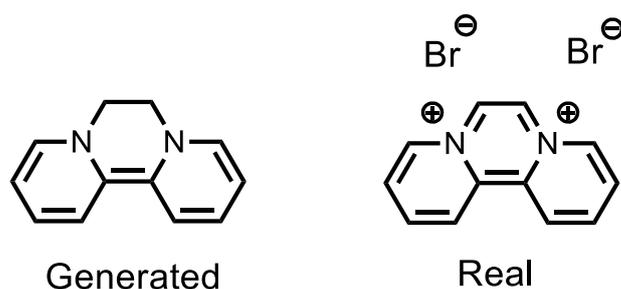
Figure 4.15: An example of an in-silico predicted guest not matching what is available in the real world.

The second difficulty that arose was those compounds whose reactivity with water and air, made them too difficult/dangerous to measure, this lead to a further reduction to a screen of ca. 100 molecules.

The third difficult is that many of the molecules are extremely strong fluorophores and so will interfere with the fluorescent signature of the coumarin used in the displacement assay, preventing the method from working.

Because of these limitations, we decided to just pick a selection of 15 compounds showing a wide range of binding affinities, and measure their equilibrium constants using

either the displacement assay as outlined in chapter 3, or NMR titrations, or by using the cage's ability to quench the fluorescence of the compound in question.

The chosen 15 are given in **figure 4.16**, with experimental log$K$ values.

6.80	8.00	7.26	6.06
6.09	5.73	5.50	2.88
4.45	4.18	4.20	4.11
	1.11	3.40	3.60
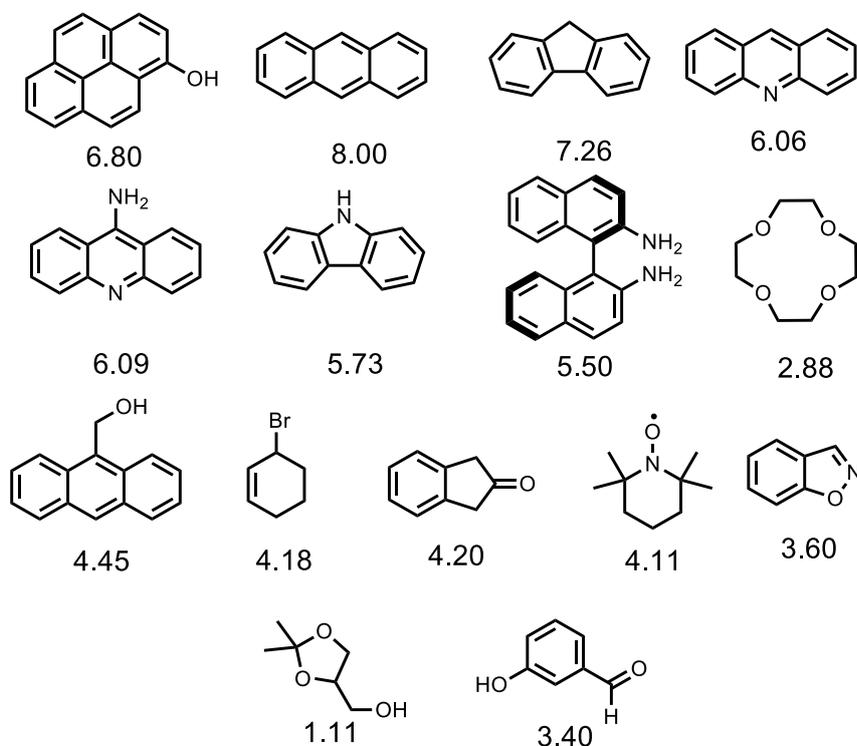
Figure 4.16: The 15 chosen compounds from the virtual screen showing the logK_{expt} values
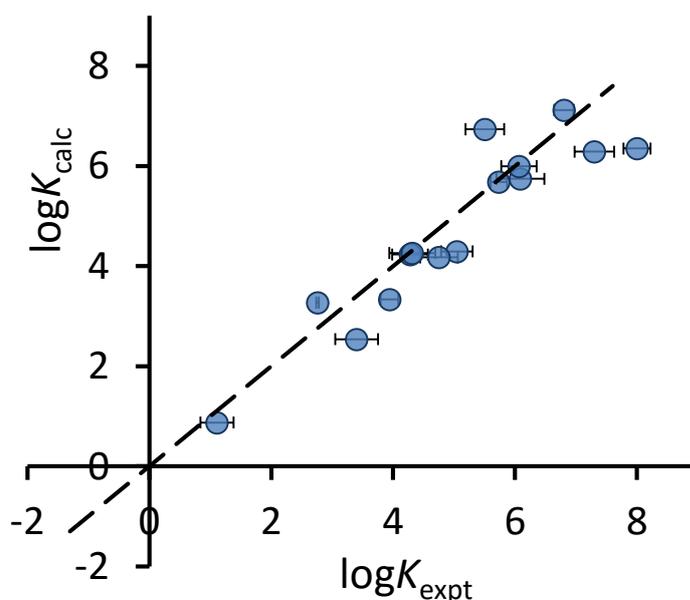
Figure 4.17: Comparison of experimental binding constants logK_{expt} for the 15 new guests in **figure 4.16** identified using GOLD with binding constants calculated logK_{calc} using **Eq. 3** The dotted line corresponds to $y = x$ (RMSD = 0.79).

The correlation between predicted and measured binding constants for this set of 15 guests (**figure 4.17**) is very good and clearly shows the predictive value of GOLD for identifying new guests. The RMSD for the training set of 54 known guests (0.79) is identical to the RMSD for the new set of 15 guests. This is particularly encouraging, because the new guests include classes of compound that were not present in the original training set: several polycyclic aromatics, and compounds with no polar groups. Several of the new guests identified by GOLD in this single screen bind more strongly than our previous best guest (cycloundecanone, log$K$ = 6.1) which was the culmination of hundreds of experimental measurements. The new guests include classes of compound that we had not previously considered, and include several well-known fluorophores; a stable radical (TEMPO); and a crown ether which is itself a host for metal ions – all of which suggest interesting new avenues for exploration in the physical properties of supramolecular assemblies.

We can use this new data to further tune the scoring function (a refinement of sorts) to generate a slightly better scoring function (**Eq. 4**) that fits all data (below) which gives a slightly better fit with all the data (RMSD 0.78) in **figure 4.18**.

$$\log K_{calc} = -4.47\ f(\text{ligand\_clash}) + 0.23\ f(\text{part\_buried}) - 0.11\ f(\text{non-polar})$$
$$+\ 0.62\ f(\text{ligand\_torsion}) - 0.92 f(\text{ligand\_flexibility}) \qquad \textbf{Eq. 4}$$
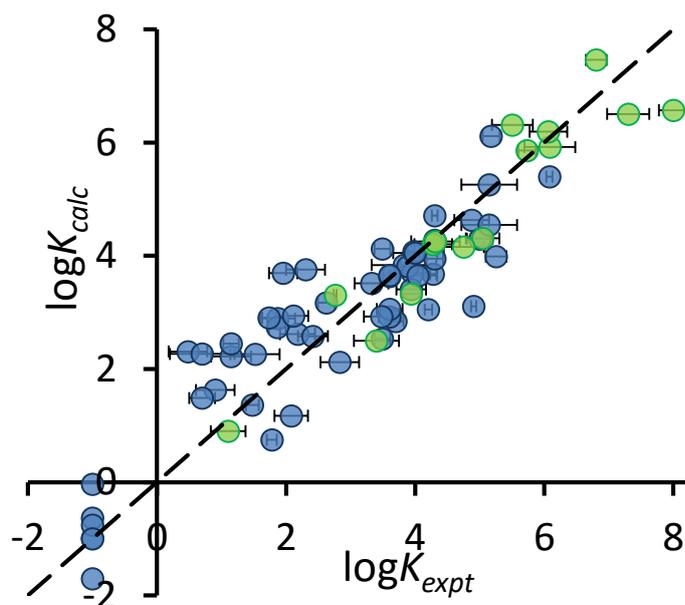
Figure 4.18: Comparison of logK_{expt} with logK_{calc} using **Eq. 4** Training set (blue); new 15 guests (red); The dotted line corresponds to $y = x$ (RMSD = 0.78).

4.2.7 Investigating hits

We wanted to investigate further some of the interesting 'hits' from the screen which might give complexes with unusual properties. The two guests we chose for further study are benzisoxazole (**figure 4.19, left**; see chapter 6), and TEMPO (**figure 4.19, right**).

Benzisoxazole   TEMPO
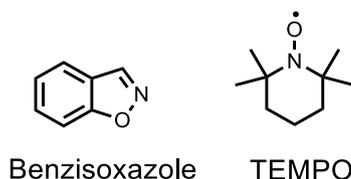
Figure 4.19: Benzisoxazole and TEMPO

TEMPO is a stable radical, and we wondered what effect its confinement in the cage cavity might impose, particularly since the cage (assembled with Co(II)) is paramagnetic, and therefore has some unpaired electrons that might interact with this radical. To do this we asked for aid from the national EPR service at the University of Manchester.

X-band EPR experiments were performed on powder samples at 160 K for the empty cage, the cage-TEMPO host-guest complex, and for free TEMPO.

The empty cage gives only a very weak feature, at low fields, under these conditions (**figure 4.20**).
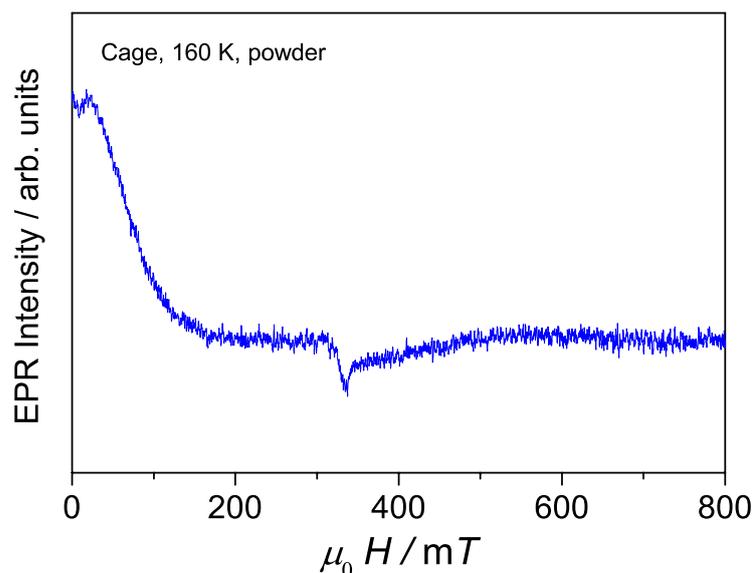
Figure 4.20: X-band EPR spectrum of powder sample of empty cage at 160 K

The cage-TEMPO host-guest species (**figure 4.21, top**) gives an EPR spectrum of a magnetically dilute TEMPO radical, with characteristic ^{14}N hyperfine coupling. For comparison the solid-state (neat) TEMPO sample (**figure 4.21, bottom**) gives a single line (i.e. broadened by intermolecular interactions).
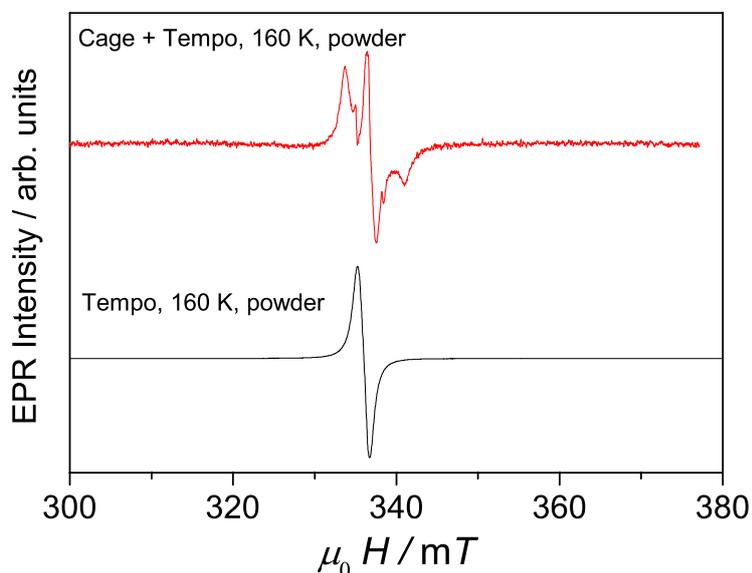
Figure 4.21: X-band EPR spectra of powder samples of {Co$_8$}-TEMPO host-guest species (top) and TEMPO (bottom) at 160 K.

This is consistent with the radical being inside the cage, keeping it well isolated from other TEMPO molecules and thus it's fine structure can be observed due to absence of exchange broadening. Surprisingly the TEMPO spectrum is nicely resolved, indicating that because there is a large miss-match in T_1 relaxation values (TEMPO being much slower than the cage), it doesn't 'feel' the effect of the Co(II) ions at all.

There are a couple of extra features in the solid-state host-guest spectrum at 160 K (the sharp lines flanking the central line in **figure 4.21, top**). These are characteristic of some translational degree of freedom of the radical in the cage. The spectra were modelled in the slow motion regime (i.e. where the spectra result from incomplete averaging on the spectroscopic timescale), using the aniosotropic (rigid limit) parameters from much lower temperature frozen solution measurements (below), and including an isotropic rotational correlation time, τ_{corr}. A $\tau_{corr} \approx 8$ ns gave a close fit. (The spectroscopy was run at ca. 10 GHz, these measurements are very sensitive to motion on the ns timescale.)

4.2.8 ZINC screen

Now armed with a good scoring function for predicting guest binding, we decided to run another virtual screen with ZINC, a library of 35 million of biologically and medicinally interesting purchasable molecules, ready to dock.[14]

The top 20 guest structures are shown on the next page (**figure 4.22**) in rank order (strongest predicted binder first). The logK_{calc} values for these guests range from 4.77 to 3.25. The strengths of these binding might not reach as high as the in-house library due to most of the compounds in the ZINC library containing multiple polar groups, but the strength of binding is still considered strong (almost K = 10^5 M^{-1}).
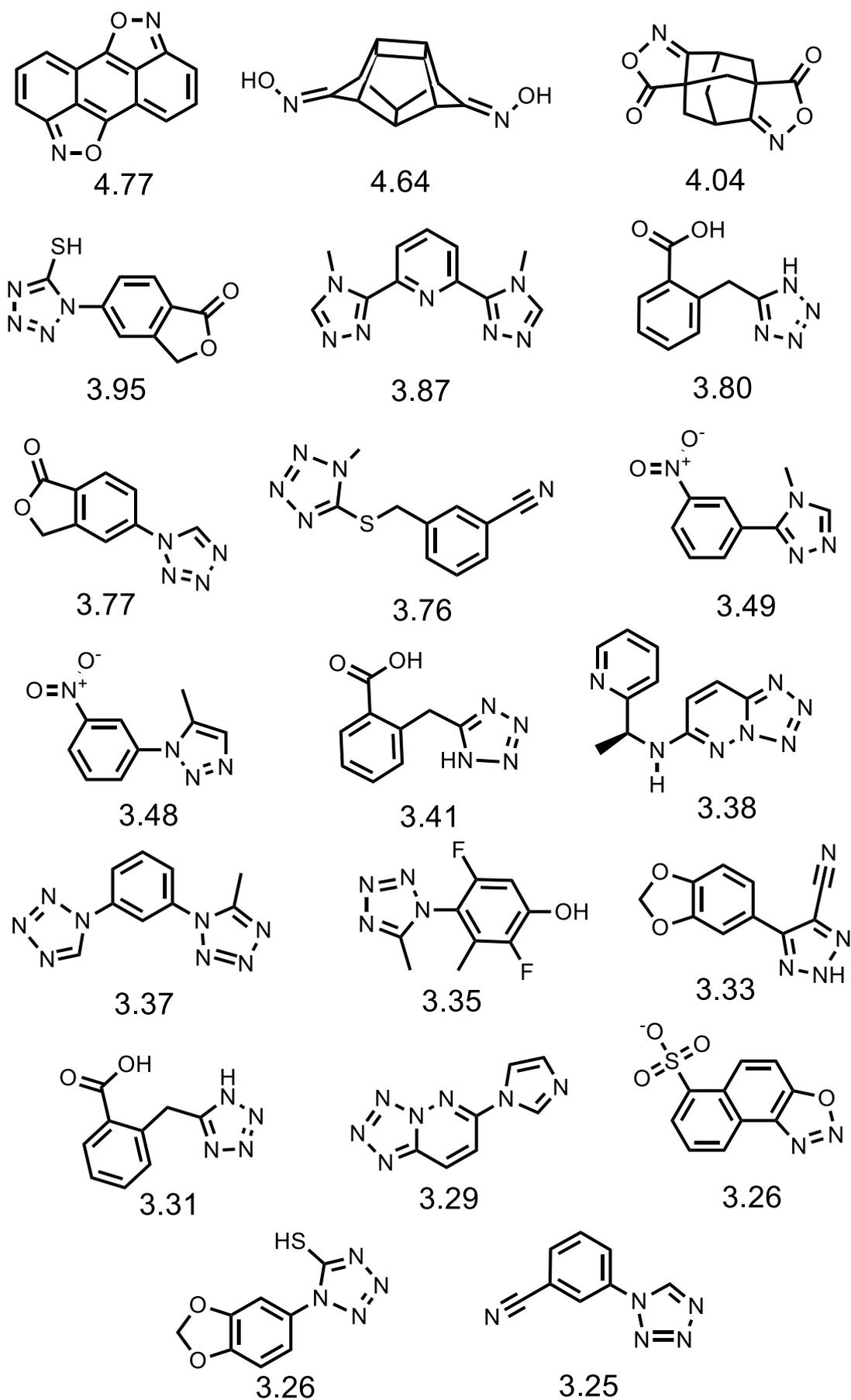
Figure 4.22: Top 20 predicted guests from the ZINC screen with logK_{calc} values from **Eq. 3**

4.3 Conclusions

In conclusion, we have demonstrated for the first time that docking software, developed for the analysis of protein / ligand interactions in drug discovery, can be used to identify new guests for a synthetic supramolecular receptor and accurately predict binding constants to within an order of magnitude. A training set of 54 guests was used to optimise a GOLD scoring function, which included a new term to account for the loss of conformational mobility when flexible guests bind. The scoring function is unique to this host, but the process of developing a scoring function is sufficiently straightforward that, given enough known guests to provide an initial training set, a scoring function specific to any synthetic receptor can be developed in the same way.

This methodology creates the possibility for guest binding in artificial molecular containers to be predictable and for new guests to be identified with confidence by virtual screening. The ability to reliably predict host-guest interactions will in turn open the door to a massive expansion of possible types of functional behaviour that can be developed with molecular containers and allow synthetic hosts to achieve their full potential.

4.4 Experimental techniques and procedures

Chemicals

All chemicals were purchased from commercial sources and used as supplied unless otherwise stated.

The host cage $[Co_8(L^{1,5-naphOH})_{12}](BF_4)_{16}$ was prepared according the published method.[12]

Nuclear magnetic resonance spectra

All NMR data were collected using a Bruker Avance III 400 MHz NMR spectrometer at 298 K with the parameters: 3072 scans, 284 ppm sweep width, O1p value of 0 ppm, D1 0.5 s, aq 0.5 s. The NMR data was processed using Bruker Topspin 3.1

Fluorescence spectra

Fluorescence data was collect either using a Horbia Jobin Yvon Fluoromax-4 spectrofluorometer and a quartz cuvette of 1 cm path length; or using a BMG FLUOstar Omega plate reader and either a 300 μl Hellma 96 well quartz microplate or a 100 μl Griener Bio-one μClear black 384 well plate. All spectra were collected at 298 K.

4.4.1 Calculating logK_{calc} using GOLD

Host ('Protein')

We used the previously-derived crystal structure of the cage (see main text) as the 'protein' by importing the coordinates into GOLD as a .mol2 file. Solvent molecules and anions were removed such that only the cage cation was considered.

Guests ('Ligands')

We first created the SMILES strings that describe the guest molecules by using ChemCell[1] (a Macro that enables Microsoft Excel to convert columns of chemical names and CAS Numbers into SMILES strings). We then used the program TORCH[2] to generate the 3D minimised structures for each of the guest molecules. These molecules were exported as a combined .mol2 file for use in GOLD, and as a combined .sdf file for use in

XedeX.[13] The structures were visually checked to ensure they had been created successfully.

Running GOLD

The Host and Guest sets were then imported to GOLD as the 'Protein' and 'Ligands' respectively. The binding site was defined by using the cobalt atoms to locate the centre of the cavity; the scoring function was selected (in this case ChemPLP); and the .conf file was created ready to be run in GOLD. This was all done by following GOLD's built-in wizard.

A positioning constraint for the guest was added to locate H-bond acceptors such as carbonyl groups in one of the two H-bonding pockets in the corners of the cage cavity. This was achieved by including (as a .mol2 file) the two solvent molecules (MeOH in this case) that occupied these binding sites in the crystal structure. We added to the end of the GOLD .conf file the line:

"constraint similarity acceptor C:/location/solvent.mol2 10"

The .conf file was subsequently run through GOLD saving one solution per ligand, and the outputs were exported as a .csv file for use in Microsoft Excel.

Calculating the number of rotors (our 'ligand_flexibility term')

Using the .sdf file produced by TORCH for the set of guests, we used Babel[4] to separate the molecules into individual .pdb files which were subsequently run through XedeX (using an in-house Linux computer cluster) to calculate the number of independent rotors for each guest. A script was written to take the individual output files into a combined output .txt file. The number of rotors for each guest was copied from this file into Microsoft Excel and used as the 'ligand_flexibility' term along with the other terms output by GOLD (see Table 1, final column).

Generating the Scoring function

We used Microsoft Excel's 'solver' add-on to do a non-linear least-squares regression analysis, in which the weightings of the different contributions to the scoring function

(see Table 1) were varied to minimise the sum of the squares of the errors between the calculated and experimental log$K$ values.

4.4.2 Measuring the logK_{expt} values

Depending on the guest measured three different methods were used:

1) **Fluorescence displacement assay** (see Chapter 3)

The host solution was made up of cage (5.5 x 10^{-5} M) and coumarin **12** (1 x 10^{-5} M) with water.

The guest solution was made up of guest (1 - 10 mM) using the host solution.

To each well, different amounts of host and guest were added to a total volume of 300 μl (96 well plate, for a 384 well plate 100 μl) from all host solution to all guest solution using 12-24 wells per titration. The plate was then heated to 35 °C for 20 minutes to allow for mixing to occur, and then cooled to 25 °C and equilibrated for 20 minutes before the fluorescence emission at 450 nm (using a 400 nm excitation wavelength and the receiver gain was set to a well containing free coumarin **12**) of each well was measured.

Each titration was repeated twice and by using Microsoft Excel, the concentration of guest was plotted against fluorescent intensity and the resulting curve fitted (using an algorithm to take into account the second equilibrium between cage and fluorophore) to produce a value for the binding constant, K.

2) **Fluorescence quenching of guest**

The 'host' solution was made up of fluorescent guest (concentrations at around 1/K) with water.

A solution of cage (concentrations at around 10/K) was made up using the stock solution of guest.

To each well, different amounts of cage and guest were added to a total volume of 300 μl (96 well plate, for a 384 well plate 100 μl) from all cage/guest solution to all guest solution using 12-24 wells per titration. The plate was then heated to 35 °C for 20 minutes to allow

for mixing to occur, and then cooled to 25 °C and equilibrated for 20 minutes before the fluorescence emission at the wavelength for the fluorescent guest of each well was measured.

Each titration was repeated twice and by using Microsoft Excel, the concentration of cage was plotted against fluorescent intensity and the resulting curve fitted to produce a value for the binding constant, *K*.

3) NMR titrations

The host solution of cage (0.2 mM) were made up in D_2O.

The guest solutions were made up with guests dissolved using the host stock solution.

12 NMR tubes were made up of varying host and guest solutions ranging from all host to all guest solution, and the NMR spectra were recorded.

Each titration was repeated twice and by measuring the integrals between he free host, and bound host signals, a value for the binding constant, K was obtained.

4) Fluorescence screening using the plate reader

The host solution was made up of cage (5.5×10^{-5} M) and coumarin **12** (1×10^{-5} M) with water.

The guest solution was made up of guest (30 mM) using the DMSO.

To each well, 300 µl (96 well plate, for a 384 well plate 100 µl was added) host solution was added and the fluorescence emission at 450 nm (using a 400 nm excitation wavelength and the receiver gain was set to a well containing free coumarin **12**) of each well was measured. 1 µl of each guest solution (0.3 µl for a 384 well plate) was added to three separate wells (to allow for an average value to be calculated). The plate was then heated to 35 °C for 20 minutes to allow for mixing to occur, and then cooled to 25 °C and equilibrated for 20 minutes before the fluorescence emission was measured.

The average value of the fluorescence read-out was taken and using Microsoft Excel, the value for the corresponding binding constant, log*K*, could be obtained.

4.5 References

1) D. B. Kitchen; H. Decornez; J. R. Furr; J. Bajorath, *Nat. Reviews*, **2014**, *3*, 935-949

2) C. Beato, PhD Thesis, University Degli Studi Di Parma, **2013**

3) G. Jones; P. Willett, *Curr. Opin. In Biotech.*, **1995**, *6*, 652-656

4) H. Sun; D. O. Scott, *Chem. Biol. Drug. Des.,* **2010**, *75*, 3-17

5) L. Berg; C. D. Andersson; E. Artursson; A. Hörnberg; A. K. Tunemalm; A. Linusson; F. Ekström, *PLoS ONE*, **2011**, *6*, 1-12

6) G. Jones; P. Willett; R. C. Glen, *J. Mol. Biol.,* **1995,** 245, 43-53

7) G. Jones; P. Willett; R. C. Glen; A. R. Leach; R. Taylor, *J. Mol. Biol.*, **1997**, *267*, 727-748

8) M. L. Verdonk; J. C. Cole; M. J. Hartshorn; C. W. Murray; R. D. Taylor, *Proteins*, **2003**, *52*, 609-623

9) O. Korb; T. Stützle; T. E. Exner, *J. Chem. Inf. Model.*, **2009**, *49*, 84-96

10) S. Turega; M. Whitehead; B. R. Hall; M. F. Haddow; C. A. Hunter; M. D. Ward, *Chem. Commun.,* **2012**, *48*, 2752-2754

11) S. Turega; M. Whitehead; B. R. Hall; A. J. H. M. Meijer; C. A. Hunter; M. D. Ward, *Inorg. Chem.,* **2013**, *52*, 1122-1132

12) M. Whitehead; S. Turega; A. Stephenson; C. A. Hunter; M. D. Ward, *Chem. Sci.,* **2013**, *4*, 2744-2751

13) http://www.cresset-group.com/products/xedtools/

14) J. J. Irwin; T. Sterling; M. M. Mysinger; E. S. Bolstad; R. G. Coleman, *J. Chem. Inf. Model.,* **2012**, *52*, 1757-1768

Chapter 5

The use of pH changes to control guest binding in water

4.1 Introduction

The ability to alter guest binding to a host is of huge interest in the supramolecular chemistry world as this allows for a much greater level of control to the binding event. If refined enough, this control could be utilised to deliver/uptake a target load at will. Despite stronger and stronger host-guest complexes being formed, controlling the uptake and release of a guest using some external stimulus remains relatively undeveloped. In most cases a guest can be displaced from the host by the addition of a competing (usually more strongly binding) guest; or by the destruction of the host itself, thus releasing the guest in the process. These examples, although they involve release of the guest at will, offer very little control over the process and are essentially irreversible.[1-4] It is possible to alter the concentration of the solution, thus altering the relative speciation between the free and bound species, however this is very limited.

