
# COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.

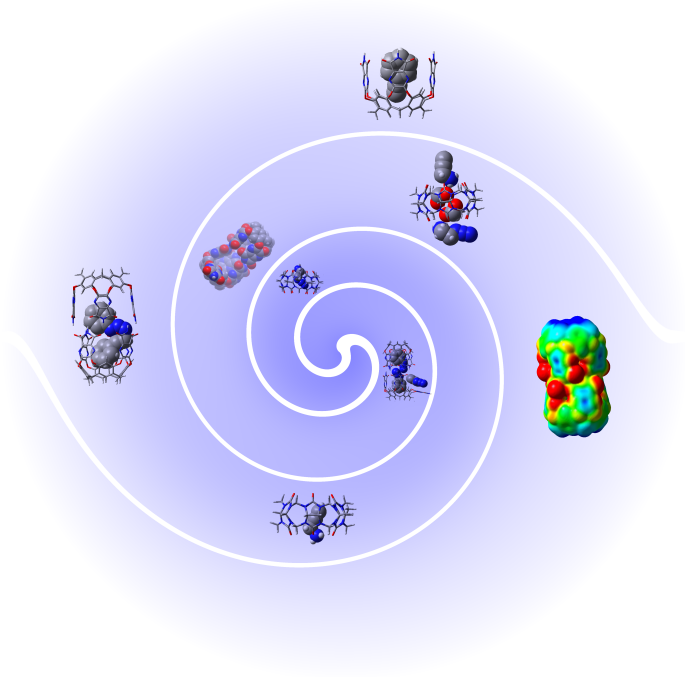
**Charles Goehry**

**Dipòsit Legal: T 1545-2014**

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Computational studies on host-guest catalysis

**Charles GOEHRY**



**Doctoral Thesis**

Institute of Chemical Research of Catalonia (ICIQ)
Rovira i Virgili University

Tarragona, May 2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Computational studies on host-guest catalysis

## Doctoral thesis

**Charles GOEHRY**

Supervised by Prof. Feliu Maseras

Institute of Chemical Research of Catalonia (ICIQ)



Rovira i Virgili University
Tarragona

May 2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014


ICIQ - Institut Català d'Investigació Química
Avgda. Països Catalans 16,
43007 Tarragona (Spain)

WE STATE that the present study, entitled "Computational studies on host-guest catalysis", presented by Charles GOEHRY for the award of the degree of Doctor, has been carried out under our supervision at the Institute of Chemical Research of Catalonia ICIQ, and that it fulfills all the requirements to be eligible for the European Doctorate Award.

Tarragona, 21 May 2014

Doctoral thesis supervisor — Doctoral thesis co-supervisor

Prof. Dr. Feliu Maseras — Dr. Maria Besora

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Acknowledgements

This work has been carried out in the Institute of Chemical Research of Catalonia (ICIQ) in Tarragona (September 2010 - June 2014) under the supervision of Prof. Feliu Maseras and Dr. Maria Besora.

I express my warmest thanks to Prof. Feliu Maseras who provided me the opportunity to me come in Tarragona to undertake this thesis. He gave me a lot of freedom and was always supportive. Without his supervision and constant help this dissertation would not have been possible.

I would like thank Dr. Maria Besora for the many things she taught me and her patient help throughout my PhD.

I want to thank Prof. Nuria Lopez and Prof. Carles Bo for their kindness, for welcoming students from all over the world, and for taking part in organizing the life of the research center.

I would like to thank Dr. Jochen Blumberger for his hospitality in his group in UCL (University College London) during the summer of 2012. This was a very important experience to me. I could use new computational tools and to explore a new research field.

Special thanks go to Nuria Vendrell. She makes a remarkable job and smoothens all the administrative processes. I received uncountable direct and indirect support from her.

I also want to salute my numerous coworkers and former coworkers for the good times spent together and what we learned

from each other.

Finally my most important acknowledgments are towards my immediate family who have been the single biggest motivating force in my life. I thank my mother Mrs Michele Goehry and my father Dr. Christian Goehry for their encouragement, advice and for the many sacrifices they have made over a number of years. I also thank my sister Miss Pauline Goehry for her support, advice and the joy she brings around herself. Last but not least, I also want to thank my wife Mrs Hanqing Li Goehry. There are uncountable reasons why I should thank her, but there is not enough space for that here. Moreover, I welcome our baby girl Lætitia Goehry, whose birth brought a whole new light in our lives. This dissertation is dedicated to them.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Contents

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Chapter 1

# Introduction

## 1.1 Supramolecular Chemistry

Supramolecular chemistry is a dynamic, interdisciplinary and relatively young field that has been defined by Jean-Marie Lehn as "the chemistry of the intermolecular bond" and "the chemistry beyond the molecule".[1] A great potential for applications is emerging from the creative design of interacting molecules. Controlling the non-covalent bond, understanding and predicting the behavior and interplay of large molecular assemblies is the goal of the researchers in the field.

In the 1990s and 2000s supramolecular chemistry and catalysis really blossomed and became a mature field. New classes of supermolecules appeared. Among them stand porphyrin-based supermolecules,[2,3] self-assembled nanotubes,[4] supramolecular electronic polymers,[5] bio-inspired foldamers,[6] supramolecular gels,[7] hydrogen-bonded assemblies[8] and metal-based self-assembled clusters.[9,10]

In this chapter, we will go back to the genesis of the field, and then review the different types of interactions that are involved in supramolecular chemistry. Finally, we will see the capabilities of such systems to catalyze reactions by focusing on several successful examples.

### 1.1.1 The Origins of Supramolecular Chemistry

Much progress has been done[11] since the synthesis of the first macrocyclic polyethers[12–14] and cryptants,[15,16] which have the property to selectively trap alkali cations and halide anions, as displayed Figure 1.1.

In this case, a specific compound binds to the host which excludes other potential guests on specific criteria such as size, shape or charge. This is

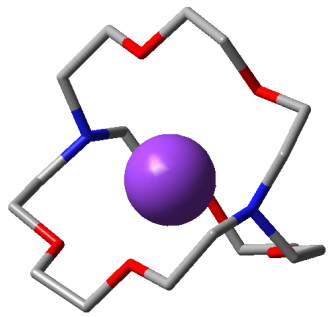


Figure 1.1: *A cryptand trapping a potassium cation.*

called supramolecular recognition and is a key concept used for supramolecular systems. It soon became clear that larger hosts were needed to catch larger guests molecules.

A molecular "pocket" is a concave structure that exhibits the advantage of sterically constrain binding. The guests that can bind to it will see their conformational freedom reduced. This binding can be achieved by various strategies, such as van der Waals interactions, hydrogen-bonds, $\pi$-stacking or electrostatic interactions. Just as in biochemistry, those forces can be combined. Also, a cavity presents the advantage of using three dimensions and extending the surface of contact between the host and the guest, which enhances their interactions.

### 1.1.2 Covalent Structures

Catenanes[17,18] and rotaxanes[19] were exploited, as well as concave structures: the calixarenes.[20,21] Inclusion compounds such as cyclodextrines[22] or cucurbiturils,[23,24] showed particular abilities. They were found to hold specific molecules or even to promote reactions or favor a reaction pathway.

Each type of macrocycle is also a class in the sense that they are typically built from a single pattern, like a frieze or a chain that would be connected at its extremities. Therefore a new molecule can be obtained just by varying the amount of repetitions.

Cram studied extensively the synthesis and properties of calixarenes. He notably synthesized a dimeric capsule made of two calix[4]arenes units covalently bound, therefore capable of permanently trapping a guest.[25]

### 1.1.3 Self-Assembly

Preorganization is a central notion in the design of new assemblies. In the framework of supramolecular chemistry, self-assembly is a spontaneous process where disordered and separate molecules aggregate and organize themselves under the influence of non-bonded interactions.

Although intrinsically much weaker than covalent bonds, intramolecular or intermolecular forces can be accumulated, as DNA strands do for example. Together, they can become strong enough to overcome entropic effects arising during the process of self-assembly.

By foreseeing their dimensional arrangement, one can design simple building blocks that will spontaneously aggregate with each other to construct complex objects such as nanowires[26] and helicoidal structures.[27]

### 1.1.4 Self-assembled Molecular Flasks

Unlike covalent hosts, self-assembled molecular flasks, are obtained from the attraction of small components non-covalently bound. They provide a unique localized microenvironnement by partially or totally isolating the guests from the solvent.[28,29]

Several hydrogen bonded supramolecular capsules have been proposed by Rebek, who gave them names according to their shapes: softball,[30] tennis ball,[31] capsule.[32]

The softball is relatively small, and made of two identical subunits. They self assemble through N-H – O hydrogen bonds. The curvature of the system, necessary to have a nearly spherical, empty cage is provided by two glycoluril units on each monomer, see Figure 1.2.

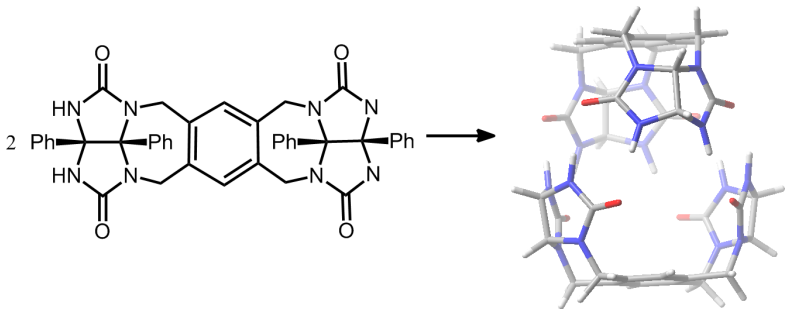



Figure 1.2: *Rebek's supramolecular self-assembly "tennis ball". Phenyl groups pointing outside the assembly have been replaced by hydrogen atoms for clarity.*

Calix[4]pyrroles were exploited by Ballester et al. for the preparation of a

hydrogen-bonded capsule.[33] Eight urea groups interact as the two monomers are intricated, see Figure 1.3

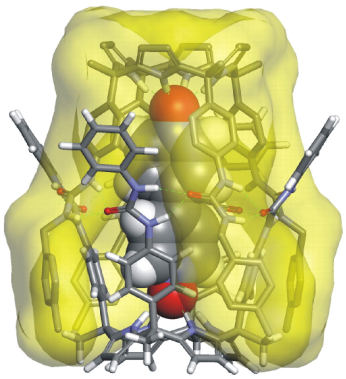


Figure 1.3: *Ballester's self-assembled calix[4]pyrrole capsule*

Fujita proposed an octahedral cage made of four organic triangular and planar ligands held together by six metal ions,[34] see Figure 1.4. It displays a large inner core in which guests can be encapsulated.

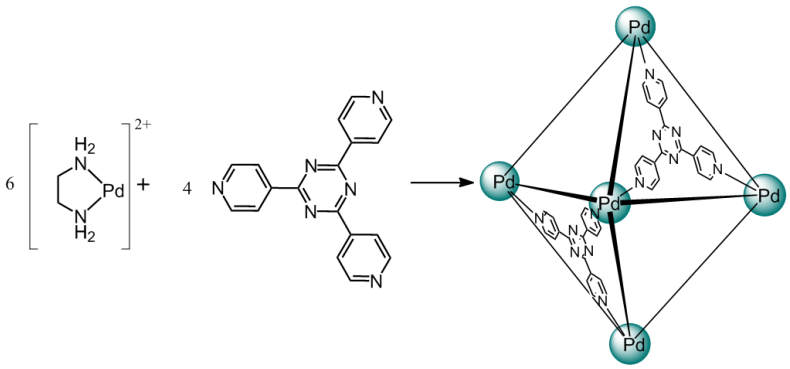



Figure 1.4: *Fujita's self-assembled molecular flask*

A similar strategy exploiting metal-ligand interaction used by Raymond[35] resulted in the synthesis of a pyramidal host (Figure 1.5) that exhibits the ability of exchanging guests dynamically and selectively.

## 1.2 Non-covalent Interactions

Non-covalent intermolecular forces characterize supramolecular chemistry. These reversible, weak interactions belong to several categories enumerated

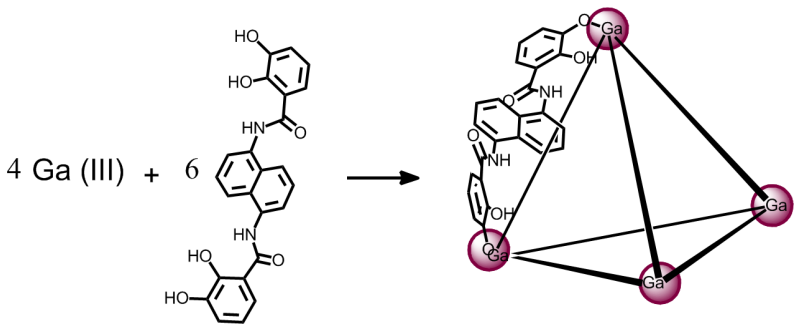



Figure 1.5: *Raymond's self-assembled supramolecular cage*

here.

They open to researchers the road for precise interactions that occur in biology such as molecular recognition, substrate binding, enzyme catalysis or encoding, decoding and storing of information.

### 1.2.1 Electrostatics

Electrostatics interactions form a major contribution to non-covalent binding. The electronic distribution of a supermolecule contains positively and negatively charged regions that can interact favorably with a region of opposite charge. Two groups containing charges of the same sign will on the other hand repel each other.

Formally, one can split such interactions in three categories that are charge-charge ("ionic"), charge-dipole and dipole-dipole (Keesom forces) interactions, but they all are fundamentally of the same nature.

Going further, a permanent dipole can induce a dipole in an otherwise neutral region (Debye forces) and dipoles can appear instantaneously (London dispersion forces).

Together, London, Debye and Keesom forces form the van der Waals forces, see Figure 1.6.

As we will see in the next chapter there are several ways of modeling such interactions.

### 1.2.2 Dispersion

Dispersion forces, also referred to as London forces are universal middle and long-range forces resulting from dynamical correlations between fluctuating charge distributions.[36] Therefore, their nature is quantum-mechanical. It

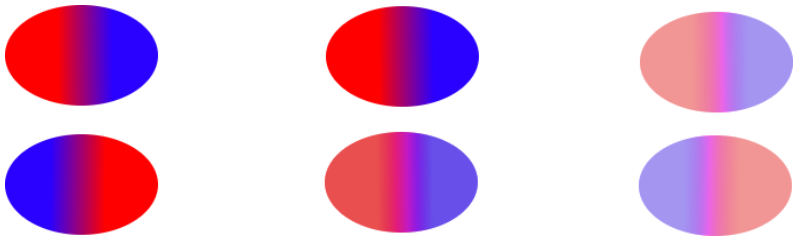


Figure 1.6: *Three types of interactions involving dipoles. Left: alignment of two permanent dipoles. Middle: Interaction of an induced dipole (bottom) with a permanent dipole (up). Right: two instantaneous dipoles interacting. Red and blue represent regions of opposite charge.*

represent the weakest electrostatic force but gains in magnitude as interacting molecules become larger and more keen to stick to each other.

Considering two molecular fragments, the electronic motion of an electron of a first fragment results in an uneven distribution at a given time, which produces an instantaneous dipole. This dipole will affect the electronic cloud of the other fragment, creating an attraction. This attraction varies as the inverse sixth power of the distance between the fragments, but higher-order electric moments lead to other terms such as induced quadrupole-dipole, induced quadrupole-quadrupole, etc. which vary as $1/\mathrm{R}^8$, $1/\mathrm{R}^{10}$.

Although their magnitude is relatively weak compared to other non-bonded interactions, solely dispersion forces may liquefy rare gases in absence of permanent dipole. Dispersion forces grows stronger as the electron density, the polarizability and the size of the system become more important.

### 1.2.3 $\pi - \pi$-Stacking

Despite extensive experimental and theoretical investigations[37–41] the nature of aromatic stacking still suffers a lack of clear definition, although consensus is in favor of its existence. The difficulty relies in the isolation of "pure" stacking effects, since interaction energies are mixtures of electrostatic and dispersion effects. Additionally, the term $\pi-\pi$-stacking is misleading, since it does not refer to p electrons or $\pi$ orbitals, but rather to aromaticity, which is an exclusive characteristic of delocalized, planar rings following the Huckel's rule.

The system usually exploited to study the characteristics of $\pi - \pi$ interactions is of the benzene dimer, or a combination of benzene and/or substituted benzene rings. There are three representative (stable) configurations of benzene dimers: Sandwich, T-shaped and Parallel-displaced, see

Figure 1.7.

Sanders[42] proposed a model which allows to estimate which configuration would be favored, based on the substituent. In most cases, the T-shaped and parallel-displaced geometries are predicted to be favored, while the sandwich geometry only occurs when one ring is electron-rich and the other electron-poor.

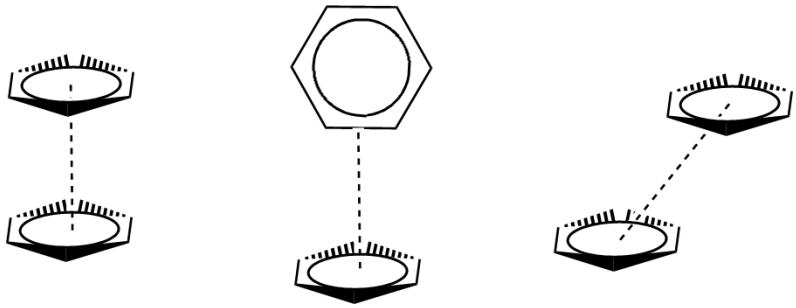


Figure 1.7: *The three main configurations of the benzene dimer, characteristic of $\pi - \pi$ stacking. From left to right: sandwich, T-shaped and parallel-displaced.*

### 1.2.4 Hydrogen Bonds

Hydrogen bonds are weak, reversible, non covalent bonds. Their strengths typically ranges from 1 to 4 kcal/mol. If a hydrogen bond exist between say $\mathbf{A}$–$\mathbf{H}$ and $\mathbf{B}$, then $\mathbf{A}$–$\mathbf{H}$ is the hydrogen bond donor and $\mathbf{B}$ is the acceptor.

$\mathbf{B}$ withdraws electronic density from $\mathbf{H}$, due to their different electronegativities. This leaves the proton partially unshielded and the heteroatom $\mathbf{B}$ can play the role of the hydrogen bond acceptor, by sharing the electronic density of a lone pair with the depleted hydrogen.

The hydrogen bond is a strong dipole-dipole interaction.[43] Therefore, as a $\mathbf{C}$–$\mathbf{H}$ bond is not polar enough, it is a poor hydrogen bond donor. On the other hand, $\mathbf{O}$–$\mathbf{H}$ and $\mathbf{N}$–$\mathbf{H}$ are good donor groups while $sp^2$ hybridized oxygen are good acceptors for instance.

Known examples of intermolecular hydrogen bonds are found in water, carboxylic acids, and nucleotides in DNA strands, see Figure 1.8.

Supramolecular polymers and polymeric networks can be obtained using hydrogen bonds, generally by using several of them on two or more remote sites of a unique molecule.[44]

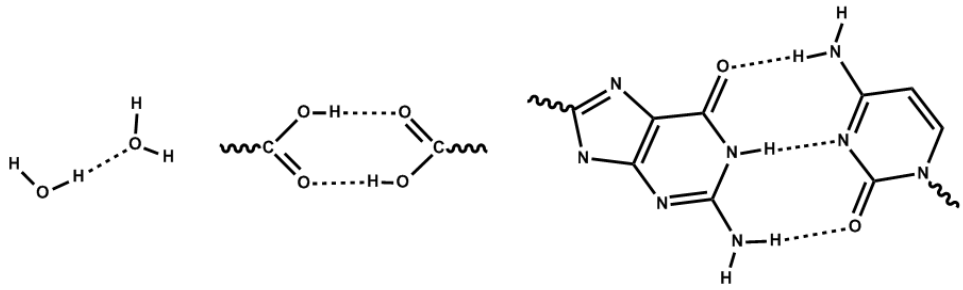


Figure 1.8: *Typical examples of H-bonding. From left to right: water dimer, carboxylic acid dimer and guanine-cytosine dimer.*

### 1.2.5 Reversible Metal Bonds

Coordination chemistry can also be used in supramolecular systems. Exploiting the reversibility of some metal-ligand bonds, Sanders created supramolecular assemblies[45,46] between a metal complex substrate with a $sp^2$ Nitrogen, see Figure 1.9. The Zn complex is constituted by an organic framework of four porphyrins arrange circularly and connected by carbon linkers with a Zn atom in the middle of each porphyrin. An extra porphyrin substituted by pyridine groups is placed at the center. The pyridins reversibly connect to the four surrounding zinc centers.

This system is different from other hydrogen-bonded of supramolecular complexes as metal-ligand interactions offer more strength and rigidity than hydrogen bonds.

## 1.3 Catalysis With Supramolecular Systems

Molecular pockets may be used to promote chemical reactions by holding reactants together in a favorable orientation. Complexation resembles the lock-key model in enzymatic catalysis. Similar strategies may be applied to enhance the rate of reactions. Selective non-covalent binding requires compatibility of size, shape and electrostatic surfaces between the substrate and the catalyst.

Catalysis requires substantial lowering of activation barriers as well as regeneration of the catalyst. The increase in rate of the formation of the desired product has to be larger than for the by-products.

Enzymes take advantage of strong substrate binding, see Figure 1.10. The green curve symbolizes the catalyzed reaction, whose activation energy is lower. The first and last transition state account for substrate binding and

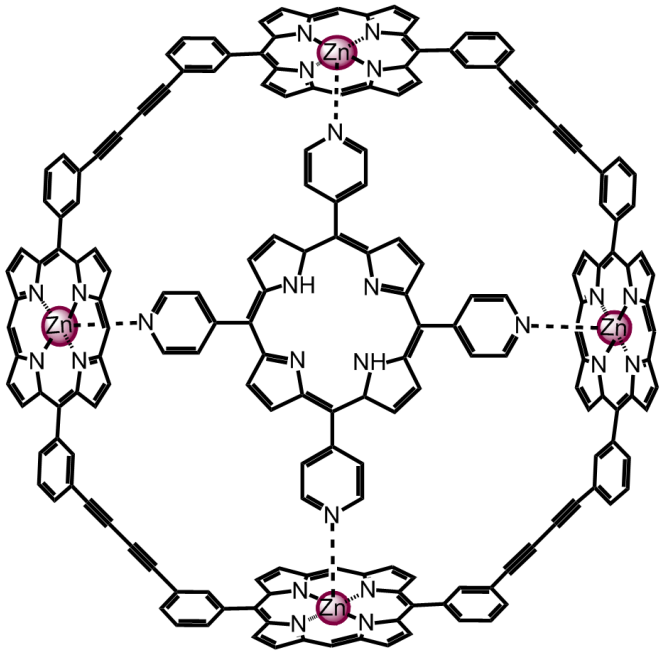



Figure 1.9: *Sanders' Supramolecular assembly.*

substrate release.

A commonly used strategy is to use simultaneously favorable interactions to capture a given substrate and strain to bring it closer to the geometry of the transition state. A drawback resulting from the application of this principle is product inhibition, which occurs when the product remains strongly bound to the supramolecular catalyst, as the geometry of the product is often similar to the geometries of reactive species.

### 1.3.1 Successful Examples of Host-Guest Catalysis

Similar effects have been observed in supramolecular catalysis. In this section, we will present some of the work that has been accomplished by the researchers in the last three decades. We chose some of the most successful examples, although they remain limited in number. Additional examples can be found in the literature.[28,47–49]

#### Deacylation With Modified Cyclodextrines

In 1979, Breslow, inspired by the work of Bender[50] used a modified cyclodextrin as an enzyme model,[51,52] see Figure 1.11.