An example of this is the Crowley group's encapsulation of cis-platin into their cage host (see chapter 1) which can be released by the addition of Cl[-] ions to destroy the cage.[4] The Clever group made a cage that undergoes a structural rearrangement when exposed to light, which resulted in a change in the guest binding affinity (**figure 5.1**).[5]
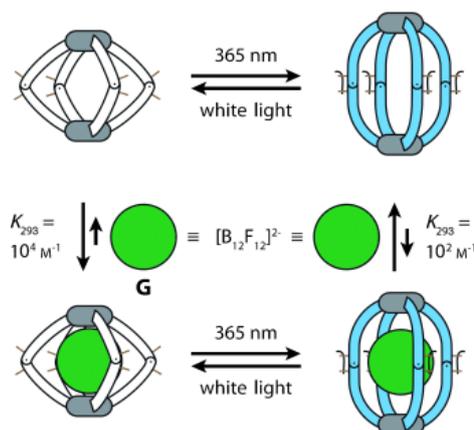
Figure 5.1: Schematic representation of altering the cage structure by using light as a guest uptake and release mechanism[5]

There have been some isolated attempts at reversible uptake and release of guest molecules by changing the property of the guest rather than the cage. Such an example is the Fujita group's example using the redox properties of ferrocene to change its charge,

and thus altering its binding affinity with the cage host (**figure 5.2, top**).[6] Also the Nitschke group have shown that in a mixture of acetone and THF, a pyridine molecule binds inside a cage cavity, however when trifluoroacetic acid was added, the pyridine molecule (now charged) is displaced by the solvent molecules. This process can be reversed by the addition of NaHCO₃ (**figure 5.2 bottom**).[7]
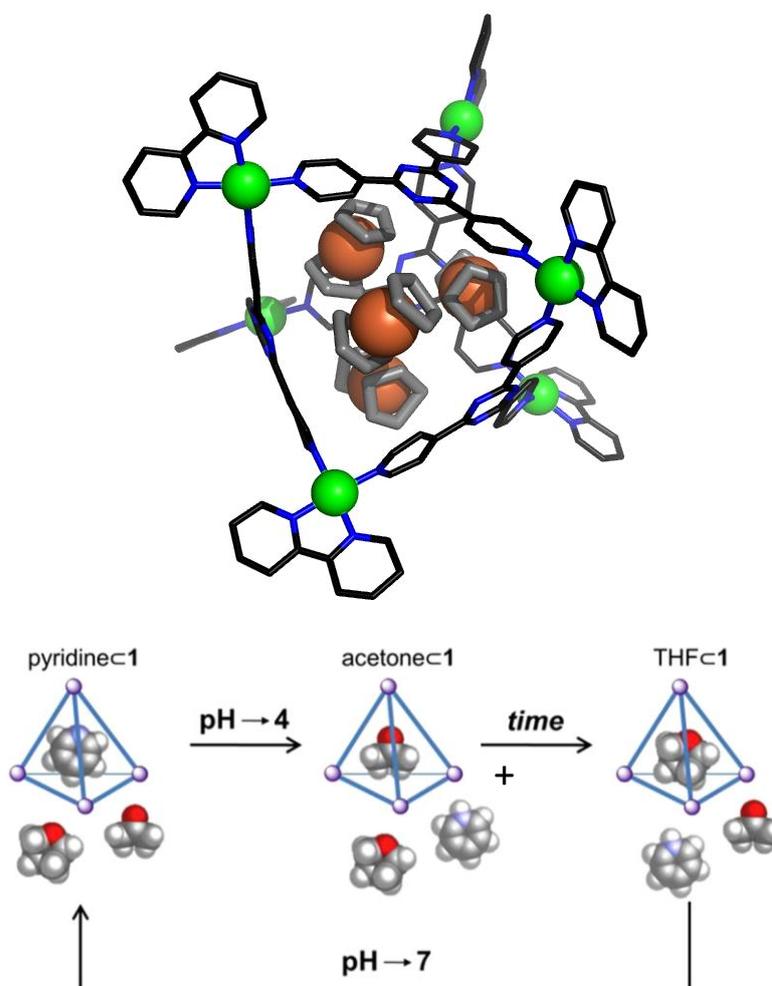
Figure 5.2: X-ray crystal structure of ferrocene encapsulated in Fujita's Pd_6L_4 cage (top)[6]; uptake and release of pyridine from a cage by changing pH (bottom)[7]

5.2 Results and Discussion

In aqueous media, the equilibrium between an acid and its base is known to have a dramatic effect on many chemical reactions from positive aspects such as the pH-switchable molecule acting as a catalyst, or removal of a protecting group, to negative aspects such as degradation of substances during a reaction. Thus investigating the effect that pH has on guest binding in the cage is an interesting possibility given that many of the guests have protonatable or deprotonatable functional groups.

Using a pH swing as a guest uptake and release mechanism will be highly advantageous for drug delivery as it doesn't require the destruction of the host molecule itself and the pH varies within the body. Thus, if the design of the host guest system is sophisticated enough, the targeted drug delivery using a cage host could be achieved.

5.2.1 Amantadine

Amantadine (Symmetrel®, **38**) (**figure 5.3, left**), is a commonly used drug to treat Parkinson's disease in its early stages, it is also used to treat other diseases such as Influenza A and pain from shingles.[8,9] When binding was attempted with the M_8L_{12} water soluble cubic cage in D_2O, no binding appeared to occur. This was unexpected since the binding constant of a similar guest (1-adamantanecarboxylic acid, **39**) (**figure 5.3, right**) was very high (8×10^4 Mol^{-1}).
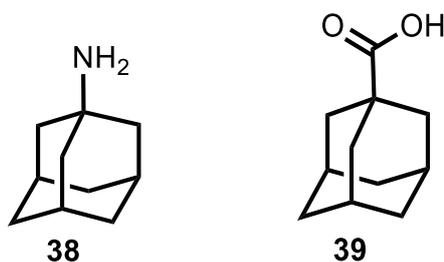
Figure 5.3: Compounds amantadine, **38**, and 1-adamantanecarboxylic acid, **39**

It was apparent that in aqueous media **38** would be protonated (NH_3^+) and potentially, since the cage itself was positively charged (16^+), this would prevent binding. To test this hypothesis, the pH of the solution was altered enough to deprotonate the amine group (pH 11) at which point, strong binding of **38** in slow exchange was observed. Then, by the

addition of some acetic acid, the solution was returned back to its starting pH and the binding was fully reversed (**figure 5.4**).
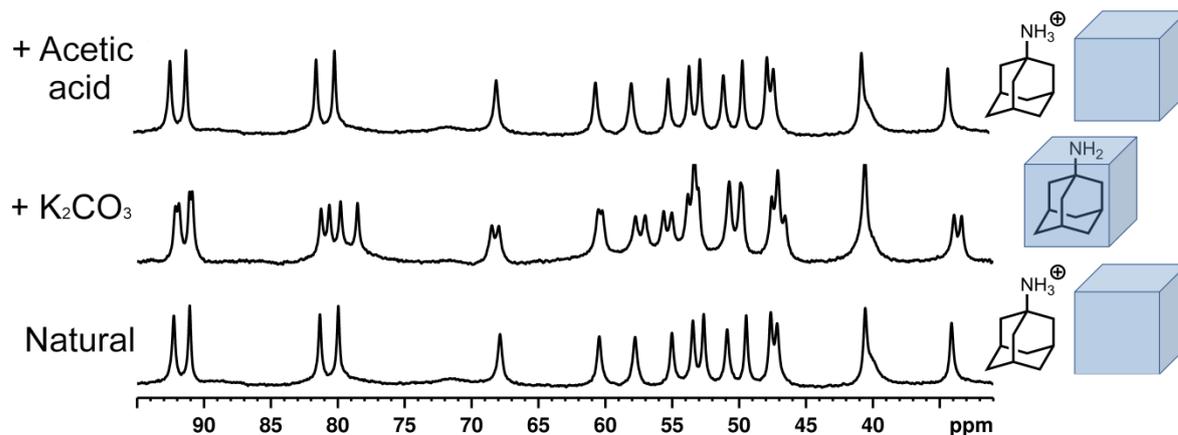
Figure 5.4: ^{1}H NMR spectrum of amantadine **38** (0.2 mM) and cage (0.2 mM) in D$_2$O at 298 K with nothing added, K$_2$CO$_3$ added until pH 11, and then acetic acid added afterwards

To test this hypothesis further a complete pH titration was conducted of amantadine (1.26 mM; pKa 10.9) in the presence of cage (0.2 mM). The pH was altered by the addition of NaOD, and the ^{1}H NMR spectrum at each pH value was measured (**figure 5.5**)
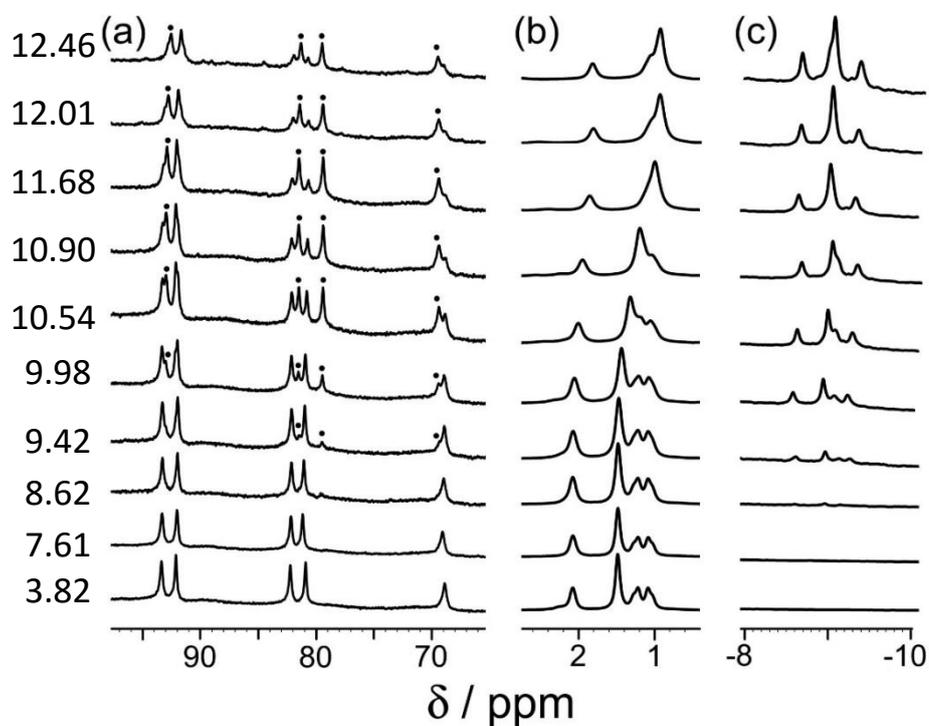
Figure 5.5: Partial ^{1}H NMR spectra with increasing pH from bottom to top showing (a) the host peaks (bound host is marked by '●' ; (b) the free guest peaks changing during pH titration; and (c) grow-in of peaks for bound guest

At the beginning of the titration (pH 3.8) the spectrum in **figure 5.5 part (a)** is showing just empty cage (the guest is protonated and so doesn't bind). As the pH is increased the emergence of new peaks (●) can be observed. These correspond to cage with guest bound inside, i.e as the guest becomes neutral it begins to bind inside the cage. Using the ratio between bound and free cage, the speciation at the corresponding pH can be calculated. Looking at **part (b)** the free guest peaks show a shift which, when plotted against the corresponding pH for the spectrum in question, matches the pH curve for deprotonation of the guest. Finally **part (c)** shows the area of the NMR spectrum where the bound guest (that is paramagnetically shifted when inside the cage) appears as the guest becomes deprotonated, and matches fully with the speciation behaviour seen in part (a).

The degree of occupancy of the cage (red line); as well as the change in chemical shift for the free guest with pH (blue), are plotted in **figure 5.6**.
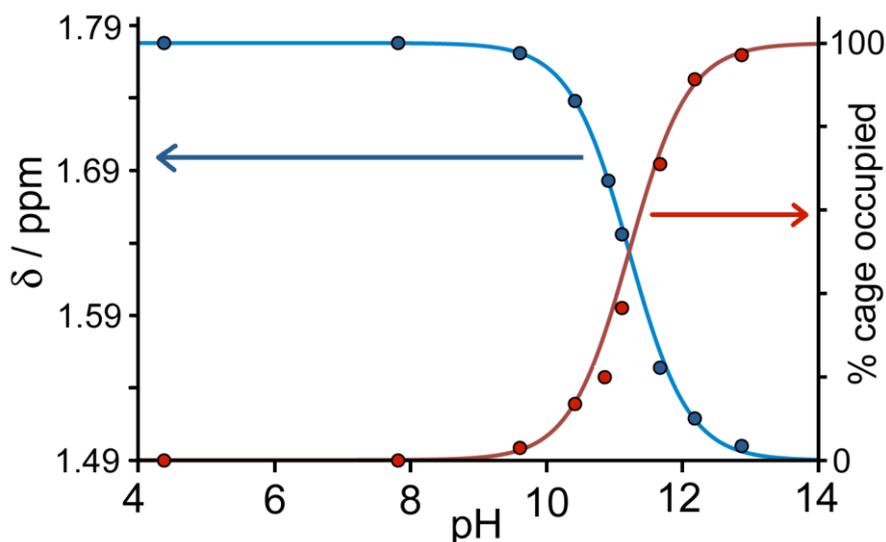
Figure 5.6: pH titration results of amantadine in D_2O at 298 K showing chemical shift of free guest against pH (blue); and occupancy of the cage as a function of pH (red)

The two curves intersect each other at the pKa of the free guest molecule indicating the uptake and release of the guest from the cage is driven purely by the deprotonation/protonation of the guest. The strength of binding (-ΔG) for the neutral form is 22.8 kJ mol^{-1} (K = 1 x 10^4 M^{-1}), and for the charged form is 6.0 kJ mol^{-1} (K = 13 M^{-1}), this shows that the cationic form binds around 4 times weaker than its neutral form.
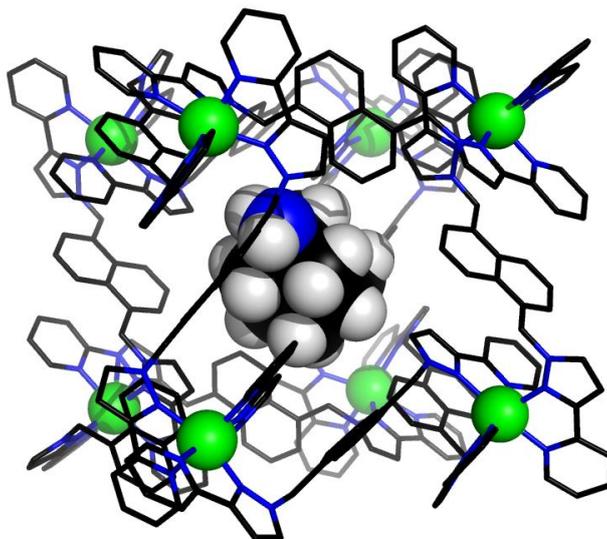
Figure 5.7: Molecular model of amantadine **38** inside the cage

The molecular model[10] of **38** within the cage (**figure 5.7**) shows how the almost spherical guest molecule **38** neatly fills the cavity.

5.2.2 The effect of pH on the cage itself

The cage begins to show signs of decomposition at a pH of 11 and above. At pH 12, decomposition occurs over a 12 hour period and at pH 13, decomposition happens in a matter of moments. This is most likely due to the high concentration of ⁻OH forming a complex with the labile Co(II) ions, thus precipitating out the free ligand.

To understand the behaviour of the cage throughout the whole pH range, a pH titration was performed using just cage on its own (0.2 mM) (**figure 5.8**). The cage itself is relatively naturally acidic with a pH of around 3.8 (at this concentration) in D_2O. This is attributed to the large number of OH protons on the cage's exterior and because of the very high positive charge (16$^+$) associated with the cage, the CH_2O-H protons might more readily dissociate in solution to help reduce this high charge. The plots of NMR chemical shift vs pH showed shifts in chemical shift occurring up to around pH 4, with an estimated pKa value of 1.38. This change in chemical shift of the cage protons with pH may well be due to the numerous OH groups covering the exterior of the cage. It is also worth noting that at this lower end of the pH scale (down to pH 1), no decomposition was detected, perhaps indicating this shift is not due to the pyridine nitrogen atoms since protonation

at this point would most likely disassemble cage. There are also some smaller shifts occurring at around pH 10 and above, which could be due to further deprotonation of the OH groups on the exterior, or perhaps due to the onset of cage decomposition.
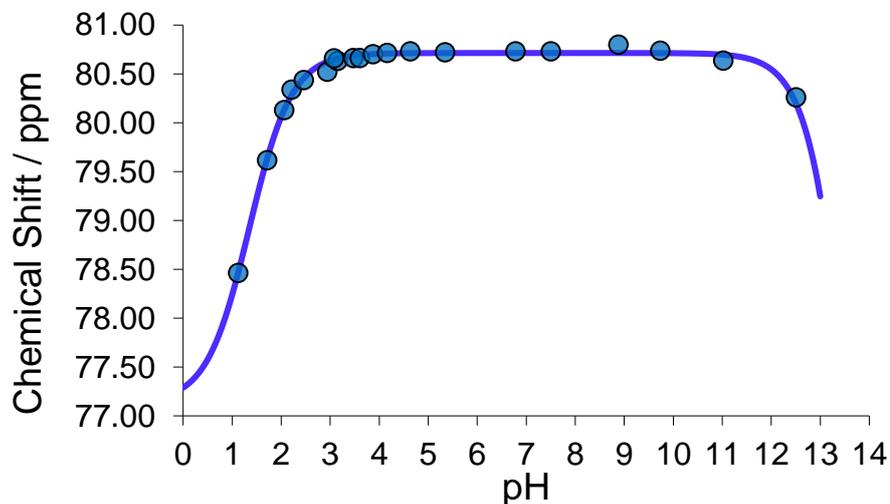
Figure 5.8: Plot of free cage (0.2 mM) signals in ^{1}H NMR with pH in D_2O at 298 K

A second titration was carried out where instead of observing the change in chemical shift vs pH, the pH of the solution itself was measured compared to the amount of ^-OH added (**figure 5.9**).
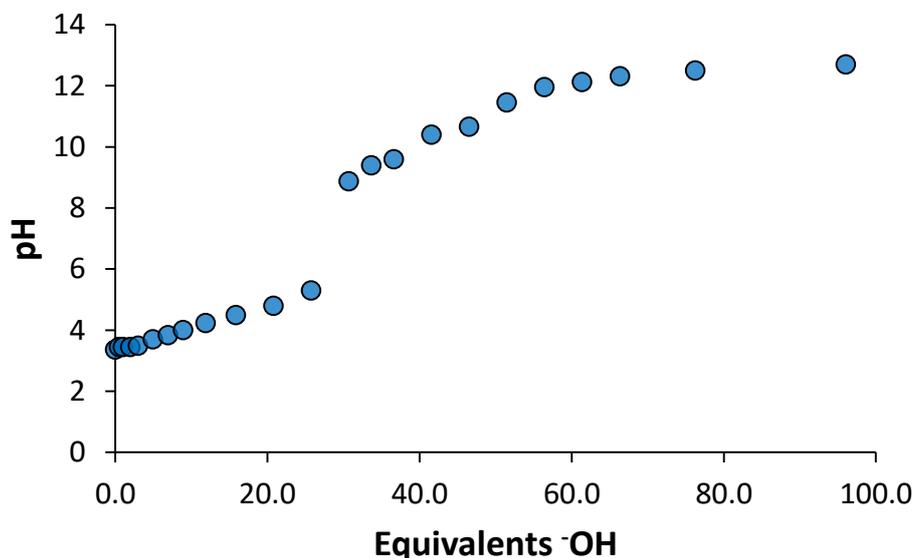
Figure 5.9: Plot of measured pH of a 0.2 mM cage solution with addition of NaOH

This plot shows that around 30 equivalents of hydroxide are required to neutralise the cage, which is surprising since the cage has a charge of 16+. One hypothesis is that there

are 24 CH$_2$OH groups on the surface, so if all were to be deprotonated, this could account for 24 equivalents of the 30 hydroxide added, but there is still 6 equivalents that need to be accounted for.

It is known that BF$_4$ anions (used to counter balance the cages 16+ charge) can hydrolyse in water, and in turn can be decomposed by hydroxide. So this is potentially where this discrepancy could be accounted for.

A much more ideal value for a pH swing to trigger guest uptake/release would be around pH 5-7, since the cage is completely stable, and this range is of much more use for practical applications associated with drug delivery. Compound **39**, mentioned earlier, has a carboxylic acid functional group which should have a pka in the lower range of the pH spectrum. This guest binds strongly to the cage when neutral (K = 8 x 10^4 M^{-1}) however it would be interesting to see if the binding would change when the acid group is deprotonated (pKa 5.1) since the guest would be negatively charged and we wondered if it would bind more strongly to the cationic cage than the neutral form for electrostatic reasons.

5.2.3 Adamantanecarboxylic acid
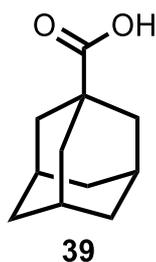
Figure 5.10: 1-adamantanecarboxylic acid, **39**

A pH titration for **39** in the presence of cage was carried out in the same way as before (**figure 5.11**), however due to the larger equilibrium constant compared to the amine, an almost equimolar concentration of host and guest were used (0.2 mM H; 0.24 mM G) rather than an excess of guest. This guest, like the amine derivative, was also in slow exchange on the NMR timescale so monitoring its binding during the experiment was simple.
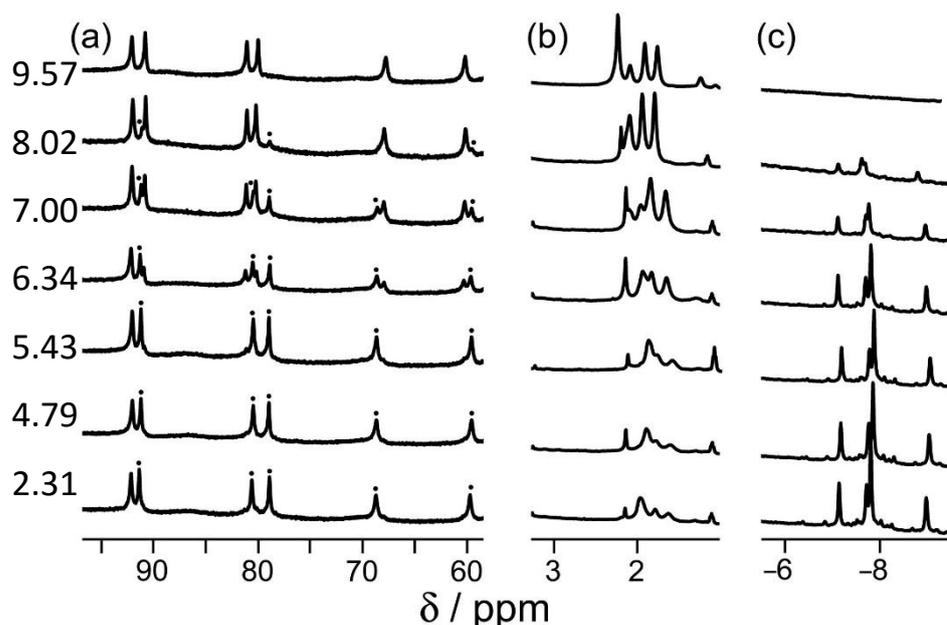
Figure 5.11: Partial ^{1}H NMR spectra with increasing pH from bottom to top showing (a) the host peaks (bound host is marked by '●' ; (b) the free guest peaks changing during pH titration; and (c) grow-in of peaks for bound guest

The interesting thing to note with the carboxylic acid guest compared to the amine guest, is that as the pH is increased (and the guest becomes deprotonated, i.e anionic in this case) the guest is expelled from the cage cavity despite the presence of a negative charge. At the beginning in its neutral form (**figure 5.11**, bottom spectrum, pH 2.31) the guest is fully bound, as can easily be seen by the characteristic parmagnetically shifted bound guest peaks in part (c), and also by the lack of free guest in part (b). As the pH increased and the guest becomes anionic, the bound guest peaks in (c) disappear (along with the bound cage peaks (*) in (a)), and the free guest peaks in (b) grow in (along with the free cage peaks in part (a)).

This implies that it is not the sign of the charge that is determining the binding, but just simply that there is a charge present i.e. electrostatic effects are not as significant compared to solvation effects. If the guest is charged, whatever the sign, it would much rather be solvated in the bulk water than inside the hydrophobic cavity of the cage.

The same plot of change in chemical shift of the free guest (blue) and the % occupancy of the host cavity (red) with pH can be plotted (**figure 5.12**). Just like its amine analogue, the two curves crossed at the pKa of the guest molecule, again indicating that it is the change in protonation state of the guest that controls uptake and release.
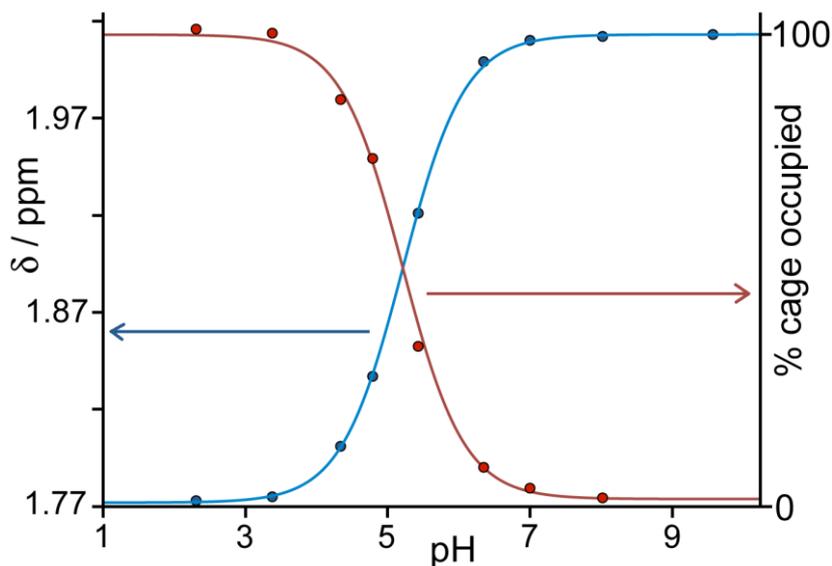
Figure 5.12: pH titration results of 1-adamantanecarboxylic acid **5.2** and cage in D_2O at 298 K showing chemical shift of free guest against pH (blue); and binding constant against pH (red)

The strength of binding (-ΔG) form the neutral form is 28.0 kJ mol^{-1} (K = 8 x 10^4 M^{-1}), and for the charged form is 17.0 kJ mol^{-1} (K = 9 x 10^2 M^{-1}), this shows that the anionic form binds around half as strongly compared to its neutral form.

The change in binding energy ($\Delta\Delta G$) for the amine **38** to **38•H$^+$** is 17 kJ mol^{-1}, and for the carboxylic acid **39** to **39•$^-$** is 11 kJ mol^{-1}, so there clearly is an electrostatic effect with the cationic cage and the cationic guest (repulsive) and the anionic guest (attractive). This effect is still dominated by preferred solvation of the charged guest, so is likely small. However, if we assume the electrostatic repulsion between the cage (+) and the charged acid (-) is equal and opposite to the electrostatic attraction between the cage (+) and the charged amine (+), then this leaves us with 14 kJ mol^{-1} for purely solvation of the charge, and +/- 3 kJ mol^{-1} for the electrostatic repulsion/attraction.

We wanted to see what would happen if we used a guest which could form a doubly charged species. For this we chose 1,3-adamantane dicarboxylic acid , **40** (successive pKas of 4.8 and 5.9) (**figure 5.13**).
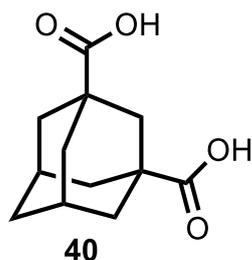
Figure 5.13: 1-adamantanecarboxylic acid, **40**

We measured the extent of binding for 1,3-adamantane-dicarboxylic acid over the same pH range. The strength of binding (-ΔG) for the neutral diacid form at pH 3 is 30.6 kJ mol^{-1} (K = 2.3 x 10^5 M^{-1}). The binding strength for the dianionic form at pH 8 was too low to measure (If the concentration of the dianion becomes too high, the cage is destroyed; however even at this point, no binding was observed, so we can say that it binds < 14 kJ mol^{-1} (K < 3 x 10^2 M^{-1}) but most likely binds far less.

5.2.4 Adamantane carboxylic acid crystal structure

X-ray quality crystals of **39** inside the cage were obtained (**figure 5.14**) using the method outlined in chapter 3, however since **39** was a solid (unlike the oil of cycloundecanone), the cage crystals were soaked in a saturated solution of **39** in n-hexane for 1 day. Given the fact that the guest was administered in its neutral acid form we assume that it is in this form in the host cavity, and not as the adamantane-1-carboxylate anion, which has a much lower binding affinity. As is normal for cage complexes of this type, weak scattering resulted in a relatively high R1 value of 16%, which means that detailed analysis of structural minutiae is not appropriate, but the formation of the complex and its key structural features are clear.
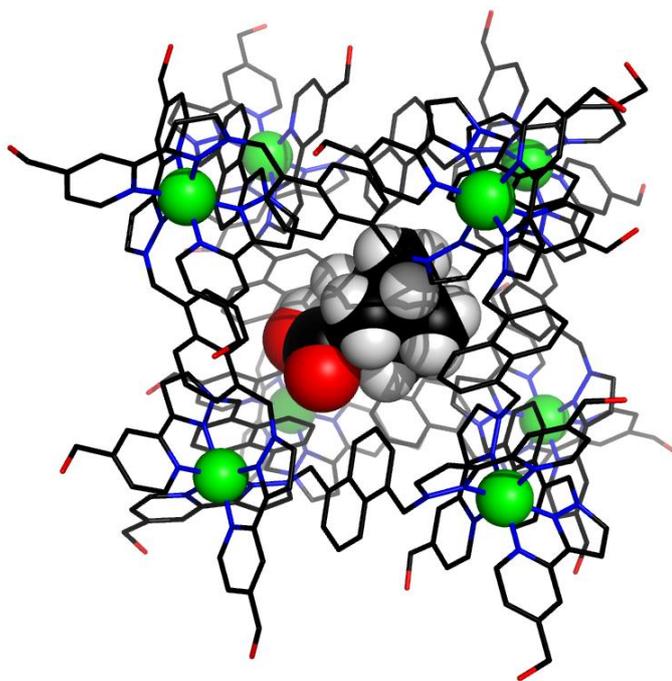
Figure 5.14: X-ray Crystal structure of the adamantine carboxylic acid **39** - cage complex

The adamantyl unit lies centrally in the cavity with the COOH group projected towards one of the two *fac* tris-chelate metal vertices which lie at either end of the long diagonal, with short CH•••O contacts (2.7 – 2.9 Å) between the carboxylic acid oxygen atoms and some of the naphthyl and methylene protons on the interior surface of the host (**figure 5.15**) (the associated non-bonded O•••C separations are in the range 3.5 – 3.8 Å).
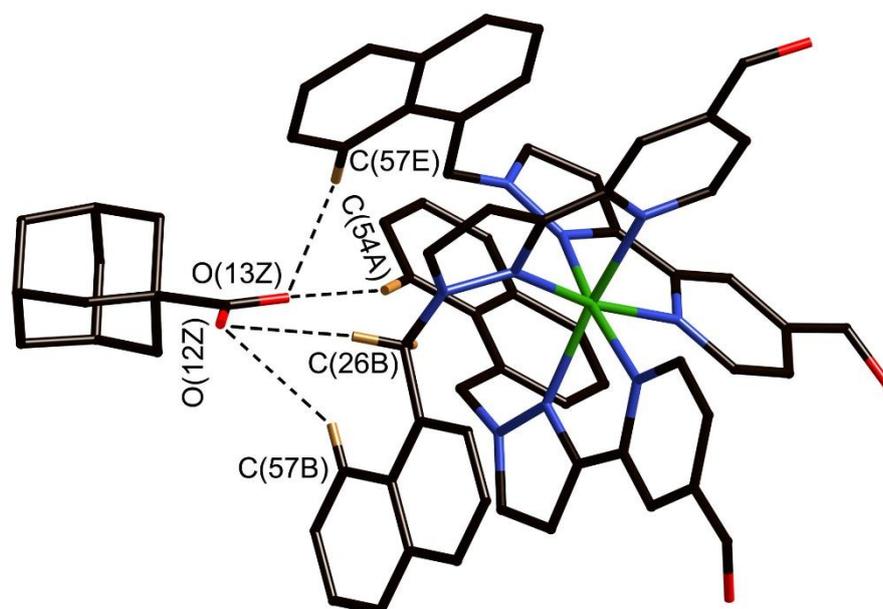
Figure 5.15: Close-up view from the crystal structure of the four closest contacts between the oxygen atoms of the guest and some of the naphthyl and methylene CH protons of the host in the binding pocket

The guest is disordered over two symmetry-equivalent positions with 50% site occupancy in each: one orientation is shown in **figure 5.14**, and the alternative orientation (related by inversion) has the COOH group oriented towards the symmetry-equivalent opposite corner of the host. The two *fac* tris-chelate sites in the cage each provide a convergent group of CH protons in a region where the electrostatic potential on the internal surface is most positive, thus resulting in an H-bond donor pocket which is responsible for guest binding in organic solvents and which also provides an anchoring point for the polar part of the guest.

5.2.5 Other guests

The same experiment was completed on a number of different guests (**figure 5.16**): aspirin, **41**; isoquinoline, **42**; detomidine, **43**; and (-)-nicotine, **44**; all with a range of different functional groups and pKa values. They all show a difference in binding strengths between the protonated and deprotonated forms and they all bind most strongly at a pH where the guest is neutral, and not when the guest is predominantly charged, which fits nicely with what was seen in the previous examples.
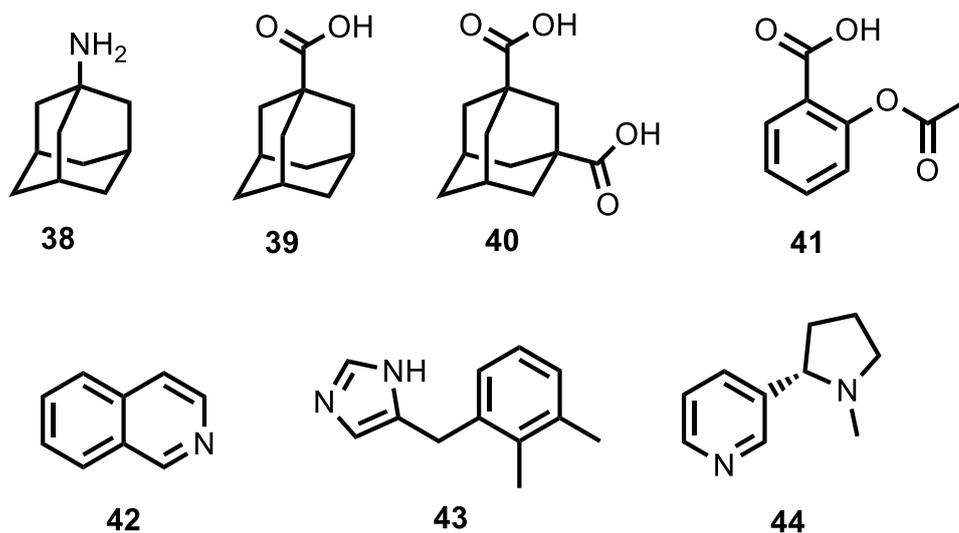
Figure 5.16: Guests used to investigate pH-dependent binding in this chapter; from left to right top to bottom: amantadine, **38**; 1-adamantanecarboxylic acid, **39**; 1,3-adamantanedicarboxylic acid, **40**; aspirin, **41;** isoquinoline, **42**; detomidine, **43** and (-)-nicotine, **44**

A summary of the results of all the guests is given in **table 5.1** and **figure 5.17**.

Guest	pKa	Neutral form		Charged form	
		K / M^{-1}	-ΔG / KJ Mol^{-1}	K / M^{-1}	-ΔG / KJ Mol^{-1}
38	10.9	1.0(3) x 10^4	22.8(7)	13(7)	6(1)
39	5.1	8.0(2) x 10^4	28.0(1)	9.0(5) x 10^2	17.0(3)
40	4.8, 5.9	2.3(4) x 10^5	30.6(4)	<300	<14
41	3.5	1.2(3) x10^2	11.9(6)	<3	<3
42	5.5	1.2(5) x 10^4	23.3(8)	10(2)	5.7(5)
43	7.2	70(30)	10.5(8)	<3	<3
44	8.1	81(20)	10.9(5)	<4	<3

Table 5.1. Summary of the pH titrations with all the guests

If we look at the ΔΔ*G* values from neutral to cationic of the isoquinoine **42** is 18 kJ mol^{-1}, which agrees very well with the amine **38** at 17 kJ mol^{-1}.

GH$_2$= 230000 M^{-1}
G$^{2\ominus}$ < 300 M^{-1}

GH = 82000 M^{-1}
G$^{\ominus}$ = 900 M^{-1}

GH = 120 M^{-1}
G$^{\ominus}$ < 3 M^{-1}

G = 12000 M^{-1}
GH$^{\oplus}$ = 10 M^{-1}

G = 81 M^{-1}
GH$^{\oplus}$< 4 M^{-1}

G = 70 M^{-1}
GH$^{\oplus}$ < 3 M^{-1}

G = 10000 M^{-1}
GH$^{\oplus}$ = 13 M^{-1}

pH
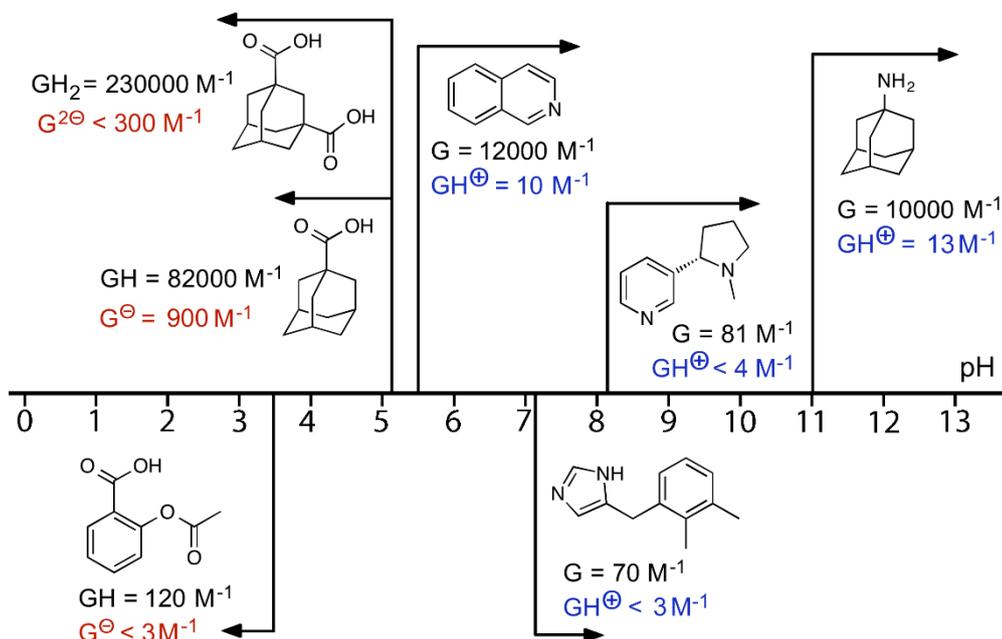
Figure 5.17: Graphical summary of association constants for guests in neutral and charged states

5.2.6 Selective uptake and release of different compounds from a mixture

Using what we know about at what pH a particular guest molecule will bind, and how strongly they bind, we initially devised a system where two guests present in the same solution along with the cage, could be selectively uptaken and released, depending solely on the pH of a solution. Using a switching pH either in the acidic region, or the basic region.

The three guests chosen were 1,3-adamantane dicarboxylic acid (**40**), amino adamantane (**38**), and cyclononanone (**21**). Their structures and binding constants are summarised in **figure 5.18** and **table 5.2**.
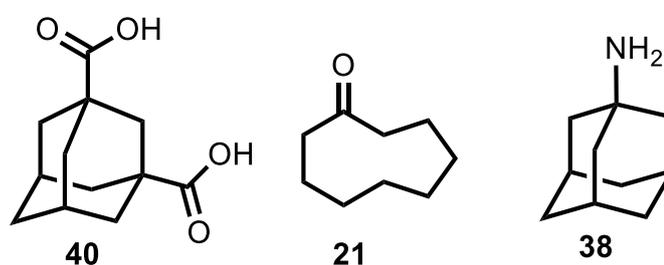
	40	21	38
Cation	-	-	13
Neutral	2.3×10^5	1.1×10^4	1.0×10^4
Anion	<300	-	-

Figure 5.18 / Table 5.2: Structures and equilibrium constants (M^{-1}) of **40**, **21**, and **38**

The diacid **40**, was chosen due to its high binding constant >10^5 M^{-1} below its pKa (pH <5). The amine **38**, was chosen because it binds strongly above its pKa (pH >11) with K = 10^4 M^{-1}. The cyclic ketone **21**, was chosen because its binding is pH independent, so should bind when the acid and amine do not (pH between 5 and 11). In addition this ketone guest has a lower K than the diacid, and so at pH <5, when the diacid is neutral, should be displaced. It does however have a similar K to the amine, so a large excess of amine will need to be added to displace the ketone at pH >11. The weaker binding 8 carbon cyclic ketone was also considered, however due to its weaker binding, more would be required to observe binding with the cage under the experimental conditions.

A two component scenario

We initially tested this idea with the two extremes (first switching between the carboxylic acid, **40** and the ketone, **21**; then separately switching between the amine, **38** and the ketone, **21**). This leave us with the two scenarios outlined in **figure 5.19**.
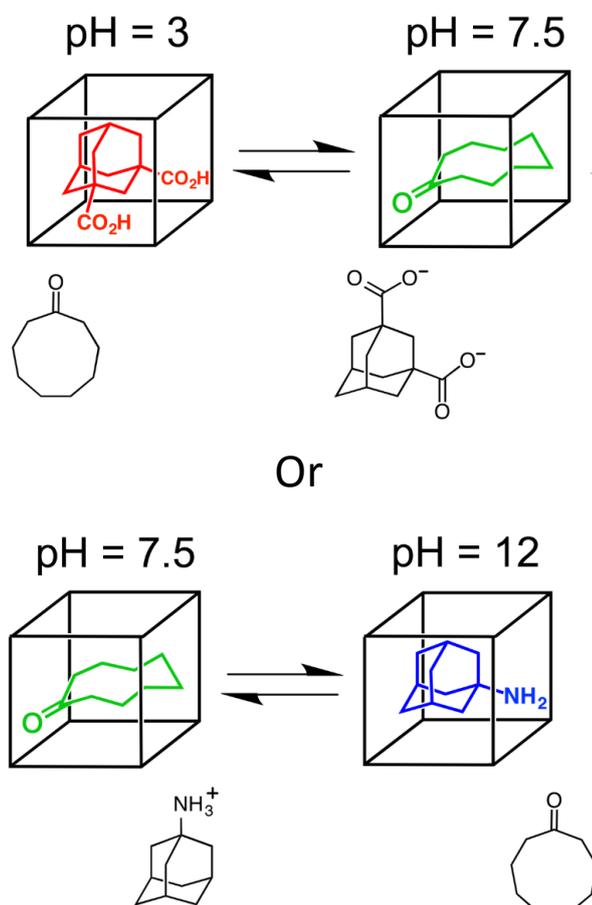
Figure 5.19: Schematic representation of the pH dependence of binding based on the combination of either the acid/ketone (switch at pH 5) (top); or the amine/ketone (switch at pH 11) (bottom).

Just by looking at the region of the ^{1}H NMR spectrum where the bound guest can be observed (around -5 to -11 ppm), it becomes simple to observe which guest is binding during the experiment, since each set of guest signals is unique. **Figure 5.20** shows what the NMR signature of each of the individual guests look like when bound inside the cage, along with the spectrum of the empty cage for comparison.
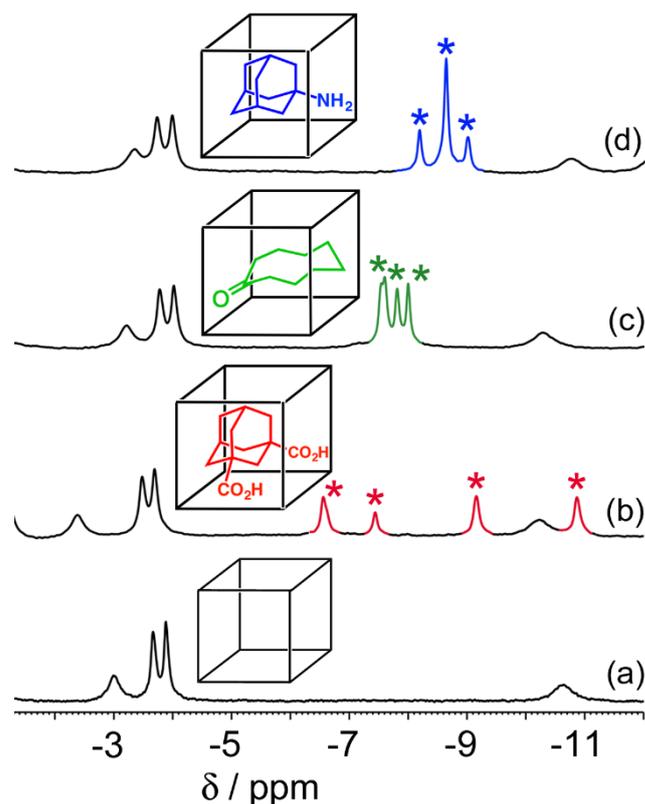
Figure 5.20: ^{1}H NMR spectra showing (a) free cage; (b) bound diacid **40**; (c) bound ketone **21**; and (d) bound amine **38**

The pH titrations were conducted with 0.2 mM cage and 0.2 mM ketone **21** with either 0.7 mM diacid **40** for the 2 component acid based pH switch; or 2.0 mM amine **38** for the 2 component base based pH switch.

The NMR spectra for each experiment are in **figure 5.21**.
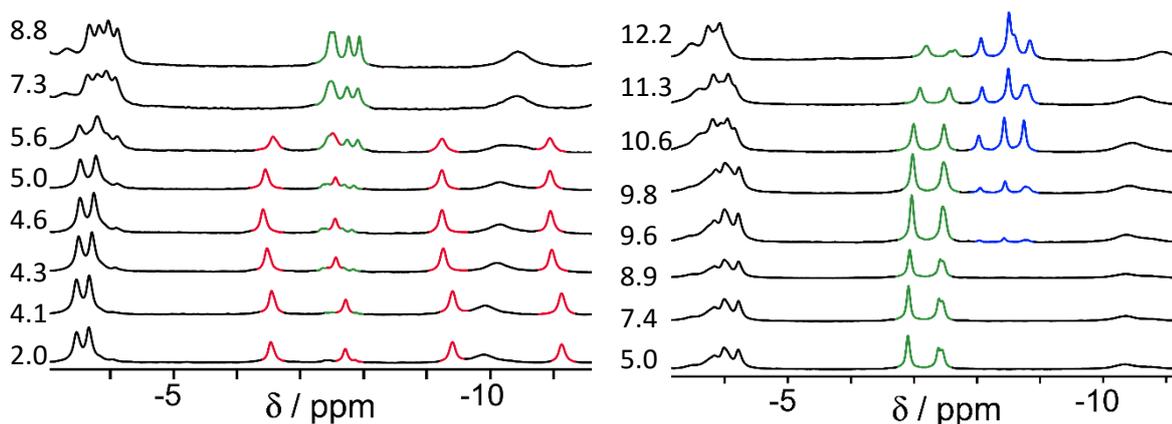
Figure 5.21: The bound guest region of the ^{1}H NMR during the pH titration with the 2 component diacid/ketone (**40/21**; left) and amine/ketone (**21/38**; right)

The 2 component switching for the diacid/ketone experiment operates in full conversion between the two guests. Essentially at the start (pH 2.0) the diacid is fully neutral, binds very strongly (>> ketone) so is the only species bound. As the pH increases, the diacid becomes charged, binds very weakly (<< ketone), so the ketone replaces the acid to bind preferentially to the cage cavity.

This same switching is observed for the amine/ketone case, however not to full conversion. This calls for a greater excess of the amine guest to fully kick out the similarly binding ketone guest.

The 3 component scenario

We then took this even further to see if we could switch between 3 guests in solution by just changing the pH to show an unprecedented degree of control in switchable guest uptake and release. For this we combined the two 2 component experiments to generate a 3 component one (**figure 5.22**).
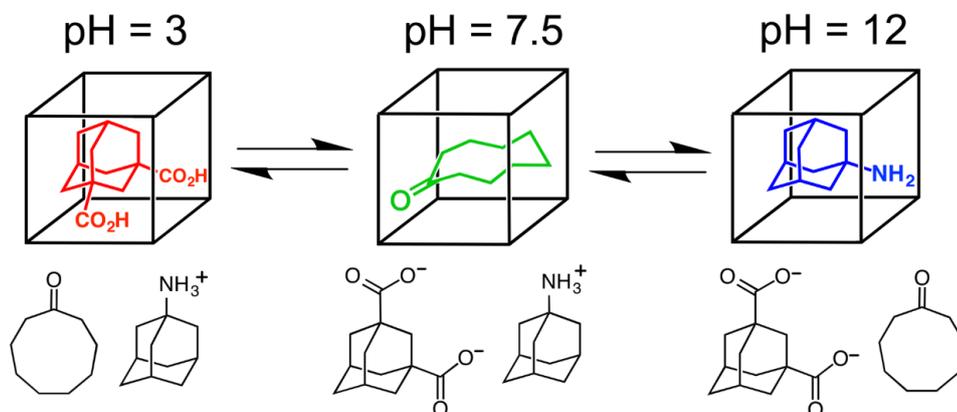
Figure 5.22: Schematic representation of the pH dependence of binding based on the combination of the 3 guests.

The pH titration was conducted with 0.2 mM cage, 0.75 mM diacid, 0.2 mM ketone, and 7.1 mM amine; and the NMR spectra obtained along with the speciation curve of the three guests is in **figure 5.23**.
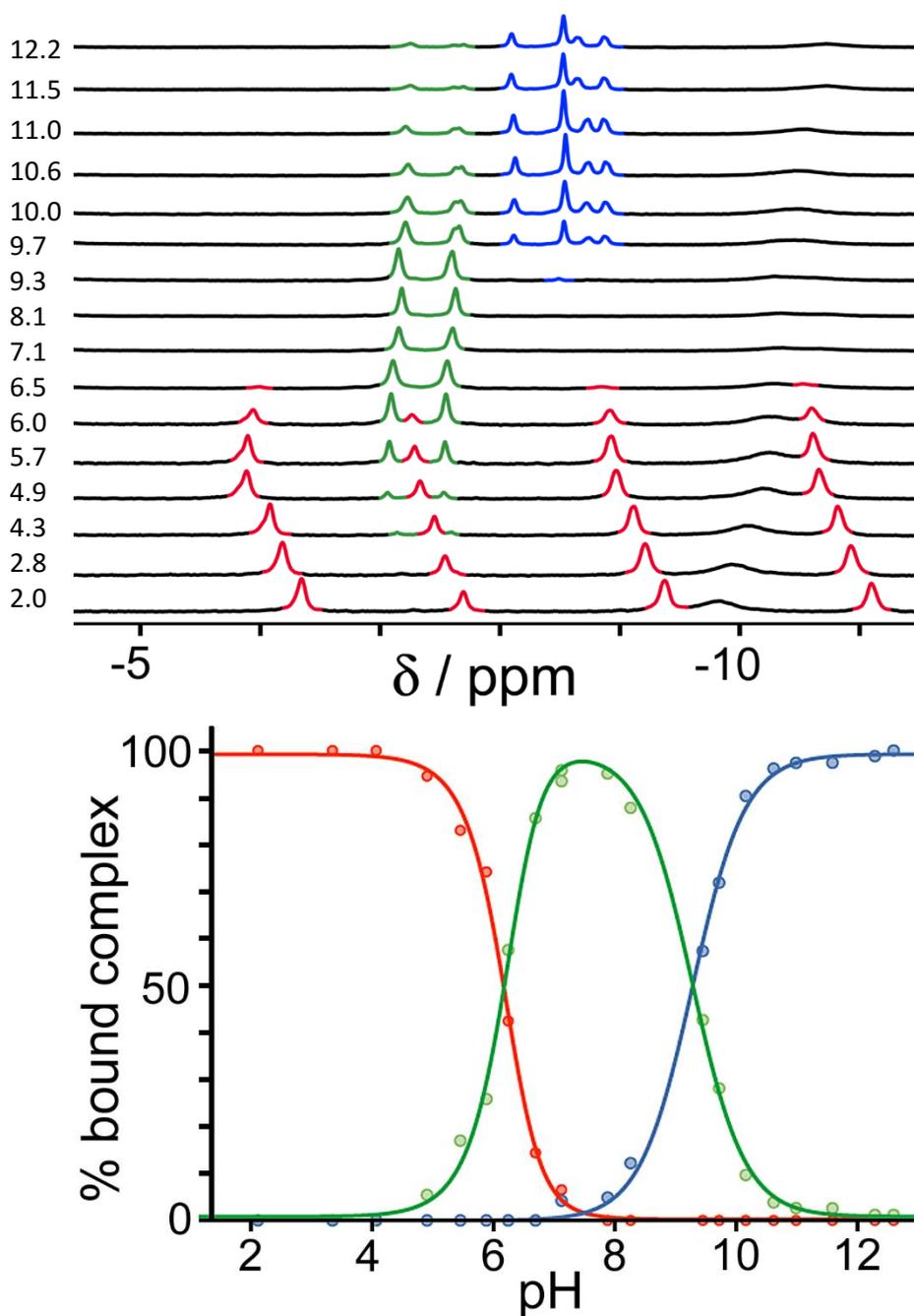
Figure 5.23: The bound guest region of the ^{1}H NMR during the pH titration going from pH 2.0 (bottom) to pH 12.2 (top)

At pH 2, the only guest that binds is the diacid. As the pH increases, and the diacid begins to become charged it leaves the cavity and some of the ketone begins to bind. At neutral pH, the only guest to bind is the ketone. As the pH increased further to the basic end, the amine guest begins to become neutral and so starts to bind, and finally at pH 12, when the amine guest is basically fully neutral, around 97% of the cage is occupied by the amine. 100% conversion could not be achieved due to the limited solubility of the amine.

5.3 Conclusions

The effect of pH on the host guest equilibrium with the M_8L_{12} cubic cage in water has been investigated. The protonation and deprotonation of the guest molecules changes their hydrophobicity and so changes the binding affinities with the cage. It turns out that the sign of the charge, whether it be positive or negative, appears to have little to no effect on the binding. The change in binding affinity is largely down to increased polarity of the cationic or anionic forms, and therefore the guest prefers to be solvated by the bulk water solvent rather than be encapsulated in the hydrophobic cavity of the cage.

This behaviour has been demonstrated on a number of guests with different functional groups throughout the entire pH range, including some drug molecules.

This principle of using pH changes to control guest binding was extended to demonstrate how a host cage can select one of three possible guests from a mixture using a single external stimulus (a pH change) – an unprecedented degree of control over guest binding. For any potential applications of molecular containers in which stimulus-responsive guest binding is an important factor, this ability to switch reversibly between any one of *multiple* bound states using a *single* stimulus represents a new level of sophistication and control in host guest chemistry which will expand the range of functions that can be developed.

5.4 Experimental techniques and synthetic procedures

Chemicals

All chemicals were purchased from commercial sources and used as supplied unless otherwise stated.

Nuclear magnetic resonance spectra

All NMR data was collected using a Bruker Avance III 400 MHz NMR spectrometer at 298 K with the parameters: 3072 scans, 284 ppm sweep width, O1p value of 0 ppm, D1 0.5 s, aq 0.5 s. The NMR data was processed using Bruker Topspin 3.1

5.4.1 Measurements

pH measurements

All pH measurements were made using a Hamilton Spintrode pH combination electrode at 298 K and calibrated with calibration standards at pH 4.01, 7.0 and 10.01

Host solution - The water soluble M_8L_{12} cubic cage (16 mg) was dissolved in 10 ml D_2O to make a 0.2 mM stock solution.

Guest solution – The guests were individually dissolved in 5 ml of host solution (to keep the host concentration constant during the titration). The mass used varied with the desired concentration and the molecular masses of the guests themselves.

All experiments were repeated twice to give sets of data, and the values are quoted as the average with an error of two standard deviations from the mean.

pH titrations – slow exchange

The guest solutions were equilibrated using a water bath thermostated at 298 K. The pH was measured and then adjusted to the desired value by addition of NaOD or DCl (1 M). The ^1H NMR spectrum for each addition was measured.