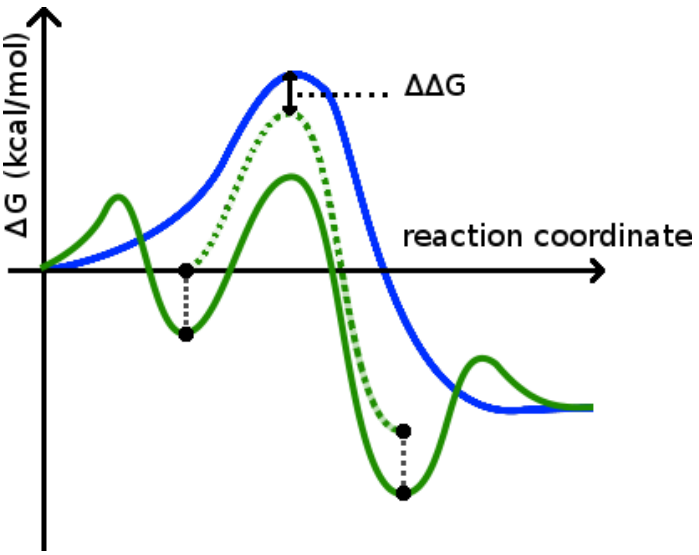



Figure 1.10: *Simple representation of enzymatic catalysis. Blue: uncatalyzed reaction. Green: enzyme-catalyzed reaction.* $\Delta\Delta G$ *represents the catalytic energy gain*

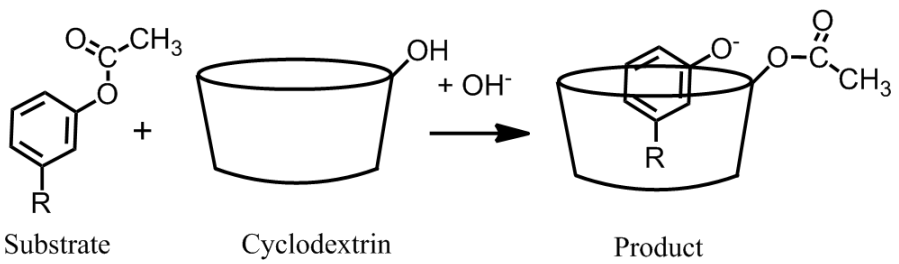



Figure 1.11: *Deacylation of the substrate by a cyclodextrin, as studied by Breslow.*

Their approach consisted of adjusting both the structure of the cyclodextrin and the structure of the substrates. The rate of acylation of $\beta$-cyclodextrines was found to be accelerated by as much as $7.5*10^5$ times compared to the background reaction.

Remarkably, not only confinement effects, but also enthalpic stabilization is reported, however the authors state it could come from solvation effects. The best substrate ferrocene-2-acrylic acid completely fills the $\beta$-cyclodextrin cavity, while other guests do not meet this criterion.

For this guest, acceleration is reported to be two orders of magnitude lower when using a smaller $\alpha$-cyclodextrin cavity. Modifying further the $\beta$-

cyclodextrin by "capping" it resulted in an even larger acceleration rate (*c.a.* $1*10^6$-fold).

### Catalysis by Cucurbit[6]uril

Another successful example is provided by Mock and coworkers, who used a cucurbit[6]uril to promote a 1,3 dipolar Huisgen cycloaddition. The strategy here consists of bringing the positively charged guests in close contact in the receptor, which is triggered by the oxygen-fringed portals, see Figure 1.12. Comparison of bimolecular reaction results in a calculated acceleration of $5.5*10^4$, but product inhibition prevents catalysis.

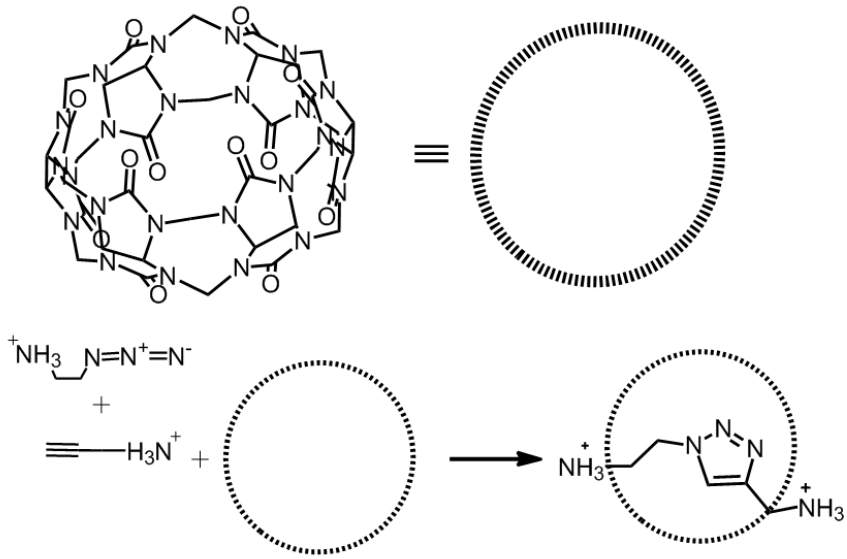



Figure 1.12: *A cucurbit[6]uril enhances a Huisgen cycloaddition.*

This system will be at the center of in Chapter 3. A previous study carried in our computational laboratory is available, but newer and better models are available since then so we decided to come back to it. Also, this system displays the highest effective molarity as reported by Di Stefano[47] in 2004. We refer the reader to section 1.5 for an introduction to effective molarities.

### Diels-Alder With Linked Porphyrines

Sanders used a porphyrin trimer to catalyze a Diels-Alder reaction with up to a 1030-fold acceleration of a Diels-Alder reaction,[2] see Figure 1.13. Control

over the selectivity has been achieved by varying the lengths of the three linkers.

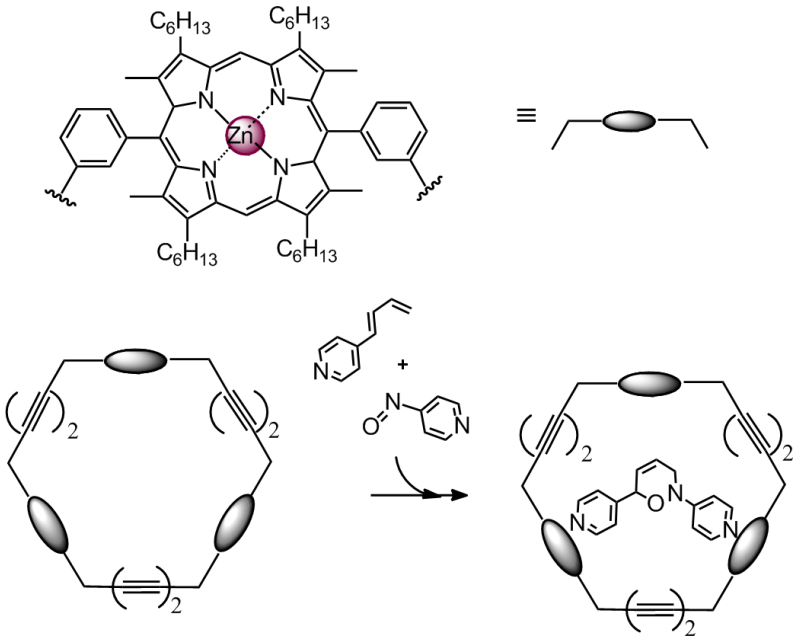



Figure 1.13: *Diels-Alder acceleration with metalloporphirin trimers*

It appears also that the largest acceleration rates are correlated to the largest binding constants of the reactive guests. As it is often the case, the products binds strongly to the receptor. A strategy used to open the door to catalytic turnover was to rearrange the product through a pericyclic reaction.

#### Alcohol Oxidation With Cyclodextrin

Benzyl alcohol oxidation has been reported to be catalyzed by a bridged cyclodextrin by Bols,[53] see Figure 1.14. A series of substrates were screened and accelerations up to $6*10^4$-fold were found under mild conditions. A cyclic mechanism is proposed. The process would be favored by hydrogen bonding between the proton of the benzylic alcohol and a keto hydroxy group of the host.

#### Self-Assembled Non-Metallic Nanocages as Nanoreactors

Rebek and his coworkers designed several molecular cages reversibly formed by self-assembly of two identical substructures.[30–32]

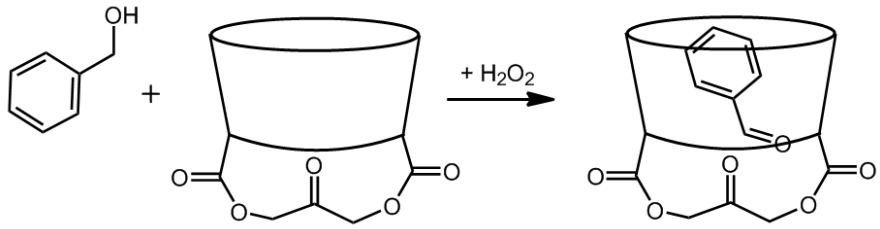



Figure 1.14: *A cyclodextrin catalyzing the transformation of benzyl alcohols to aldehydes*

Those cages have been used to accelerate bimolecular reactions such as Diels-Alders[54] and Huisgen cycloaddition.[55] Early calculations by Houk suggested that "doors" could open to let the guests enter.[56] Several hydrogen bonds of the assembly can be broken simultaneously to a conformational change in the molecule leaves behind an open route to the core of the cage. This way not all favorable hydrogen bonds would need to be broken for a molecule to enter or leave.

For instance, a Diels-Alder in a supermolecular "Softball" is reported to be accelerated 200 times respect to the background reaction, see Figure 1.15. A common problem that arises is the lack of turnover, as the products remain trapped in their respective hosts.

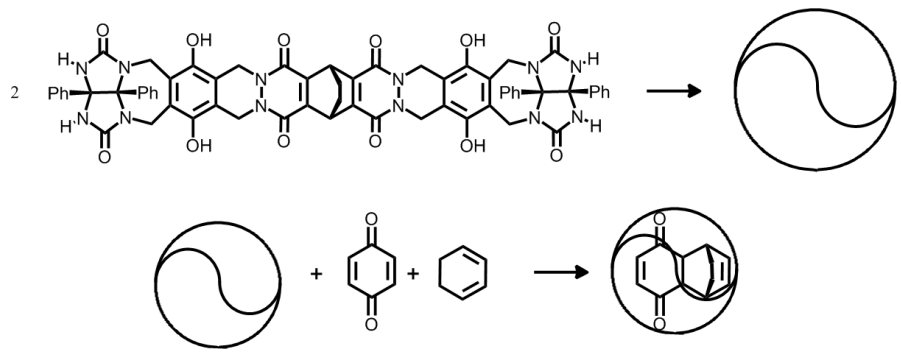



Figure 1.15: *Rebek's self-complementary "Softball" promotes a Diels-Alder reaction.*

A similar strategy has been used by Rebek: using synthetic routes developed by Gutsche[57] and Cram,[58] he designed a self-complementary capsule. It displays a large inner space and has a great ability to co-encapsulate aromatic guests. It has been found to enhance a Huisgen cycloaddition,[32,55] see Figure 1.16. This system will be studied in detail in Chapter 4, since

it is self-assembled, displays a large inner space and entirely forbids direct interaction of the guests with the solvent. Effective molarities displayed by this system are also large, but two orders of magnitude below those of the cucurbit[6]uril[47] system.

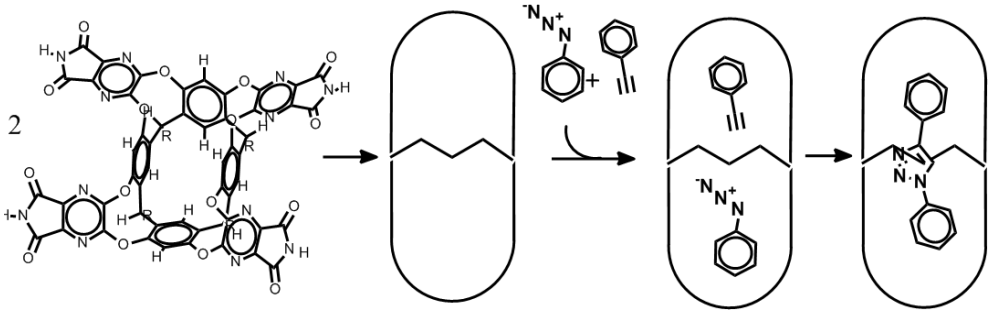


Figure 1.16: *Rebek's self-assembled "Capsule" promotes a Huisgen cycloaddition.*

**Self-Assembled Nanocages Using Labile Metal-Ligand Interaction Enhancing a Nazarov Cyclization**

Raymond et. al demonstrated the ability of a $Ga_4L_6^{12-}$ self-assembled nanocage that exploits labile metal-ligand interactions[59] to accelerate by a million-fold a Nazarov cyclization,[60] see Figure 1.17. The turnover issue was circumvented as Sander did, by making the product of the encapsulated reaction subsequently react again, resulting into turnovers numbers up to 160. The product of the second reaction is polar and soluble in water. Also it is larger than the product of the first reaction, so that the host-guest complex can be destabilized. The product is then released, leaving the nanocage ready for the following cycle.

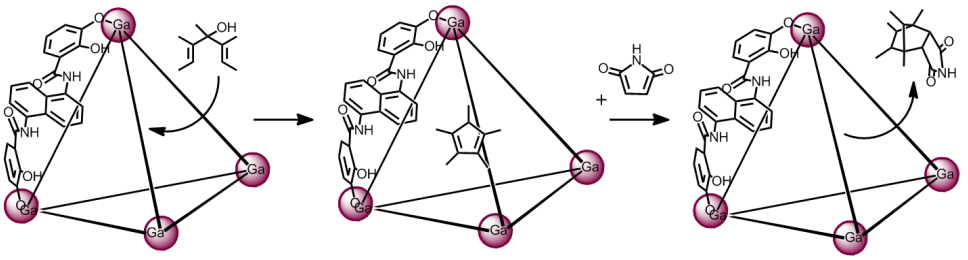


Figure 1.17: *Raymond's Self-Assembled supramolecular host catalyzing a Nazarov cyclization*

## 1.4 Click Reactions

The concept of "click" chemistry was created in 2001 by K. Barry Sharpless,[61] and describes a method to quickly and efficiently synthesize target molecules from smaller units, in an attempt to mimic nature. Practical and reliable reactions are used and the philosophy focuses on easy-to-make building blocks.

The characteristics of a "click" reaction have been defined and include modularity, high yields, simple conditions, availability of starting materials, absent or benign solvent, easy product isolation, high thermodynamic driving force and atom economy. Also, completion should be fast and lead to a single product.

### 1.4.1 Organic 1,3-dipolar Cycloaddition

1,3 dipolar cycloaddition takes place between a 1,3 dipole such an azide, ozone or a nitrous oxide and dipolarophile such as alkene or a alkyne. The 1,3 dipole reacts with the dipolarophile in a concerted way, and form a five-membered ring, see Figure 1.18. Two regioisomers are obtained, in amounts that depend on the nature of the reactants.

This reaction is ofter called Huisgen reaction, after the name of R. Huisgen, who investigation the kinetics and mechanism of such reaction.[62]

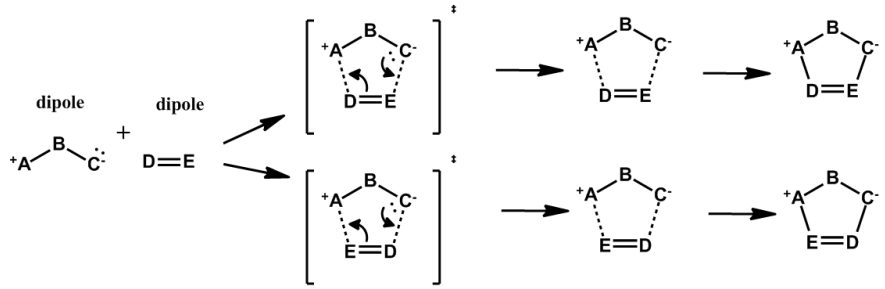



Figure 1.18: *Huisgen 1,3-dipolar cycloaddition*

These Huisgen reactions have been studied experimentally[62–65] but also computationally by several groups.[66–68] In the absence of strain or copper catalyst, they are not regio-selective, very slow, and require high temperatures, and therefore cannot be considered as "click".

### 1.4.2 Huisgen 1,3-dipolar Cycloaddition Catalyzed by Copper

Huisgen 1,3-dipolar cycloaddition are usually performed either thermally or catalyzed by copper,[69–71] and the reaction can be enhanced by microwave heating.[72]

The copper catalyst binds to the terminal alkyne and renders the reaction very efficient and regioselective, which entitles it for the "click" qualificative, see Figure 1.19.

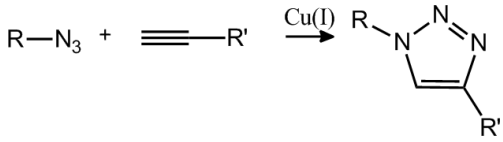



Figure 1.19: *Huisgen 1,3-dipolar cycloaddition catalyzed by copper*

## 1.5 Effective molarities

One of the aspects of host guest-catalysis that can be analyzed is that of entropic versus enthalpic effects, as is our concern in this project. Supramolecular receptors are able to trap more than one guest. As a consequence, not only intramolecular, but also intermolecular reactions can occur. In other words, a bimolecular reaction occurs within a bigger molecule. Or is it really so, and to what extent? Is the reaction still really intermolecular?

Some tools can be used to quantify this effect. Effective Molarity (EM),[47] equation (1.1), has the dimension of concentration. $\mathrm{k_{inter}}$ and $\mathrm{k_{intra}}$ respectively represent the rates for the intramolecular and intermolecular reactions.

$$EM = k_{intra}/k_{inter} \tag{1.1}$$

The value obtained corresponds to the theoretical concentration that one of the reactants in solution would need to achieve the same rate as the accelerated, intramolecular reaction.

Another tool commonly used the effective concentration. It is the concentration achieved by the guest inside the host, considering the volume available. When both effective molarities and effective concentrations are used at the same time, one can Analise the contribution of the concentration in the reaction rate, increased because of the lowered volume surrounding the molecule available, and the enthalpic/entropic modifications of the reaction profile. If the effective molarity is higher than the effective concentration, it means that the reaction is favored by entropic effects, not by concentration effects.

## 1.6 Computation Studies on Host-Guest Catalysis

Synthetic supramolecular catalysts such as the ones presented section 1.3.1 are challenging for computational chemistry, mainly because of the size of the systems. Only a handful of studies on host guest catalysis are available. We will present them in this sections.

### 1.6.1 Cucurbit[n]uril in Diels-Alder Catalysis

Ganguly et al. studied the potential role of two different cucurbit[n]urils in a Diels-Alders reaction between cyclopentadiene and methyl acrylate,[73] see Figure 1.20. They used a pure DFT approach as structures and energies were obtained using M05-2X. Solvent was reproduced using an SMD model. Surprisingly, this reaction has been investigated experimentally in solution and with cyclodextrines hosts, but not with cucurbiturils. The authors demonstrated the role of cucurbit[7]uril in facilitating this cycloaddition. On the other hand, a cucurbit[6]uril was opted out, as the host-guest interaction was poor due to steric repulsion. Finally, they predicted that the end-product was the major product, also for steric reasons.

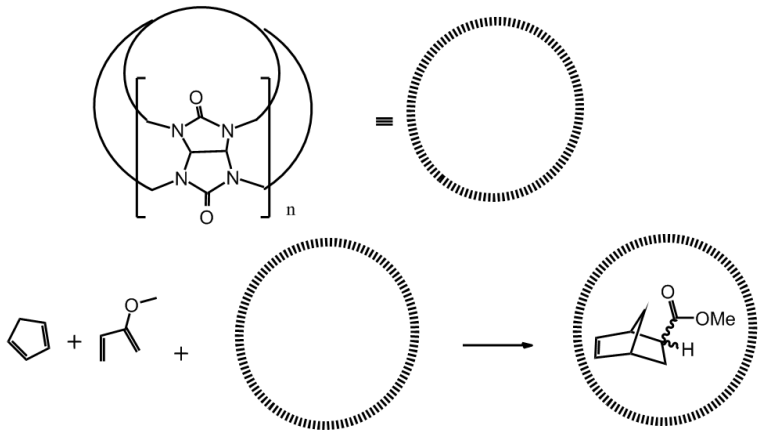



Figure 1.20: *Representation of the catalytic reaction theoretically studied by Ganguly et al.*

### 1.6.2 Cucurbit[6]uril Promoting a Click Reaction

A theoretical study by Maseras et. al is available in the literature.[74] It concerns the acceleration of a Huisgen reaction between an azide and an acetylene by a cucurbit[6]uril host,[75] as presented Figure 1.12. The study

proposed to review the main steps of the cycloaddition using DFT methodology, with a B3LYP functional and a 6-31G* basis set. It reproduced the regioselectivity and acceleration of the reaction experimentally observed. It concludes that the stabilization of the TS plays a minor role in the acceleration and the major effect is the removal of entropic constraints.

In this thesis, we want to come back on this communication to improve the results obtained on this system. We will include a methodology survey as well as a refined mechanistic and kinetic study.

### 1.6.3 Self-Assembled Capsule Promoting a Diels-Alder Reaction

Shuah Li et al. studied the Diels-Alder reaction between a p-quinone and a cyclohexadiene, catalyzed by a self-assembled capsule.[76,77] The reaction is presented Figure 1.15. The authors use a combination of Monte Carlo, force field and DFT simulations. Monte Carlo calculations are used to generate a large amount of conformations, evaluated with the Amber Force Field. Among the conformers generated, 20 are selected for subsequent evaluation with DFT method M06-2X/6-31G**. Further on, transition states and product are found with the same DFT method. Finally, results are compared with the background reaction.

The results show the complexes are bound with hydrogen bonds and $\pi$-stacking. Also, the relative position of the reactants inside the capsule influences the energy of activation required, and induces endo-selectivity. Finally, the free energy of reaction is lowered by the capsule, which allows to reduce entropic constraints.

### 1.6.4 Enhanced Reactivity of a Triazole Formation by a Molecular Flask

The azide-alkyne reaction enhanced by Rebek's capsule[32,55] has been studied by Palacios et al.,[78] see Figure 1.16. Using ONIOM methodology, they investigated the stability of the capsule and several intermediaries along the encapsulated reaction path. Finally, they compared the results to the reaction in solvent. The supramolecular capsule was treated at the PM6 (semi-empirical) level, while reactive parts where assigned to a M06-2X/6-311G** treatment. Solvent effects appear to be missing, as well as free energies, as only gas-phase enthalpies are discussed.

This work could reproduce the regioselectivity induced by the supermolecule. As no stabilization of the transition state by the capsule was

found, the acceleration of the reaction could not be rationalized. Finally, the lack of turnover could not be quantitatively explained.

We selected this experimental study[55] for this thesis, and will intend to improve the results obtained theoretically.[78]

### 1.6.5 Acceleration of the Hydrolysis of Orthoformates by a Self-Assembled Supramolecular Cage

Warshel et al. used a empirical valence bond (EVB) model to reproduce the hydrolysis of formate, accelerated by a nano cage developped by Raymond,[35,79] see Figure 1.5 and 1.21.

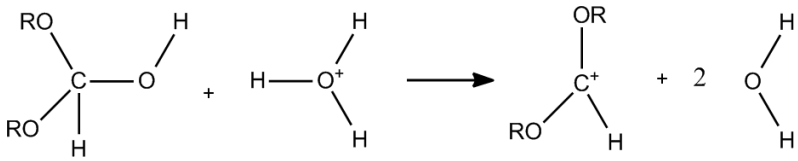



Figure 1.21: *The formate hydrolysis reaction*

These calculations intend to reproduce the electrostatic effects to rationalize the stabilization of the transition state of the reaction by the host. Substrates were treated at the MPW1PW91/6-311++G** level, and solvent effects were evaluated with the COSMO model. From this work, it is concluded that the rate accelerations are due to an electrostatic stabilization of the transition state. The oxonium cation needed for the reaction, as well as the charged transition state appear very well stabilized by the negatively charged nanocage, while the reactant is neutral. The role of the entropy does not seem predominant.

The authors point out the relevance of theoretical approaches in determining the origin of catalytic effects. They can help understanding the different contributions (electrostatic, entropic, etc) to a given reaction, particularly in the case of host-guest catalysis. Finally, the authors point out that unlike it is observed in enzymatic catalysis, the inner shape of the nanocage does not stabilize the transition state. The role of the nanocage seems more limited to constrain the entry of guest by rejecting the largest ones.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Objectives

Aiming at making specific reactions more efficient, as it occurs in living organisms with enzymes, a detailed understanding of existing man-made systems is necessary to improve them further. Doing so, we hope to guide the researchers in the field in the conception, design and innovation concerning new large chemical entities which require careful adjustments and specific structural features, as we know from proteins.

The main purpose of this doctoral thesis is to learn about host-guest catalysis. A detailed understanding of energetic and mechanistic features of aggregation and reaction involving supramolecular systems is necessary. To this end we identified four objectives:

- Finding the adequate DFT-based computational method for a balanced description of all interactions in a specific host-guest catalysis system.
- Explore the possibilities of alternative methods (QM/MM, tuned MM) for an efficient description of host-guest catalysis
- Application of those theoretical concepts to the computational study of a Huisgen reaction inside a cucurbituril and resorcinarene capsule
- Calculation of complete reaction networks for the systems analyzed and quantitative evaluation of the key steps of the process.
- Put the first steps towards the formulation of a new paradigm in the computation of host-guests systems.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Chapter 2

# Computational Methods and Models

Using theoretical chemistry, one can obtain insights on a selected molecule properties. A profusion of techniques may be used, which have a point in common: they use computer resources to tackle numerous mathematical equations in order to provide a model. Also, they have a common central purpose: estimating the relative energy of a given chemical structure. Other properties can be computationally measured such as spectra, charge distribution, etc.

Different classes of methods can be classified by the way they describe chemical entities. There are five main classes: MM, semi-empirical, HF-based, DFT and in addition, so-called hybrid models can combine them. We will present them briefly along this chapter.

A theoretical chemist has to pick a method that will give him accurate results in an acceptable time. The choice consists of balancing the two by deciding what approximations can be done. The results obtained will be highly dependent on that initial choice and one has to bear in mind that none of them can give physically exact results for systems larger than the hydrogen atom.

## 2.1 Methods

### 2.1.1 Molecular Mechanics

In Molecular Mechanics (MM), there are no explicit electrons, no orbitals, no Schrödinger equation. In some cases, formal or partial charges are used. The objects of molecular mechanics are point-like atoms of specific types and various classes of virtual springs are parametrized to govern their interplay.

The atom types or classes intend to bypass the physics behind the chemistry to focus only on the commonly observed behavior. For instance an $sp^2$ carbon is planar and has 3 neighbors which are ideally separated by 120 degrees. This ideal shape does usually not occur, deformed by cyclic strains or non-bonded interactions for instance. This results in the energy of the system rising, for instance following an harmonic potential.

It is inherently difficult for force fields to reproduce the energetics of bond formation or breaking. For example hydrogenation of ethylene ($sp^2$ carbons) gives ethane ($sp^3$ carbons). At the same time, a H-H bond is broken and two H-C bonds are created. This represents a discontinuity and therefore an inconsistency results from these atom type changes.

An illustrative example with oxygenated water is provided Figure 2.1, showing quantum chemistry intends to reproduce precisely the behavior of electrons, while molecular mechanics does not know electrons. Instead, they see atom types, springs and other empirical parameters.

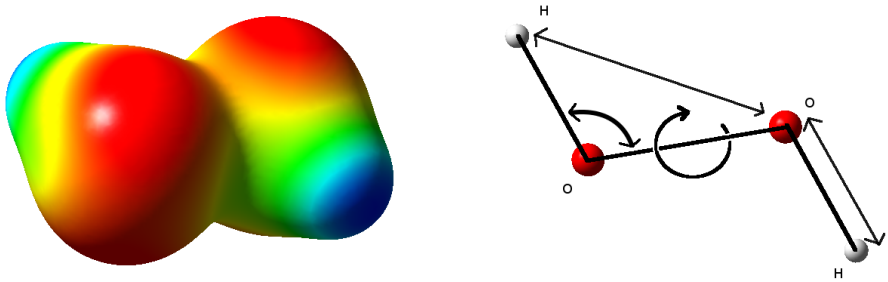



Figure 2.1: *Illustrative example with oxygenated water showing that molecular mechanics does not know electrons, it sees atom types, springs and parameters. Left: Electrostatic potential mapped on density surface calculated with density functional B97D. Right: point-like atoms and their relations: bond stretching, angle bending, torsion and van der Waals interactions.*

The force field provides the energy of a given configuration, based on a usually large set of parameters, which are empirically derived using data sets. Depending on the nature of the data set used to parametrize the force field, a given system may or may not be properly described. The chemical "family" of the data set is therefore important in the choice of a force field.

All force fields are not strictly equivalent, and use different terms and equations in the potential energy function (equation (2.1)) that is applied on a given configuration. Generally, it is divided in bonded (equation (2.2))

and non-bonded (equation (2.3)) terms:

$$E_{total} = E_{bonded} + E_{non-bonded} \tag{2.1}$$

$$E_{bonded} = E_{bonds} + E_{angles} + E_{dihedral}(+E_{crossterms}) \tag{2.2}$$

$$E_{non-bonded} = E_{Electrostatics} + E_{vdW}(+E_{pol}) \tag{2.3}$$

On the right hand of equation (2.2), one can find expressions of the type:

- $$E_{bonds} = \sum_{bonds} k_{bond}(r - r_0)^2 \tag{2.4}$$

$\mathrm{k_{bond}}$ is a constant assigned to each bond type, $\mathrm{r}$ is the bond length, while $\mathrm{r_0}$ is the equilibrium distance. This potential makes each bond behave like a classical spring.

- $$E_{angles} = \sum_{angles} k_{angle}(\theta - \theta_0)^2 \tag{2.5}$$

$\mathrm{k_{angle}}$ is a constant assigned to each angle type. $\theta$ is the value of the angle formed by the 3 atoms considered, and $\theta_0$ is the equilibrium value. This potential intends to simulate the effect of orbitals. For instance a $sp^3$ carbon such as that in methane will adopt a H-C-H angle of *c.a.*. 109 degrees.

- $$E_{dihedrals} = \sum_{dihedrals} k_{dihedral}(1 + cos(n\psi - \psi_0)) \tag{2.6}$$

$\mathrm{k_{dihedral}}$ is a constant assigned to each dihedral type. $\psi$ is the value of the dihedral and $\psi_0$ is the equilibrium value. $\mathrm{n}$ is the multiplicity of the dihedral, which depends on the hybridization of the central atoms. For example, in the ethene molecule, the dihedral constant for $\mathrm{H_2C = CH_2}$ rotation will be high, there will be two maxima and two minima, and the multiplicity will be 2.

Also, a variety of cross terms can be introduced, to reproduce for example out of plane bendings, etc. We do not present them here for clarity, since they are not always used.

On the right side of equation (2.3), two or three terms are usually found:

- $$E_{Electrostatics} = \sum_{ij, i \neq j} \frac{q_i q_j}{r_{ij}} \tag{2.7}$$

Equation 2.7 describe charge-charge interactions, which are considered to be punctual.

There are various formulations for the calculation of empirical van der Waals energies. Here is the example of oplsaa,[80] which uses a Lennard-Jones potential:

- $$E_{vdW} = \sum_{ij,i\neq j} 4\epsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} + (\frac{\sigma_{ij}}{r_{ij}})^{6}] \tag{2.8}$$

The van der Waals term crudely represents weak forces that can be attractive or repulsive, and are very important to maintain cohesion.

Optionally, a non-additive polarizable component can be used to reproduce the effects of induced dipoles, sometimes called "many body effects".

- $$E_{pol} = -2\sum \mu_i E_{i0} \tag{2.9}$$

A threshold (e.g 10 Å) may be applied to remove long-distance interactions that become very numerous as the size of the system increases. Force fields represent a very cheap, tunable solution for modeling molecular structures. A molecule of several hundreds or even thousands of atoms may be optimized on a personal computer rather quickly.

In conclusion, force fields are a very fast way of producing structures, that may be refined using other methods. If the force field is developed to be applied on specific systems or family of systems, the results will be accurate. The major drawback of force fields is the intrinsic difficulty to describe chemical reactivity: bond formation and breaking.

### 2.1.2 Hartree-Fock Theory

Hartree-Fock theory is the basis of electronic structure theory and molecular orbital (MO) theory. Hartree-Fock equations[81] represent the formally simplest level of approximation to solve the many-body Hamiltonian. It states that the wave function that describes quantum systems can be obtained from a single Slater determinant, see equation (2.10). Columns refer to spin-orbitals, while lines refer to electrons.

$$SD = 1/\sqrt{N} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_{\mathrm{n}}(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_{\mathrm{n}}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathrm{N}) & \phi_2(\mathrm{N}) & \cdots & \phi_{\mathrm{n}}(\mathrm{N}) \end{vmatrix} \tag{2.10}$$

In a variational scheme, a trial wave function that is has to be injected in the HF equations, as in equation 2.11. Molecular orbitals (MO, or spin-orbitals) are a linear combination of a set of chosen basis functions $\chi_{\mathrm{p}}$. The basis functions determine the quality of the solution, see section 2.1.3. The variational parameters $\mathrm{c_{pi}}$ have to be optimized respect to each spin-orbital and each basis function.

$$\phi_i = \sum_{p=1}^{M} c_{pi}\chi_p \tag{2.11}$$

For each set of $c_{pi}$ coefficients, one calculates the Slater determinant and the corresponding energy. This implies the calculation of monoelectronic integrals (depending on the coordinates of one electron, but two basis functions) and bielectronic integrals (two electrons, four basis functions).

In Schrödinger formalism (equation 2.12), the hamiltonian operator $\mathbf{H}$ relates the wave-function $\Psi$ to the energy E of the quantum system.

$$\mathbf{H}\Psi = E\Psi \tag{2.12}$$

In the Hartree–Fock method, the Fock matrix is defined by the Fock operator and approximates the hamiltonian, see equation 2.13. The effects of electron-electron repulsion are only averaged.

$$\mathbf{F}_i = \mathbf{h}_i + \sum_{j}^{N}(\mathbf{J}_j - \mathbf{K}_j) \tag{2.13}$$

$\mathbf{h}$ is the one-electron hamiltonian that describes the motion of an electron in the field of all nuclei of the system. $\mathbf{J}$ is the coulomb operator that represents the electron-electron repulsion between the j-th and i-th electrons in the system. $\mathbf{K}$ is the exchange operator estimates the effect of exchanging two electrons. It can be demonstrated that the optimal HF function has to verify the equation:

$$\mathbf{FC} = \mathbf{SC}\varepsilon \tag{2.14}$$

Where the Fock matrix $\mathbf{F}$ contains the Fock matrix elements.$\mathbf{C}$ is the matrix containing the coefficients $c_{pi}$. The overlap matrix $\mathbf{S}$ contains the overlap elements of the basis functions. $\varepsilon$ is the diagonal matrix of orbital energies. This equation fits very well in a variational scheme for its solution.

One then looks for a new set of molecular orbitals to minimize the energy of the system, by diagonalizing Fock's operator applied on MO. At this point, an iterative procedure is required to converge to the solution.

Hartree-Fock formalisms takes into account Pauli formalism but each electron only perceives an *averaged* non-local electron field from other electrons. Although sufficient for a qualitative description of structures, it is not considered sufficient for quantitative description of molecular systems and is therefore a branching point in the family tree of theoretical methods.

Simplifications can be done leading to semi-empirical methods. Refinements can also be done to introduce correlation energy, either in a perturba-

tional (MPn family[82]), or variational (configuration interaction[83]) form, see Figure 2.2 for a graphical depiction of HF-based methods.

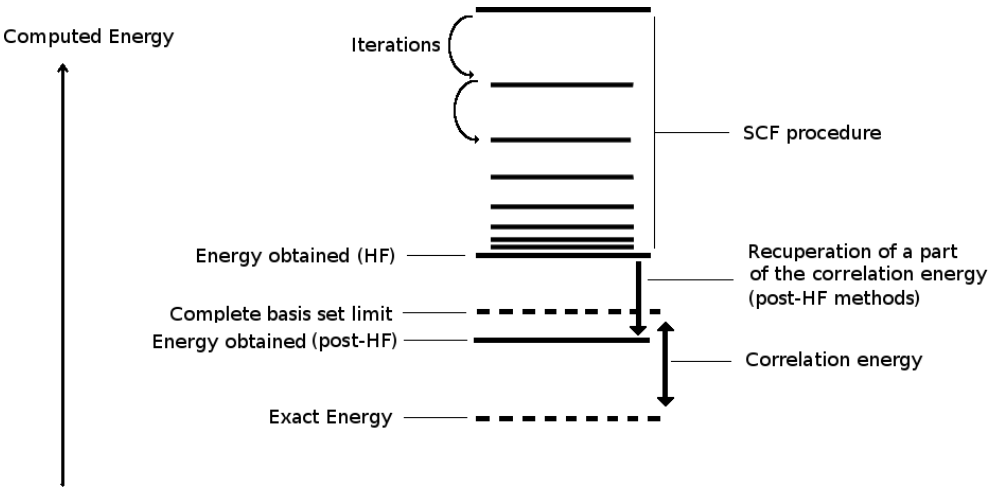



Figure 2.2: *Illustration of the variational principle. The objective is to get the lowest possible energy.*

### 2.1.3 Basis Functions

Basis functions are typically atom-centered functions that form the basis of molecular orbitals.

The very use of a basis set in an approximation known as linear combination of atomic orbitals (LCAO). It states that molecular orbitals can be expressed as a combination of atom-based basis functions. The largest and more numerous they are, the more precise will be the solution and the more demanding will be the calculation.

For this reason, they are required to be physically meaningful, have a shape that permits an efficient calculation, and to be in minimal amount to keep calculations affordable, since the computation cost scales as $\mathrm{N}^3$ in the best case.

The most commonly used basis functions fall in two categories: Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO). STO take the following form of exponentially decaying functions:

$$\chi^{STO}_{a,b,c}(x,y,z) = Nx^a y^b z^c e^{-\xi r} \tag{2.15}$$

N is a normalization constant, a+b+c reproduce the angular momentum. $\xi$ is related to the effective charge of the nucleus, as in well-known empirical

Slater rules and r is the electron distance to the nucleus. This type or orbitals are known to be accurate but the integrals are more demanding to compute than GTO.

In an attempt to make calculation of integrals easier, Gaussian Type Orbitals proposed:

$$\chi_{a,b,c}^{GTO}(x,y,z) = Nx^a y^b z^c e^{-\xi r^2} \tag{2.16}$$

The Gaussian Product Theorem states that using GTO simplifies the integrals and results in speed-up of calculations by several orders of magnitude. Unfortunately, long-range behavior of GTO is incorrect. This problem is solved by combining GTO to mimic STO:

$$\chi(CGTO) = \sum_{i}^{k} c_i \chi(PGTO) \tag{2.17}$$

Equation 2.16 and the right-hand of equation 2.17 refers to primitive gaussian type orbitals. CGTO refers to contracted gaussian type orbitals. The CGTO scheme is adopted to reduce the number of orbitals by linear combination. This procedure is called contraction. A minimal basis set is set with one basis function per orbital. A double-$\xi$ (or higher order) basis set will use two (or more) basis functions per atomic orbital.

A split valence basis set intends to increase accuracy, keeping as low as possible the computational cost. The strategy used is to focus on valence electrons, that are more relevant in chemistry than core electrons, since they are less tightly bound to the nucleus. A double-$\xi$ split valence basis set will have two contracted functions for each valence atomic orbital for instance.

Basis sets can be supplemented by polarization and diffuse functions. Polarization functions add a basis function of higher molecular momentum so if the valence electrons are p, one d function will be added to let the p orbital Polaris's. This can be important when using methods that take into account electronic correlation. Finally, diffuse functions can be added to describe more accurately the "tail" of the orbitals, that is the electrons far from the nucleus. Diffuse functions are particularly important when dealing with anions.

A basis set is made of basis functions and defines the space in which a quantum chemical problem has to be solved. When dealing with interaction energies, a molecular fragment (e.g. a water molecule in a water dimer) may use the basis functions of the other fragments. As a result, the energy of the dimer will act as if it had more basis functions. This create a small shift in the energy that has to be corrected, usually by the counterpoise method.[84,85] This errors is called basis set superposition error (BSSE) and is

routinely calculated in theoretical chemistry. BSSE becomes smaller as the basis set increases, because the new basis function from other fragment play a minimal role.

### 2.1.4 Semi-empirical Methods

Semi-empirical methods represent a simplification of the Hartree-Fock formalism. The cost and number of electronic integrals in the Fock matrix is reduced in semi-empirical methods by considering only valence electrons explicitly together with a minimal basis set and simplifying some 4-electron integrals. The amount of the reduction of the integral and the number of parameters define the method.

Parameters intend to "repair " the quality of the reduced amounts of integrals performed on small basis sets. Popular methods such as AM1[86] and PM3[87] methods are based on the Neglect of Differential Diatomic Overlap (NDDO) approximation.

Of course the training set of molecules influences the result and semi-empirical methods are viewed as crude methods, especially since Density Functional Theory methods became more easily accessible. Semi-empirical calculations are more intensive than molecular mechanics, but can be carried of for very large systems as well and do not suffer the drawbacks of atom types. They can be used to model reactions, can treat hybridization schemes, but are generally considered as less accurate than DFT.

### 2.1.5 Density Functional Theory

The theoretical grounds of Density Functional Theory (DFT) where set by Hohenberg and Kohn.[88] The idea at the origin of the theory is that the ground state energy of a molecular system depends only on its electron density. Ironically the relation between them relies in a function, which expression remains unknown. The density is defined as the square or the wave-function and is centered around atoms and vanishes at large distances:

$$\rho(\vec{r}) = |\psi_i(\vec{r})|^2 \tag{2.18}$$

$$\rho(\vec{r} \to \infty) = 0 \tag{2.19}$$

By definition, integrating the density corresponds to the number of electrons of the system:

$$\int_{space} \rho(\vec{r}) d\vec{r} = N \tag{2.20}$$

Different families of functionals intend to reproduce the relation between density and energy:

$$E_{DFT} = F[\rho] = F[f(x, y, z)] \tag{2.21}$$

$$\rho = f(x, y, z) \tag{2.22}$$

The energy calculated by the means of DFT takes the following form:

$$E_{DFT} = T[p] + E_{ne}[p] + E_{ee}[p] + E_{nn}[p] \tag{2.23}$$

$\mathrm{T[p]}$ is the kinetic energy, $\mathrm{E_{ne}[p]}$ is the nuclei-electron interaction, $\mathrm{E_{ee}[p]}$ is the electron-electron interaction, and $\mathrm{E_{nn}[p]}$ is the nuclei-nuclei interaction. $\mathrm{E_{ne}[p]}$ is treated classically while $\mathrm{E_{ee}[p]}$ is split into a classical coulombic interaction $\mathrm{J[p]}$ and $\mathrm{E_{xc}[p]}$, whose form is not exactly known. In the Born-Oppenheimer approximation, $\mathrm{E_{nn}[p]}$ is a constant that only depends on the atomic positions.

We obtain:

$$E_{DFT} = T[p] + E_{ne}[p] + J[p] + E_{xc}[p] + E_{nn} \tag{2.24}$$

#### Exchange and Correlation

DFT uses Kohn-Sham equations[89] to generate densities. Electrons are considered to interact with each other indirectly, through the potential they generate. Of course this is physically incorrect and one must add the $\mathrm{E_{xc}[p]}$ term which symbolizes exchange and correlation.

Actually, as the form of $\mathrm{E_{xc}[p]}$ is not known, it contains everything that has not been described elsewhere. This includes notably electron exchange and correlation. The problem is that without this term, electrons would not see of each other more than a shared potential. In reality, instantaneous movements of electrons influences the movement of nearby electrons. This is reflected by what is known as Fermi and Coulomb holes.

The Fermi hole is larger in intensity than the coulomb hole and arises from the fact that electrons do not interact with themselves. An illustration of this error is easily made since even in $\mathrm{H_2^+}$, $\mathrm{J[p]}$ will be positive where it should be zero. The electron sees its own potential or density, which is conceptually wrong.

The Coulomb hole presents somehow similar aspects: an electron moves quickly through the density of other electrons. Repulsion is maximized if another electron is moving also close to the first one. Eventually they will have to avoid each other, otherwise the energy would rise.

In the framework of non-interacting electrons, these Fermi and Coulomb hole are omitted and an extra contribution needs to be added to come back to more meaningful physics.

There are different approximations meant to deal with this problem. This results in different "DFT families". All of them approximate the $\mathrm{E_{xc}}$ term.

#### DFT families

Understanding the nomenclature of functionals may be difficult since correlation and exchange functions are not necessarily developed together and may be combined, resulting in a vast range of functionals. An overview of the various Density Functionals is shown in Table 2.1.

$\rho$ represents the density. It denotes that the energy in DFT is derived from it. $\nabla\rho$ represents the gradient of the density. When $\nabla\rho$ is used (and not only $\rho$), the functional considers a non-uniform electron gas in the estimation of correlation and exchange energies. The same idea is applied for $\nabla^2\rho$, which is the second derivative of the density. These increasingly complex schemes are also more and more costly from a computational point of view. Hartree-Fock exchange (noted HF ex. in Table 2.1) denotes that hybrid methods use all or part of the Hartree-Fock exchange.

#### Dispersion

As dispersion takes place far from the nucleus, these correlations appear at vanishing densities, making them particularly difficult to model. An efficient, yet simple strategy is to use empirical Becke-Jonson damping functions[108] of the form :

$$E_{Dispersion} = -\sum_{ij} f_{damp} * \left( \frac{C_{6,ij}}{R_{ij}^6} + \frac{C_{8,ij}}{R_{ij}^8} + \frac{C_{10,ij}}{R_{ij}^{10}} + ... \right) \qquad (2.25)$$

Where $\mathrm{f_{damp}}$ is a dumping function, which role is to let the empirical dispersion appear only in the middle and long-range region, so to fit interaction energy curves.[109] Usually only the first term $(\mathrm{C_6/R^6})$ is used and parameters are the geometrical mean of atomic $\mathrm{C_6}$ parameters.

A dispersion scheme can easily be added on any DFT calculation. In the context of DFT, the DFT-D (DFT plus dispersion) energy is then simply expressed as expressed in equation 2.26.

$$E_{DFT-D} = E_{DFT} + E_{Dispersion} \qquad (2.26)$$

It should be noted that the functional M06 is not DFT-D, since is not explicitly corrected for dispersion with an empirical scheme. Instead, it accounts for dispersion it by a heavy parametrization procedure.