The pKa was calculated by plotting pH against the chemical shift of the free guest peaks, and fitted using the Microsoft Excel add-on, Solver, by minimising the sum of the errors between calculated (**Eq. 1**) and measured chemical shift.

$$\delta_{Calc} = \alpha_0 \times \delta(A) + \alpha_1 \times \delta(AH) \qquad \textbf{Eq. 1}$$

Where δ is the chemical shift of the peaks in question (ppm), δ(A) is the chemical shift of the fully deprotonated species; δ(AH) is the chemical shift of the fully protonated species and α_0 and α_1 follow the relationship outlined in **Eq. 2**.

$$\alpha_0 = \frac{10^{-pKa}}{10^{-pH} + 10^{-pKa}} \text{ and } \alpha_1 = \frac{10^{-pH}}{10^{-pH} + 10^{-pKa}} \qquad \textbf{Eq. 2}$$

For the 2 pKa guest (1,3-adamantane dicarboxylic acid) the equations are as follows:

$$\alpha_2 = \frac{10^{-pKa1} \times 10^{-pKa2}}{(10^{-pH})^2 + (10^{-pKa1} \times 10^{-pKa2}) + (10^{-pKa1} \times (10^{-pH}))}$$

$$\alpha_1 = \frac{10^{-pH} \times \alpha_2}{10^{-pKa2}}$$

$$\alpha_0 = 1 - \alpha_1 - \alpha_2 - \alpha_3$$

$$\delta_{Calc} = \alpha_0 \times \delta(AH_2) + \alpha_1 \times \delta(AH^-) + \alpha_2 \times \delta(A^{2-})$$

For the 3 guests in a mixture curve in fitting the speciation of cyclononanone; it indirectly has 3 pKa values (2 for the acid side, and 1 for the amine side); the equations become:

$$\alpha_3 = \frac{10^{-pKa1} \times 10^{-pKa2} \times 10^{-pKa3}}{(10^{-pH})^3 + (10^{-pKa1} \times 10^{-pKa2} \times 10^{-pKa3}) + (10^{-pKa1} \times 10^{-pKa2} \times 10^{-pH}) + (10^{-pKa1} \times (10^{-pH})^2)}$$

$$\alpha_2 = \frac{10^{-pH} \times \alpha_3}{10^{-pKa3}}$$

$$\alpha_1 = \frac{10^{-pH} \times \alpha_2}{10^{-pKa2}}$$

$$\alpha_0 = 1 - \alpha_1 - \alpha_2 - \alpha_3$$

$$\delta_{Calc} = \alpha_0 \times \delta(AH_3) + \alpha_1 \times \delta(AH_2^-) + \alpha_2 \times \delta(AH^{2-}) + \alpha_3 \times \delta(A^{3-})$$

The binding constant at each pH point was calculated by measuring the integral of the free host and host-guest peaks relative to each other using Topshim's deconvolution feature. This was done for several pairs of peaks and the average value was plotted.

pH titrations – fast exchange

The host and guest solutions were equilibrated using a water bath thermostated at 298 K. The pH of both solutions were measured and then adjusted to the same desired value by the addition of NaOD or DCl (1 M). 12 samples of varying host and guest concentrations were made up going from pure host solution to pure guest solution, to a total volume of 600 ml per sample. The samples were ran using an 8 inch NORELL 507-HP NMR tube, sealed with pressure caps to ensure no solvent loss occurred over the course of the titration. The ^{1}H NMR spectra were run for each sample and the concentration of guest was plotted against change in chemical shift (ppm) and fitted to obtain a value for the binding constant. The titration was completed at two pH values (fully protonated and dully deprotonated) to get the binding constants for the two extremes.

5.4.2 X-ray crystallography

The crystal structure data collection of the complex $[Co_8L_{12} \bullet (5.2)](BF_4)_{16}$ was performed at the EPSRC National Crystallography Service at the University of Southampton, UK. Data were corrected for absorption using empirical methods (SADABS) based upon symmetry-equivalent reflections combined with measurements at different azimuthal angles. The structure was solved and refined using the SHELX suite of programs. The asymmetric unit contains one half of the molecule which lies astride an inversion centre. A combination of disorder of anions / solvent molecules and solvent loss resulted in weak scattering, necessitating use of extensive geometric and displacement restraints to keep the refinement stable.

The asymmetric unit contains one half of the cage complex which lies astride an inversion centre, as well as one complete guest molecule whose atoms all have site occupancies of 0.5. Thus, the complete complex contains one guest molecule disordered over 2 symmetrically equivalent (and spatially overlapping) orientations with the O atom pointing towards diagonally opposite corners Co(1) and Co(1A). The usual disorder of anions / solvent molecules and solvent loss characteristic of cage complexes of this type resulted in weak scattering, necessitating use of extensive geometric and displacement restraints to keep the refinement stable: these are described in detail in the CIF. We could locate and refine four of the expected eight $[BF_4]^-$ anions in the asymmetric unit; all show disorder of the F atoms. Large regions of diffuse electron density which could not be modelled, accounting for the remaining anions plus solvent molecules, were eliminated from the refinement using of the 'SQUEEZE' function in the PLATON software package.

We cannot use charge balance considerations to determine whether or not the guest is protonated since the distinction is not crystallographically obvious as extensive disorder of the tetrafluoroborate anions in the structure means that not all of them could be located. Although the two C—O bond distances of the carboxylic acid (or carboxylate) group appear to be approximately equivalent, *i.e.* there is no obvious short (double) and long (single) distinction between the C—O bonds, the presence of disorder of the entire guest over two orientations – plus the additional possibility of C=O / C–OH disorder within each orientation – means that we cannot draw any conclusion from the bond lengths.

Crystallography data table

Complex	[Co$_8$L$_{12}$](BF$_4$)$_{16}$•(adamantane carboxylic acid)
Formula	C$_{371}$H$_{328}$B$_{16}$Co$_8$ F$_{64}$N$_{72}$O$_{26}$
Molecular weight	8071.43
T, K	100(2)
Crystal system	Monoclinic
Space group	$C2/c$
a, Å	27.3936(7)
b, Å	39.1227(10)
c, Å	42.964(3)
α,˚	90
β,˚	107.152(8)
γ,˚	90
V, Å^3	42973(4)
Z	4
ρ, g cm^{-3}	1.248
Crystal size, mm^3	0.23 x 0.18 x 0.1
Data, restraints, parameters	49124, 2560, 1889
Final R1, $wR2^{b}$	0.1630, 0.4367

a   These formulae (and consequently the crystal densities) are necessarily approximate given that large amounts of diffuse electron density in solvent-accessible voids was removed from the refinements using the OLEX 'Solvent Mask' function. See CIFs, and comments in experimental section, for details.

b   The value of R1 is based on 'observed' data with $I > 2\sigma(I)$; the value of wR2 is based on all data.

5.5 References

1) P. Mal B. Breiner; K. Rissanen; J. R. Nitsche, *Science*, **2009**, *324*, 1697-1699

2) M. M. J. Smulders; J. R. Nitschke, *Chem. Sci.,* **2012**, *3,* 785-788

3) B. Therrien, *Chem. Eur. J.,* **2013**, *19*, 8378-8386

4) J. E. M. Lewis; E. L. Gavey; S. A. Cameron; J. D. Crowley, *Chem. Sci.,* **2012**, *3*, 778-784

5) M. Han: R. Michel; B. He; Y. S. Chen; D. Stalke; M. John; G. Clever, *Angew. Chem. Int. Ed.,* **2013**, *52*, 1319-1323

6) W. Y. Sun; T. Kusukawa; M. Fujita, *J. Am. Chem. Soc.,* **2002**, *124*, 11570-11571

7) M. M. J. Smulders; S. Zarra; J. R. Nitschke, *J. Am. Chem. Soc.,* **2013**, *135*, 7039-7046

8) http://www.netdoctor.co.uk/seniors-health/medicines/symmetrel.html

9) http://www.parkinson.org/Parkinson-s-Disease/Treatment/Medications-for-Motor-Symptoms-of-PD/Amantadine-(Symmetrel).aspx

10) MacroModel, version 9, Schrödinger Inc., New York, NY, 2011

Chapter 6

Highly efficient catalysis of the Kemp Elimination in the cage cavity

6.1 Introduction

One very promising area of host-guest chemistry is supramolecular catalysis (chapter 1), since artificial container molecules have the ability to rival the selectivity and rate enhancements for reactions on bound substrates than those seen in biology. Container molecules provide relatively rigid central cavities that may mimic binding pockets in enzymes, and the mechanisms by which accelerations in rates are achieved are similar to those of enzymes too. Catalysis requires stabilising (and so lowering the energy) of the transition state; or an increase in effective molarity by bringing two or possibly more reactants close together in the same confined space.[1-4]

The similarities between artificial container molecules and enzyme binding sites have allowed studies into the mechanisms by which reactions occur in biology, since the relative simplicity of the artificial systems allows easy probing of the reactions as they occur.[1-4] One instance where this has been applied is with the tail-to-head terpene (THT) cyclisation, to prepare complex terpene natural products.[5] The THT cyclisation mechanism is not very well understood due to the complex array of terpenes that are synthesised in nature in the very complex binding site in the cyclase enzyme. The THT cyclisations are often referred to as the most complex chemical reactions that occur in nature! By using a resorcinarene capsule as the host (**figure 6.1 top**), the Tiefenbacher group showed evidence that direct isomerisation (**figure 6.1, blue pathway**) was the mechanism for the cyclisation, as opposed to the indirect mechanism (**figure 6.1, red pathway**).[5]
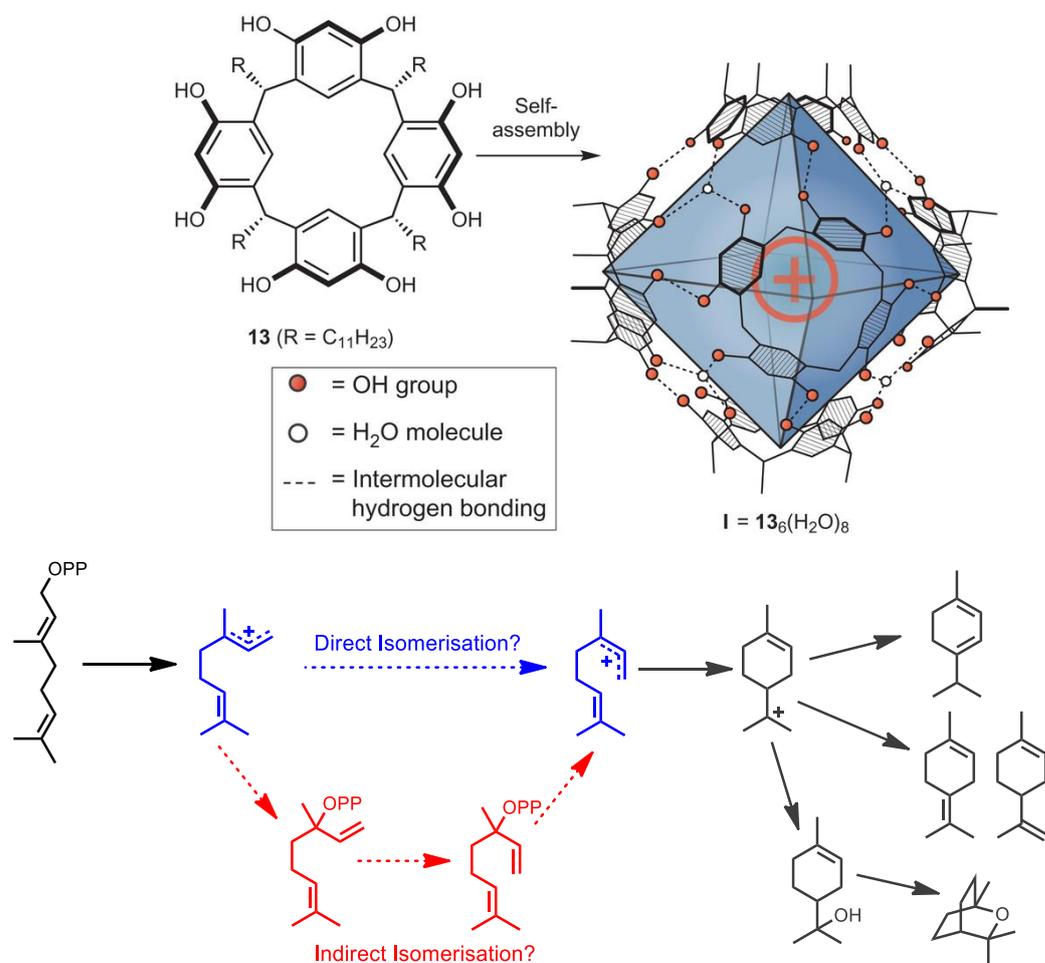
Figure 6.1: The resorcinarene capsule host (top) and the THT cyclisation mechanism, showing either direct (blue) or indirect (red) isomerisation as one of the key mechanistic steps (bottom).[5]

In terms of sheer rate enhancements, currently the biggest rate enhancement based on a reaction inside a synthetic host is for the Nazarov cyclisation which goes 2.1 x 10^6 times faster in a specific cage than the background reaction. This remarkable enhancement of the reaction rate, rivals those seen in nature (**figure 6.2**).[6]
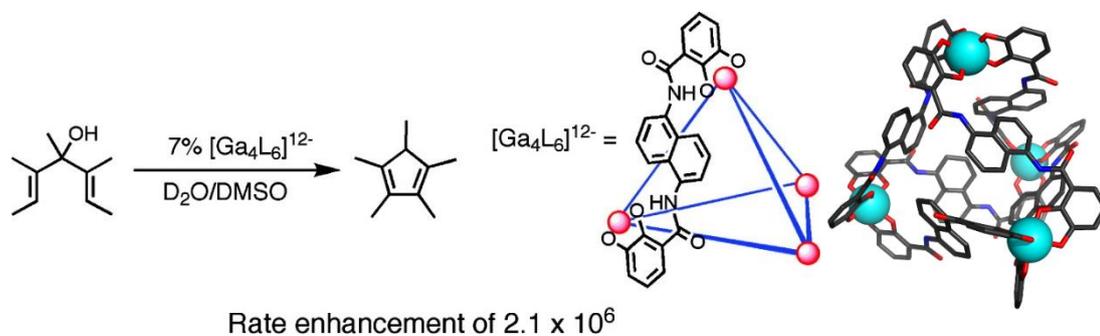
Figure 6.2: The Nazarov cyclisation - the current biggest rate enhancement.[6]

One of the biggest challenges with artificial capsules as catalysts is product inhibition. Often the favourable interactions that stabilise the transition state also stabilise the product, and so the product often binds to the cavity too, thus inhibiting the reaction and preventing catalytic turnover.[1-4] This is a big limitation with the efficiency of such catalytic systems: they can be fast but have few cycles.[1-4,7]

Some strategies to combat this have been designing reactions where the product has a different geometry from the starting material which creates steric clashes with the cavity and so it is expelled (**figure 6.3**).[8]
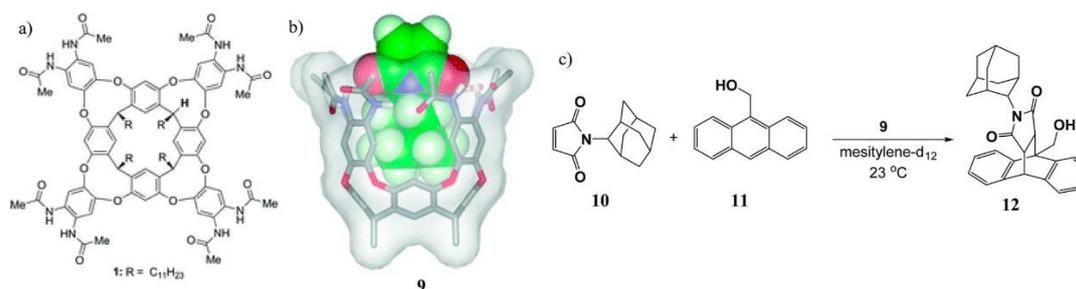
Figure 6.3: Catalytic turnover achieved by shape mismatch between product and cavity.[8]

Another strategy was employed by Raymond and co-workers where the strongly binding product formed from the catalytic step undergoes a second reaction (in the case in **figure 6.4**, is hydrolysis) to form a weakly binding product, and so the catalyst can turnover.[9]
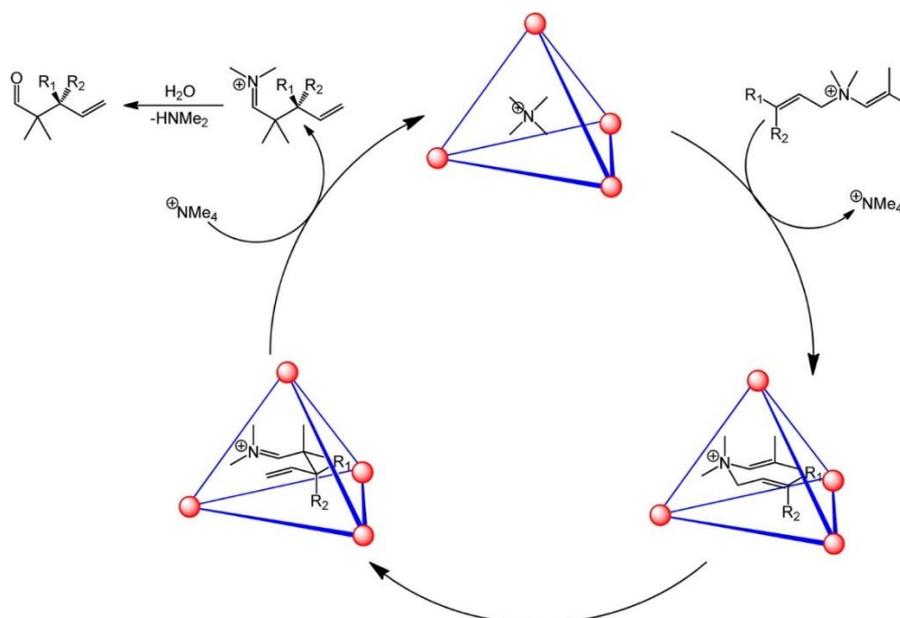
Figure 6.4: Catalytic turnover achieved by hydrolysis of the initially-generated product to remove the strongly binding catalysis product.[9]

6.2 Results and Discussion

6.2.1 Background reaction

The reaction we chose for this chapter was the Kemp elimination (**figure 6.5**).[10] This reaction follows the base catalysed decomposition of 1,2-benzisoxazole (a hit from the GOLD screen). We chose this reaction because the starting material is neutral and binds strongly in the cage cavity (log$K$ = 3.6) but the product is charged (pKa 6.8, therefore anionic under prevailing conditions), and from the work outlined in chapter 5, charged molecules do not bind as they prefer to be solvated. Thus the conversion from neutral strongly binding species to charged non-binding species will be the basis of product release, and thus catalytic turnover (**figure 6.6**).
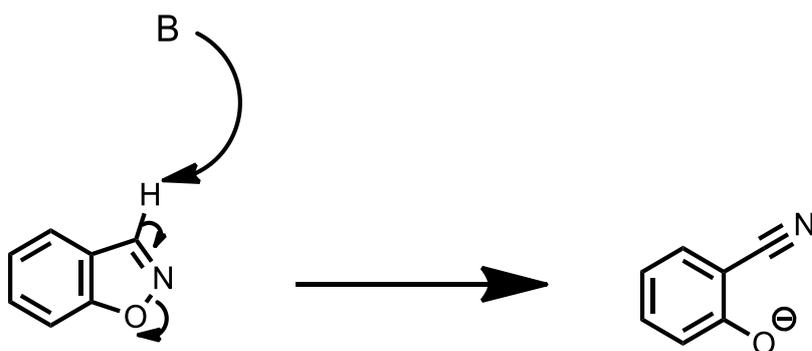
Figure 6.5: Base catalysed decomposition of 1,2-benzisoxasole
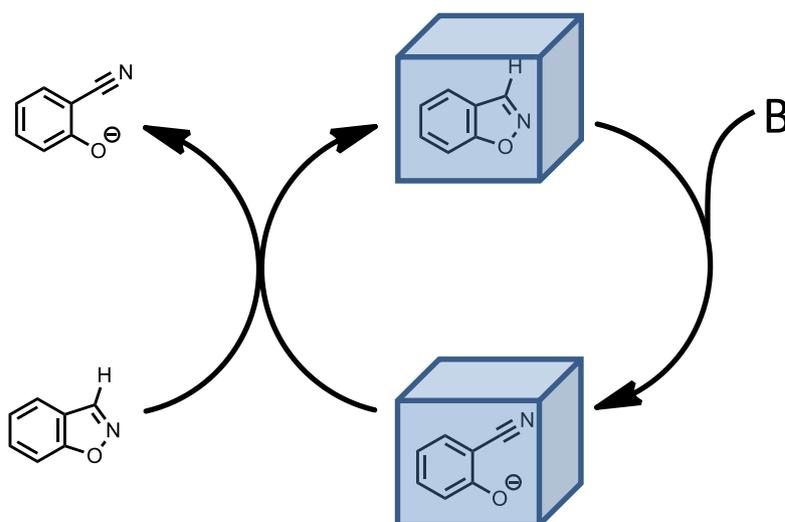
Figure 6.6: Schematic representation of the catalytic cyclic of the Kemp elimination in the presence of cage, showing catalytic turnover

The Kemp elimination is an extremely well-studied reaction,[10,11] particularly as an E2 elimination that can be adapted to a wide range of reaction rates, and as a sensitive probe for catalytic systems, both biological and artificial. It is an important reaction to study because it involves a C-H proton transfer step that is a key process in many biological systems.[10,11] The reaction is first order with ⁻OH under basic conditions and reaches a minimum rate around pH 6 where water rather than hydroxide becomes the base.[10,11] The background reaction (i.e. with no catalyst present) was measured at a range of pD values (**figure 6.7**) using UV/vis spectroscopy to monitor the appearance of the product (easily observable by the strong absorbance at around 320 nm), and our observed measurements of reaction rate (blue dots) match well with the literature (black line).[10,11]
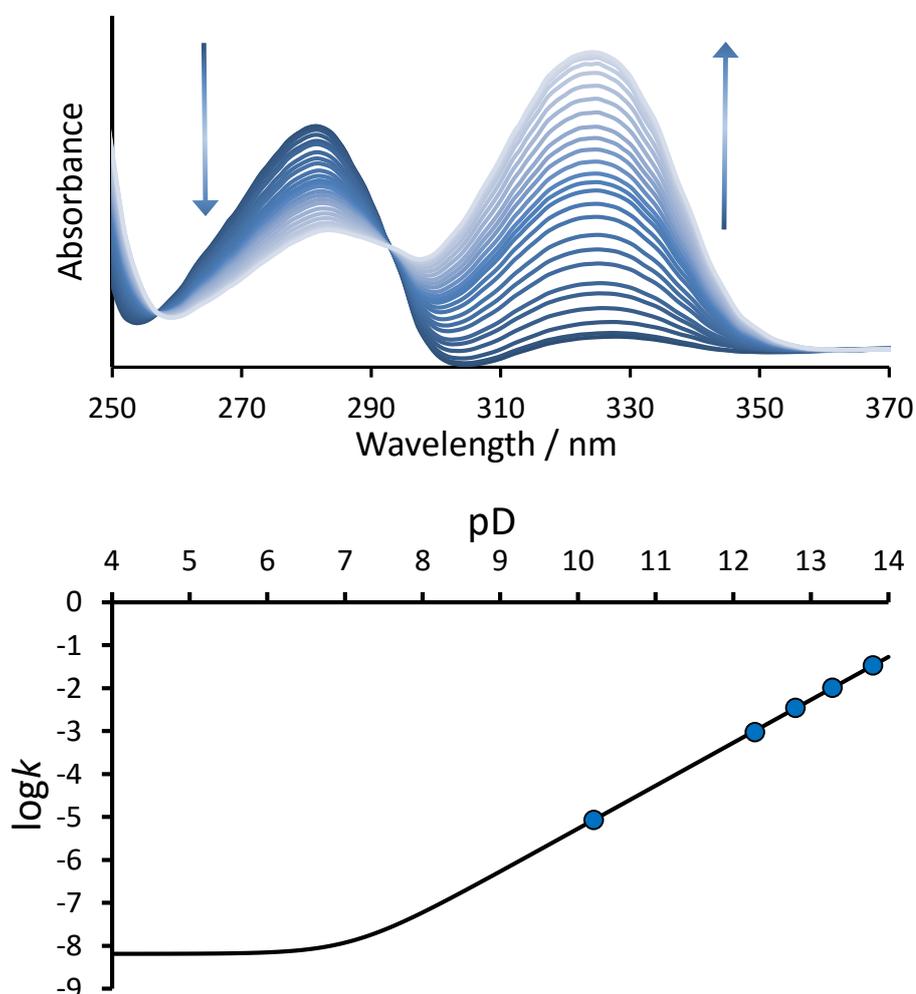
Figure 6.7: Background reaction showing the UV/Vis absorbance spectra of disappearance of starting material and emergence of product with time (top); the first order rate dependence with ⁻OD, with our measurements (blue dots) and literature measurements (black line) (bottom)

Note the use of pD rather than pH, since we are using D_2O rather than H_2O. The pH meter readings can be corrected by the addition of +0.4 pH units to give the pD values.

The blue dots have a gradient of one, which when on a log plot, means that as the pD decreases by one, the logk also decreases by one, and therefore indicates first order with base.

6.2.2 Caged reaction

Now that the background behaviour was understood, we moved on to the reaction in the presence of cage. Initially we attempted this by UV/vis spectroscopy, identical to the background reactions. This proved to be problematic due to the extremely strong absorbance of the cage masking both the starting material and product absorbance bands. However in the presence of very small amounts of cage at very low concentrations, initial rate data could be obtained. **Figure 6.8** shows the emergence of product at cage loadings of 0, 5 and 10 mol% compared to starting material.
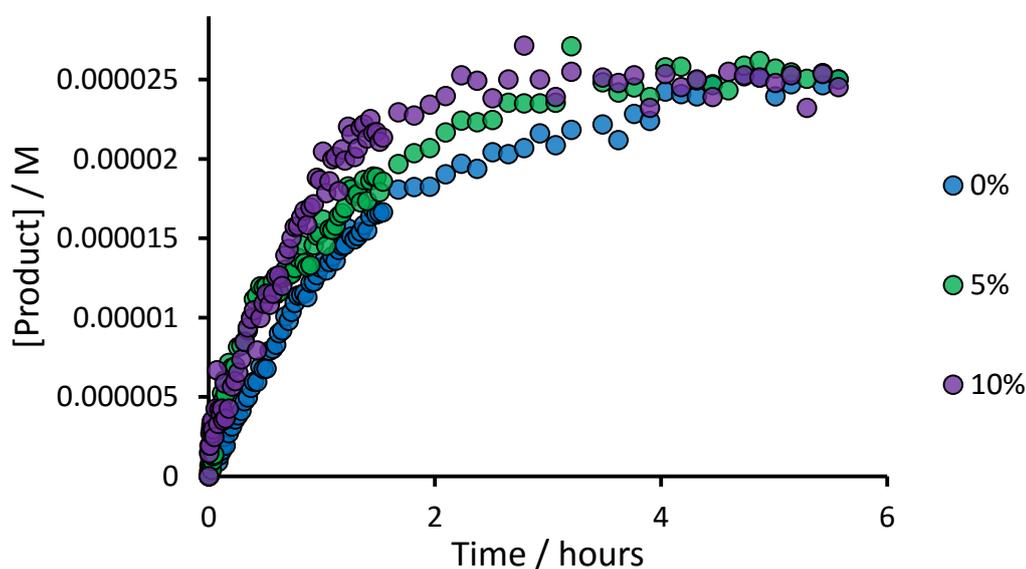
Figure 6.8: the emergence of product at cage loadings of 0, 5 and 10 mol% compared to starting material

The change in observed rate is not very significant, since at these concentrations (2.5 x 10^{-5} M benzisoxazole; 0, 1.25 x 10^{-6} and 2.5 x 10^{-6} M cage) only 11% of the cage will be bound, with only 1% of the benzisoxazole bound. So only 1% of the initial substrate undergoes the catalytic pathway. To get the catalysed rate (k_{cat}) the full curve was fitted

using Michaelis-Menten kinetics. At this pD (11.2) the k_{cat}/k_{uncat} (so the rate enhancement of the catalysed rate versus the uncatalysed rate) is 200. The apparent rate of product formation is doubled, and at this 100:1 ratio of free:bound guest, there is a 200 times rate acceleration for the 1% of guest that is bound.

The small changes in observed rate are not optimal, and we wanted to work in the domain where we just observe the catalysed pathway. In order to do this we needed to work at high cage concentrations, and low guest concentrations, so that there is little 'free' guest, and therefore the background uncatalysed pathway is minimalised. This is often referred to as sub-saturation kinetics, and is used regularly with enzyme kinetics since it greatly simplifies the experiment. This means that turnover numbers can't be measured but that is done separately. The advantage is that we can assume that (almost) all of the observed reaction is going through the catalysed pathway. Operating at such high concentrations of cage makes it impossible to monitor with UV/vis spectroscopy (**figure 6.9**).
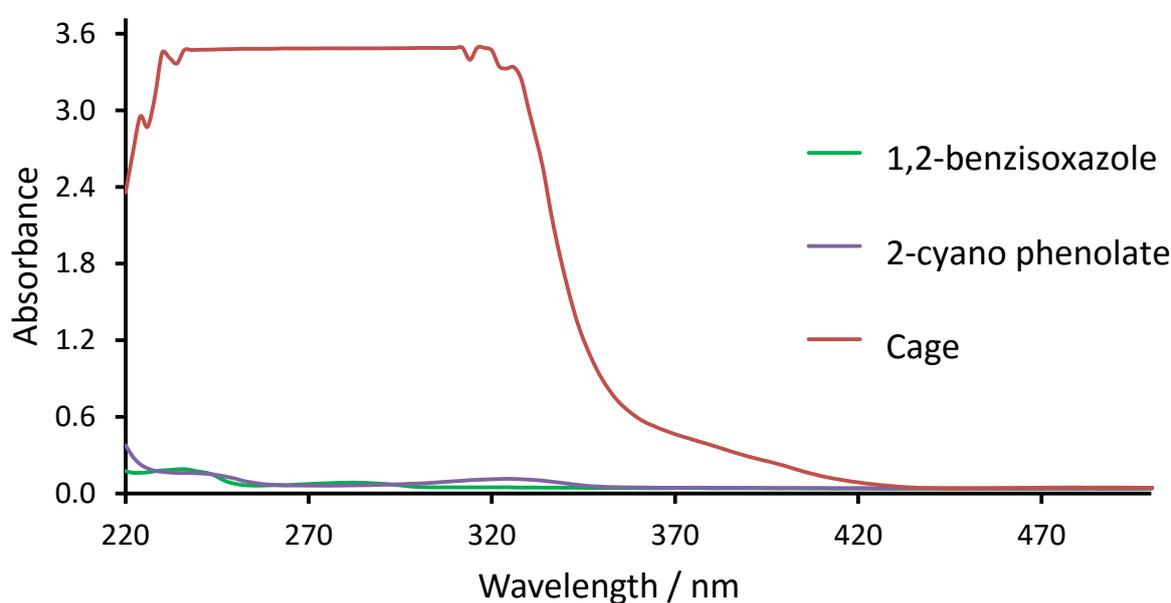
Figure 6.9: UV/vis spectra of the cage (red, 0.1 mM), benzisoxazole starting material (green, 0.01 mM), and cyano phenolate product (purple 0.01 mM).