Table 2.1: *The density functional families*

| Type | $\rho$ | $\nabla\rho$ | $\nabla^2\rho$ | HF ex. | examples | comments |
|---|---|---|---|---|---|---|
| LDA | YES | NO | NO | NO | SVWN5,[90] PW92.[91] | Also known as uniform gas model. Incorrect band gap and binding energies. Only "moderate" chemical accuracy |
| GGA | YES | YES | NO | NO | PB86,[92,93] PBE,[94] revPBE,[95] B97,[96] BLYP,[93,97] OLYP,[97,98] PWLYP. | Adds gradients: also known as semi-local. Better treatment of electronic structure and quickly varying densities. Better results than LDA. |
| Meta-GGA | YES | YES | YES | NO | TPSS,[99] M06-L.[100] | Laplacians of spin densities or orbital kinetic energy density. |
| Hybrid-GGA | YES | YES | NO | Some | B3LYP,[97,101] PBE0.[94,102] | Inclusion of some HF exchange. |
| Hybrid-Meta-GGA | YES | YES | YES | Some | TPSSh,[103] M06.[104] | Combination of Hybrid and Meta - GGA |
| Double Hybrid | YES | YES | YES | More | MC3BB,[105] B2PLYP,[106] B2K-PLYP,[107] mPW2K-PLYP[107] | Double-Hybrid HF (HF exchange + SAC correlation). Roughly doubled computer time. Allegedly fifth step in Jacob's ladder for DFT |

### 2.1.6 ONIOM Partition Scheme

The size of the system strongly influences the choice of the modeling method. In certain cases, the system is too large for electronic structure methods. On the other hand, force field methods are unable to reproduce bond formation and bond breaking processes that require a proper description of electrons and the rearrangements they undergo.

In a large molecule, one can consider certain regions to be more important than others. For example, in a protein, the active site is of primary importance. The same stands for large organometallic catalysts for instance. To this end, hybrid QM/MM models have been developed to assign a specific method to a specific region of a molecule.[110–112]

One part of the molecule will be treated with a force field (or any method

considered affordable computationally speaking). The rest of the molecule is assigned to a more demanding method. A third layer may be used but we will focus our explanations to a two-layer scheme for simplicity.

In the ONIOM approach, the QM/MM energy is obtained in the following way:

$$E_{tot}(QM+MM) = E_{MM}(Real) + E_{QM}(model) - E_{MM}(model) \quad (2.27)$$

The subscripts refer to the level at which the calculation is carried out. The labels in parenthesis correspond to the region of the molecule where the method is applied. "Real" refers to the whole molecule and "model" refers to the selected QM region, see Figure 2.3.

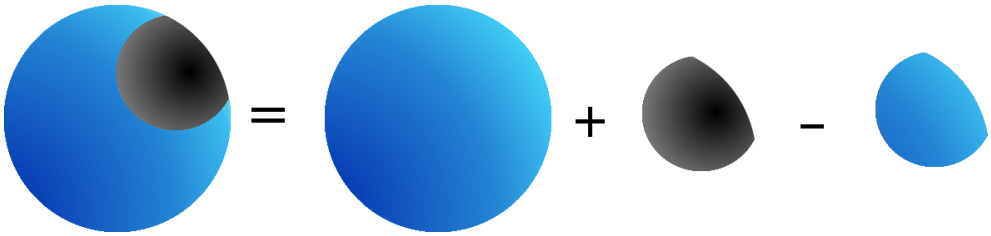


Figure 2.3: *Representation of the different regions of an ONIOM calculation and how is computed the ONIOM energy.*

Mechanical embedding presented here requires only the MM energy, calculated on the whole molecule. Force field parameters will therefore take into account van der Waals and electrostatic interactions between the QM and MM parts.

A difficulty relies in the definition of the partition and of the atoms at the interface. One strategy for this is the use of "link" atoms. The most reliable way to do this is to avoid splitting in the middle of aromatic rings and other delocalized regions. As a rule of thumb, a single bond should be used. Typically, a C-C single bond is chosen for the interface. One carbon is assigned to each region. In the QM region, the MM carbon will be replaced by an hydrogen. For the MM system, the position of the carbon is mapped on this hydrogen. Two calculations at the MM level will be carried: one for the whole molecule, and one for the QM region.

### 2.1.7 Solvation

The solvent often plays an important role in reactions, or in the stability of chemical species.

One can rely on a discrete model that implements explicit solvent molecules. Of course, this approach is very demanding computationally. Also, the large

amounts of possible configurations makes a global or local minimum difficult to reach.

Another approach is not to consider discrete solvent molecules but instead, the solute is immersed in a dielectric medium that represents the solvent polarity. A cavity is this continuum is created around the solute. It is generated by overlapping nuclear-centered spheres. The electron density of the solute induces the polarization of the continuum. The solute and the continuum self consistently interact and adjust to each other. The reaction field is the field generated in response to the presence of the continuum. It depends on the dielectric constant, the size and shape of the cavity, the electron density of the solute.

This strategy is cheap and works well for reactions in which the solvent does not actively participate into the chemical process (*i.e.* coordination, catalysis by solvent, etc). Among modern methods for continuum solvation, one can mention PCM[113,114] and SMD.[115] An intuitive view is presented Figure 2.4.

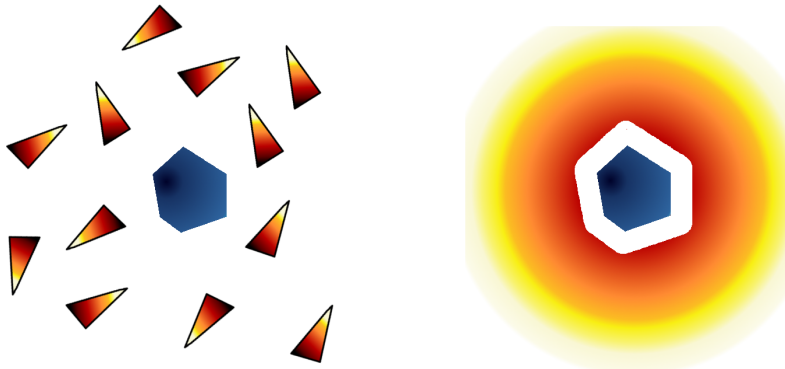


Figure 2.4: *Illustration of the solvent Models. Left: Explicit model. Right: Implicit Model.*

## 2.2 Applicability to Supramolecular Systems and Limits

Modern computational chemistry takes advantage of the dramatic increase of computer resources that has occurred in the last few decades. Typically, chemical precision can be reached for mechanisms concerning molecules of up to about 200 atoms in a realistic time. Supramolecular chemistry suffered however of the lack of correct description of non-covalent (notably middle-range, attractive) effects. In the recent years, the addition of van der

Waals corrections to popular DFT methods quickly became the reference for supramolecular systems.[116]

## 2.3 Methods Used in This Study

All calculations presented in this work are carried with the Gaussian 09 package.[117] The same benchmark of methods is used for both supramolecular studies.

We screen several methods to find the best match for our systems. Force fields are very likely to give distinct results depending on their parametrization. Among them, we test **UFF**[118] and **MM3**.[119] Given that MM3 is not available in Gaussian, we used a set of scripts written in our laboratory[120] that provides an interface between Tinker[121] and Gaussian. The next choice is **AM1**,[86] which belongs to the class of semi-empirical methods.

In the attempts to find a method that would be at the same time inexpensive and precise enough, we include in the benchmark three ONIOM models:
ONIOM(B3LYP:UFF), ONIOM(B3LYP:MM3) and ONIOM(B3LYP:AM1) are respectively identified in this thesis as **O-UFF, O-MM3** and **O-AM1**. The supramolecular hosts are assigned to the "low layer", with respectively UFF, MM3 and AM1. The guests are allocated to the "high layer", for which we use the density functional B3LYP.[97,101] For the QM part, a valence double-zeta polarized basis set 6-31G* is used.[122] The systems included in this work do not require "link" atoms, since both layers do not share a covalent bond, see Chapters 3 and 4.

These choices arise from previous studies: in certain cases where long range interactions are particularly important, an ONIOM model can overcome the results of a non-dispersion corrected calculation.[123]

Full QM calculations were also performed. Four density functionals are incorporated in the benchmark together with a 6-31G* basis set. Among them is the popular hybrid-GGA **B3LYP**.[101] Then comes the GGA PBE1PBE[102] functional, we note **PBE**. More recent functionals are also chosen such as the hybrid meta GGA **M06** belongs to the Meta-Hybrid GGA class and was developed by Truhlar and Zhao.[104] It contains a large number of parameters which are semi-empirically fit on various data, in order to minimize errors, notably on non-covalent interactions.

And finally, we also test the **B97D** functional from Grimme,[124] a semi-empirically corrected GGA-type functional. The strategy used here relies on the correction of the existing B97 functional[96] and is explicitly parametrized by including atom-pairwise dispersion corrections, together with a damping

function that corrects the response for small radii.

### 2.3.1 Solvent Effects

The solvent effects have been accounted in three different ways depending on the methods considered.

#### Optimization In Gas Phase

In most cases, all geometries are fully optimized in gas phase without symmetry restriction. Solvent effects are incorporated in the results by adding a SMD[115] single-point on the final geometry obtained after the optimization.

Frequency calculations are carried out as well in all cases, in order to access free energies and to confirm the type of extremum reached during the optimization (minimum or transition state). The Gibbs free energy in solvent $\mathrm{G_{sol}}$ can be obtained with the following formula:

$$G_{sol} = G_{gas} + E_{sol} - E_{gas} \tag{2.28}$$

In equation 2.28, $\mathrm{G_{gas}}$ is the free energy in gas phase, $\mathrm{E_{gas}}$ is the potential energy in gas phase, and $\mathrm{E_{sol}}$ is the potential energy in solution.

#### Methods Involving Force Fields

For **UFF**, **MM3**, **O-UFF** and **O-MM3**, the solvation energy is not directly accessible. An estimate is obtained using equation 2.29.

$$E_{sol} - E_{gas} = E_{sol(B3LYP)} - E_{gas(B3LYP)} \tag{2.29}$$

#### Methods Optimizing With an Implicit Solvent Model

We propose a variation on **B97D** and **M06** were the optimizations are performed in solvent. Those two methods are respectively referred to as $\mathbf{B97D_{sol}}$ and $\mathbf{M06_{sol}}$. In those two cases, the value of $\mathrm{G_{sol}}$ is obtained directly.

### 2.3.2 Basis Set Superposition Error

Basis Set Superposition Errors (BSSE)[84,85] have been calculated for all DFT methods. They represent a correction on the energy calculated on the fully converged geometry. To compute BSSE, fragments have to be defined. Each covalent molecule is defined as a fragment. An exception is of course made

for bimolecular reactions. To stay consistent, we keep the initial fragmentations, see Figure 2.5. The BSSE can be calculated on a larger assembly and is generalizable to any amount of fragments.

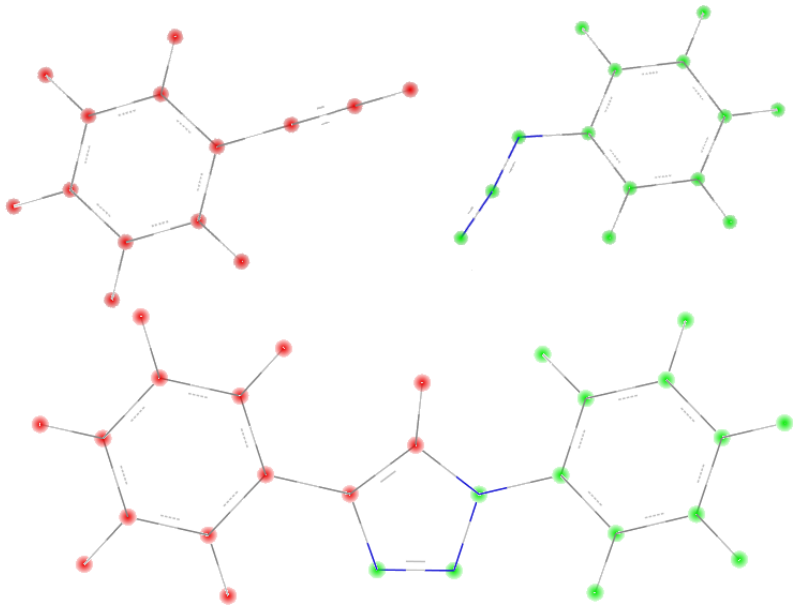


Figure 2.5: *Illustration of the fragmentation scheme used. Assignations in red and green are kept before(top)/after(bottom) creating bonds.*

### 2.3.3 Standard States

In our calculations, an ideal gas of 1 atm. is assumed as the reference state. To compute solvation free energies, a correction of 1.9 kcal/mol has to be added in all calculations to denote the transfer 1 atm. $\rightarrow$ 1M.[125] For unimolecular reactions, the correction is self-compensating. However when the number of moles changes such as in a bimolecular reaction, it is not the case. All the free energy values presented here include this correction.

### 2.3.4 Geometry Optimization

Obtaining the energy of a given system is not the end of the road. In the framework of Transition State Theory developed by Eyring,[126] relevant structures are minima and saddle points. For a given mechanism, they are extrema across a unique Potential Energy Surface in a dynamic equilibrium.

#### Minima

Real-life chemical systems are never quite in *the* configuration that minimizes their energy, even at 0K, since the zero-point energy cannot be canceled, some structural fluctuations will then always remain, due to the quantum nature of chemical systems and the oscillation of their wave-functions. As a theoretical object a minima is nonetheless very useful, precisely because the *actual* structures oscillate around the minimal energy structure.

Algorithms are in charge of finding the closest relevant minimum, based on a guess or starting point that is generally obtained either by X-ray for example or drawn "by hand". Algorithms do not know anything about electronic structures or molecules. They play with parameters (cartesian or internal coordinates: bond lengths, angles, and numerous dihedrals) and intend to minimize a associated property (in our case the energy). Analyzing the curvature of the PES region around the point to optimize (with first and sometimes seconds derivatives), they extrapolate the next "step" and the new value to give to each parameters. The procedure is repeated until convergence. Internal coordinates are more performant than cartesian coordinates since they are chemically more meaningful and reduce significantly the amount of parameters to optimize.

#### Saddle Points

Saddle-point optimization resemble minima optimization in all aspects, except for a set of coordinates (e.g. bond distances and angles) grouped together in an eigenvector. For example in a $S_N2$ reaction, the particular set of coordinate of interest will be the nucleophile-electrophile distance and the electrophile-leaving group distance. As one gets shorter, the other gets larger and the transition state (TS) of the reaction is located where the energy is maximum.

Several techniques may be used to locate a TS, which is generally a difficult task. The TS "region" may be lost by the algorithm, which "falls": it maximizes non-relevant coordinates while minimizing the coordinates of interest.

One difficulty often relies in obtaining a good Hessian (second derivative matrix) since its evaluation is costly and relies on the quality of the guess structure provided to the algorithm as well as the method used, and eventually the size of the basis set.

It is possible to avoid the first calculation of the Hessian. To do so, one has to guide the optimizer by feeding it with set of structures, to realize an intuitive "motion-picture" movie of the TS crossing. In the example of

the $S_N2$ reaction, the user will provide three structures in a well-defined sequence. The optimizer it will convert the proposed motion: (i) attack of the nucleophile, (ii) transition structure, (iii) departure of the leaving group, into a intelligible eigenvalue that has to be negative.

The curvature of the TS may represent another difficulty. The optimizer evaluates the region around the TS. If this region of the PES is almost flat, the program may discard the corresponding eigenvalue, assuming that it is of little importance. Rather, the algorithm will look for other negative eigenvalues, which may be numerous when the structure is far from a well-defined, local extremum. In this case, the optimizer will be wrong and the user may have to constantly reset the calculation with a proper starting structure and eventually to recompute the costly Hessian, in order to get more meaningful curvature estimates and attempt to eliminate "parasite" eigenvalues.

## 2.4 Kinetic Model

Chemical kinetics study the rates of chemical reactions. Many factors can affect it, such as experimental conditions (pressure, temperature, medium, concentration, presence of a catalyst, etc.) as well as the structure of chemical species along a given mechanism.

For simple reaction mechanism, it is easy to have an idea of the reaction kinetics without the need of sophisticated tools. However, for processes with several ramifications, loops, or when the reactants are not in equal initial amounts, the use of specialized software is strongly recommended.

To perform a kinetic analysis, one has to break down into elementary reactions the overall process. We will illustrate the problem with an example: chlorinated compounds such as CFC are known for their role in ozone depletion. The overall reaction is written:

$$O_3 + O \longrightarrow 2\,O_2$$

In atmospheric conditions, this reaction is very slow. However, CFCs can act as catalysts. To describe ozone depletion, one cannot rely on this equation only, although it still stands globally:

$$\begin{array}{c} Cl + O_3 \longrightarrow ClO + O_2 \\ ClO + O \longrightarrow Cl + O_2 \\ \hline O_3 + O \longrightarrow 2\,O_2 \end{array}$$

To accurately describe this process, one should take into account the role of other CFC and contaminants, photons, variation of atmospheric conditions and so on. Taking everything into account may result in a very difficult exercise and sometimes one has to neglect one or many factors and contributions.

In the framework of this study, we expect kinetics to be a very valuable tool to help us describing the relative importance of chemical species and processes. Modeling comes very handy for kinetics to provide information on reactions. To do so, one just has to use Eyring equation 2.30 to convert free energy of activation calculated by a model into a rate of reaction.

$$k_{reaction} = \frac{k_B T}{h} e^{-\Delta G^{\ddagger}/k_B T} \tag{2.30}$$

Thus, to describe the kinetics of the following elementary reactions:

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C} \rightleftharpoons \mathrm{D}$$

One has to obtain by modeling a free energy ($\Delta \mathrm{G}$) of activation that corresponds to the difference of free energy between A+B and the transition state that permits the transformation to C. The same stands for C to D. Also, as the double arrow suggests, one also needs the activation energy to transform D back into C. Figure 2.6 displays the problem in a simple manner.

In this example we consider that the transition of C back into A and B can be neglected (simple arrow), due to the height of the barrier.

At this stage, we have to use three $\Delta \mathrm{G}$ that correspond to three processes:

process 1 : $\Delta \mathrm{G}^{\ddagger}_{\mathrm{A+B\rightarrow C}} = 12.0$ kcal/mol.

process 2 : $\Delta \mathrm{G}^{\ddagger}_{\mathrm{C\rightarrow D}} = 20.0$ kcal/mol.

process 3 : $\Delta \mathrm{G}^{\ddagger}_{\mathrm{D\rightarrow C}} = 25.0$ kcal/mol

Using Eyring equation, we obtain three rates constants.

$\mathrm{k}_1$=9.7E+03 $\mathrm{M}^{-1}\mathrm{s}^{-1}$

$\mathrm{k}_2$=1.3E-02 $\mathrm{s}^{-1}$

$\mathrm{k}_3$=2.8E-06 $\mathrm{s}^{-1}$

The units are different since the first process is a bimolecular reaction while the second and third are unimolecular. These rates fully describe the variation of concentration of the various species present:

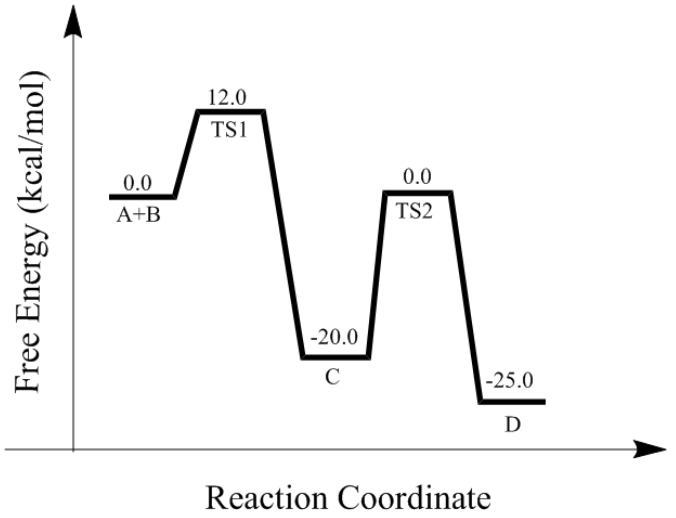



Figure 2.6: *Energy diagram of hypothetical species modeled with arbitrary values assigned.*

$$rate_1 = k_1 * [A] * [B] = -\frac{d[A]}{dT} = -\frac{d[B]}{dT} \tag{2.31}$$

$$rate_2 = k_2 * [C] = -\frac{d[C]}{dT} \tag{2.32}$$

$$rate_3 = k_3 * [D] = -\frac{d[D]}{dT} \tag{2.33}$$

The quantities of interest being concentrations of A, B, C and D and their rate of creation/destruction, we therefore have to solve ordinary differential equations of the form:

$$\frac{d[A]}{dT} = -k_1 * [A] * [B] \tag{2.34}$$

$$\frac{d[B]}{dT} = -k_1 * [A] * [B] \tag{2.35}$$

$$\frac{d[C]}{dT} = k_1 * [A] * [B] - k_2 * [C] + k_3 * [D] \tag{2.36}$$

$$\frac{d[D]}{dT} = k_2 * [C] - k_3 * [D] \tag{2.37}$$

To solve these equations, one usually relies on algorithms. We use Acuchem,[127] a free program that implements them. It takes parameters such as initial concentrations and rates of reaction as an input.

In this example, we put at $[A]_{t=0}$=0.2 mol/L and $[B]_{t=0}$=0.5 mol/L. The output (Figure 2.7) shows the variation of the different species along time. We clearly see the concentrations of A and B equally decreasing as C is generated in the first tens of milliseconds, until A is entirely consumed. It is only much later that C starts turning into D.

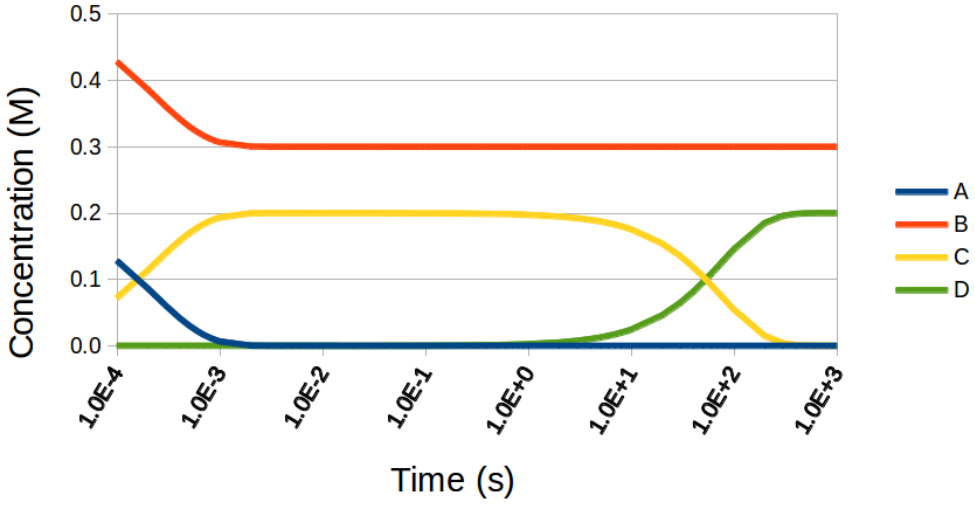



Figure 2.7: *Energy diagram of hypothetical species. These values can be used in a kinetic model.*

This example allows us to illustrate how we are going to proceed with real systems. First, we aim at exploring the most relevant parts of the energy surface concerning supramolecular systems. Then, the data obtained on each system will be integrated in a reaction network. At this stage, we should have obtained computationally all the relative stabilities of all the compounds of interest, as well as the transition structures linking them. Further, free energies of transitions can be converted into reaction rates, which will be used in a kinetic model such as the one we presented. With this protocol, we aim at providing a meaningful description of the dynamical behavior of two supramolecular systems and their guests along time, without having to undertake time-expensive Molecular Dynamics (MD).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Chapter 3

# Cucurbituril

## 3.1 The Host

The cucurbit[6]uril (noted CB6) supermolecule labeled **1** is a macrocycle that was synthesized for the first time in 1905,[23] see Figure 3.1. Its molecular structure was first determined through crystallography by Freeman et al. in 1981.[24]

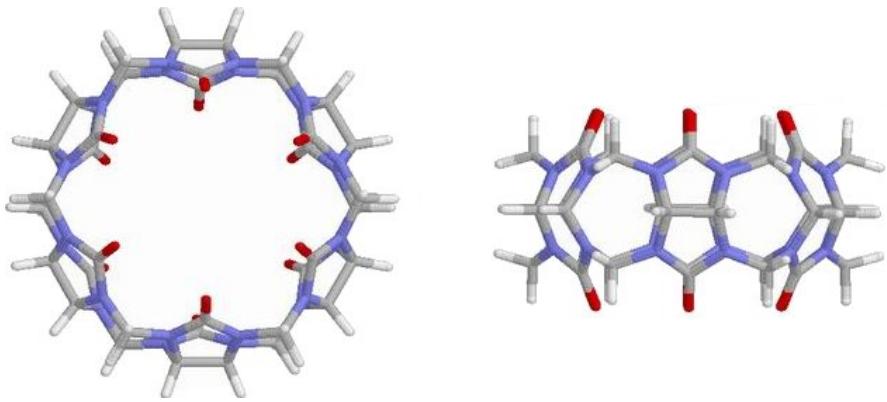


Figure 3.1: *Molecular model of the cucurbit[6]uril macrocycle. Left: top view. Right: side view.*

It is *ca.* 10 Å wide and 6 Å high. Its torus shape allows small solvent or solute molecules to enter and leave the center of the macrocycle. The interior is hydrophobic while the two carbonyl-fringed portals are negatively charged, ready to accept cations through charge-dipole interactions.

## 3.2 An Efficient Host for a Cycloaddition

A cycloaddition reaction between azidoethylamine **2** and propargylamine **3** (Figure 3.2) was experimentally found to be regioselective[75] and a kinetic acceleration (by comparison of bimolecular reactions) of 5.5 * $10^4$ was calculated to be induced by the presence of **1**. Calculation of effective molarity led to a value of 1.6 * $10^4$M,[47] which is particularly high and denotes an effectiveness superior to the simple "concentration" of reactants in the cucurbituril host.

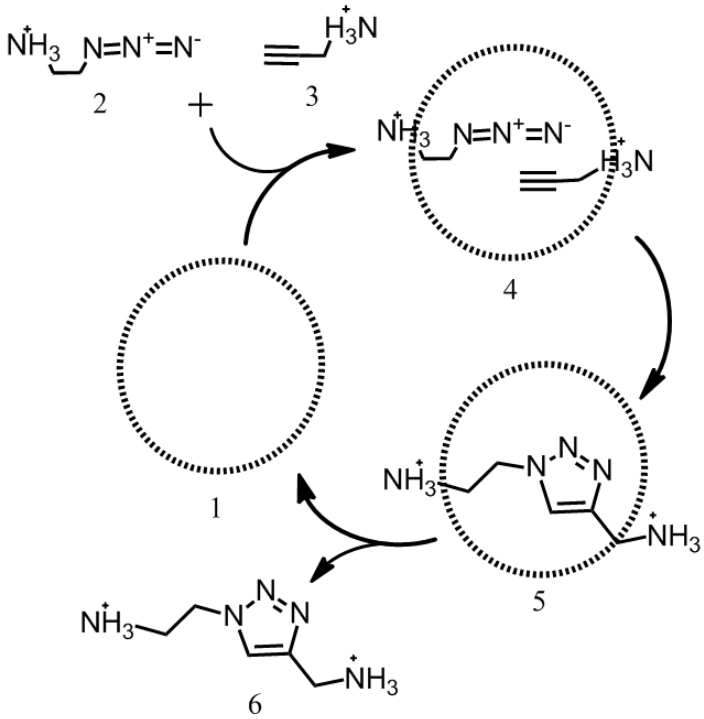



Figure 3.2: *The catalytic cycle under study. The macrocycle is represented by bold dashed lines.*

Thanks to mechanistic and kinetic investigations,[75] Mock et al. noted a saturation phenomena: all the macrocycles were occupied by the guests. This indicates that the capture of the reactive species is exothermic.

The catalytic cycle works as follows. In the first step, the complex **4** is formed when the reactants **2** and **3** reach the center of **1**. In the second step, the cycloaddition reaction proceeds, leading to structure **5**. Finally the product **6** has to be released from the macrocycle so that a turnover can take place to complete the catalytic cycle.

### 3.2.1 The Turnover Issue

Noticeably, the reaction between the azide and the alkyne in CB6 is not catalytic since product release was experimentally found to be the rate-limiting step (step $\mathbf{5} \rightarrow \mathbf{1}+\mathbf{6}$) rather than the reaction itself[75] (step 4 $\rightarrow$ 5). This can suggest two different, yet not incompatible hypotheses that slow down the reaction as the macrocycle is saturated.(i) Product release is endothermic, (ii) a significant barrier has to be crossed to release the product, which is more bulky than separated reactants and may have some more difficulty to go out of the macrocycle.

### 3.2.2 Identifying a Suitable Theoretical Method

A computational study for this reaction was already published in our group,[74] and confirms the active role of $\mathbf{1}$ in regioselectivity. This study concluded that the key role of the supermolecule is the partial removal of entropic barriers promoting catalysis. In this publication, DFT method B3LYP/6-31G* was employed.

In this study, we have carefully revised the previous results, with the objective of enhancing their quality and broadening the spectrum of the results obtained. Details about the methods applied in this work are given in Section 2.3

First we benchmark the methods (Section 3.3), afterwards we provide a more detailed mechanistic and kinetic description of the reaction promoted by the macrocycle (Section 3.5).

## 3.3 Results

We have calculated the free energies of the species forming the catalytic cycle, considering 12 methods ranging from Molecular Mechanics to Density Functional Theory. The results are summarized in Table 3.2. Some relevant structural parameters specified Figure 3.4 are used to compare the results produced by the different methods, in complement to the energies. They are presented in Table 3.1.

### 3.3.1 Reactive Molecules Captured

The trapping of the species $\mathbf{2}$ and $\mathbf{3}$ by $\mathbf{1}$ leads to $\mathbf{4}$ through weak bonding: H-bonds and van der Waals interactions. Complex $\mathbf{4}$ is presented Figure 3.4. We present here an comparative analysis of the packing of the reactive

species by the macrocycle. On Figure 3.4 are superimposed several parameters selected for the analysis.

In Table 3.1 **a**, **b** and **c** are given as an average and the value in parenthesis corresponds to the standard deviation over the different measurements. If this value is different from zero, it indicates an elliptic character of the host, as shown Figure 3.3 for an example. On the left the standard deviation is zero as all the measurements give the same value, so the host is not elliptic. On the right, the two measurements of the diameter indicate a different value so the host has an elliptic character.

All the parameters are measured center-to-center, based on atomic coordinates. **a** is the opening of the portals. **b** is the height of the macrocycle. **c** is the diameter of the macrocycle. Finally, d and e are selected to represent the "packing" of the reactants by a given method, *i.e.* how close are the reactants when trapped by the macrocycle.

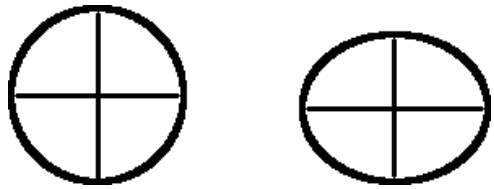


Figure 3.3: *Illustration of the measurement of the elliptic character of the host. Left: non-elliptic host. Right: the host has an elliptic character, as both segments have different lengths.*

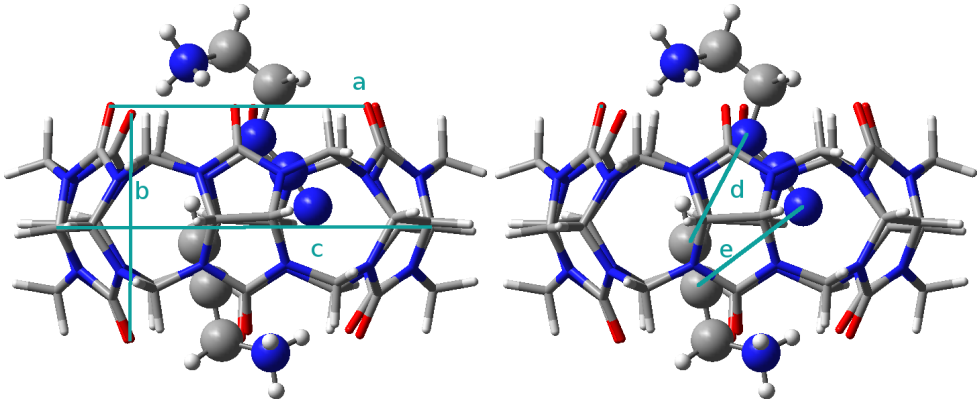



Figure 3.4: *Chosen set of parameters for structural analysis. See text for details.*

As we can see Table 3.1, all the methods agree on similar geometric parameters: the opening (a) slightly shrinks upon inclusion of guests while

Table 3.1: *Measured values of the diameter (***a** *and* **c** *and height* **b** *of the cucurbituril before and after aggregation, according to different computational methods.* **e** *and* **d** *give a measure of the packing of the reactants. Values given are atomic center-to-center distances in Ångströms, together with the standard deviation in parenthesis (e and d are not averaged). "Outlier" values are set bold for clarity.*

| | Compound **1** | | | Compound **4** | | | | |
|---|---|---|---|---|---|---|---|---|
| | a | b | c | a | b | c | e | d |
| **UFF** | 6.7 (0.0) | **5.9** (0.0) | 10.4 (0.0) | 7.1 (0.5) | 6.1 (0.0) | 10.4 (**0.6**) | **5.3** | **4.3** |
| **MM3** | 6.9 (0.0) | 6.3 (0.0) | 10.1 (0.0) | 6.8 (0.5) | 6.2 (0.0) | 10.1 (0.3) | 3.4 | **2.9** |
| **AM1** | 7.0 (0.0) | 6.2 (0.0) | 10.4 (0.0) | 6.8 (0.6) | 6.0 (0.1) | 10.4 (0.3) | 3.4 | 3.5 |
| **O-UFF** | 6.7[a] (0.0) | **5.9**[a] (0.0) | 10.4[a] (0.0) | 6.7 (0.4) | 6.1 (**0.4**) | 10.4 (**0.1**) | **5.9** | **4.8** |
| **O-MM3** | 6.9[b] (0.0) | 6.3[b] (0.0) | 10.1[b] (0.0) | 6.9 (0.4) | 6.2 (0.0) | 10.1 (0.3) | 3.2 | **2.9** |
| **O-AM1** | 7.0[c] (0.0) | 6.2[c] (0.0) | 10.4[c] (0.0) | 6.7 (0.6) | 6.0 (0.1) | 10.4 (0.3) | 3.4 | 3.7 |
| **B3LYP** | 7.2 (0.2) | 6.2 (0.0) | 10.2 (0.0) | 6.9 (0.5) | 6.1 (0.1) | 10.3 (0.3) | 3.4 | 3.5 |
| **PBE** | 7.1 (0.3) | 6.2 (0.0) | 10.2 (0.0) | 6.9 (0.5) | 6.0 (0.1) | 10.2 (0.3) | 3.4 | 3.6 |
| **M06** | 7.0 (0.0) | 6.2 (0.0) | 10.2 (0.0) | 6.7 (0.6) | 6.0 (0.1) | 10.2 (0.5) | 3.2 | 3.5 |
| **M06**$_{sol}$ | 6.9 (0.1) | 6.1 (0.0) | 10.2(0.1) | 6.8(0.6) | 6.0 (0.1) | 10.1 (0.5) | 3.1 | 3.5 |
| **B97D** | 7.1 (0.0) | 6.2 (0.0) | 10.2 (0.0) | 6.9 (0.6) | 6.1 (0.1) | 10.2 (0.4) | 3.2 | 3.5 |
| **B97D**$_{sol}$ | 7.0 (0.0) | 6.2 (0.0) | 10.3 (0.0) | 7.0 (0.7) | 6.2 (0.1) | 10.2 (0.4) | 3.1 | 3.4 |

[a] Refers to the full **UFF** calculations.
[b] Refers to the full **MM3** calculations.
[c] Refers to the full **AM1** calculations.

the height **b** and the internal diameter **c** are quite constant, see Figure 3.4. Finally the macrocycle gains some elliptic character, as demonstrated by the statistical deviation of **a** and **c**, increasing from **1** to **4** on the 6 measurements made for each method.

Concerning the packing of guests (measurement of d and e), the geometric parameters are again quite constant. The notable exceptions are: **MM3** and **O-MM3** that over-aggregates the guests compared to other methods, while with **UFF** and **O-UFF** none of the guest does properly fit in the macrocycle so guest-guest distances of the reactive parts are larger by *c.a.* 1-2 Å than for other methods.

We discuss now the reaction **1**+**2**+**3** → **4** that corresponds to the inclusion of the reactive guests in the core of the cucurbituril host. We will use the enthalpy as a synonym for $E_{sol}$ although it is not formally exact. Let us recall that **O-UFF**, **O-MM3** and **O-AM1** correspond to ONIOM models with B3LYP for the reactants, while respectively UFF, MM3 and AM1 are used to model the host.

The aggregation energies presented in Table 3.2 show unrealistic results

with the **UFF** and **O-UFF** approaches, which do not appear to be able to describe the complexation of the supermolecule and its guests. We will consider possible explanations in the discussion.

Table 3.2: *Computed relative energies for the basic steps of the cycloaddition within the host. Values are in kcal/mol, relative to the separated species.*

| | **UFF** | | **MM3** | | **AM1** | |
|---|---|---|---|---|---|---|
| | $\Delta E_{sol}$ | $\Delta G_{sol}$ | $\Delta E_{sol}$ | $\Delta G_{sol}$ | $\Delta E_{sol}$ | $\Delta G_{sol}$ |
| 1+2+3 → 4 | 87.2 | **108.4** | -19.2 | **17.0** | -2.6 | **27.9** |
| 1+2+3 → 5 | 148.9[a] | **187.6**[a] | -0.7[a] | **34.1**[a] | -45.8 | **-8.5** |
| 2+3 → 6 | 36.5[a] | **54.5**[a] | 14.6[a] | **24.4**[a] | -42.4 | **-27.5** |
| | **O-UFF** | | **O-MM3** | | **O-AM1** | |
| | $\Delta E_{sol}$ | $\Delta G_{sol}$ | $\Delta E_{sol}$ | $\Delta G_{sol}$ | $\Delta E_{sol}$ | $\Delta G_{sol}$ |
| 1+2+3 → 4 | 87.5 | **102.3** | -6.0 | **27.9** | -2.1 | **29.1** |
| 1+2+3 → 5 | 23.7 | **63.0** | -92.0 | **-50.7** | -77.7 | **-38.8** |
| 2+3 → 6 | -69.0[b] | **-52.0**[b] | -69.0[b] | **-52.0**[b] | -69.0[b] | **-52.0**[b] |
| | **B3LYP** | | **PBE** | | **M06** | |
| | $\Delta E_{sol}$ | $\Delta G_{sol}$ | $\Delta E_{sol}$ | $\Delta G_{sol}$ | $\Delta E_{sol}$ | $\Delta G_{sol}$ |
| 1+2+3 → 4 | -14.7 | **20.8** | -19.7 | **15.0** | -38.6 | **0.3** |
| 1+2+3 → 5 | -87.4 | **-45.3** | -103.2 | **-61.8** | -114.5 | **-69.5** |
| 2+3 → 6 | -69.0 | **-52.0** | -81.1 | **-64.2** | -71.1 | **-54.3** |
| | **M06$_{sol}$** | | **B97D** | | **B97D$_{sol}$** | |
| | $\Delta E_{sol}$ | $\Delta G_{sol}$ | $\Delta E_{sol}$ | $\Delta G_{sol}$ | $\Delta E_{sol}$ | $\Delta G_{sol}$ |
| 1+2+3 → 4 | -31.1 | **-3.2** | -40.2 | **-8.4** | -36.4 | **-6.7** |
| 1+2+3 → 5 | -107.2 | **-69.9** | -111.3 | **-70.6** | -108.3 | **-72.0** |
| 2+3 → 6 | -72.7 | **-55.2** | -64.1 | **-47.2** | -65.9 | **-47.7** |

a: Inconsistency due to the chemical reaction stemming from atom type change. See Section 2.1.1
b: refers to a full **B3LYP** calculation.

All other methods find a marked enthalpic stabilization ranging from -2.1 kcal/mol for **O-AM1** to -40.2 kcal/mol for **B97D**.

In the case of **MM3**, **AM1**, **O-MM3**, **O-AM1** as well as **B3LYP**, **PBE** and **M06** the enthalpic stabilization exists but is not sufficient to overcome the entropy loss that always comes along with aggregation processes. The resulting free energies in solution are positive and range from 0.3 kcal/mol for **M06** to 29.1 kcal/mol for **O-AM1**. This is not consistent with experimental results: the reaction becomes regioselective, and with sufficient amount of guests becomes independent of the concentration. This indicates that all the macrocycles are occupied by the guests, therefore binding to it.

Only **M06$_{sol}$**, **B97D** and **B97D$_{sol}$** are able to find a proper stabilization

of the aggregate, respectively of -3.2, -8.4 and -6.7 kcal/mol. This denotes that the guests are able to bind to the host, in agreement with experimental results.

### 3.3.2 Reaction Inside the Macrocycle

We will not discuss the energetics of the reaction $\mathbf{1}+\mathbf{2}+\mathbf{3} \rightarrow \mathbf{5}$, also in Table 3.2. It is particularly interesting to compare the numbers from $\mathbf{1}+\mathbf{2}+\mathbf{3} \rightarrow \mathbf{4}$ to those of $\mathbf{1}+\mathbf{2}+\mathbf{3} \rightarrow \mathbf{5}$. The difference tells us about the transformation from $\mathbf{4}$ to $\mathbf{5}$. As the reaction occurs, $\mathbf{4}$ is turned into $\mathbf{5}$, so we expect a strong stabilization associated to the cycloaddition, which is irreversible. This should mask any variation of the non-covalent energies.

**UFF** and **MM3** are force fields which use "atom types" to model energies and structures. Consequently, they are inherently unable to reproduce the energetics of bond formation or breaking.

**AM1** stabilizes $\mathbf{5}$ over $\mathbf{4}$ by 36.4 kcal/mol, while all DFT methods predict a larger change of energy, ranging from -62.2 kcal/mol for **B97D** to -76.8 kcal/mol for **PBE**, thus all the methods except **UFF**, **MM3** and **AM1** seem to describe properly the transformation from $\mathbf{4}$ to $\mathbf{5}$.

### 3.3.3 Product Release

We saw section 3.2.1 that the product release has been experimentally determined as the rate-limiting step, which can be explained by kinetics and/or thermodynamics. The latter is the easiest to address since it does not require to acquire TS information on product release. We will examine kinetics later on, in section 3.5.

To obtain the energies of product release, we have to substract the values of the third and second line for each method. For instance, **B3LYP** considers the product more stable outside CB6 by 6.6 kcal/mol.

For this final step, excluding the results obtained with **UFF** and **O-UFF**, an exothermic process is suggested by **MM3, AM1, O-MM3, O-AM1, B3LYP** and **PBE**. The values are ranging from -1.3 kcal/mol for **O-MM3** to -19.0 kcal/mol for **AM1**.

On the contrary, a large destabilization in the range 14.8 - 23.4 kcal/mol is computed with dispersion-corrected DFT methods.

### 3.3.4 Basis Set Superposition Error

Basis set superposition errors were computed for the catalytic cycle, see Table 3.3. Those corrections are included in the values given Table 3.2. The errors

Table 3.3: *Basis Set Superposition Errors computed for compounds* **4** *and* **5**. *Values in kcal/mol.*

| Compound | **B3LYP** | **PBE** | **M06** |
|---|---|---|---|
| **4** | 16.5 | 14.6 | 15.2 |
| **5** | 19.7 | 17.7 | 18.5 |
| Compound | $\mathbf{M06}_{\mathrm{sol}}$ | **B97D** | $\mathbf{B97D}_{\mathrm{sol}}$ |
| **4** | 15.6 | 16.4 | 15.1 |
| **5** | 18.5 | 17.5 | 18.1 |

are computed in the narrow range 14.6-16.5 kcal/mol for **4** and 17.7-19.7 kcal/mol for **5**.

This values are very large and should be always taken into account to obtain accurate interaction energies.[128–131] It would be desirable to eliminate these errors by using very large basis sets, however the counterpoise method is much faster.

## 3.4 Discussion

We have performed a comprehensive benchmark on various computational models applied to a cycloaddition which kinetics are enhanced by a cucurbit[6]uril macrocycle. A graphical summary of the results is proposed Figure 3.5 for ease of understanding. Although we saw in Table 3.2 that the geometric description was consensual for the most part, the predicted energies are very much dependent on the method used. We review and discuss in this section the particularities, achievements and weaknesses of each model applied on this system.

### 3.4.1 Force Fields

The very fast **UFF** method uses a parametrized atom-atom 6-12 Lennard-Jones potential.[118] It appears that this force field is not functioning well with this system, probably due to the numerous hydrogen bonds are not properly described in **UFF**. As a consequence, the guests bind poorly inside the macrocycle, see Table 3.1.

**UFF** does not correctly reproduce the geometry of the azide, since it has not been parametrized for it, while **MM3** does. The $\widehat{\mathrm{NNN}}$ angle of the azide is *c.a.* 120 degrees for **UFF** and *c.a.* 172 degrees for **MM3**, similarly to results obtained from DFT.

The different contributions to the total free energy are shown Table 3.4. The problem appears to arise from the **UFF** force field energy, since the

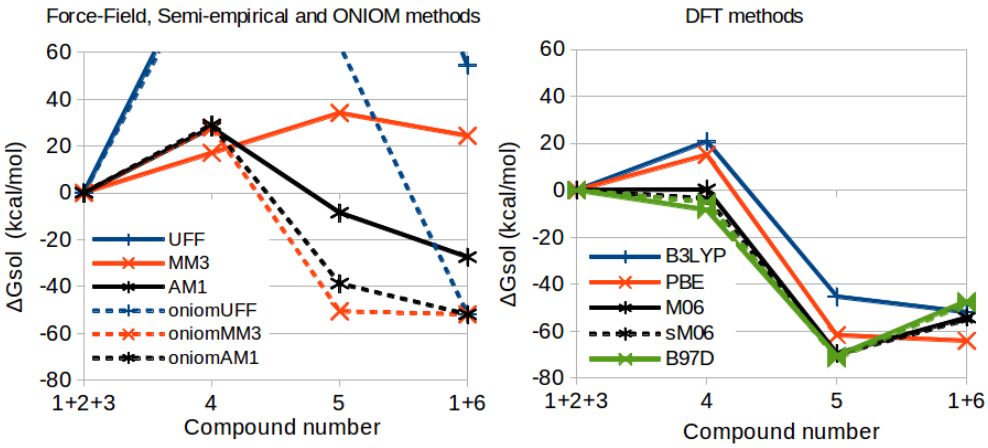



Figure 3.5: *Computed energetics of the catalytic cycle of the cucurbit[6]uril in kcal/mol.*

other components are very similar with **MM3** and **B97D**.