Because of this, we decided to move to using ^{1}H NMR spectroscopy to monitor the catalysed reaction. Ideally we wanted to work at cage concentrations at 10/K (2.5 mM) however the solubility of the cage only permitted a 1 mM cage concentration. We then chose the guest concentration at 0.85 mM since this allowed majority of benzisoxazole

>60% to be bound, whilst maintaining enough substrate to be reliably monitored by NMR (lower benzisoxazole concentrations, whilst allowing for less 'free' guest, would require more NMR scans, and so each data point acquisition will be too long for the kinetic scale of the experiment).

The experiment was initially run in D_2O at pD 10.2, and the NMR spectra of this as a function of time are shown in on the next page in **figure 6.10 top**. The disappearance of the 'free' starting material (green) and appearance of product (blue) can be observed. The effect of the cage on the reaction is immediately apparent when comparing NMR spectra at the same pD in the absence of cage (**figure 6.10, middle** – product red), and the combined rate plot (**figure 6.10, bottom**). At this pD the k_{cat}/k_{uncat} is 4500.
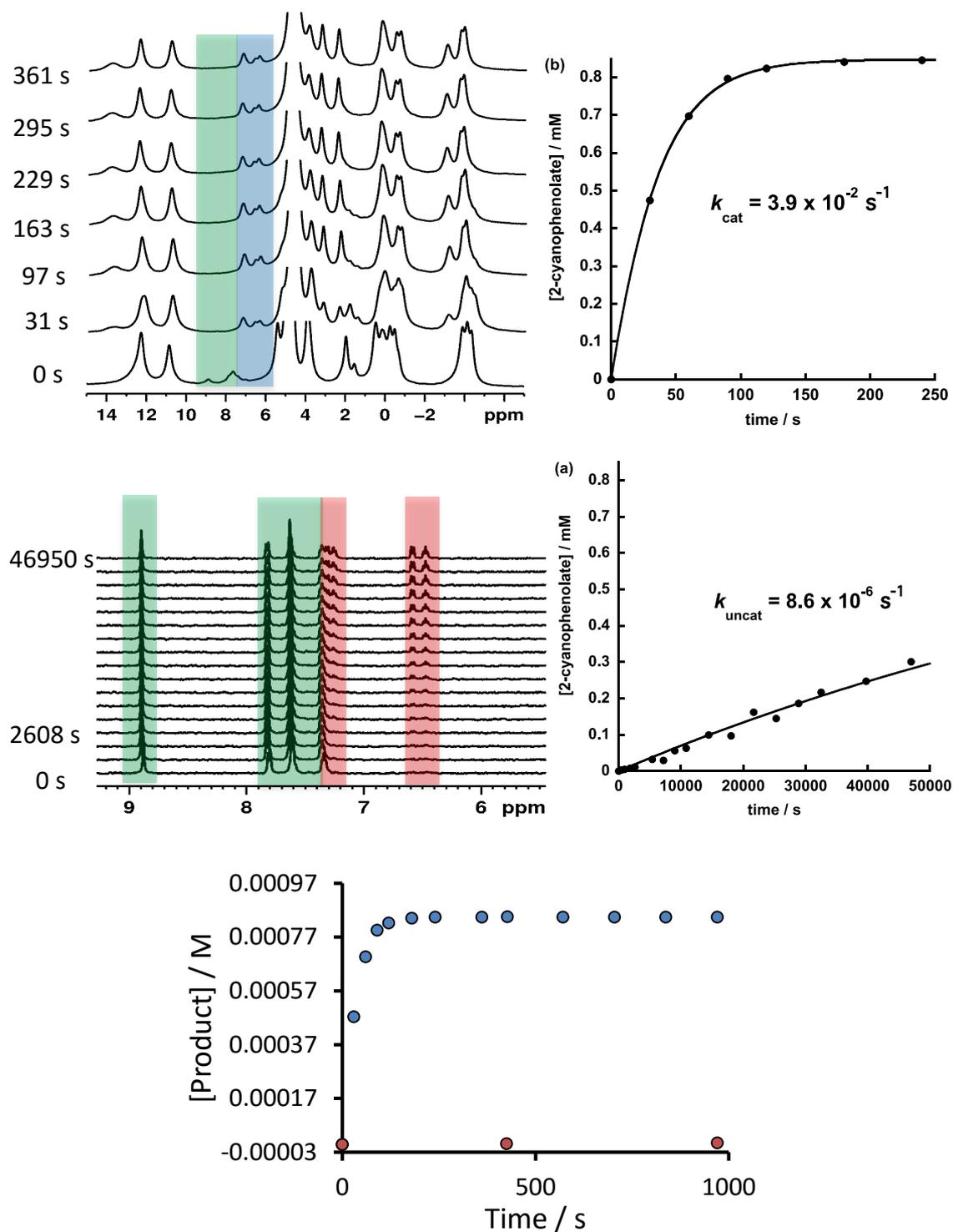
Figure 6.10: ^{1}H NMR spectra of the cage catalysed reaction and its fit (top, SM – green, P – blue); the background reaction and its fit (middle, SM – green, P – red); and the combination of the two rate profiles on the same scale (bottom)

Examination of at the cage ^{1}H NMR signals at around 60-100 ppm (**figure 6.11**), shows that at the beginning of the reaction most of the cage is bound (green), and as time progresses, the bound cage peaks disappear and the free cage peaks (blue) appear. This shows that the charged product does indeed not bind to the cage cavity.
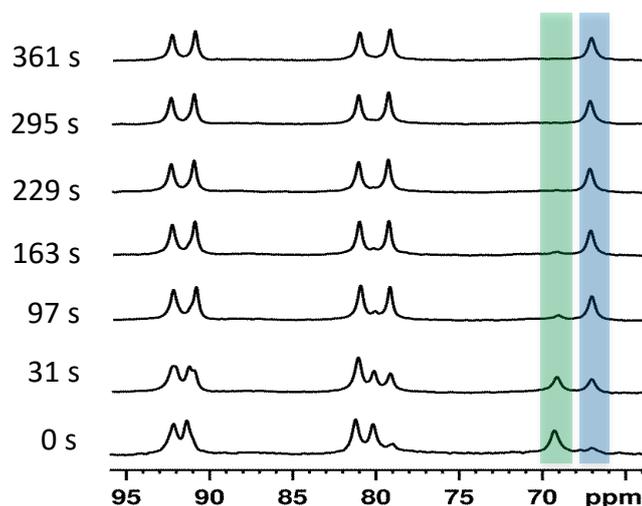
Figure 6.11: ^{1}H NMR spectrum of the cage peaks showing bound cage (green) becoming free cage (blue) as the reaction progresses.

The reaction in the presence of cage was performed at multiple pD values to see the dependence on hydroxide on the catalysed reaction rate. Surprisingly it was found to be effectively pD independent (**figure 6.12**) with the same reaction rate over the pD range 8.5-11.5 (at higher pD, the cage starts the decompose; at lower pD, the product is not released as it is protonated and binds to the cage with K = 2200 M^{-1}). The biggest rate enhancement of k_{cat}/k_{uncat} measured (pD 8.5) is 2 x 10^5 (green arrow), and the biggest theoretical rate enhancement by comparing the pD independent regions is 6 x 10^6 (orange arrow), which is greater than the previous rate enhancement record of Raymond's 2.1 x 10^6. This rate enhancement is not reachable in practice, since the catalysed rate begins to drop below a pD of 8; however the rate enhancement of 2 x 10^5 at pD 8.5 is the 2nd best so far reported.
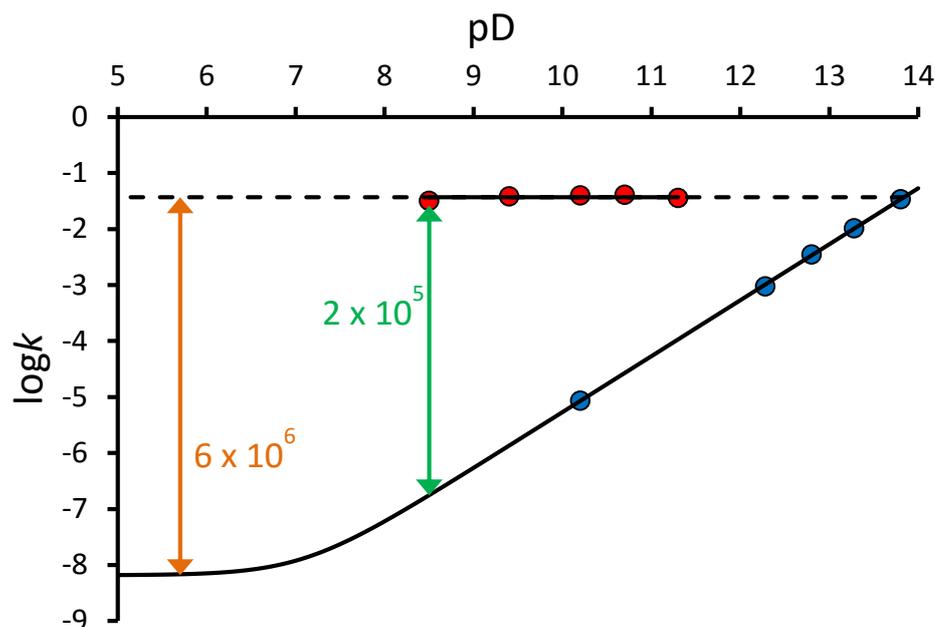
Figure 6.12: log of the rate vs pD showing the pD dependence of the background reaction (blue) and the cage catalysed reaction (red); the maximum measured rate enhancement (green arrow) and the maximum theoretical rate enhancement (orange arrow)

6.2.3 The mechanism of catalysis and crystallography

Two questions arise from these observations. Firstly, how does the cage act as a catalyst for this reaction? As the reaction proceeds the transition state involves a build-up of negative charge on the O atom, which could be stabilised by adjacent H-bond donors. However, we know that the H-bond donor pocket inside the cage cavity is less effective at stabilising H-bond acceptor sites than water: a carbonyl-containing guest in this pocket is *destabilised* relative to solvation by water by about 7 kJ mol^{-1} (see chapter 3). This penalty must be even larger for a negatively charged transition state. The interior of the cage therefore provides a *poorer* medium than water for the reaction because of preferential solvation of the developing negative charge by water. In these terms, binding the substrate is anti-catalytic.

Secondly, what is the reason for the pD-independence of k_{cat}? As there are no basic sites associated with the cage, the rate invariance in the pD 8.5 – 11.4 range could be explained by the use of water as the base for the reaction rather than hydroxide. However

if the cage interior does not stabilise the developing negative charge on the product, it is unlikely that a weak base (water) could replace the hydroxide involved in the solution reaction and give the high rates observed for the catalysed reaction. Thus neither the origin of the catalysis, nor the pD dependence of its rate, can be explained just by consideration of the environment inside the cage cavity.
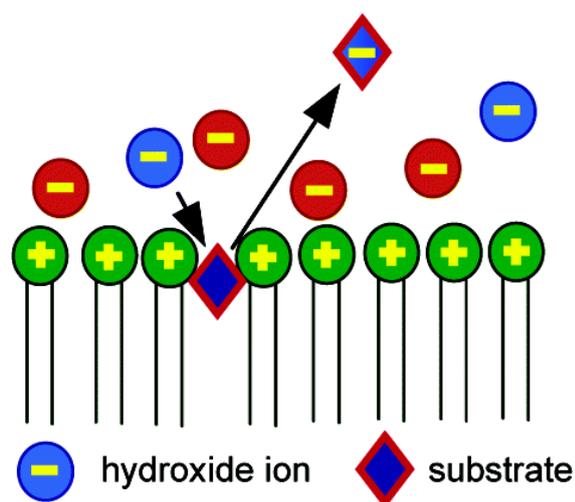
Figure 6.13: Cartoon of how micelles and vesicles accelerate reactions by attracting the negatively charged hydroxide to the positively charged surface, resulting in increased local concentration of partially desolvated (aka activated) hydroxide[12]

The high reactivity and region of pD independence are both consistent with a model that has been developed for catalysis by micelles and vesicles.[12-15] It is proposed in this model that ion-pairing effects result in accumulation of hydroxide ions around the positively charged surfaces of the micelle or vesicle, resulting in both a high local concentration of hydroxide ions and partial desolvation of the hydroxide ions which increases their reactivity (**figure 6.13**). For cationic vesicles, these effects lead to a maximum observed rate acceleration for the reaction of benzisoxazoles of about 800 fold.[15]

We propose that the surface of the highly positively charged cage catalyst (16+) acts in a similar way, concentrating partially desolvated hydroxide ions around its surface. We know from numerous structural studies that the windows in each face of the cage are invariably occupied by anions in the crystal structures (**figure 6.15, left**), which would position hydroxide close to the CH of the substrate constrained in the cavity. The crystal structure of benzisoxazole encapsulated in the cage cavity (**figure 6.14, left**; obtained by

soaking empty cage crystals in the benzisoxazole oil) shows this CH proton pointing towards the pockets (**figure 6.14, right; yellow H**)
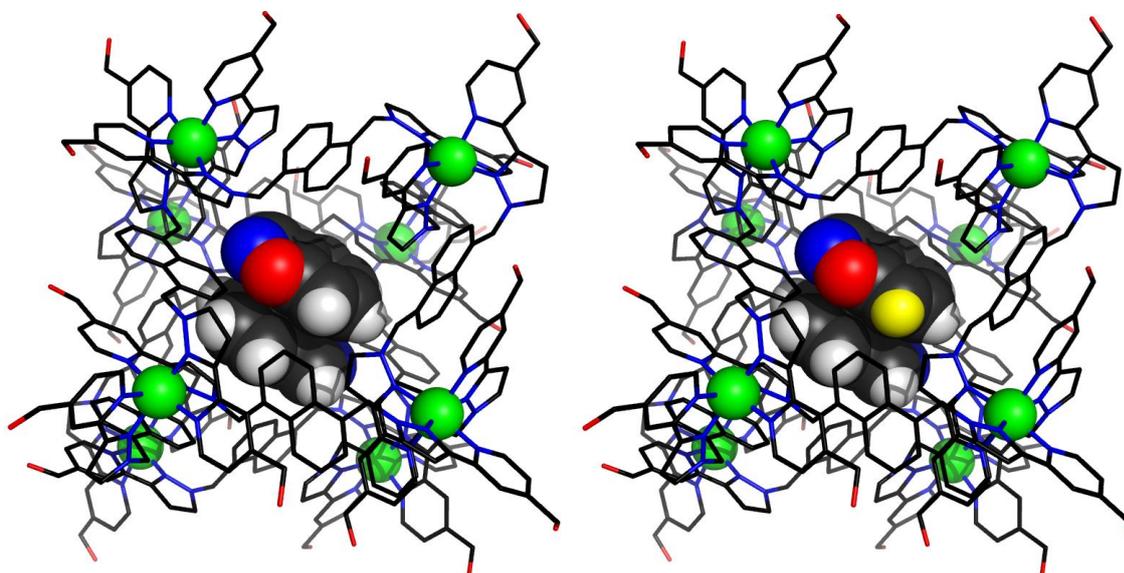
Figure 6.14: X-ray crystal structure of benzisoxazole inside the cage cavity (left) and the CH proton (yellow) involved in the Kemp elimination pointing towards the window.
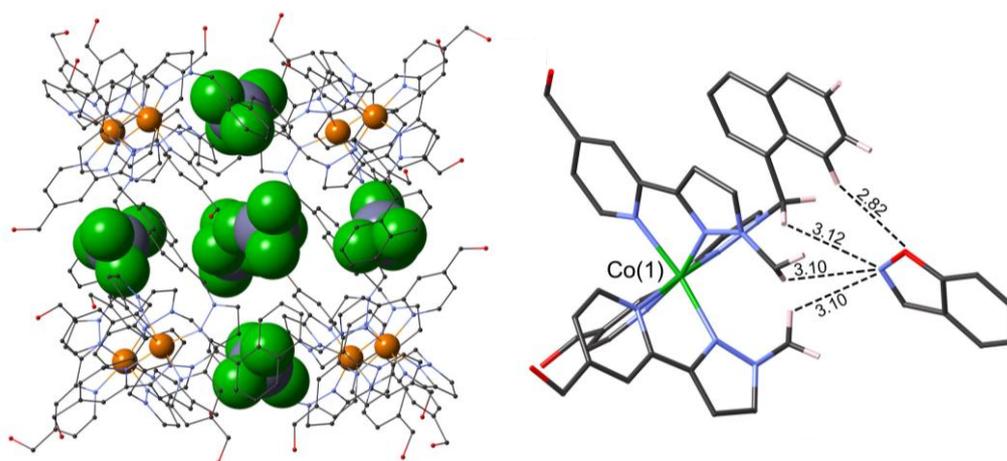
Figure 6.15: X-ray crystal structure the BF_4^- counter ions in the windows (left); and the benzisoxazole in the binding pocket showing various H-N, H-O interactions with the fac H-bond pocket and the naphthyl protons (right).

If these sites around the cage are saturated with hydroxide ions at pD 8.5 due to the high positive charge, increasing the pD to 11.4 will not result in an increase in the *local* hydroxide concentration and the rate of the reaction should therefore be independent of pD in this range. The point where the two lines (catalysed and uncatalysed reaction rates)

cross is pD 13.8 (pH 13, so hydroxide concentration of 100 mM) (**figure 6.12**), could be thought to be the effective local concentration of hydroxide, since in the presence of the cage, the rate of reaction occurs at the same rate as the background reaction occurs at pD 13.8.

This is better than other cage catalysts in that the cage has two orthogonal binding sites for the two components of the reaction, bringing both components into proximity via a hydrophobic interaction (guest) and a polar interaction (hydroxide)

A cartoon illustration of what we think is the mechanism is in **figure 6.16**.
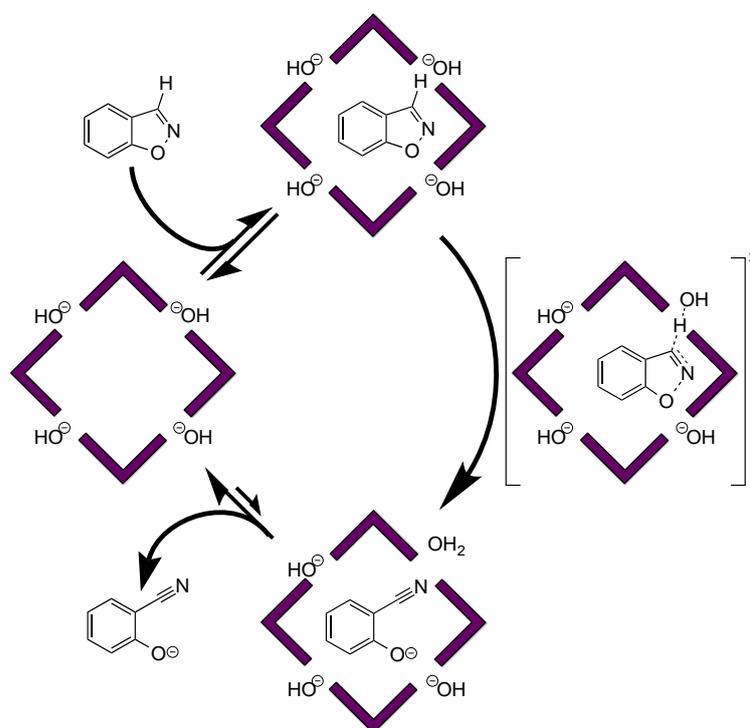
Figure 6.16: Cartoon of the catalytic reaction cycle, showing the role of the cage in bringing the benzisoxazole substrate and the hydroxide ions into close proximity

6.2.4 Control experiments

1) Is there an increase in local hydroxide concentration around the cage?

To test this, we added a large excess of chloride ions (47 mM) to the solution to compete for the sites on the cage surface: this reduced the observed rate of reaction to that of the background rate (**figure 6.17, green point**). This cannot just be an effect of changing the medium as the Kemp elimination is known to be insensitive to ionic strength[10,11] and the addition of chloride has no effect on the rate of the background reaction.[10,11] As well as being present in ~180 fold excess relative to hydroxide in solution at the pD of the experiment, chloride ions are preferentially bound to the interface region of cationic micelles (typically 10 fold[12,13]) as they are less strongly solvated by water than hydroxide, so the reduction in concentration of the catalytically active cage is greater than the ratio of the anion concentrations.

This experiment confirms that the catalytic effect is associated with the accumulation of [⁻OH] ions around the cage surface.
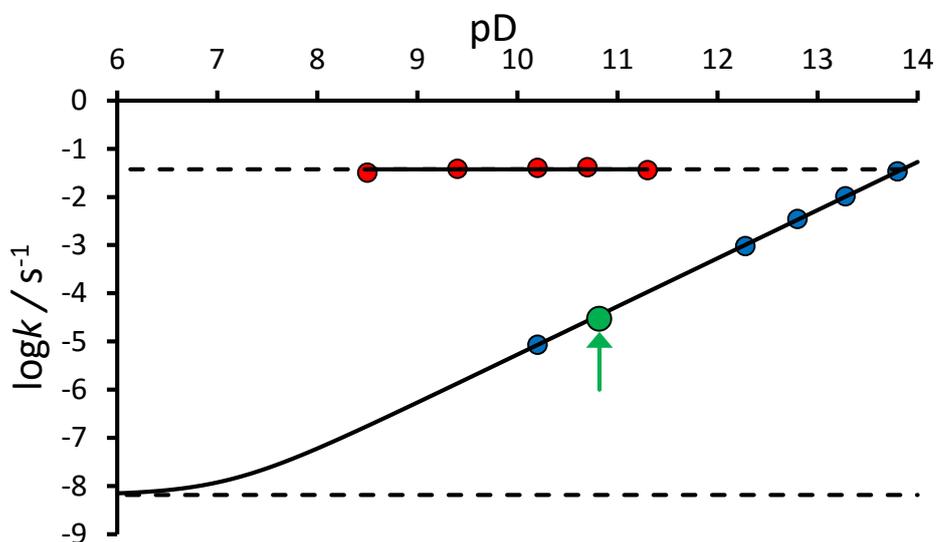
Figure 6.17: The control experiment of adding Cl⁻ anions to the cage reaction, to compete for hydroxide binding, and thus slowing the catalysed rate back to the background rate (green dot)

2) Does the reaction occur inside the cavity?

To test this, we added an excess of a strongly-binding competing guest (20 mM cycloundecanone, K = 1.2 × 10^6 M^{-1}, see chapter 3). With this inhibitor present, the reaction rate dropped to that of the uncatalysed reaction (**figure 6.18, purple dot**), because the competing guest prevents substrate binding in the cage cavity. This demonstrates that the rate acceleration does not occur due to some interaction between the cage exterior and the substrate but definitely requires the substrate to be inside the cage cavity.
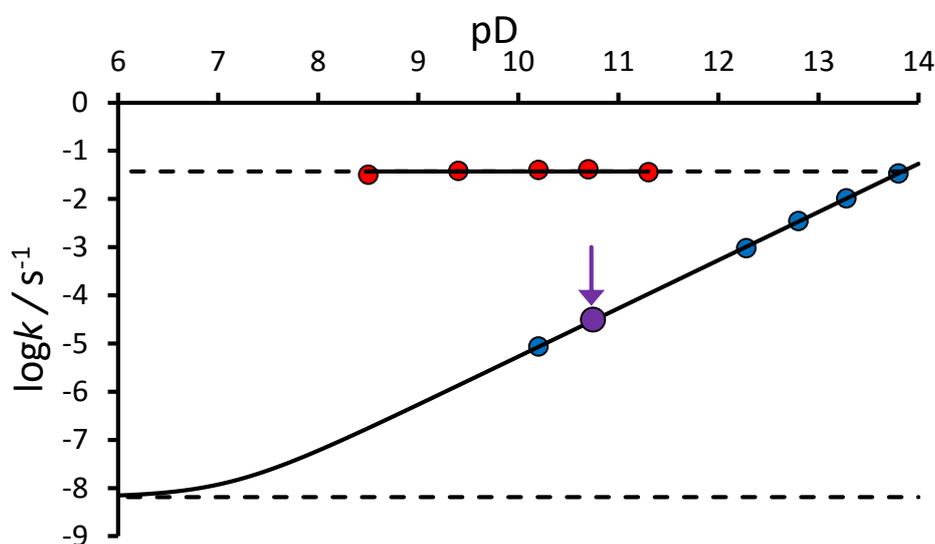
Figure 6.18: The control experiment of adding an inhibitor (cycloundecanone) to the cage reaction, to see if the rate enhancement occurred inside the cavity or not (purple dot)

1) Does it turnover?

We have already shown that the product does not bind during the catalysis experiment, but we wanted to see how many turnovers the cage could handle. To do this we added several successive portions of benzisoxazole (0.85 equivalents per portion) to a 1 mM solution of the cage in water at pD = 10.2, waiting until each aliquot had completely reacted before adding the next. Under these conditions the uncatalysed reaction does not contribute significantly, partly as it is slow at this pH and partly because under these conditions almost all of the added guest is bound to the cage.