A force field adds various contributions, which all intend to compensate for each other errors. Therefore, analyzing components independently may be misleading. The stretching, bending, torsion and out-of-plane energy contribution is only modest in both force fields. What makes **UFF** a poor competitor in front of other methods is the van der Waals energy, which is only -14.6 kcal/mol and we expect it to be a more strongly stabilizing component.

Van der Waals description in UFF is insufficient, and hydrogen bonding is not well reproduced. The characteristics of the system, *i.e.* large surface for non-bonded interaction and multiple hydrogen bonds, makes the force field energy unrealistic (see also Figure 3.5).

The guest inclusion is promoted by the dispersion interactions and the multiple hydrogen bonds between the ammonium groups of the guests and the $sp^2$ oxygen atoms of CB6.

**MM3** describes the van der Waals potential with an angle-dependent exponential 6-equation, derived from the Hill equation.[132,133] The approach and parameters of this force field seem to be somehow successful on the system tested, particularly when comparing with **UFF**.

If we consider **B97D** correct, 25.5 kcal/mol are missing, which is not negligible, but is surely amplified by the size of the system and the large interacting surface of the different fragments. We will compare force fields again with other methods in the next chapter with another supramolecular system.

Table 3.4: *Energy breakdown for the complexation step (**1**→**4**) with* **UFF**, **MM3** *and* **B97D**. *Values in kcal/mol.*

| Contribution | **UFF** | **MM3** | **B97D** |
|---|---|---|---|
| **Force Field** | | | |
| Coulomb | 28.3 | 27.1 | |
| Van der Waals | -14.6 | -130.9 | |
| Stretching | -0.2 | 0.9 | |
| Bending | -0.8 | -1.7 | |
| Torsion | -1.5 | -2.5 | |
| Out-of-plane | 0.1 | 1.0 | |
| *Subtotal* | *11.2* | -106.2 | |
| **Functional** | | | |
| Energy | | | -88.5 |
| Dispersion | | | -51.5 |
| BSSE | | | 16.4 |
| *Subtotal* | | | *-123.6* |
| **Other** | | | |
| Solvation energy [a] | 76.0 | 87.0 | 83.5 |
| -TΔS | 21.8 | 31.1 | 30.5 |
| Thermal correction | 3.2 | 8.8 | 5.0 |
| Standard state correction | -3.8 | -3.8 | -3.8 |
| *Subtotal* | *97.2* | *123.2* | *115.2* |
| **TOTAL** | **108.4** | **17.0** | **-8.4** |

[a] implicit solvent (SMD model), single points performed with B3LYP. See equation 2.29, Section 2.3

### 3.4.2 Semi-empirical method AM1

Semi empirical calculations with methods such as **AM1** can be completed rather quickly on any modern personal computer, even for the large aggregates presented here. Unfortunately, the complexation energies obtained for this system are positive. However, the structural parameters we presented Table 3.1 are acceptable as they resemble the results produced by density functionals. The setup required for an **AM1** calculation is minimal and does not require any parametrization. This method could be used to perform a quick clean-up of a hand-drawn structure for instance, before using more expensive models. Still, one should be particularly careful with non-bonded interactions.

### 3.4.3 ONIOM methods

ONIOM calculations can be a method of choice for supramolecular systems, as the size of the high layer is reasonable: 23 atoms for our system. The

computational effort for the MM part almost negligible. Comparing the resources needed to run an ONIOM and a full DFT calculation, the advantage of the first are considerable.

In this work, we split the system in two regions. The guests are treated with B3LYP while the host is modeled with a lower-level method, respectively MM3, UFF, and AM1 for **O-UFF**,**O-MM3** **O-AM1**

With **O-UFF**, comparing with **UFF**, the description of the cycloaddition reaction (within step **4**→ **5**) is fixed, but the results are still largely erroneous.

With **O-MM3**, The results are qualitatively similar with the **MM3** method. With a significant improvement in the description of the reactivity, as expected due to the failure of MM methods to describe bond formation and breaking. In parallel, the enthalpy of complexation (-6.0 kcal/mol) is not as negative as with **MM3** (-19.2 kcal/mol), while entropic contributions appear similar. If we consider DFT or DFT-D as a reference, the **O-MM3** energies are very promising, see Figure 3.5. Indeed, the similarity with **B3LYP** results indicates that the description provided by **O-MM3** has reached an important milestone.

**O-AM1** model failed as **AM1**, as complexation energies ($\Delta E_{Sol}$) are only weakly negative. We discourage its use for host-guests systems, particularly because of the insufficient description of non-bonded interactions: van der Waals and hydrogen atoms bonds. The results are quantitatively similar to **O-MM3**, for a limited computational cost which is mostly borne by the high layer.

### 3.4.4 DFT Methods

DFT methods are by far computationally more expensive, compared to all the other methods used in this study. They can however produce valuable results and be used as a reference point. Moreover we saw in Section 2.1.5 that they are not all equivalent.

Dispersion effects become considerable during the complexation: hosts and guests are in close contact with each other on a vast surface provided by the reaction chamber. Entropy works against constraints and organization, while non-bonded interactions work in favor of aggregation. The relative intensities of those two main contributions creates a subtle equilibrium that will ultimately decide whether aggregation takes place or not.

The main problem lies in the treatment of exchange and correlation, that are mathematically unknown for real systems. Unfortunately, the subtle interplay of (i) description of exchange and (ii) correlation and finally (iii) the specific needs of the system makes it difficult to predict which functional is going to perform best for a given system and why.

According to our results the popular **B3LYP** as well as **PBE** can reproduce reactivity, and hydrogen bonds appears to be well described. Besides, the structures found in DFT and DFT-D are essentially the same. These two functionals suggest that aggregation processes are unfavored in terms of free energy in solution, in contradiction to experimental results that report full occupancy of the hosts. For this reason, we have to discard them for this system. We will address this question again in the next chapter with another system.

The strategies deployed by **B97D** and **M06** to recover the dispersion energy are different, but both obtain a large stabilization ($\Delta E_{sol}$) compared to other functionals (see Table 3.2 and Figure 3.5). Aggregation energies ($\Delta G_{sol}$) are negative for **B97D** and **B97D**$_{sol}$ and **M06**$_{sol}$, but slightly positive with **M06**.

All key aspects of supramolecular chemistry mentioned in this paper are correctly rendered by the DFT-D functional **B97D**: (i) structure of the supermolecule, (ii) inclusion of guest(s) that depends on hydrogen bonds and dispersion interactions (iii) molecular deformations, and (iv) reactivity.

Grimme's **B97D** functional is about two times faster than **M06** for SCF procedures, frequency and solvent calculations. Also, convergence is generally easier with **B97D**. We will then use **B97D** from now on to study this system in more details.

### 3.4.5 Optimizations in Solvent

Compared to their gas phase analogs, solvent optimized geometries (methods **B97D**$_{sol}$ and **M06**$_{sol}$) show small changes, both in energy and structure.

Optimizing in solution has an extra-cost that is not negligible, considering the already large computational effort required by full-DFT calculations.

An explicit model would be more costly, although more accurate, in particular to render entropy-related contributions in (de)solvation and (dis)aggregation. Indeed water molecules may have a contribution: the $sp^2$ oxygen on the rims of the CB6 molecules are candidates for hydrogen bonding.

To confirm or invalidate our implicit model, we tested water and formic acid (the mixture used as solvent in the experiment) as guests for the CB[6] molecule. It turns out that only one (-2.4 kcal/mol) or two (-4.8 kcal/mol) molecules of formic acid interact favorably with the macrocycle. Adding more formic acid appears to work against thermodynamics. CB[6] is known not to be water-soluble[134] and water displays positive free energies of complexation. This suggests that the entropic role of solvent is limited and validates the "implicit" solvent approach, but one or two formic acid molecules may have to be included explicitly in the core of CB6.

### 3.4.6 Basis Set Superposition Error

Computing the BSSE leads to large values. In this context, it is mandatory to take them into account. For the study of smaller systems or with very large computing facilities, one may consider the use of larger basis functions to minimize BSSE. Unfortunately, this approach is difficult since it drastically increases the computational cost associated.

### 3.4.7 Conclusion on the Methods

Having selected the best method from the benchmark that is **B97D**, we propose Figure 3.6 the structures of the optimized geometries at this level of theory. In addition we propose a close look at the binding mode of **4** in Figure 3.7

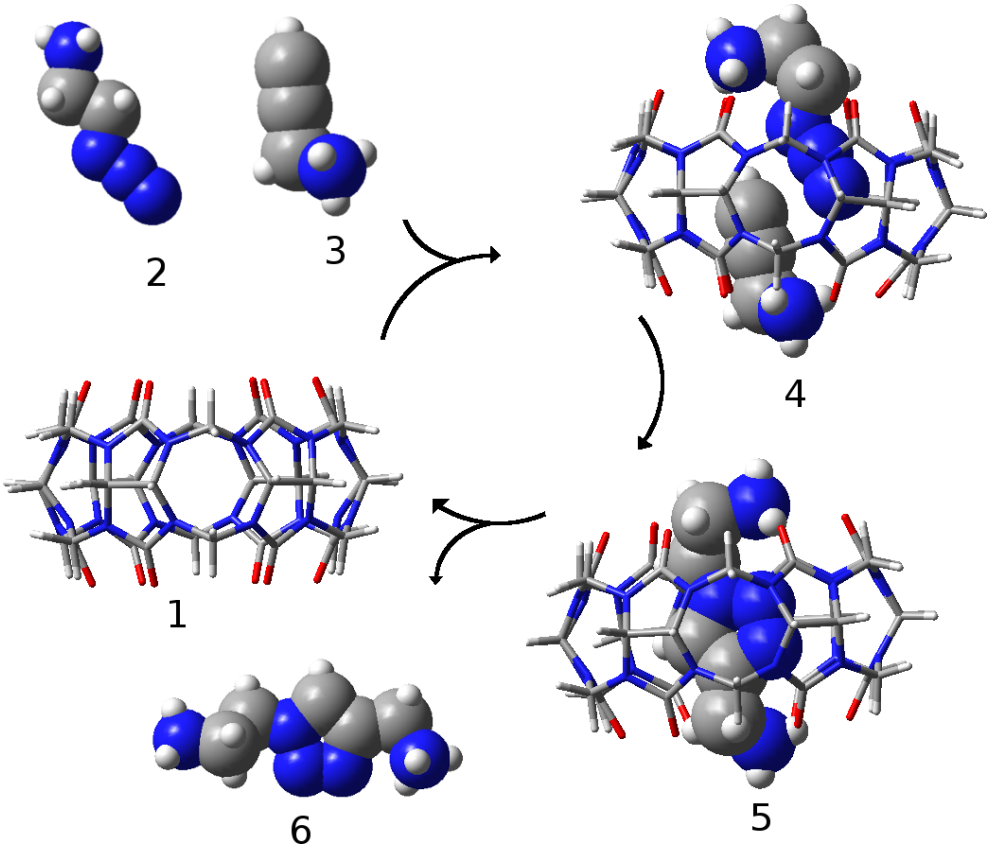



Figure 3.6: *Geometries obtained for the catalytic cycle at the B97D level of theory. Host: stick model. Guests: space-filling model.*

Although the final complexation energy is relatively modest with -8.4 kcal/mol, it is important to underline that the components forming this energy are large and of opposite sign, see Table 3.2 and 3.4.

The major components that work against of aggregation are:

**Entropic:** +31.7 kcal/mol.
**Solvation Effects:** +83.5 kcal/mol.

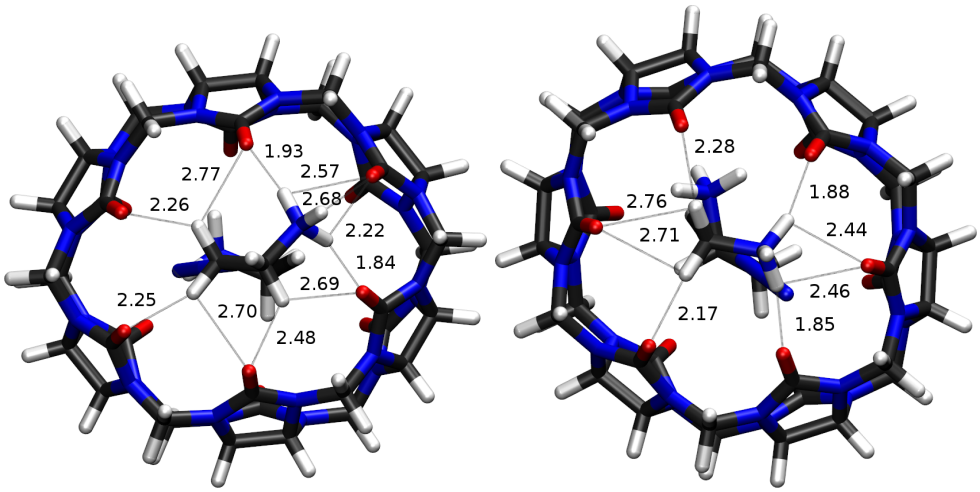



Figure 3.7: *Geometry obtained for the adduct* **4** *at the* **B97D** *level of theory. Hydrogen bonds are visible and the numbers refers to their lengths. Left: view from the side of the bound azide. Right: view from the side of the bond acetylene.*

The large decrease of entropy is primarily explained by the number of aggregated molecules rather than by their nature. The loss of solvation energy occurs as the implicit solvent models considers the hosts and guests partially unsolvated as they start interacting. Also, each guest holds a positive charge that is stabilized by the solvent.

The major component that work in favor of aggregation is the potential energy:

**Potential energy:** -123.6 kcal/mol.

The potential energy is also very large, and forms the stabilizing component. Out of the -123.6 kcal/mol , -51.5 kcal/mol is recovered by dispersion corrections, and +16.4 kcal/mol arises from BSSE corrections. Its magnitude is explained by the large surface of contact available for van der Waals-type of interactions. Also, the positive charge of each guest interacts strongly with the oxygen fringes on each side of the macrocycle.

## 3.5 Incorporation of the Dynamic Dimension to the Model

How does the system behaves on the long run? A major concern is the description of aggregation and disaggregation processes. One should incorporate all the relevant extrema of the energy surface. Here the complication does not rely on conformations since the cucurbituril molecule as well as the aggregate are both well defined. Rather, the size of the system and the location of transition states, particularly concerning aggregation is problematic.

Enclosed solvent molecules may prevent reaction. Furthermore, non-productive binding represents alternative binding modes such as aggregation of guests with their functional groups pointing outwards. Finally, symmetrical pairing of guests may also slow down the formation of the product, as well as non-productive binding, a binding mode where guests reactive functional groups point outside the center of the macrocycle.

In this section, we aim at providing a description of processes, *i.e.* switching from static to dynamic representations.

### 3.5.1 Enhancement of the Cycloaddition Reaction

Mock and coworkers obtained a significant acceleration in the formation of the reaction product. Experimentally, the lack of turnover prevents the completion of the catalytic cycle.[75] To check that the cycloaddition step is not rate-limiting, we have to calculate the corresponding barrier, which requires the energy of the most stable intermediate.

Here we address the question of the stability of the guests in the macrocycle. Three aggregates are presented: an azide (**2**) or an acetylene (**3**) alone or simultaneously trapped by the macrocycle (**1**). The results are presented in Table 3.5, and the structures in Figure 3.8

Table 3.5: *Energies of each guest separately trapped by CB6, as well as simultaneously trapped. Values in Kcal/mol.*

| compound | $\Delta E_{sol}$ | $\Delta G_{sol}$ |
|---|---|---|
| separated species | 0.0 | 0.0 |
| complex **1**+**2** | -26.0 | -9.4 |
| complex **1**+**3** | -30.2 | -15.7 |
| **4** | -40.2 | -8.4 |

It appears that the azide (-9.4 kcal/mol) does not bind as well as the acetylene (-15.7 kcal/mol). The thermodynamic stabilization of the second guest is limited by the presence of the first one in the macrocycle. What

is remarkable is the weaker thermodynamic stability of the simultaneously trapped reactive species (-8.4 kcal/mol). When put together, they appear not to be in their optimal configuration, see Figure 3.8. When compare the stability of **4** to complex **1+3**, which is the most stable, we still get a enthalpic stabilization ($\Delta\Delta E_{sol}$) of -10.0 kcal/mol. However the entropy (($\Delta\Delta G_{sol}$)-($\Delta\Delta E_{sol}$)) plays a destabilization role: +17.3 kcal/mol. This value of +17.3 kcal/mol is larger that what is normally observed for an aggregation process, which indicates that the incoming compound **2** constrains entropically the complex **1+3** while entering CB6.

To calculate the proper activation energy of this reaction, we need to compare the energy of the complexed transition state with the energy of the most stable intermediate which is the complex **1+3**.

We calculate here the energy of the transition state inside the host and compare it with the energy of the complex **1+3** to obtain the activation free energy. We propose Table 3.6 a comparison with the reaction barrier in solvent with the barrier inside CB6.

Table 3.6: *Energetics of the 1,3 dipolar cycloaddition, with and without catalyst. Energies in kcal/mol*

| reaction | | $\Delta E_{sol}$ | $\Delta G_{sol}$ |
|---|---|---|---|
| **2+3** → TS | in solution | 16.0 | 27.5 |
| complex **1+3** → **TS 5** | inside CB6 | -0.3 | 20.6 |

It is striking how different are the components of the activation energy barriers. In the uncatalyzed reaction both the enthalpy ($\Delta E_{sol}$ = 16.0 kcal/mol) and the entropy ($\Delta G_{sol}$-$\Delta E_{sol}$ = 11.5 kcal/mol) play a role. The total free energy of activation is $\Delta G_{sol}$ = 27.5 kcal/mol. In the macrocycle, the energetics are completely altered. The activation enthalpy has vanished and even becomes slightly negative (-0.3 kcal/mol). Consequently, only the entropy ($\Delta G_{sol}$-$\Delta E_{sol}$) directs the energetics of the reaction, which is evaluated at 20.9 kcal/mol. This interesting results tells us in great details what a successful strategy can be.

In agreement with the experimental study, our results suggest a large acceleration of the cycloaddition reaction. We identified four components that weigh in the modification of the energy profile of the Huisgen reaction by CB6: (i) The potential energy component is canceled as discussed. (ii) The entropic cost rises by confinement and strain. (iii) The strain of the macrocycle induced by the guest plays a compression role. (iv) The solvent does not participate in the reaction anymore.

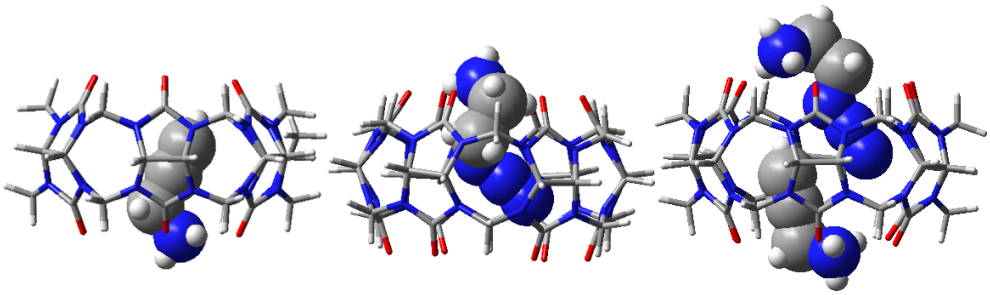


Figure 3.8: *Molecular models for each guest individually entrapped (left and middle) and entrapped together (right). Geometries optimized with B97D.*

### 3.5.2 (Dis-)Aggregation Processes

Complexation processes are delicate to describe, since the entropy plays a determinant role. The formation of a given supermolecule may be thermodynamically favored, it is still presumed to come with a small kinetic barrier. The location of the corresponding transition state is difficult, precisely because the optimizations schemes we use do not include entropic effects. Calculation of entropy requires indeed the evaluation of thermodynamic properties, notably by the costly diagonalization of the Hessian matrix.

To estimate the activation energy of an aggregation reaction, we test the following procedure: starting from the aggregate, we slowly take away one of the guests by performing a relaxed scan on a relevant coordinate. This coordinate will be the only constrain and with this method, we obtain an energy profile of the disaggregation process. A step size of 0.25 Å was used, the scan was performed between one of the oxygen atoms of the lower rim and the carbon in alpha of the azide group of the guest. One on ten points obtained along the path are picked and we perform a frequency calculation on these points to obtain free energies. We also include BSSE and solvent correction.

We present this scan Figure 3.9. On the far left of the graph is plotted the free energy of the complex, from which the guest is gradually removed from the host. On the far right is added a point, where the free energies of the guest and the host are computed separately.

The aggregation and disaggregation processes we intend to model are for instance: (i) Aggregation of **1** with **2** (ii) Aggregation of **1** with **3** (iii) Aggregation of **1**+**2** with **3** (iv) Aggregation of **1**+**3** with **2** (v) Disaggregation of **5**

Invariably, aggregation processes showed almost perfect compensation of

potential energy with solvation effects, on long to intermediate distances. This is surprising since both contributions can be very large. When the guests enter the core, enthalpic stabilization becomes significantly larger. On the other hand, entropy shows an unreliable behavior.

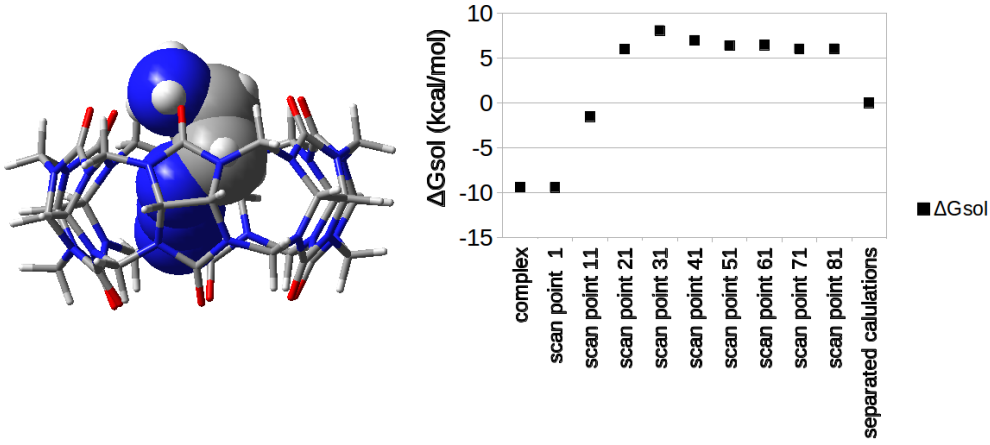



Figure 3.9: *Left: Complex used for the scan. Right: results of the scan. Energies in kcal/mol.*

On the point 81, on the far right of the graph, the O-C distance (scanned parameter) is 26 Å. The closest atom-atom distance between the host and the guests is 16 Å. If we push things further away, the behavior remains the same. Why is the entropy not increasing further (and therefore the free energy could drop to 0)?

The models we use cannot handle well the assignation of low frequencies. For a non-diatomic non-linear molecule, three vibrations will be assigned to translation modes, three to rotation modes and the rest to vibrational modes. In our cases, we will find very low vibrational modes between the host and the guests at large distances. Those low vibrations should instead be interpreted as rotations or translations, since the two molecules are too far apart. Additionally, the imprecision on vibrations sometimes results in the low-lying vibrational modes to be treated as imaginary frequencies and are consequently ignored. We could not find a solution to handle this discontinuities, even for very simple systems.

For those reasons, the contribution of the entropy appears unreliable and the identification of a TS becomes problematic. To circumvent this difficulty, we will assume that these processes do not possess a proper transition state and that their kinetics are limited by diffusion. For an arbitrary reaction $A + B \rightarrow C$, an associated rate constant can be calculated, from which we derive a pseudo-activation energy, although there is formally no transition state.

This value can be estimated, as described by the classical Smoluchowski expression,[135] equation 3.1.

$$k_{diff} = 4\pi DRN_A \tag{3.1}$$

Here, $\mathrm{k_{diff}}$ is the rate constant for a diffusion-controlled reaction. $\mathrm{D}$ is the sum of the diffusion coefficients of A and B. $\mathrm{R}$ is the sum of the radii of the each molecule, assuming they would be spherical. Finally, $\mathrm{N_A}$ is the Avogadro number.

The value we obtain for $\mathrm{k_{diff}}$ is 1.1*$10^{11}$ between one host and one guest, which corresponds to a very fast reaction with pseudo activation free energy of $\Delta\mathrm{G} = 2.4$ kcal/mol.

### 3.5.3 Building a Comprehensive Kinetic Model

To construct a pertinent kinetic model, we need to consider each relevant conformation and the transitions that links them. We have to take into account solvent, non productive binding as well as all of their possible combinations. This reveals to be a vast task, as illustrated Figure 3.10.

In this scheme, different species are represented as objects with different shapes and colors. The CB6 is represented by a circle, the azide by a blue elongated shape, and the acetylene by a red elongated shape. The 1,4 product is represented by a purple ellipsoid and the 1,5 product by a pink one. Formic acid is represented in black, while water does not appear on this scheme as a guest, since we found that its stability of complexes involving water was systematically inferior to complexes involving formic acid.

We distinguish three areas: the uncatalyzed reaction on the top, the catalytic cycle on the bottom right corner and the various possibilities and arrangements occupy the rest of the figure. In shaded colors are indicated the "non-productive" bindings.

We considered there can be up to two guests since CB6 has two openings. Also up to two solvent molecules can be incorporated.

As an approximation, we considered to retain from the computed structures only the most relevant ones, *i.e.* the most stable structures and some intermediaries required to connect them. The reaction network appears greatly simplified, see Figure 3.11.

To build our kinetic model, we still need to properly relate and connect all the intermediates. To this end, we will use diffusion processes with the pseudo activation barriers mentioned earlier. The only standard transition energy we need has been obtained for the cycloaddition step (with and without CB6).

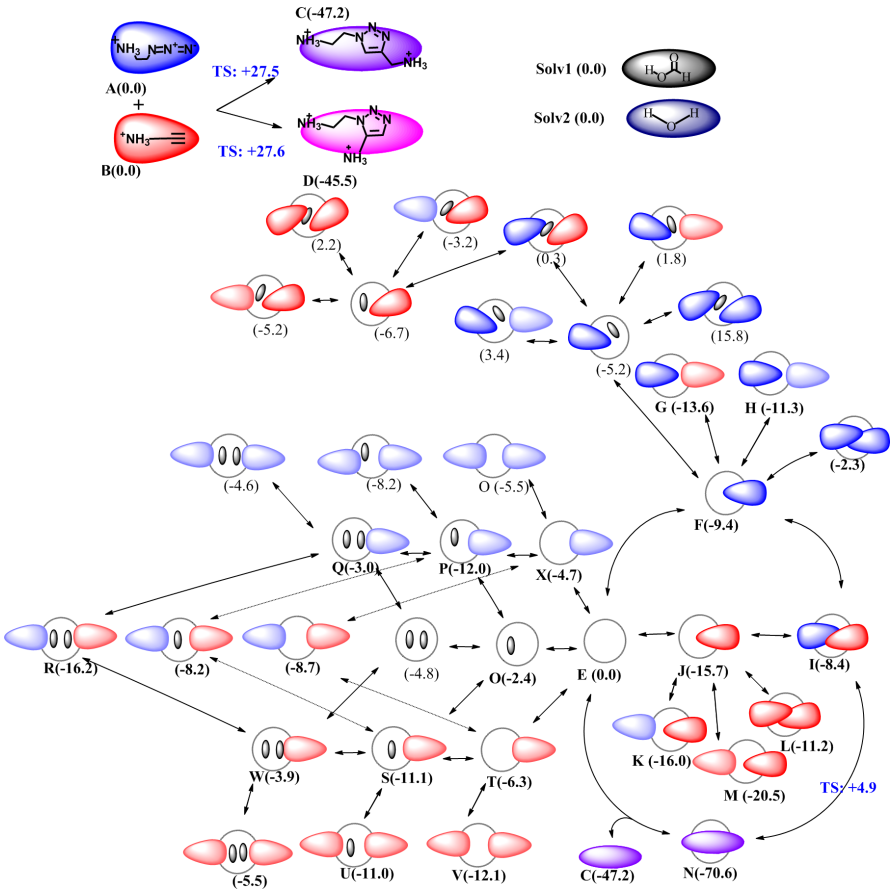



Figure 3.10: *Network of computed structures related to CB6 and its guests. Each structure is in dynamic equilibrium with its neighbors on the graph if they are related by arrows, some of which have been omitted for clarity. Energies displayed in kcal/mol.*

All the free energies are converted into rate constants. At this point, everything is ready to use a kinetic program such as Acuchem[127] or Tenua.[136] Both will give the same results, negligible accumulation of numerical errors put aside.

At t=0, the concentrations of **A**, **B** and **E** are respectively of 0.0150 M, 0.00375 M and 0.0030 M. As some complexes involve solvent (formic acid) molecules, we assign it a concentration of 11.6 M.

The results of the kinetic simulation are presented in Figure 3.12. The graph represents the evolution of the concentration of various complexes and species over time, using a log-log scale. For the guests, one can see that **B** slowly disappear, caught by the macrocycle. The concentration of **A** is

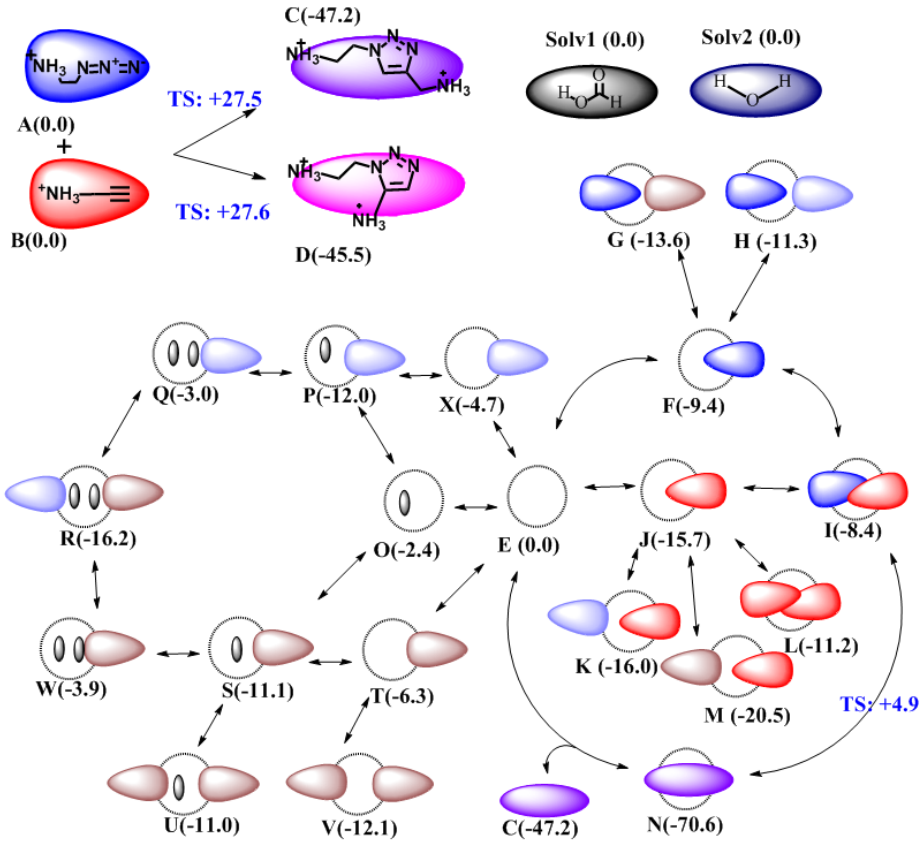



Figure 3.11: *Pruned network of structures selected for the kinetic model. Only the most stable structure and some intermediates have been selected. Energies displayed in kcal/mol.*

4 times more important than **B**, so no consumption is noticeable, and **A** is always majoritary among the solutes. Without surprise, **E** quickly disappears in favor of other complexes, such as **O** ($10^{-12}$s to $10^{-9}$s), which is in turn replaced by **P** and **S** ($10^{-9}$s to $10^{-3}$s), which are also replaced by **R** that will remain a dominating species as long as the reaction did not completely finish ($10^{-3}$s to $10^{+5}$s, *i.e.* up to about 1 day). **J** and **M** follow a similar trajectory to **R**. The preponderance of **R** over **M** is due to the presence of solvent in the complex, although **R** is not as stable as **M**.

Inside **E**, The complete conversion of **A** and **B** to **N** takes $10^6$ seconds, which corresponds to about 12 days. At this moment, **C** and **D** are found in very small amounts: about 5 orders of magnitude behind. **C** and **D** have

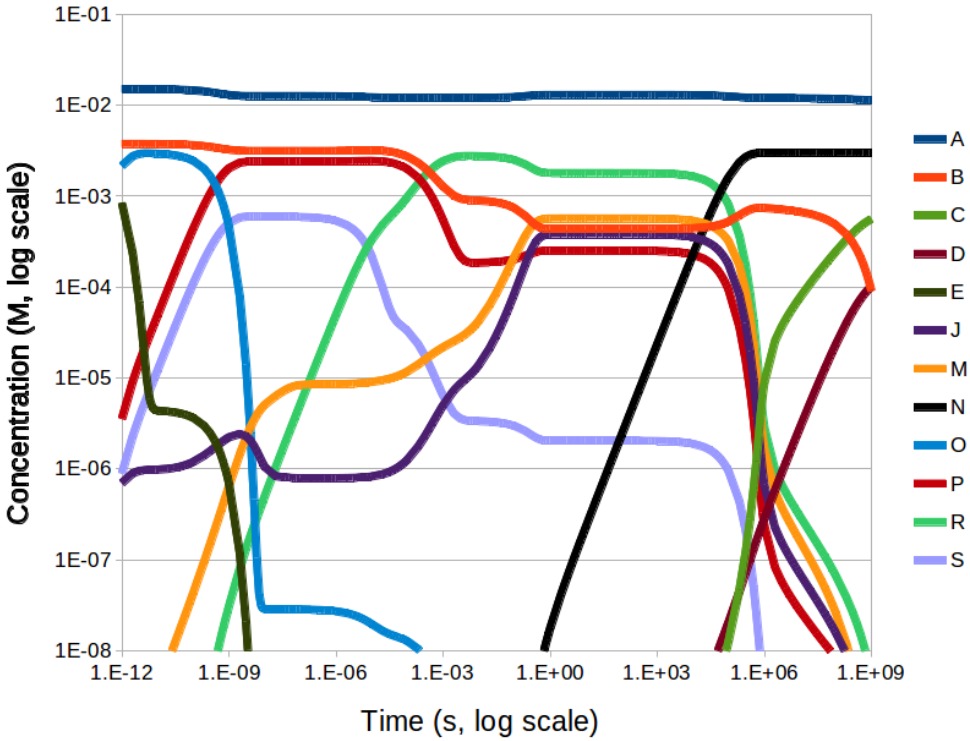



Figure 3.12: *Time evolution of the concentration of the some of the most relevant species related to CB6 and its guests.*

almost exactly the same rate of formation in solution, but at large time scales one can see the concentration of **C** taking off because it can be liberated from the host and replaced by other guests. Later, the product occupies most of the hosts and we can see the concentrations of other complexes dropping.

The species that do not appear have less interest and have been removed for clarity. Most of the absent species follow a similar evolution: the concentration increases till reaching a maximum around $10^{-7}$ seconds, then slowing vanishes after a plateau between 1 and $10^4$s, to finally disappear as the reaction completes.

The kinetic models reveals to be very helpful. Our choice to include some of the lesser stable intermediary is not critical but one can appreciate to follow precisely the evolution of the concentration of different species over time.

### 3.5.4 Confronting Experimental and Theoretical Data

The kinetic models gave us the results we expected from our data, and are in good agreement with experimental evidence. Non-productive binding seems to be a major problem in the catalytic cycle, see Figures 3.11, 3.12 and 3.13. The main species concerned are **M** and **R**. Being very stable, they hold back the reaction. Furthermore, as **R** contains solvent, the trend is even further pushed in this direction.

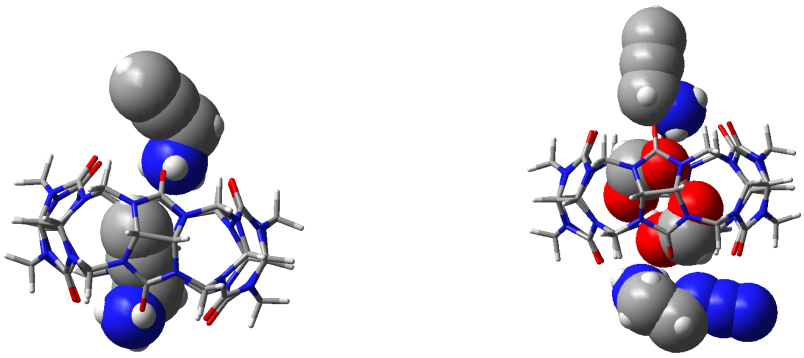


Figure 3.13: *Representation of the two the most stable complexes:* **M***(left) and* **R** *(right).*

Those stable complexes slow down the catalysis. Another factor is the stability of complex **I**, which is lower than **J** and **F**, although we saw section 3.5 that it is not entirely problematic. However, the guests will spend most of their time in a disadvantageous configuration which is a major hold back for this reaction.

To compare our results to experimental kinetic data, we gather in Table 3.7 the values obtained from the kinetic study in Mock's paper.[75] To do so, we first convert dissociation constants and kinetic constants into respectively energy differences ($\Delta G$) and activation energies ($\Delta G^{\ddagger}$), using the following formulas:

Dissociation constants:

$$K_d = e^{-\Delta G/RT} \tag{3.2}$$

Rate constants:

$$k = \frac{k_b T}{h} * e^{-\Delta G/RT} \tag{3.3}$$

In this manner, it is easy to compare the results obtained by both methods, as proposed in Table 3.7 . It is worth mentioning here, as we will see in

the next subsections, that the experimental data are not directly measured, but indirectly obtained by making some assumptions in the treatment of the measured parameters.

Table 3.7: *Comparison of experimentally reported and theoretical kinetic data.*

| | Experimental data | | | | Theoretical data | | | | |
|---|---|---|---|---|---|---|---|---|---|
| reaction | kinetic constant | → | ΔG | $\Delta G^{\ddagger}$ | kinetic constant | ← | ΔG | $\Delta G^{\ddagger}$ | react. order |
| **A+B → C** | $1.2*10^{-6}$ | → | | 25.4 | $4.1*10^{-8}$ | ← | | 27.5 | 2 |
| **A+B → D** | $1.2*10^{-6}$ | → | | 25.4 | $3.4*10^{-8}$ | ← | | 27.6 | 2 |
| **F ⇌ E + A** | $2.5*10^{-3}$ | → | 3.5 | | $1.2*10^{-7}$ | ← | 9.4 | | * |
| **J ⇌ E + B** | $6.5*10^{-4}$ | → | 4.3 | | $2.8*10^{-12}$ | ← | 15.7 | | * |
| **M ⇌ J+ A** ** | $3.0*10^{-1}$ | → | -0.7 | | $1.2*10^{-1}$ | ← | -1.2** | | * |
| **I ⇌ J+ B** ** | $3.0*10^{-1}$ | → | -0.7 | | $1.2*10^{-1}$ | ← | -1.2** | | * |
| **I → N** | $1.9*10^{-2}$ | → | | 19.7 | $1.1*10^{+3}$ | ← | | 13.3 | 1 |
| **N→ E + C** | $1.7*10^{-4}$ | → | | 22.5 | $4.6*10^{-5}$ | ← | | 23.3 | 1 |

*: Dissociation constant.

**: Averaged value: Attempt to reproduce the assumption of two identical rate constants made by experimentalists.

#### Background Cycloaddition

It appears that an agreement is obtained for the uncatalyzed reaction. Experimentally, it is straightforward to measure by UV spectroscopy, by monitoring the disappearance of the signal that corresponds to the absorbance of the azide group. No precision is given regarding the regioselectivity of the reaction. Our data slightly overestimate the experimental values, but gives two almost indistinguishable rate constants.

#### Decomplexation

For the decomplexation of **F** and **J**, experimental and theoretical data are completely different. How is this possible ? (De)-complexation experiments measure the disappearance of the absorption of a known calibrating guest, displaced by the guest to study. Kinetic data is interpreted by assuming a dissociative mechanism.[137]

To understand the values obtained, we performed a series of calculations involving the bulky chromophore guest used for calibration as well as the guests of interest in this study (**A** and **B** in the new scheme). We present

Figure 3.14 this trapped chromophore (in green) and its stability whether associated or not with guests.

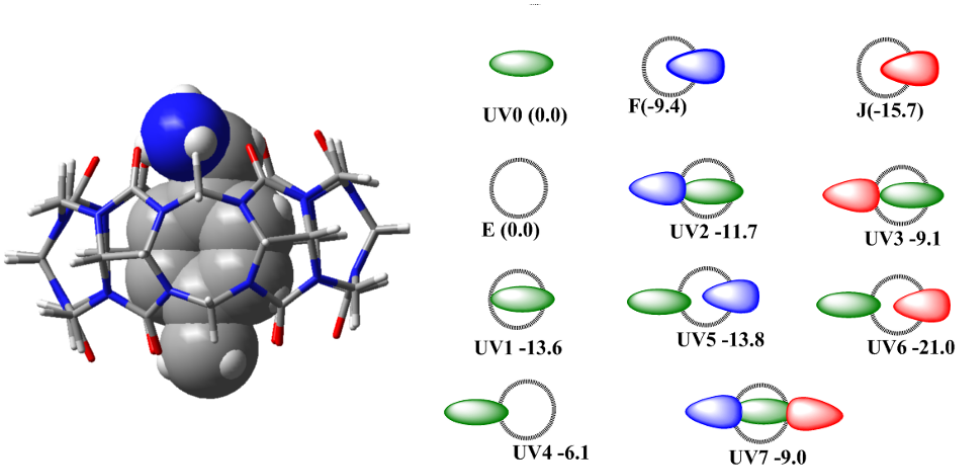



Figure 3.14: *Left: optimized structure of the guest used for reference in the dissociation study. Right: several associated structures computed.*

The results show that the chromophore does not completely dissociate, as a guest **A** or **B** can replace it in inside the host, while the chromophore would remain outside, bound to the oxygen rim of the host unoccupied by the amino group the guest **A** or **B**. The result is a more stable complex in both cases. Therefore, according to our results, a dissociative mechanism is not taking place, and the experimental data should be re-interpreted accordingly.

Let us show first what the experiments reflect in details. On Figure 3.15 is shown the first part of the experiment, involving CB6 with a chromophore displayed in green. The complexation is followed by the disappearance by increase of optical density. As CB6 is in excess, a pseudo-first order rate constant is obtained: $k_1$.

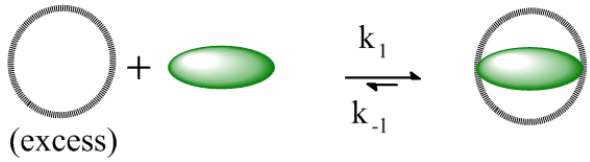



Figure 3.15: *Illustration of the first part of the experiment intended to obtain the forward rate k1.*

In the second part of the experiment, after equilibrium is obtained, a ex-

cess of displacing agent (in black) is added, see Figure 3.16. It is assumed a dissociative mechanism based on experiments with alkylamonium ions. As the binding of the displacing agent (1,6-hexanedamine) is stronger as instantaneous, the change in absorbance should reflect the backwards process: $k_{-1}$.

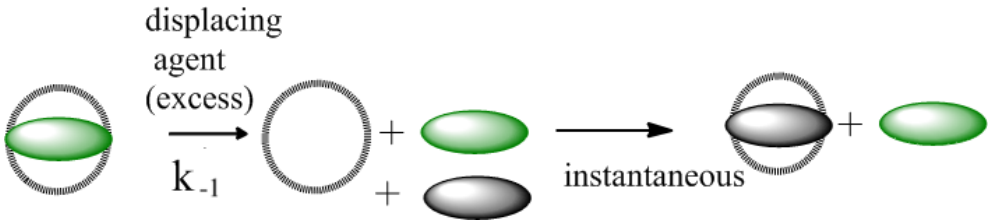



Figure 3.16: *Illustration of the second part of the experiment intoned to obtain the backwards rate k-1.*

Our results suggest that the mechanism is not dissociative for **A** and **B**. The chromophore does not necessarily need to dissociate for **A** to enter. A possibility is that the chromophore crosses the macrocycle and at the same time **A** replaces it inside. We illustrate this possibility with the example of **A** in Figure 3.17. According to the relative energies just presented Figure 3.14, this appears less energetic and thus more likely. The observed "backwards" rate constant in the experiment is then monitoring $\mathbf{UV1} \rightarrow \mathbf{UV5}$ instead of $\mathbf{UV1} \rightarrow \mathbf{E}$.

The dissociation constants calculated would then reflect the equilibrium $\mathbf{UV5} \rightleftharpoons \mathbf{F} + \mathbf{UV0}$ for guest **A** and $\mathbf{UV6} \rightleftharpoons \mathbf{J} + \mathbf{UV0}$ for guest B. For these equilibriums respectively calculate a Kd of $5.7*10^{-4}$ and $1.3*10^{-4}$. Taking into account the sensitivity of the formula used, this can be considered as good agreement. We will present the new results in Table 3.8.

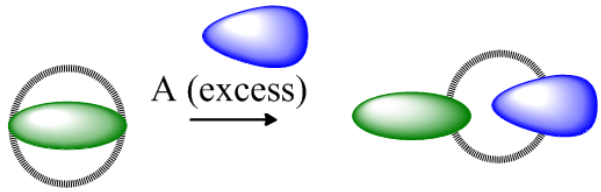



Figure 3.17: *Mechanism of replacement taking place when* **A** *is used instead of the displacing agent, according to our results*

Table 3.8: *Corrected comparison of experimentally and theoretical kinetic data.*

| | Experimental data | | | | Theoretical data | | | | |
|---|---|---|---|---|---|---|---|---|---|
| reaction | kinetic constant | → | ΔG | $\Delta G^{\ddagger}$ | kinetic constant | ← | ΔG | $\Delta G^{\ddagger}$ | react. order |
| **A+B → C** | $1.2*10^{-6}$ | → | | 25.4 | $4.1*10^{-8}$ | ← | | 27.5 | 2 |
| **A+B → D** | $1.2*10^{-6}$ | → | | 25.4 | $3.4*10^{-8}$ | ← | | 27.6 | 2 |
| **UV5 ⇌ F + UV0** | $2.5*10^{-3}$ | → | 3.5 | | $5.7*10^{-4}$ | ← | 4.4 | | * |
| **UV6 ⇌ J + UV0** | $6.5*10^{-4}$ | → | 4.3 | | $1.3*10^{-4}$ | ← | 5.3 | | * |
| **M ⇌ J + A** | ** | → | ** | | ** | ← | ** | | * |
| **I ⇌ J + B** | ** | → | ** | | ** | ← | ** | | * |
| **I → N** | $1.9*10^{-2}$** | → | | 19.7** | | | | | 1 |
| **J → N** | | | | | $4.7*10^{-3}$ | ← | | 20.6 | 2 |
| **N→ E + C** | $1.7*10^{-4}$ | → | | 22.5 | $4.6*10^{-5}$ | ← | | 23.3 | 1 |

*: Dissociation constant.