We can see from **figure 6.19** that after multiple additions of guest the reaction profile is completely unchanged, and so there is no detectable change in activity after 5 turnovers.
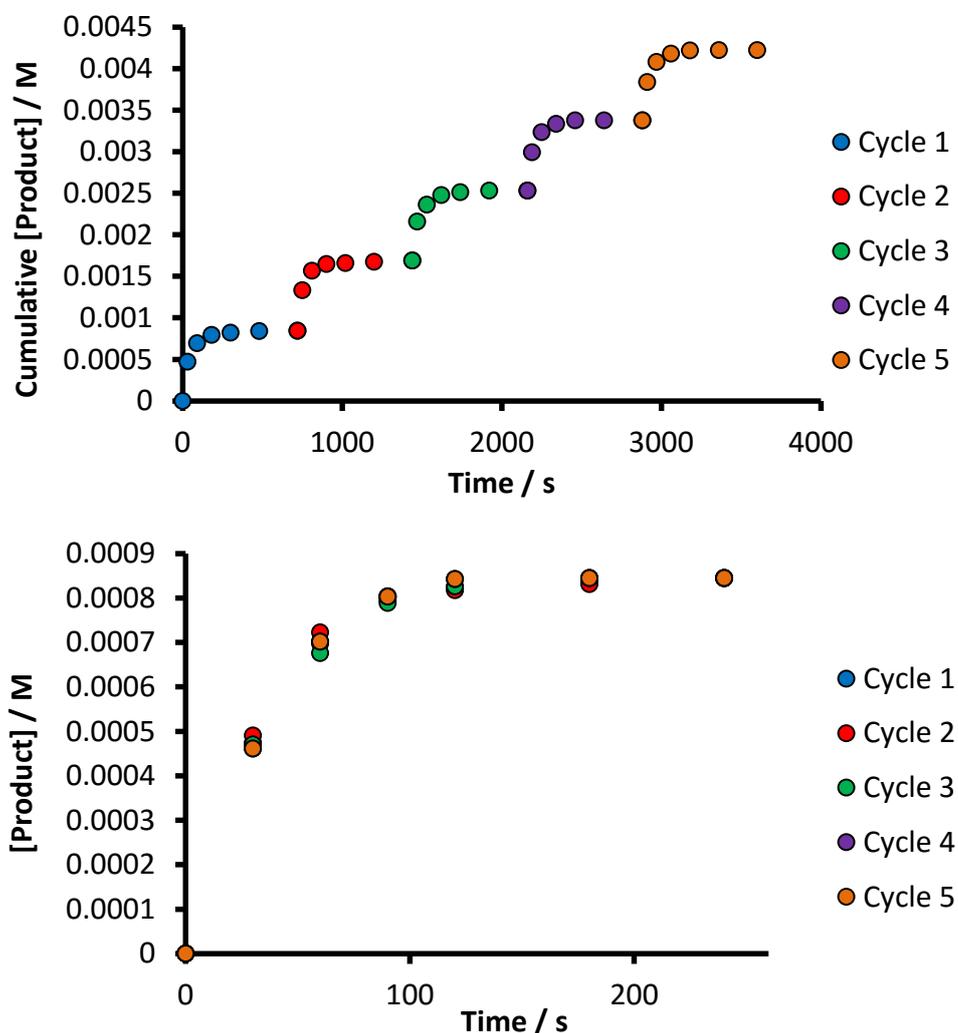
Figure 6.19: The successive addition of benzisoxazole for 5 turnovers (top) and its overlay showing that the rate is unchanged between cycles (bottom)

We then in a separate experiment added 100 equivalents of benzisoxazole to a 0.1 mM solution of cage, at pD 9.9 where the k_{cat}/k_{uncat} ratio is *ca.* 8800. After conversion of all the benzisoxazole to 2-cyanophenolate (100 turnovers), the ^{1}H NMR spectrum of the cage was unchanged (**figure 6.20**).
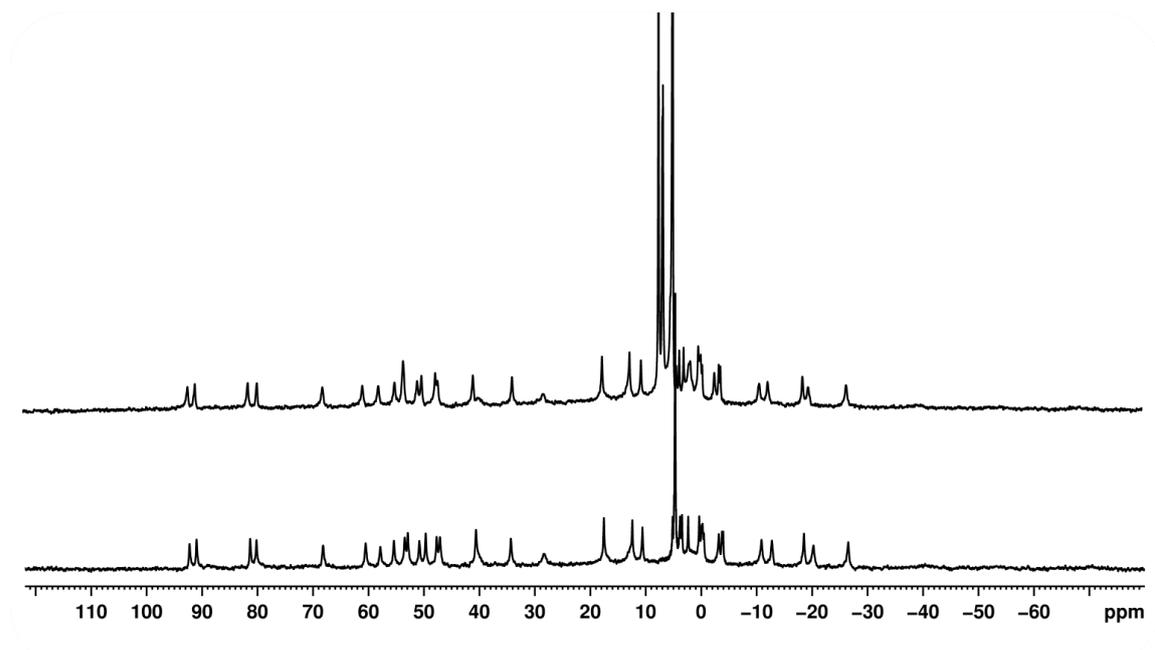
Figure 6.20: ^{1}H NMR spectrum of free cage (bottom) and cage after 100 turnovers has occurred inside its cavity

6.3 Conclusions

In conclusion we have demonstrated that the $[Co_8L_{12}]^{16+}$ coordination cage is an outstandingly effective catalyst for the Kemp elimination using benzisoxazole as substrate due to a combination of (i) a high local concentration of partially-desolvated hydroxide ions around the cavity arising from ion-pairing with the cationic cage, and (ii) localisation of the hydrophobic substrate in this cavity. Thus, the catalyst uses two different types of supramolecular interaction, associated with different recognition sites on the cage, to bring the two reacting components into close proximity. k_{cat} is independent of pD in the range 8.5 – 11.4 leading to a maximum observed rate acceleration of 2 x 10^5 fold. This is much greater than previously observed for catalysis by vesicles and micelles, and these cages accordingly present more specific binding cavities and robust structures than these weakly bound supramolecular aggregates.

Based on this, there are many interesting possibilities for catalysis with other guests (such as amides, phostphate esters) that react with base; also reactions with other anions that accumulate around the surface, such as reactions with halogens e.g. fluoride.

6.4 Experimental techniques and synthetic procedures

Chemicals

All chemicals were purchased from commercial sources and used as supplied unless otherwise stated.

Instruments

Instruments used for spectroscopic analyses were: (i) Cary 1Bio for UV/Vis spectrophotometry; (ii) Bruker AV3-400 for ^{1}H NMR spectroscopy. The cage complex was prepared as described previously. All pH measurements were made using a Hamilton Spintrode pH combination electrode at 298 K and calibrated with calibration standards at pH 4.01, 7.0 and 10.01.

6.4.1 Measurements

Monitoring the reaction.

Above pD 12, the uncatalysed Kemp elimination reaction was monitored by UV/Vis spectroscopy at 298 K at various concentrations of NaOD. The change in absorbance at 330 nm was fit to a first order rate equation to obtain the value for the observed rate constant k_{uncat}. At pD10.2, the pD was controlled using a 0.1M buffer solution containing $NaHCO_3$ / Na_2CO_3 to achieve the desired pD and the reaction monitored by ^{1}H NMR spectroscopy at 298 K.

The catalysed reaction, in the presence of the cage, could not be followed by UV/Vis spectroscopy due to the very strong absorptions of the cage in the UV/region which obscured the spectra of both of the substrate benzisoxazole and product cyanophenolate. Instead reactions were followed by ^{1}H NMR spectroscopy at 298 K, monitoring the intensity of product peaks close to 6.5 ppm which do not overlap with signals from either starting material or cage. The cage concentration was 1 mM in D_2O: in the absence of added base this solution is weakly acidic, and the pD was increased as required to a maximum of 11.4 by addition of portions of NaOD. The benzisoxazole starting material was then added to the NMR tube such that its concentration was 0.85 mM and ^{1}H NMR

spectra were recorded at regular intervals. The pH was checked at the end of the reaction to ensure that it had not changed. To obtain k_{cat}, the appearance of product over time was fit to the Michaelis-Menten reaction scheme by numerical modeling with SimFit2008 using the previously determined association constants for the substrate to the cage. Each measurement was repeated three times and an average taken.

Control experiments

(i) To determine the effect of a competitive inhibitor, the reaction in presence of catalyst was performed exactly as described above but with 20 mM cycloundecanone added; this binds in the cage much more strongly ($K = 1.2 \times 10^6$ M^{-1}) than benzisoxazole.

(ii) To examine the effect of chloride ions, the catalysed reaction was monitored as described above but in the presence of 47 mM LiCl (higher concentrations than this resulted in decomposition of the cage). We note that the observed reaction is slightly slower than predicted for the background reaction alone, which can be explained by the substrate being protected from reaction when it binds to the cage surrounded by chloride ions.

Spontaneous reaction

The spontaneous reaction of benzoisoxazole was estimated from the data in reference 27. The maximum rate constant for the spontaneous decomposition at 30 °C is reported as 1.1×10^{-8} s^{-1}. Using the activation parameters for the hydroxide catalysed to estimate the reactivity a 25 °C leads to the estimate of 6×10^{-9} s^{-1}. This is a conservative estimate of the maximum observed rate constant for the spontaneous reaction.

6.4.2 X-ray crystallography

The data collection was performed using a Bruker APEX-2 CCD diffractometer using Mo-Kα radiation from a sealed-rube source. Data were corrected for absorption using empirical methods (SADABS) based upon symmetry-equivalent reflections combined with measurements at different azimuthal angles. The structure was solved and refined using the SHELX suite of programs.

The asymmetric unit contains one half of the cage complex which lies astride an inversion centre, as well as one complete guest molecule whose atoms all have site occupancies of 0.5. Thus, the complete complex contains one guest molecule disordered over 2 symmetrically equivalent orientations with the N/O atoms pointing towards diagonally opposite corners Co(1) and Co(1A). The usual severe disorder of anions / solvent molecules and solvent loss characteristic of cage complexes of this type resulted in weak scattering, necessitating use of extensive geometric and displacement restraints to keep the refinement stable: these are described in detail in the CIF. We could locate and refine six of the expected eight $[BF_4]^-$ anions in the asymmetric unit; all show disorder of the F atoms. Large regions of diffuse electron density which could not be modelled, accounting for the remaining anions plus solvent molecules, were eliminated from the refinement using of the 'SQUEEZE' function in the PLATON software package. The structural determination is therefore of poor quality by conventional small-molecule standards although it is typical for a coordination cage. The gross structure of the cage, and the presence of the guest in the cavity and its position / orientation in the cavity, are clear and we make no claims for structural details beyond this.

Crystallography data table

Complex	$[Co_8L_{12}](BF_4)_{16} \bullet (benzisoxazole)$
Formula	$C_{367}H_{317}B_{16}Co_8F_{64}N_{73}O_{25}$
Molecular weight	8010.32
T, K	100(2)
Crystal system	Monoclinic
Space group	*C2/c*
a, Å	27.2747(12)
b, Å	38.8257(17)
c, Å	42.232(2)
$\alpha,°$	90
$\beta,°$	108.089(3)
$\gamma,°$	90
V, Å^3	42512(4)
Z	4
ρ, g cm^{-3}	1.252
Crystal size, mm^3	0.1 x 0.1 x 0.1
Data, restraints, parameters	22378, 2390, 1837
Final *R*1, *wR*2^b	0.156, 0.404

a   These formulae (and consequently the crystal densities) are necessarily approximate given that large amounts of diffuse electron density in solvent-accessible voids was removed from the refinements using the OLEX 'Solvent Mask' function. See CIFs, and comments in experimental section, for details.

b   The value of *R*1 is based on 'observed' data with *I* > 2σ(*I*); the value of w*R*2 is based on all data.

6.5 References

1) M. Yoshizawa; J. K. Klosterman; M. Fujita, *Angew. Chem. Int. Ed.,* **2009**, *48*, 3418-3438

2) M. Raynal; P. Ballester; A. Vidal-Ferran; P. W. N. M. van Leeuwen, *Chem. Soc. Rev.,* **2014**, *43*, 1734-1787

3) H. Vardhan; F. Verpoort, *Adv. Synth. Catal.,* **2015**, *357*, 1351-1368

4) C. J. Brown; F. D. Toste; R. G. Bergman; K. N. Raymond, *Chem. Rev.,* **2015**, *115*, 3012-3035

5) Q. Zhang; K. Tiefenbacher, *Nature Chem.,* **2015**, *7*, 197-202

6) C. J. Hastings; M. D. Pluth; R. G. Bergman; K. N. Raymond, *J. Am. Chem. Soc.,* **2010**, *132*, 6938-6940

7) J. Kang; J. Rebek, *Nature*, **1997**, *385*, 50-52

8) R. J. Hooley; J. Rebek, *Org. Bimol. Chem.,* **2007**, *5*, 3631-3636

9) D. Fiedler; R. G. Bergman; K. N Raymond, *Angew. Chem. Int. Ed.,* **2004**, *43*, 6748-6751

10) M. L. Casey; D. S. Kemp; K. G. Paul; D. D. Cox, *J. Org. Chem.,* **1973**, *38*, 2294-2301

11) D. S. Kemp; M. L. Casey, *J. Am. Chem. Soc.,* **1973**, *95*, 6670-6680

12) J. E. Klijn; J. B. F. N. Engberts, *J. Am. Chem. Soc.,* **2003**, *125*, 1825-1833

13) J. E. Klijn; J. B. F. N. Engberts, *Org. Bimol. Chem.,* **2004**, *2*, 1789-1799

14) Y. Zhao; C. Beuchat; Y. Domoto; J. Gajewy; A. Wilson; J. Mareda; N. Sakai; S. Mtile, *J. Am. Chem. Soc.,* **2004**, *136*, 2101-2111

15) R. Blomberg; H. Kries; D. M. Pinkas; P. R. E. Mittl; M. G. Grütter; H. K. Privett; S. L. Mayo; D. Hilvert, *Nature*, **2013**, *503*, 418-421

Appendix 1

Chapter 3 Data

1.1 Fits For those in fast exchange/done by fluorescence

17

18

19

20

21

22

23

24

29

30

31

33
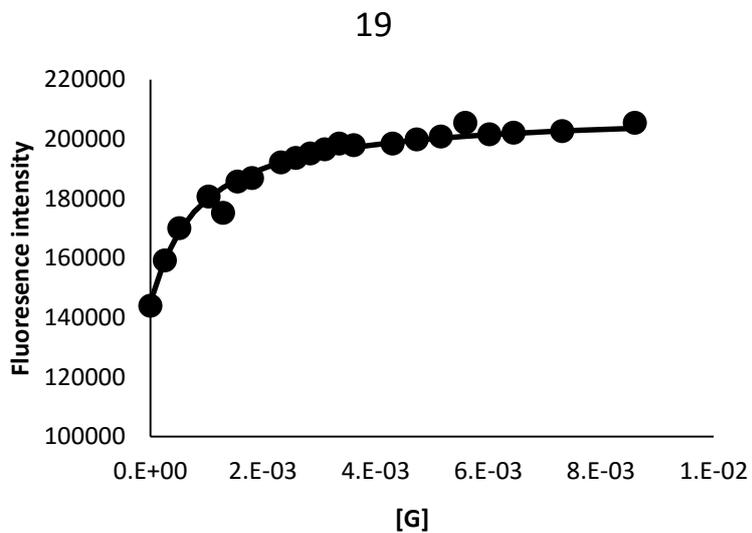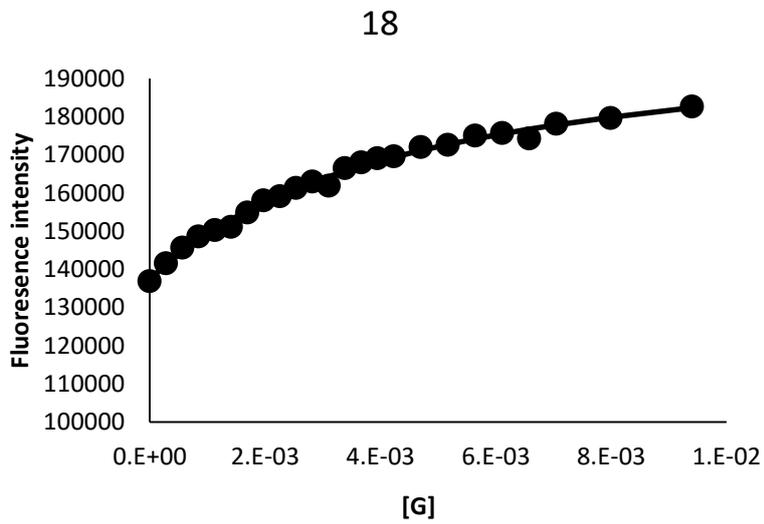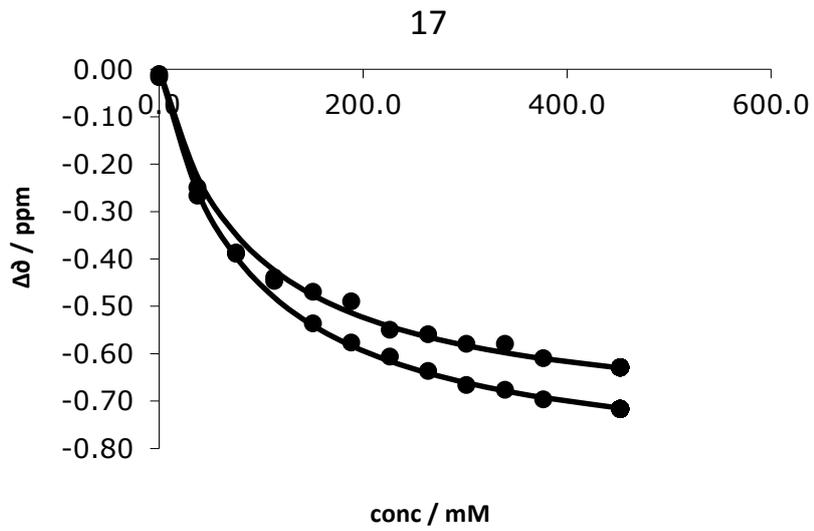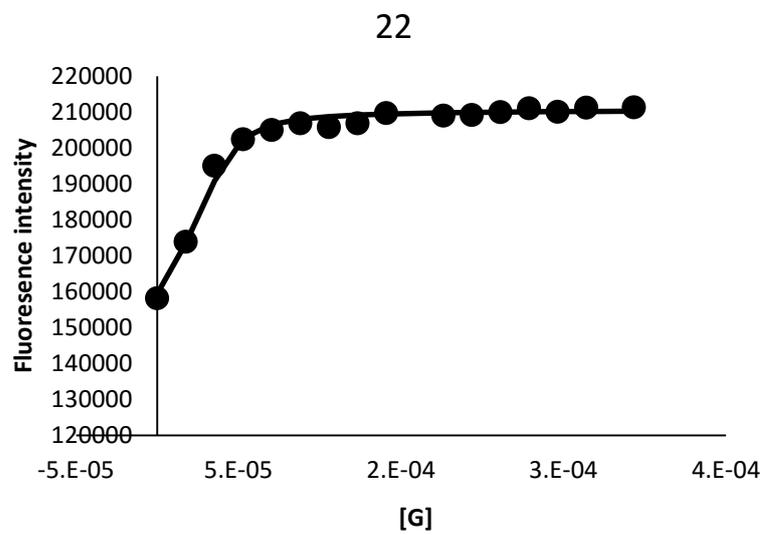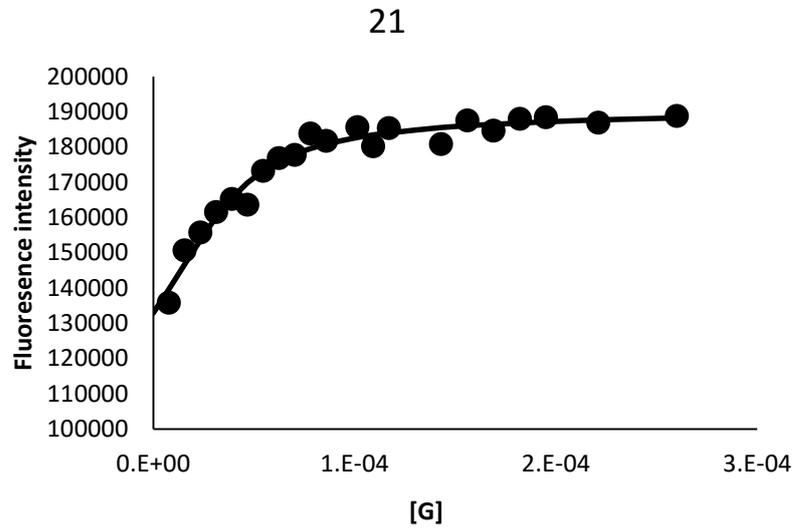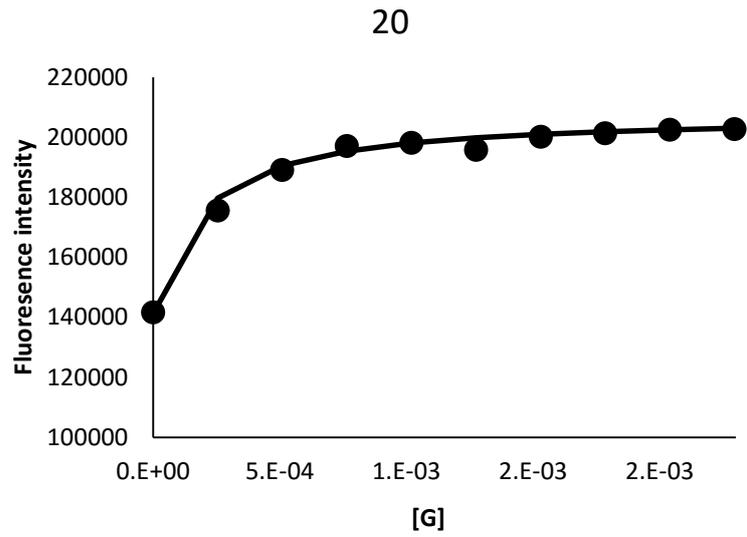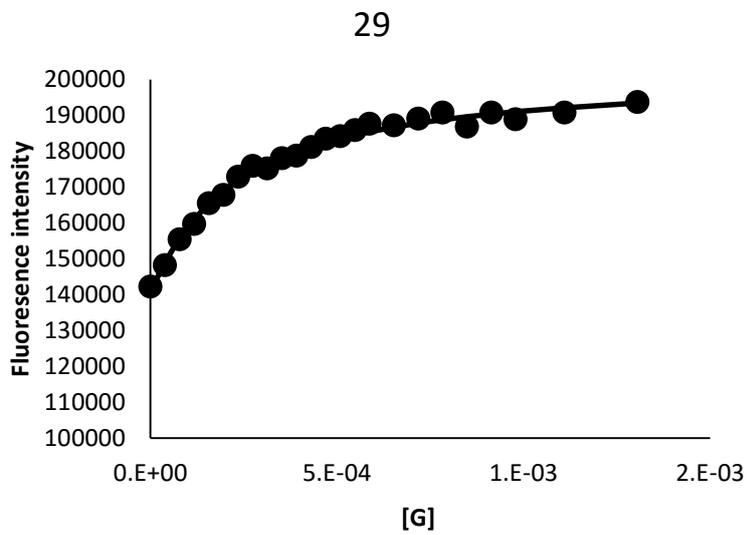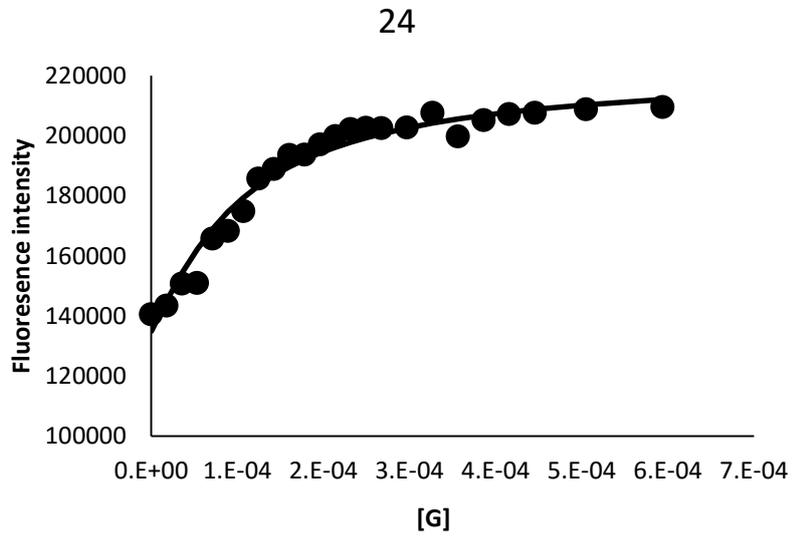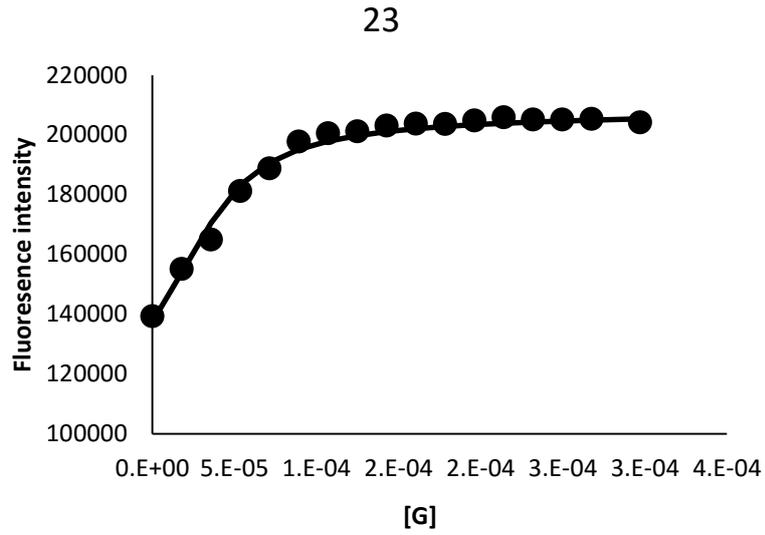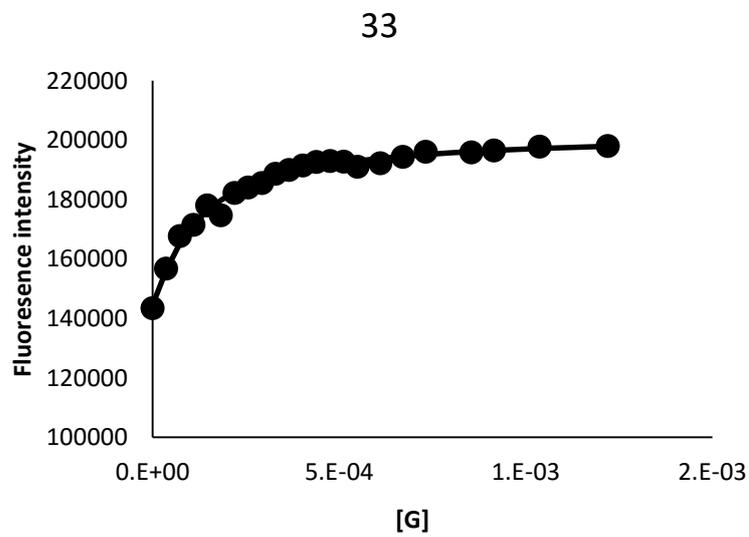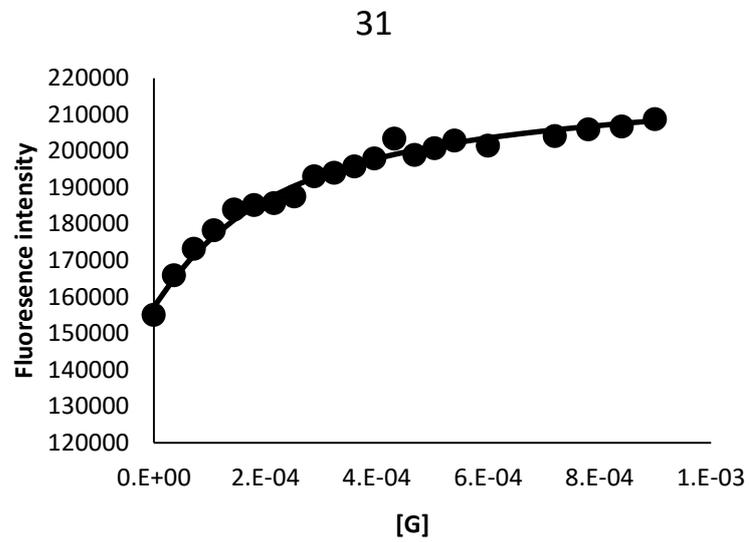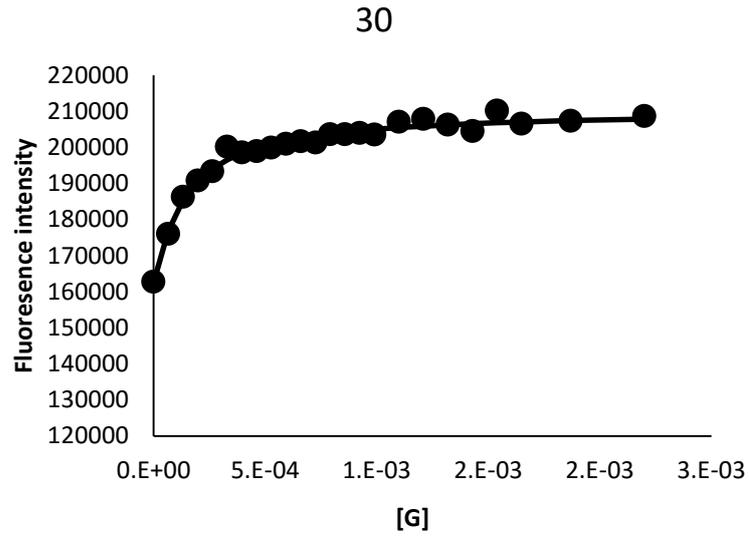
34
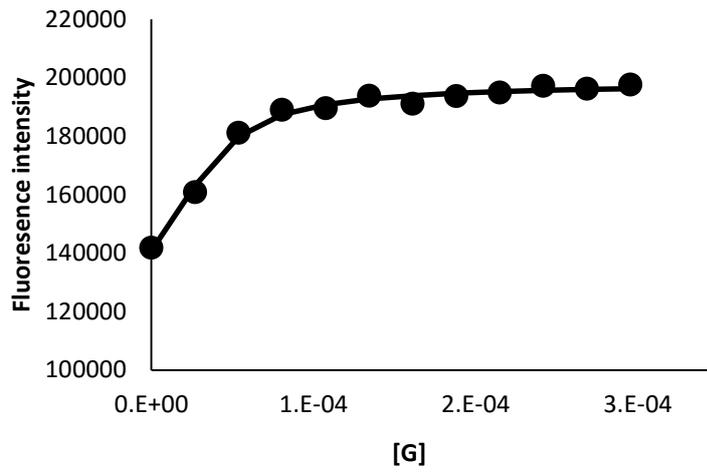
35
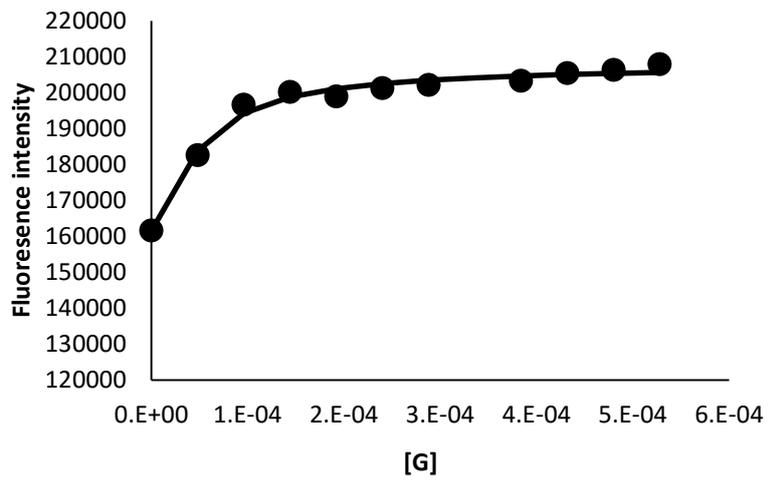
27
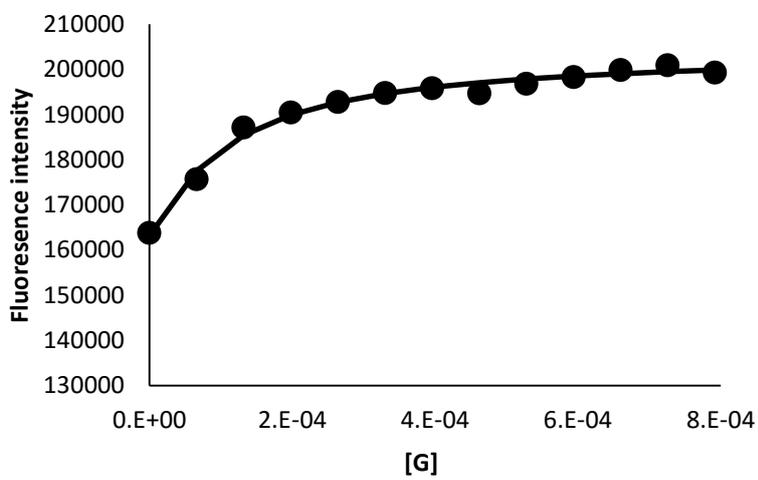
28
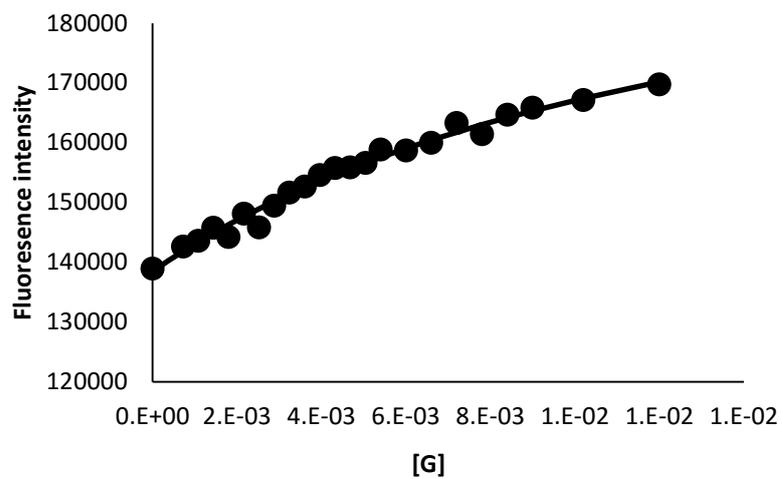
Appendix 2

Chapter 4 Data

2.1 Training Set
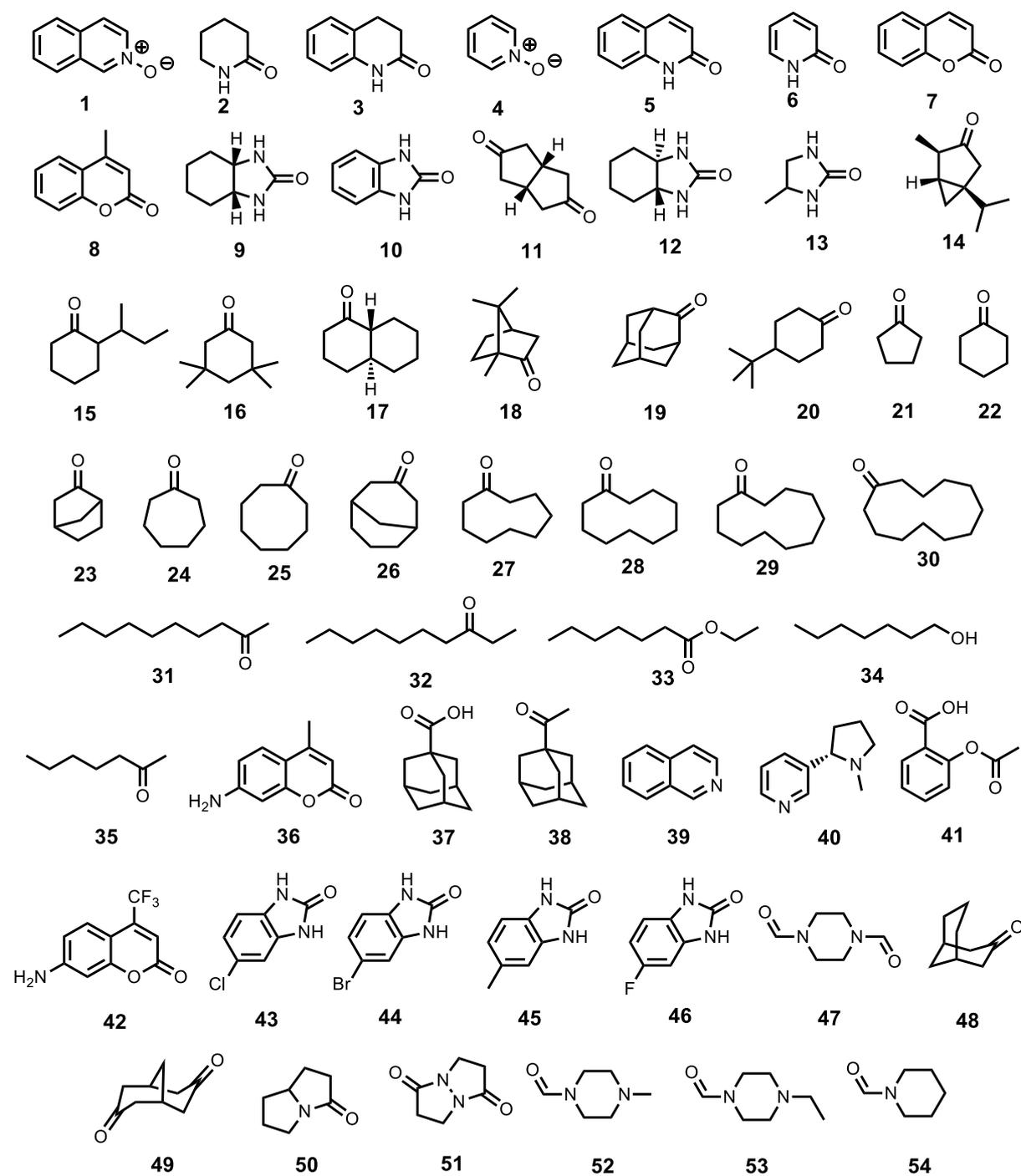
2.2 Data tables used in calculations

Table 1: Numerical values generated by GOLD for the individual terms that contribute to the scoring functions (main text, eq. 2 and 3).

Guest	Ligand_clash	Ligand_torsion	Part_buried	Non-polar	Ligand_flexibility
1	0	0	-2.109	-43.3137	0
2	0	0	-3.0465	-27.4388	0
3	0	0	-2.4099	-41.2634	0
4	0	0	-2.26	-32.0514	0
5	0	0	-2.7351	-40.6438	0
6	0	0	-3.1134	-27.9508	0
7	0	0	-2.4121	-41.0365	0
8	0	0	-2.7123	-46.2622	0
9	0	0	-3.4782	-33.1938	0
10	0	0	-2.9428	-30.1587	0
11	0	0	-4.2726	-29.1526	0
12	0	0	-3.4382	-32.0023	0
13	0	0	-4.6213	-22.7748	0
14	0	0.6359	-0.5272	-52.666	1
15	0	0.2683	-1.9122	-49.4866	2
16	0	0	-1.2999	-46.3249	0
17	0	0	-2.2249	-43.0767	0
18	0	0	-1.7939	-41.3553	0
19	0	0	-1.7048	-38.256	0
20	0	0.3499	-2.0598	-44.3617	1
21	0	0	-1.9986	-27.2969	0
22	0	0	-2.2161	-32.0919	0
23	0	0	-2.1068	-32.1512	0
24	0	0	-1.9886	-34.0851	0
25	0	0	-2.0557	-37.4516	0
26	0	0	-2.2293	-39.0059	0
27	0	0	-2.0968	-38.8338	0
28	0	0	-2.0458	-47.1174	0
29	0	0	-1.708	-54.4098	0
30	0	0	-1.1725	-60.0173	0
31	0	0.2147	-2.2221	-55.531	7
32	0	0.0484	-2.0233	-58.793	7
33	0	0.6156	-3.0882	-50.0399	6
34	0	0.0352	-2.0632	-40.2044	6
35	0	0.0866	-2.1602	-38.7078	4
36	0	0	-0.9723	-42.3058	0

37	0	0.0052	-3.5062	-45.4026	1
38	0	0.0084	-2.1935	-50.4456	1
39	0	0	-2.1185	-42.6227	0
40	0	0.015	-2.6923	-49.133	1
41	0	0.9904	-4.0205	-43.4843	3
42	1.7813	0.028	8.9078	-58.9343	1
43	0	0	-3.5162	-34.3235	0
44	0	0	-4.146	-36.4154	0
45	0	0	-2.6802	-34.5214	0
46	0	0	-4.0999	-36.3648	0
47	0	0.0005	-4.053	-33.0628	2
48	0	0	-2.2123	-39.1164	0
49	0	0	-4.1075	-33.1122	0
50	0	0	-1.7961	-39.1621	0
51	0	0	-3.4414	-29.1303	0
52	0	0.0019	-2.1997	-34.6724	1
53	0	0.0016	-3.0278	-30.4643	1
54	0	0.0108	-2.7451	-37.1465	2
55	0	0	3.0391	-72.2022	1
56	0	0	0	-61.7662	0
57	0	0	0	-61.1466	0
58	0	0	-0.74033	-59.7962	0
59	0	0	-1.35915	-58.5762	0
60	0	0	-0.9865	-57.2076	0
61	0	1.0855	3.3839	-58.3369	1
62	0	0	-4.59413	-40.9762	0
63	0	0.0005	-1.2025	-60.3498	2
64	0	0	0	-39.5214	0
65	0	0	-0.5568	-40.2497	0
66	0	0	-0.5162	-49.7484	1
67	0	0	-1.783	-35.1129	0
68	0	0.0812	-3.8245	-33.5117	2
69	0	0.0011	-3.823	-41.0994	1

2.3 Data tables used for figures

Table 2: Measured log$K$ values and scores calculated by ChemPLP for the training set of guests (see **figure 4.4**, main text).

Guest	logK$_{expt}$	ChemPLP Score	Guest	logK$_{expt}$	ChemPLP Score	Guest	logK$_{expt}$	ChemPLP Score
1	3.49	49.81	19	4.28	45.54	37	4.90	53.86
2	1.15	39.49	20	3.94	50.66	38	4.30	57.79
3	3.83	48.47	21	1.15	47.52	39	4.00	49.12
4	1.87	38.88	22	1.73	38.51	40	1.95	45.94
5	3.96	47.67	23	2.11	39.26	41	2.08	53.28
6	1.52	35.61	24	2.62	40.55	42	-1.00	54.68
7	3.88	47.96	25	3.32	44.58	43	3.70	43.31
8	5.00	53.35	26	3.60	46.56	44	3.60	44.45
9	1.86	40.64	27	4.04	46.04	45	3.60	42.13
10	3.49	38.32	28	5.15	53.94	46	3.48	45.25
11	2.83	43.45	29	6.08	59.72	47	1.78	49.63
12	2.18	40.87	30	5.18	62.76	48	3.60	46.59
13	1.48	41.77	31	-1.00	62.48	49	2.41	46.87
14	4.30	57.43	32	-1.00	64.03	50	2.30	44.66
15	4.20	54.43	33	-1.00	58.15	51	0.48	40.70
16	4.88	52.62	34	-1.00	47.30	52	0.70	42.40
17	3.98	50.59	35	-1.00	45.61	53	0.90	39.14
18	5.26	46.88	36	4.30	49.32	54	0.70	46.49

Table 3: Measured log*K* values, and log*K* values calculated by **Eq. 2** for the training set of guests (see **figure 4.5**, main text).

Guest	logK_{expt}	logK_{calc}	Guest	logK_{expt}	logK_{calc}	Guest	logK_{expt}	logK_{calc}
1	3.49	3.28	19	4.28	2.92	37	4.90	3.27
2	1.15	1.87	20	3.94	2.43	38	4.30	3.83
3	3.83	3.08	21	1.15	1.99	39	4.00	3.23
4	1.87	2.34	22	1.73	2.35	40	1.95	3.65
5	3.96	2.99	23	2.11	2.37	41	2.08	0.38
6	1.52	1.91	24	2.62	2.54	42	-1.00	-1.00
7	3.88	3.06	25	3.32	2.81	43	3.70	2.38
8	5.00	3.45	26	3.60	2.92	44	3.60	2.47
9	1.86	2.29	27	4.04	2.92	45	3.60	2.50
10	3.49	2.11	28	5.15	3.60	46	3.48	2.47
11	2.83	1.86	29	6.08	4.24	47	1.78	2.21
12	2.18	2.20	30	5.18	4.76	48	3.60	2.93
13	1.48	1.30	31	-1.00	3.69	49	2.41	2.21
14	4.30	2.52	32	-1.00	4.43	50	2.30	2.98
15	4.20	3.09	33	-1.00	2.05	51	0.48	1.96
16	4.88	3.63	34	-1.00	2.94	52	0.70	2.56
17	3.98	3.25	35	-1.00	2.67	53	0.90	2.12
18	5.26	3.16	36	4.30	3.34	54	0.70	2.67

Table 4: Measured log$K$ values, and log$K$ values calculated by **Eq. 3** for the training set of guests (see **figure 4.6**, main text).

Guest	logK$_{expt}$	logK$_{calc}$	Guest	logK$_{expt}$	logK$_{calc}$	Guest	logK$_{expt}$	logK$_{calc}$
1	3.49	4.02	19	4.28	3.58	37	4.90	3.02
2	1.15	2.20	20	3.94	3.52	38	4.30	3.82
3	3.83	3.75	21	1.15	2.40	39	4.00	3.95
4	1.87	2.83	22	1.73	2.85	40	1.95	3.58
5	3.96	3.62	23	2.11	2.87	41	2.08	1.75
6	1.52	2.24	24	2.62	3.10	42	-1.00	-1.00
7	3.88	3.72	25	3.32	3.43	43	3.70	2.81
8	5.00	4.20	26	3.60	3.55	44	3.60	2.89
9	1.86	2.70	27	4.04	3.56	45	3.60	3.00
10	3.49	2.50	28	5.15	4.42	46	3.48	2.90
11	2.83	2.12	29	6.08	5.24	47	1.78	0.71
12	2.18	2.59	30	5.18	5.93	48	3.60	3.57
13	1.48	1.39	31	-1.00	-1.06	49	2.41	2.56
14	4.30	4.95	32	-1.00	-0.83	50	2.30	3.66
15	4.20	3.08	33	-1.00	-0.51	51	0.48	2.29
16	4.88	4.50	34	-1.00	-1.83	52	0.70	2.19
17	3.98	3.97	35	-1.00	-0.10	53	0.90	1.58
18	5.26	3.88	36	4.30	4.15	54	0.70	1.41

2.4 Data tables used for figures for the 15 predicted guests

55   56   57   58

59   60   61   62

63   64   65   66   67

68   69
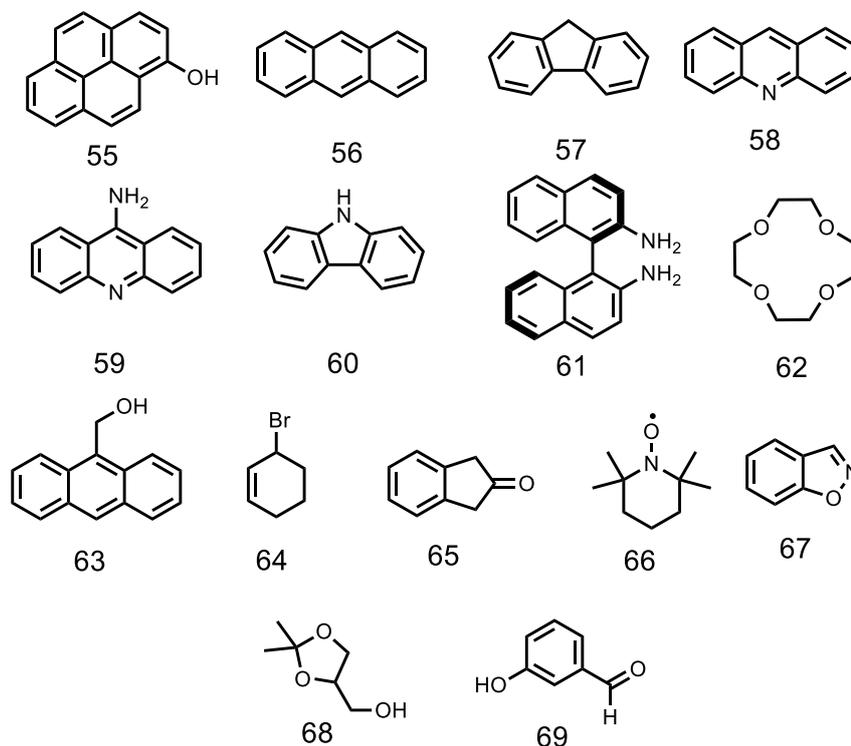
Table 5: Measured log$K$ values, and log$K$ values calculated using the final scoring function in **Eq. 3**, for an additional set of 15 guests identified by a screen of an in-house library of 3000 compounds (see **figure 4.17**, main text).

Guest	logK$_{expt}$	logK$_{calc}$	Guest	logK$_{expt}$	logK$_{calc}$
55	6.80	7.12	63	4.45	4.10
56	8.00	6.35	64	4.18	4.06
57	7.26	6.29	65	4.20	4.02
58	6.06	6.00	66	4.11	4.08
59	6.09	5.74	67	3.60	3.24
60	5.73	5.68	68	1.11	0.88
61	5.50	6.73	69	3.40	2.51
62	2.88	3.27			

References:

1   ChemCell, 2010 Collaborative Drug Discovery, Inc.

2   TorchV10, http://www.cresset-group.com/products/torch/

3   http://www.cresset-group.com/products/xedtools/

4   Open Babel, http://openbabel.org/

Appendix 3

Chapter 5 Data

3.1 pH titrations and fits

NMR spectra showing increasing pH from bottom to top and pKa curves

Adamantylamine:
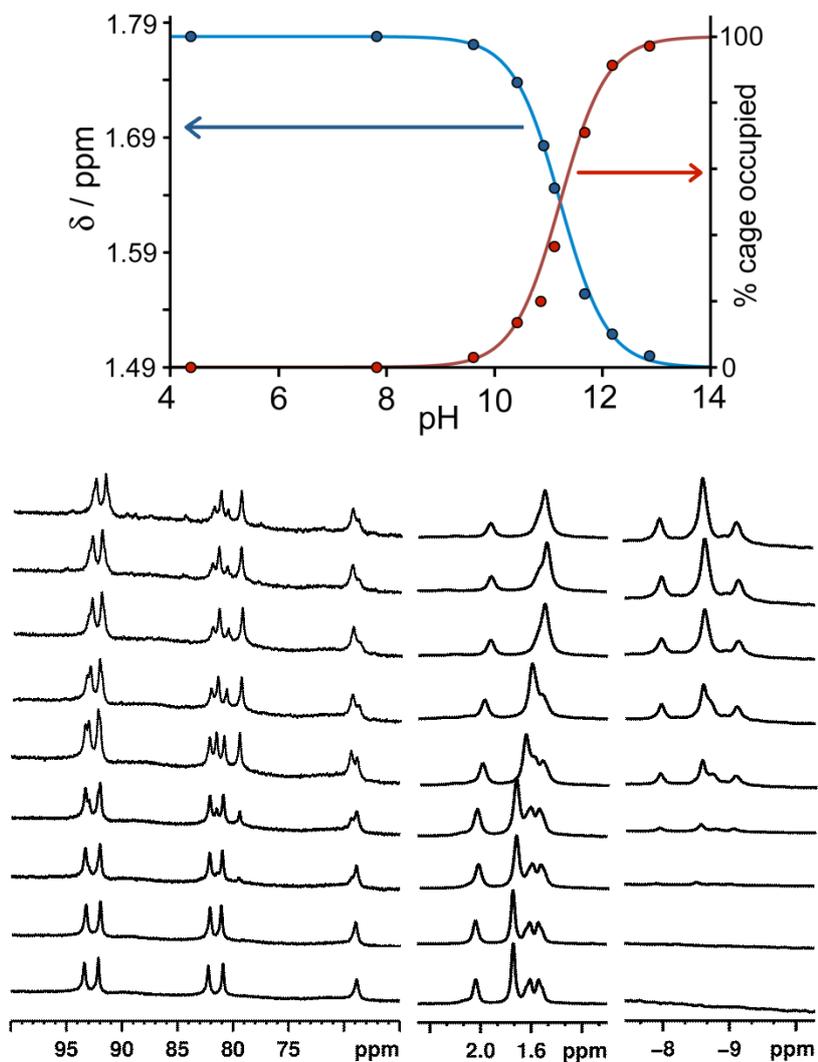
1-Adamantane carboxylic acid:
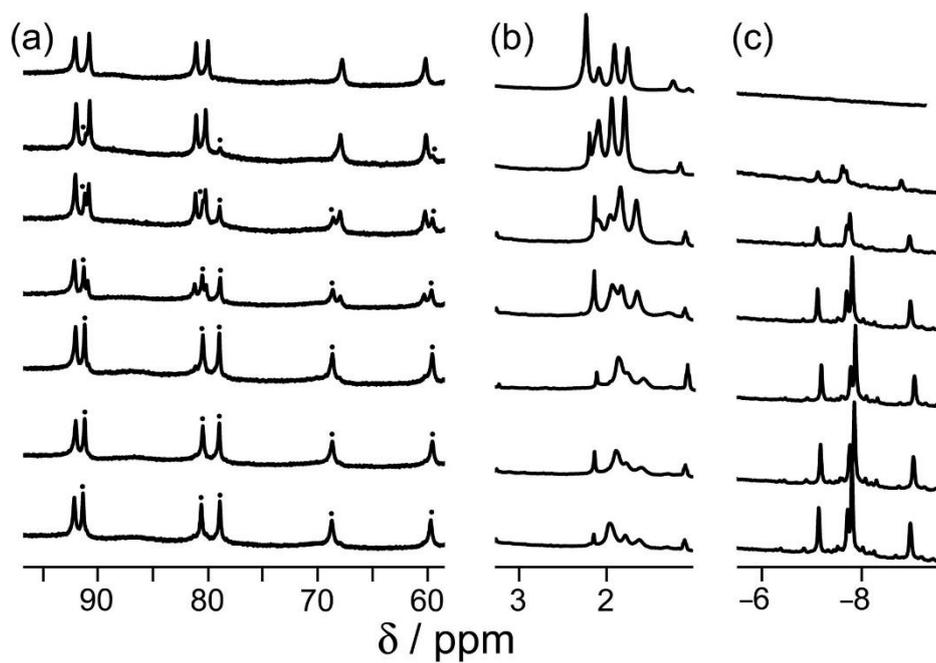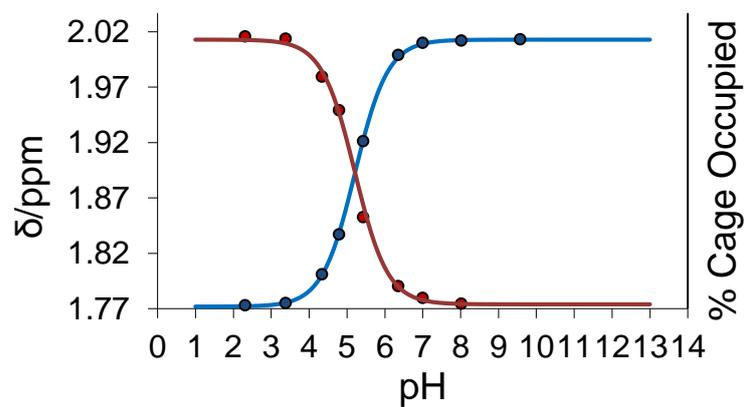
1,3-Adamantane dicarboxylic acid:
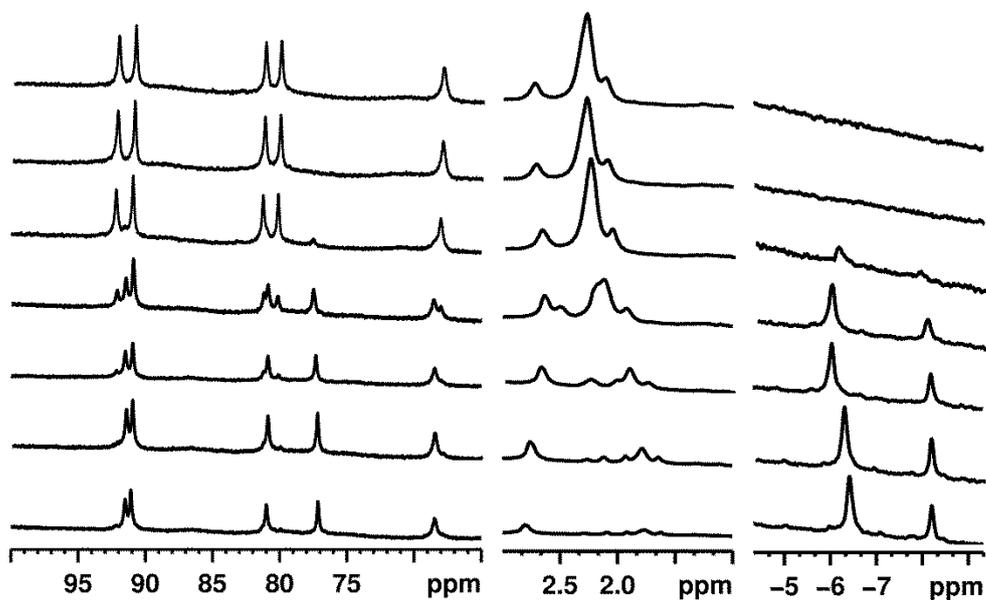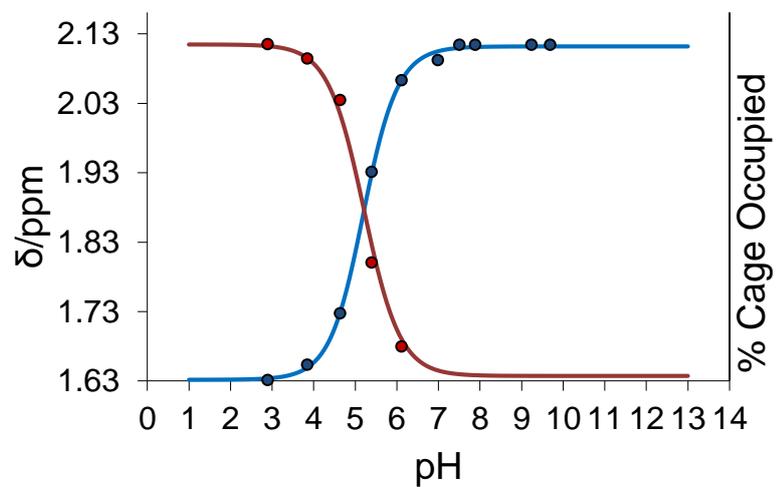
Isoquinoline:
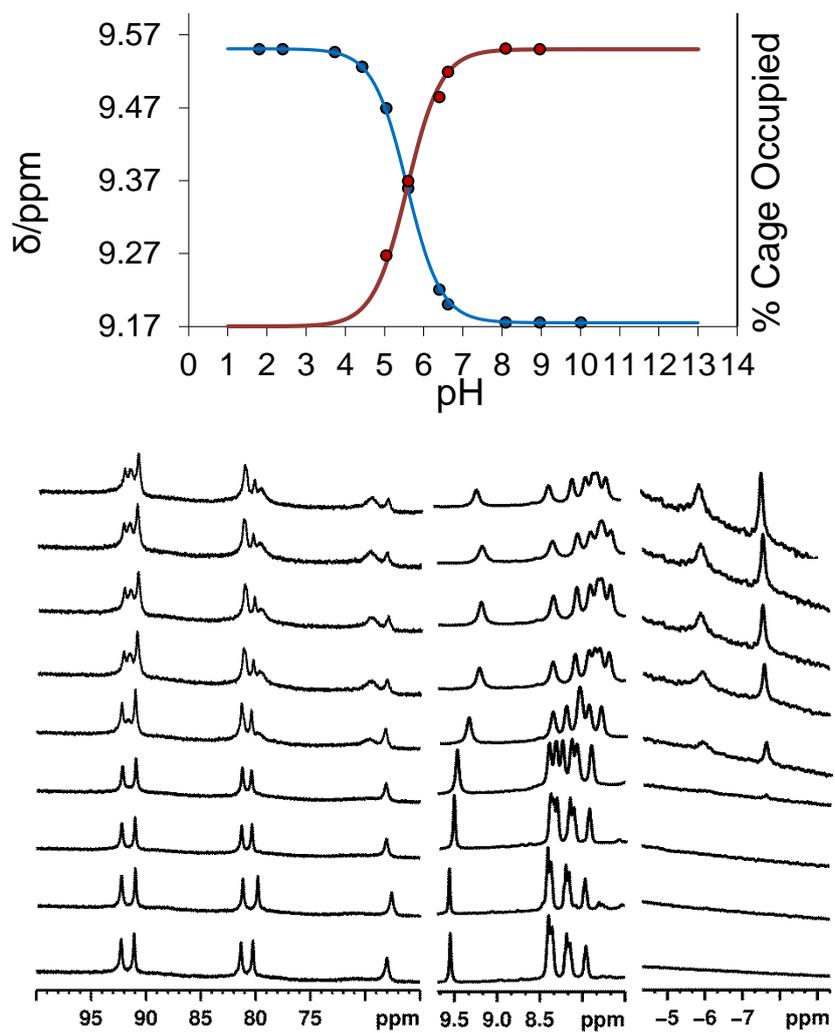
Nicotine (standard titration constant pH (9.5) increasing guest concentration from bottom to top)   (0 -> 6.3 mM):
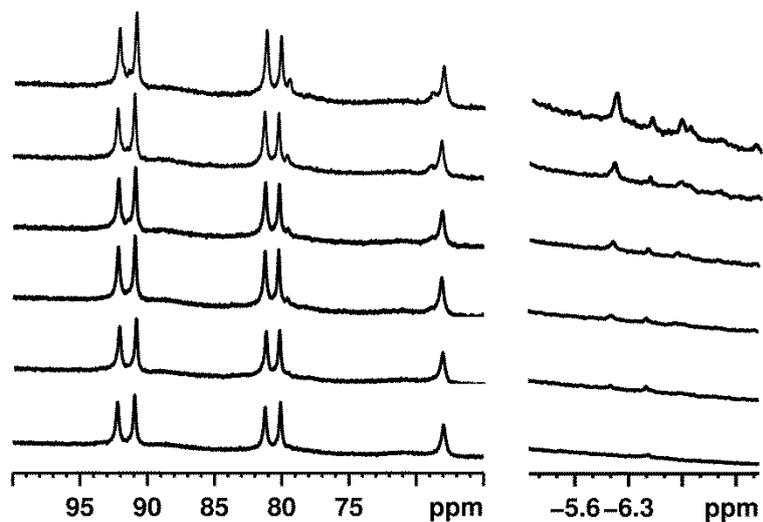
Nicotine (standard titration constant pH (3.05) increasing guest concentration from bottom to top)   (0 -> 19 mM):
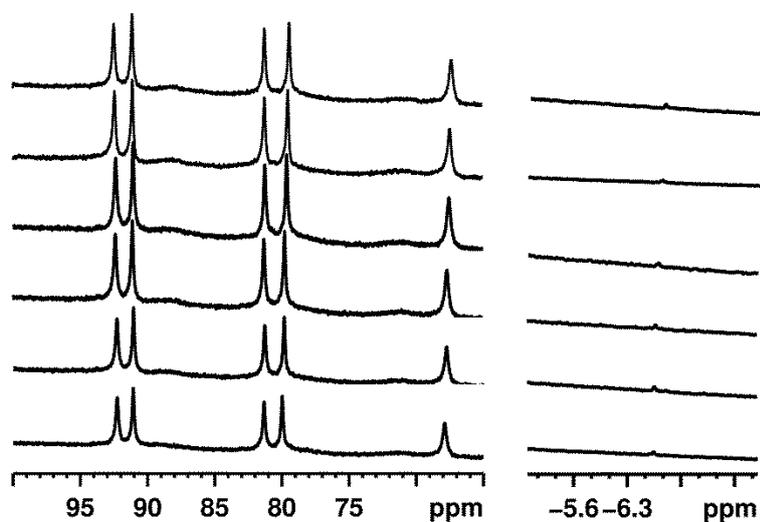
Aspirin (normal titration at constant pH (1.6) with increasing guest concentration) (0 -> 17 mM)

Top - increasing [G] from bottom to top

Bottom – increasing [G] from left to right
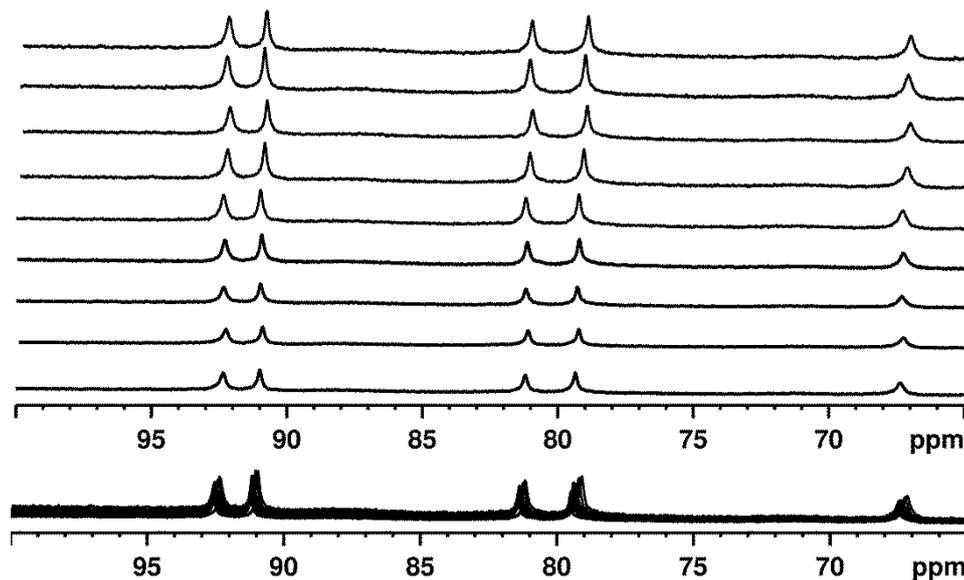
Aspirin (normal titration at constant pH (8.1) with increasing guest concentration) (0 -> 17 mM)

Top - increasing [G] from bottom to top

Bottom – increasing [G] from left to right
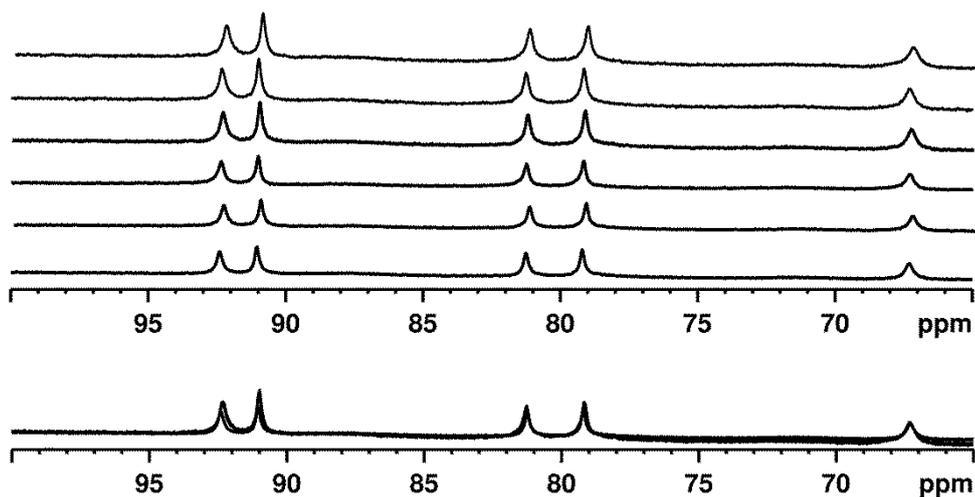
Appendix 4

Chapter 6 Data

4.1 Kinetic Data - Background reactions

pD 13.8
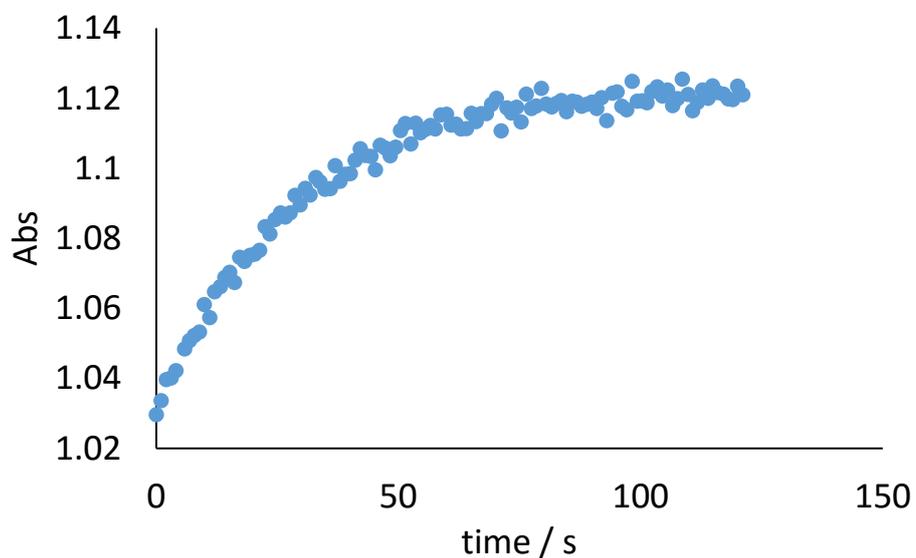
pD 13.28
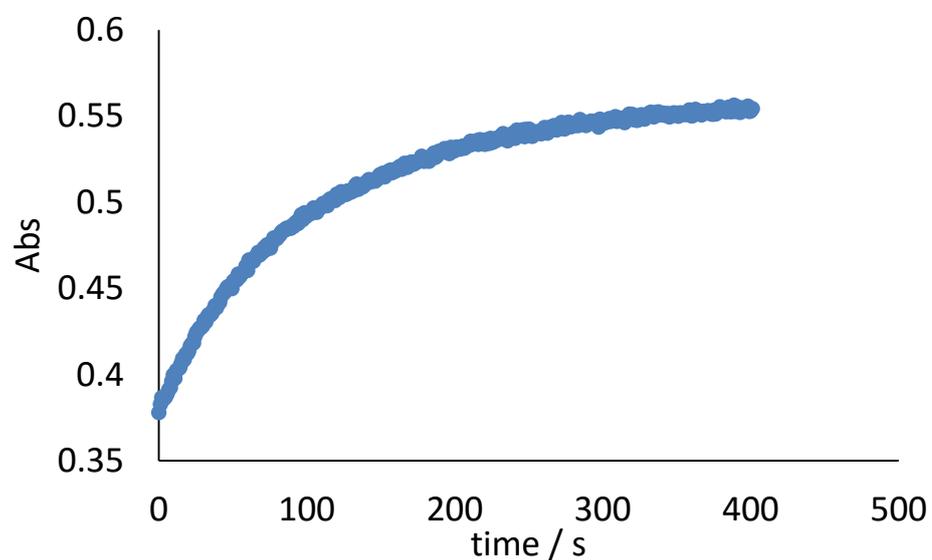
pD 12.8
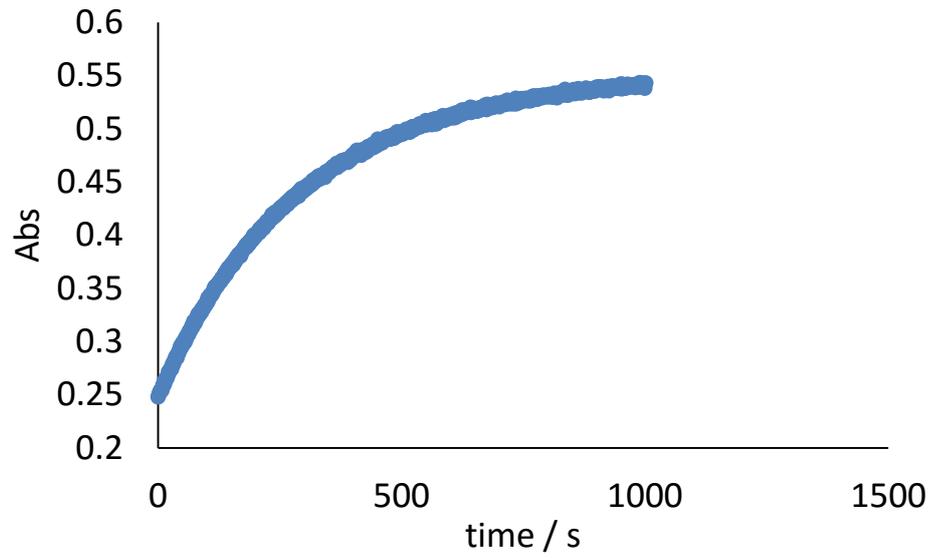
pD 12.28
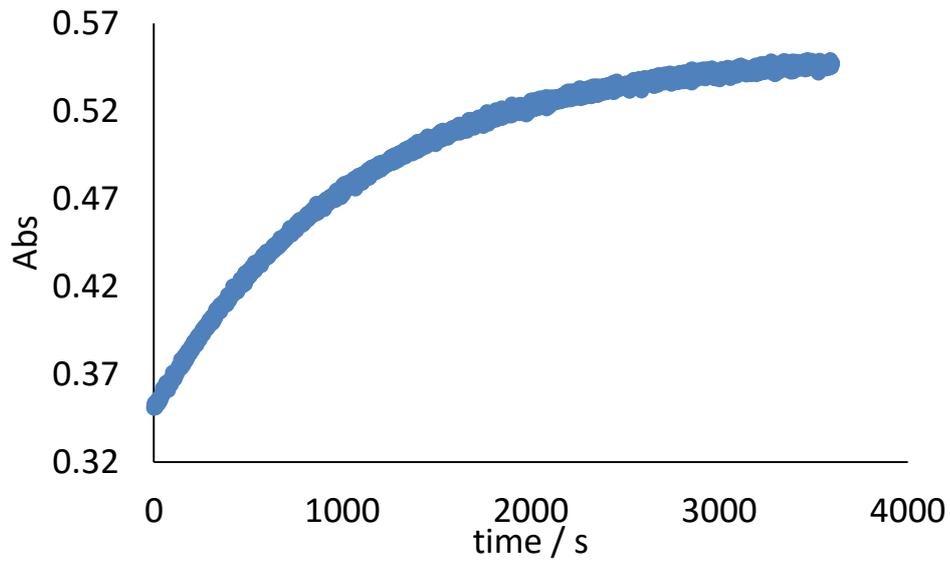
[-OD]	pD	k	logk
0.1	13.8	3.39E-2	-1.47
0.03	13.28	1.03E-2	-1.99
0.01	12.8	3.49E-3	-2.46
0.003	12.28	9.49E-4	-3.02

4.2 Kinetic Data - Catalysed reactions

pD 8.5
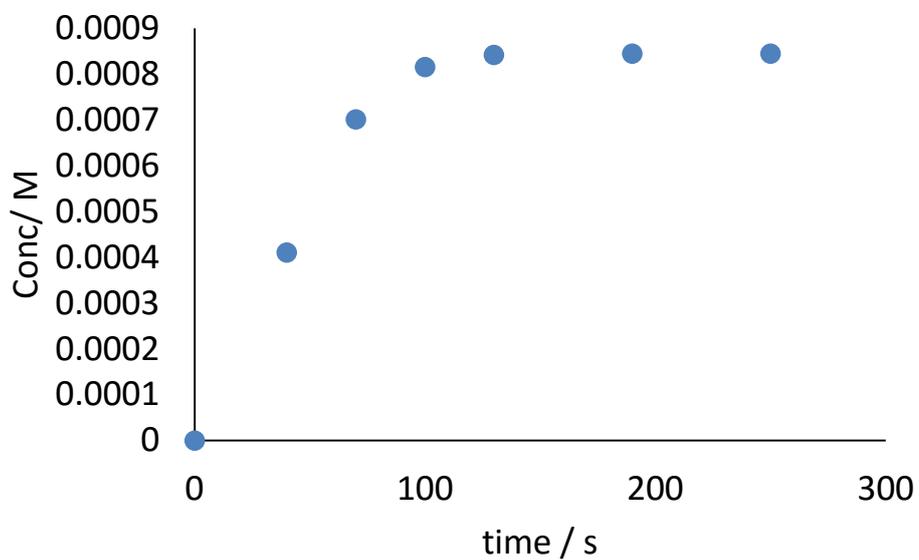
pD 9.4
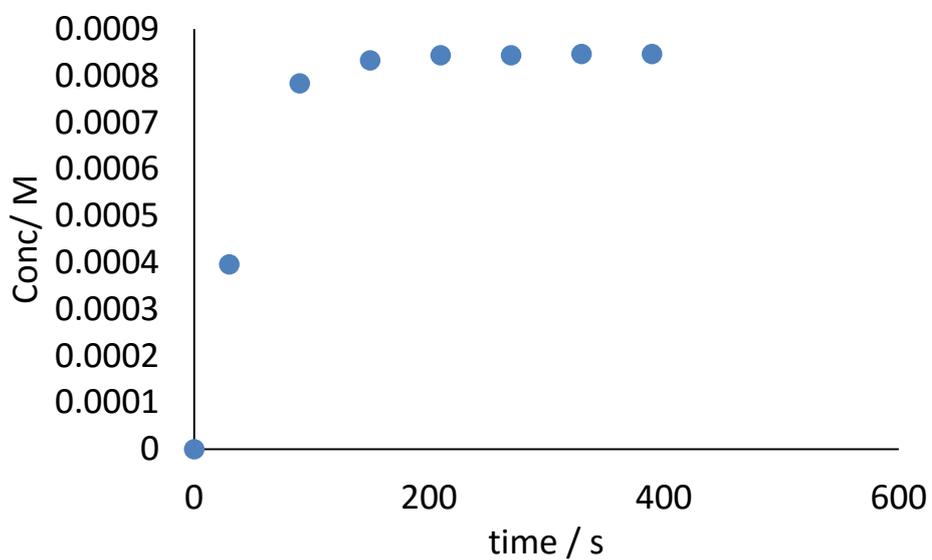
pD 10.2
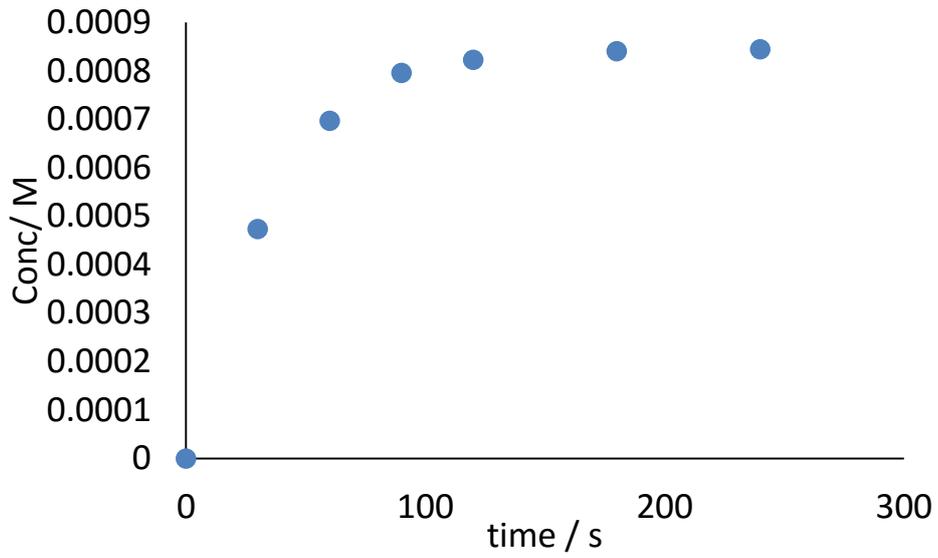
pD 10.7
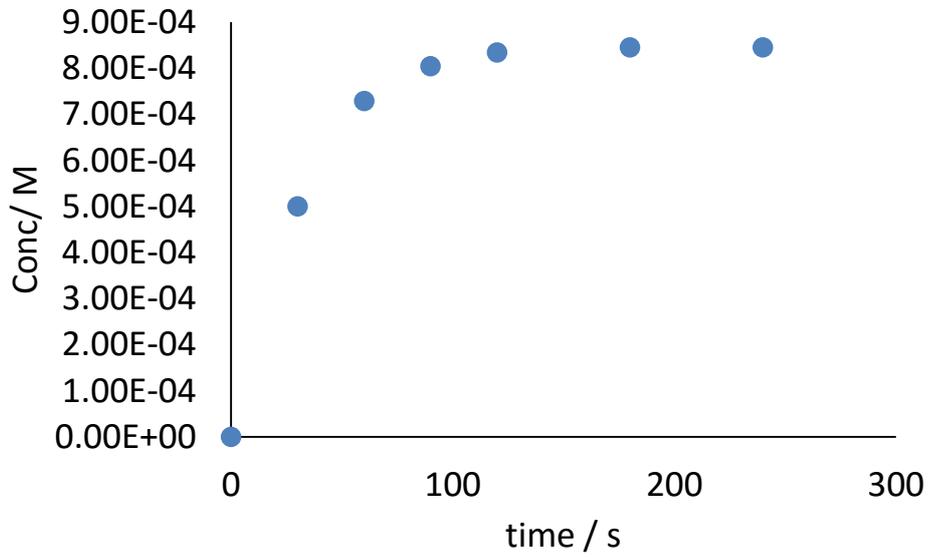
pD 11.3
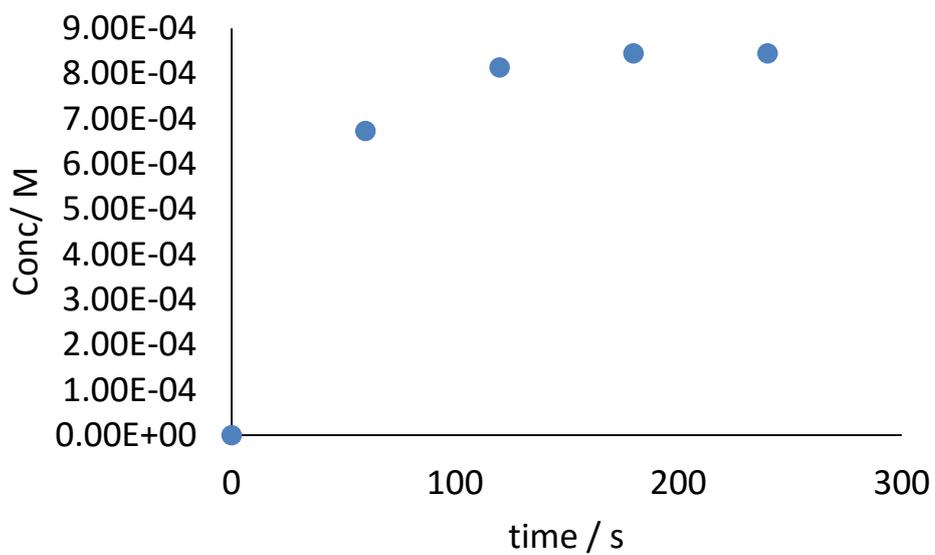
pD	kave	logkave
8.5	3.18E-02	-1.497
9.4	3.82E-02	-1.417
10.2	3.99E-02	-1.399
10.7	4.12E-02	-1.386
11.3	3.59E-02	-1.445

4.3 Kinetic Data - Control reactions

Inhibition, pD 10.75
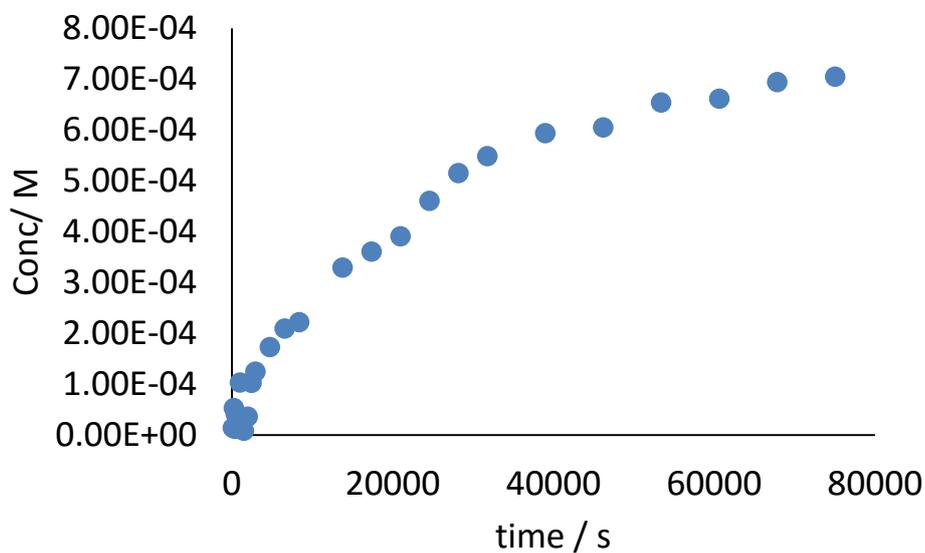
Chorlide, pD 10.82
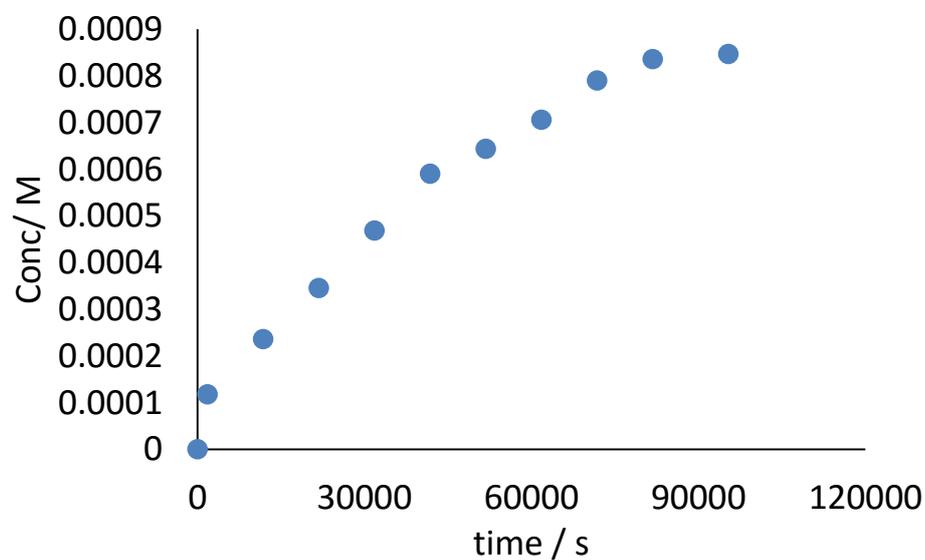
Control	pD	kave	logkave
Inhibition	10.75	3.07E-05	-4.513
Chloride	10.82	2.90E-05	-4.538

Appendix 5

Publication reprints