**: To be re-interpreted from raw experimental kinetic data. See text for details.

#### Other Dissociations

Now we discuss the dissociation constants of **M** ⇌ **J** + **A** and **I** ⇌ **J** + **B**, as shown Table 3.7. To interpret their data, the authors had to assume that both guests **A** and **B** bind equally to **J**. According to our data, it is not the case as each guest can bind inside or outside the cavity with very different affinities, as shown Figure 3.11. We reproduce the dissociation constant they observed according to our interpretation with good accuracy.

We conclude that the assumption of equal dissociation constants is not valid, and that kinetic experimental data should be re-interpreted accordingly. Furthermore, the various combinations and orientations should be taken into account.

#### The Enclosed Cycloaddition

The same thought process stands for the catalyzed reaction, as it is derived from the same data. The kinetic data is interpreted in such a way that the reaction is first order. According to our results, this is not the case, since **I** is unstable (-8.4 kcal/mol), relatively to other complexes, particularly **J** (-15.7 kcal/mol). The activation energy that corresponds to **J**+ **A** → **TS** is 20.6 kcal/mol. The theoretical activation energy is then in good agreement with the experimental value, but their data may have to be reinterpreted considering the global second order of the reaction.

#### Product release

Finally, the product release is measured by the rate of formation of the product outside the cavity. Experimental (22.5 kcal/mol) and theoretical (23.3 kcal/mol) values are in good agreement.

#### Concluding Comments on Kinetics

The lack of agreement at first sight seems to find reasonable explanations. We propose several keys to understand and reinterpret the experimental results. First, competitive complexation experiments are sensitive to non-productive binding. Second, a complex formed by a single guest with CB6 is destabilized for entropic reasons. Consequently the proper way to describe the energetics of the catalyzed reaction is not $\mathbf{I} \rightarrow \mathbf{N}$, but rather $\mathbf{J} + \mathbf{A} \rightarrow \mathbf{N}$.

## 3.6 Conclusion

We reviewed the efficiency of different methods for the modeling of a supramolecular complex. It appears that an ONIOM model has many advantages, while DFT-D completes the challenge of accuracy.

In an ONIOM model, errors arising from the low layer can be partially canceled when comparing energies. The accuracy of reactivity is ensured by the high layer (DFT). Also, mechanical embedding may reproduce the non-bonded interactions. Finally, ONIOM calculations are not computationally expensive provided that the number of atoms in the high layer remains small, which is the case here.

Yet, it appears that a large space for improvement remains with force field methods, within a QM/MM scheme. In the next chapter we will come back to this issue with another example of supramolecular system.

Choosing a full DFT description does not directly solve the problem of accuracy. We gave evidence that DFT failed to reproduce the energetics of aggregation processes, while DFT-D surpasses all the other methods in all aspects, except for the speed of calculations.

Solvent optimization does not seem to be critical here, geometries as well as energies were found to be comparable to gas-phase optimizations.

Finally, a double-zeta polarized basis set is rather limited. Computationally, choosing larger basis sets would be prohibitively demanding considering the size of the system, particularly for frequency calculations. We strongly recommend to compute BSSE, since they were found to be large.

We found that when a complex of a single guest with CB6 is formed, the second guests inclusion destabilized the complex. In other words, both guests form a stable complex individually, but not together.

To construct a reaction network, we modeled a large amount of complexes in order to take into account non-productive binding, which appears to slow down the catalyzed reaction. Assuming diffusion-limited processes, we used a kinetic model to reproduce the behavior of the system. The data required was obtained from DFT-D and converted to rate constants.

Results from the kinetic program are obtained almost instantaneously. First, the macrocycle sites are occupied by the most stable guests, and a dynamic equilibrium takes place between them. The lesser stability of guest $\mathbf{A}$ is compensated by its higher concentration relatively to $\mathbf{B}$. The reaction proceeds, enhanced by the proximity of the reactants offered in the macrocycle. Finally, the product formed occupies all the available hosts, stopping the catalyzed reaction.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Chapter 4

# Resorcin[4]arene-Based Capsule

## 4.1 Description

Calixarenes, introduced in the first chapter, may be altered in order to accommodate bulky molecules. In particular, a resorcin[4]arene can be obtained by connecting together the upper edges of the chalice to aromatic imides through four pairs of ether links, see Figure 4.1.

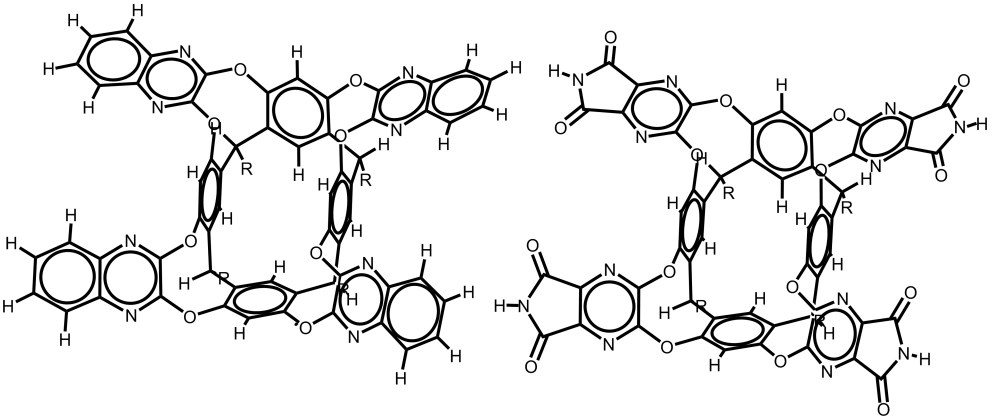


Figure 4.1: *Left: skeletal representation of Cram's host. Right: skeletal representation of Rebek's modified host.*

Rebek modified Cram's resorcin[4]arene[58, 138, 139] (Figure 4.1, left), by connecting four aromatic cyclic imides[32] (Figure 4.1, right). The new host acquired the capability to homo-dimerize. Two units will dimerize creating a seam of eight hydrogen bonds between the terminal N-H of the imide and

the vicinal carboxy groups of the other dimer, see 4.2.

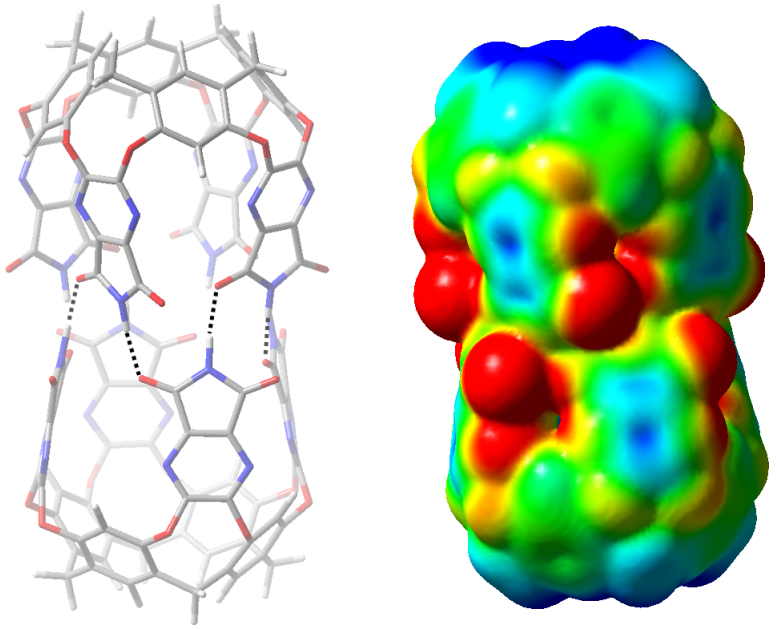


Figure 4.2: *Representations of the supramolecular capsule. Left: Stick model. Hydrogen bonds represented as dotted lines. Right: Electrostatic potential mapped onto electron isodensity, calculated with B97D.*

The resulting dimer[32] is a capsule of about *ca.* 20 Å long and 10 Å wide and has a limited lifetime of *c.a.* half a second.[140] There is an empty space inside, large enough to accommodate two benzene guests for instance. As long as a guest remains inside the core of the closed capsule, any direct interaction between the guests and the solvent is impossible. This is important since it brings a completely different environment to the guests than the one they would have in solution.

The capsule has long carbon tails at each extremity, which are essentially intended to improve solubility in non-polar mesitylene. In our calculations these tails have been pruned for computational feasibility. We assume that they will fold around each other and not perturb otherwise the behavior of the capsule.

Similarly to the case of CB6, the capsule has been used to promote in its core a Huisgen cycloaddition between two bulky guests: phenylacetylene and phenylazide.[55] Acceleration is reported to be 240-fold, but here again product inhibition prevents catalysis. We will come back to this point in Section 4.4.5.

Thanks to previous studies on similar cavitands,[58,141] we presume that a

so-called vase-kite equilibrium is very likely to take place, see Figure 4.5. Each resorcinarene monomer can adopt two different conformations, the vase-like and the kite-like , which is stable. Four eight-member rings provide for the flexibility of the substituents. Motion and stability are reported to be influenced by the temperature and the groups at the extremities of the gates. As the terminal groups were different in the experimental study, we cannot know *a priori* which of the two, vase or kite, is more stable. We include in our benchmark study both versions of the monomer, in order to test the capabilities of the different methods to handle structural distortions, see Figure 4.5.

### 4.1.1 The Most Suitable Guest

The cavity enclosed in the fully-formed capsule is elongated, not to mention quite large. It has been estimated around to be of 450Å$^3$ and may simultaneously accommodate two large molecules such as benzene or p-xylene.[55]

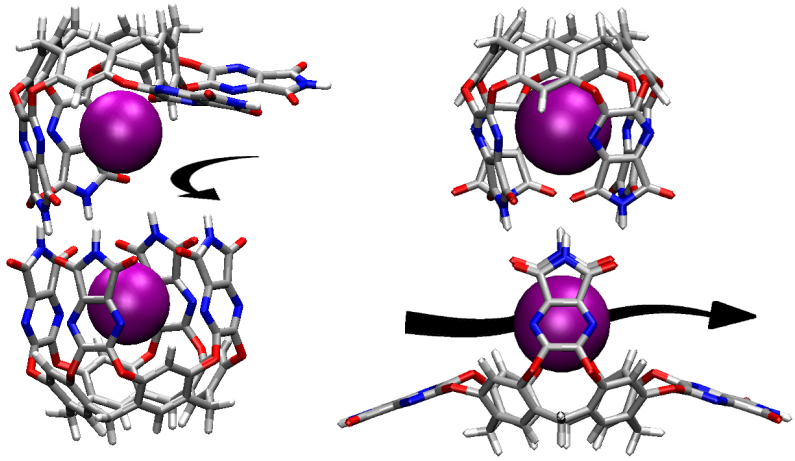


Figure 4.3: *Mechanism of guest exchange by competition, as proposed experimentally.*[142] *The purple spheres may represent any type of acceptable guest. Other mechanisms may exist.*

The conditions that a potential guest has to fulfill to be granted the access the capsule cavity are numerous. First, it needs an appropriate size to be physically able to enter and fit in. Second, it requires a good shape, in order to avoid an excessive distortion of the capsule. Third, host-guest interactions will be favored if the electrostatic surfaces match. Fourth, other non bonded interactions can weigh in such as $\pi$-stacking, London forces or hydrogen bonds. Last, if more than one guest is encapsulated, the complementarity

between the guests inside the host may also play a role in the stabilization of the whole complex.

The capsule has the ability to capture p-xylene and benzene which replace two encapsulated toluene molecules within "a few minutes".[32] This experimental result provides a wealth of insights.

It suggests strong host guests interactions, as well as an easy route towards encapsulation, while replacement of guests stays straightforward when favored thermodynamically. The reported competition takes place while the solvent (mesitylene) remains untrapped, probably because of its larger size. Also, NMR experiments indicate that guest exchange does not require dissociation of the two monomers forming the capsule.[140,142] The authors of the experimental study suggest that the guest renewal would be achieved by opening simultaneously two "gates" of the capsule, see Figure 4.3

## 4.2 Theoretical Study: Identifying the Most Well-Suited Methods Through a Comprehensive Benchmark

We studied the strengths and weaknesses of the same methods as in the previous chapter (MM, semi-empirical, ONIOM, DFT, DFT-D), in order to see their behavior on another bigger, more demanding and more flexible supramolecular system.

Being formed of two non-covalently bound subunits, the structure of the capsule is not as rigid as is the structure of CB6. Furthermore, the central resorcinarene bears four aromatic imide substituents, each one held by two ether bonds belonging to an eight-membered ring. Each of those four large substituents have a predictable, yet independent conformational freedom. For clarity, we already present the labels associated to the behavior of the capsule and his guests, as well as the solvent in Figure 4.4. On the top-left corner is shown the Huisgen reaction, and in the rest of the scheme we introduce some of the various aggregates that will be used in this chapter, as well as their relations. The guests $\mathbf{A}$ and $\mathbf{B}$ and the solvent are represented by respectively blue, red and black spheres. We will present in detail throughout this chapter each of the processes involved.

This system is certainly more demanding in terms of computational effort. Indeed, it is significantly larger than the curbit[6]uril macrocycle and includes 200 atoms, not counting the alkyl groups we pruned, the solvent no the guests.

In this section, we report the first results on the stability of various com-

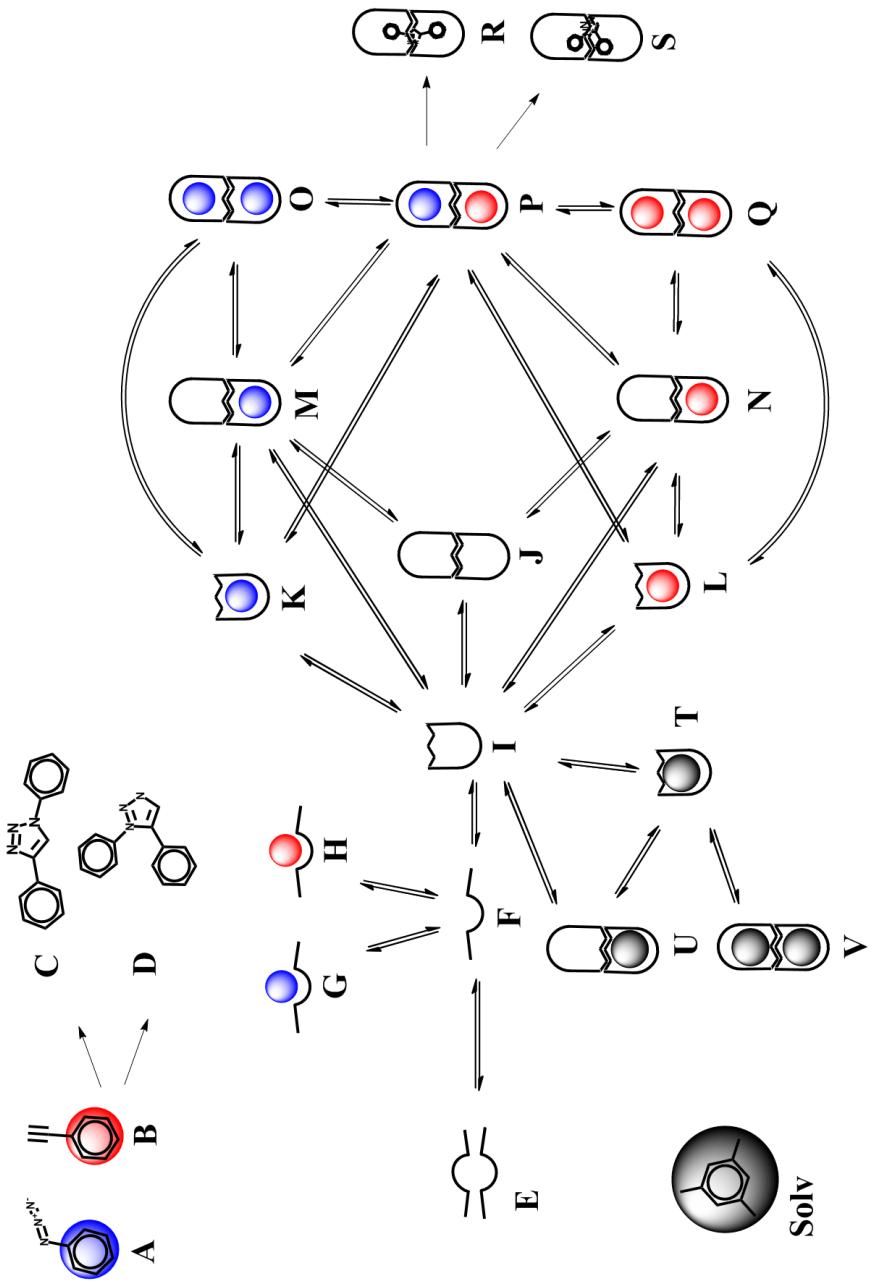



Figure 4.4: *Simplified representation of the different host-guest complexes and their relations. This scheme provides the labels use throughout this chapter.*

Table 4.1: *Stabilities of a selected set complexes involving the monomer. Labels refer to Figure 4.4. All values are in kcal/mol.*

| | **UFF** | | **MM3** | | **AM1** | |
|---|---|---|---|---|---|---|
| | ΔEsol | ΔGsol | ΔEsol | ΔGsol | ΔEsol | ΔGsol |
| **I** | 0.0 | **0.0** | 0.0 | **0.0** | 0.0 | **0.0** |
| **F** | 3.1 | **1.9** | 34.3 | **28.7** | 19.6 | **16.9** |
| **E** | -17.0 | **-21.5** | 48.5 | **56.9** | 41.2 | **51.7** |
| **K** | -24.3 | **-11.3** | -16.1 | **-5.0** | 7.6 | **19.3** |
| **L** | -25.3 | **-13.4** | -15.2 | **-1.5** | 4.6[a] | **15.3**[a] |
| **J** | 7.0 | **17.7** | -25.9 | **-4.5** | -17.4 | **10.0** |
| | **O-UFF** | | **O-MM3** | | **O-AM1** | |
| | ΔEsol | ΔGsol | ΔEsol | ΔGsol | ΔEsol | ΔGsol |
| **I** | 0.0[b] | **0.0**[b] | 0.0[c] | **0.0**[c] | 0.0[d] | **0.0**[d] |
| **F** | 3.1[b] | **1.9**[b] | 34.3[c] | **28.7**[c] | 19.6[d] | **16.9**[d] |
| **E** | -17.0[b] | **-21.5**[b] | 48.5[c] | **56.9**[c] | 41.2[d] | **51.7**[d] |
| **K** | -24.9 | **-11.2** | -15.8 | **-4.9** | 17.1[a] | **35.5**[a] |
| **L** | -22.9 | **-10.5** | -14.1 | **-3.9** | 16.6[a] | **30.5**[a] |
| **J** | 7.0[b] | **17.7**[b] | -25.9[c] | **-4.5**[c] | -17.4[d] | **10.0**[d] |
| | **B3LYP** | | **PBE** | | **M06** | |
| | ΔEsol | ΔGsol | ΔEsol | ΔGsol | ΔEsol | ΔGsol |
| **I** | 0.0 | **0.0** | 0.0 | **0.0** | 0.0 | **0.0** |
| **F** | -2.4 | **-4.3** | -2.6 | **-4.6** | 1.1 | **-5.9** |
| **E** | 10.0 | **23.8** | 4.6 | **19.4** | -9.6 | **7.8** |
| **K** | 4.7 | **15.5** | 4.7 | **16.4** | -12.1 | **5.5** |
| **L** | 2.3 | **12.5** | 3.1 | **14.4** | -9.6 | **6.9** |
| **J** | -24.8 | **0.3** | -29.1 | **-4.0** | -30.8 | **-9.8** |
| | **M06$_{sol}$** | | **B97D** | | **B97D$_{sol}$** | |
| | ΔEsol | ΔGsol | ΔEsol | ΔGsol | ΔEsol | ΔGsol |
| **I** | 0.0 | **0.0** | 0.0 | **0.0** | 0.0 | **0.0** |
| **F** | 1.1 | **-2.6** | -0.9 | **-7.2** | -0.7 | **-5.2** |
| **E** | -11.9 | **10.4** | -39.8 | **-25.4** | -41.7 | **-22.8** |
| **K** | -13.8 | **5.2** | -15.7 | **-2.9** | -16.8 | **-4.0** |
| **L** | -7.5 | **10.6** | -14.0 | **0.0** | -15.0 | **-2.7** |
| **J** | -31.0 | **-5.4** | -34.6 | **-15.6** | -34.0 | **-17.4** |

a: This value is only for comparison and corresponds to a single-point (see text).
b, c, d: only the supermolecule appears. The value respectively refers to **UFF**, **MM3** and **AM1**.

binations of hosts and guests. They are given in Tables 4.1 to 4.4.

### 4.2.1 Vase to Kite Equilibrium

The first results are focused on the vase-kite equilibrium. Vase (**I** ) and kite (**F**) only differ by the conformation of their of their eight-membered rings,

see Figure 4.5.

Vase-kite interconversions have been studied bu Cram[141] on similar cavitands. Therefore we know an equilibrium takes place, but we do not know the exact relative stabilities. We address this question with different methods, and propose the results in Table 4.1

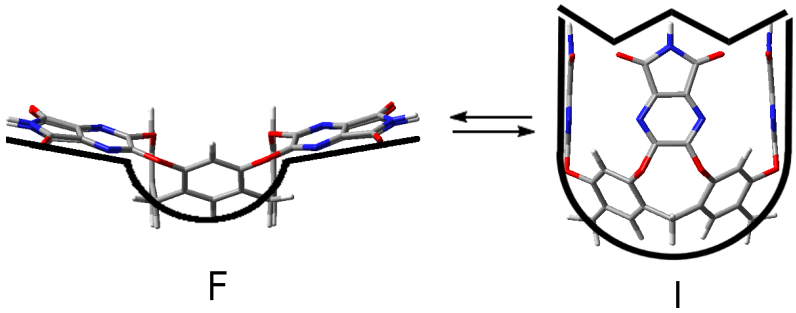



Figure 4.5: *Molecular model representing the ability of the monomer to open its gates. Schematic representations are superimposed in black bold lines. Left: kite form. Right: vase form.*

The stability of the kite monomer **F** is presented relative to the stability of a vase monomer **I** . **F** is less stable than **I** according to all non-DFT methods used. Values range from 1.9 to 28.7 kcal/mol, no agreement can be found among those methods.

On the contrary, all DFT methods tested in this study agree on a stabilization in the narrow range -2.6 to -7.2 kcal/mol for respectively $\mathbf{M06}_{\mathrm{sol}}$ and **B97D** in favor of the kite **F**. This change in free energy is mostly due to entropy gain as well as the increased surface of solvation of this conformation.

### 4.2.2 Kite - Kite interaction

Given that the kite form is more stable than the vase conformation, one could go further and propose a compound made of two interacting monomers that would not form a capsule. Instead, they would collapse in a rather flat agglomerate (**E**), see Figure 4.6. A crystal structure obtained experimentally on a related compound[143] supports this idea.

A dozen of geometries have been tested (with **B3LYP**, **PBE**, **B97D**, and **M06**, not reported here). It appears clearly that a flat, face-to-face arrangement is favored over all other possibilities we tried, see Figure 4.7. Noticeably, the crystal structure also displays the same arrangement, although the system was slightly different. This organization has been kept and subsequently tested with every method across the benchmark.

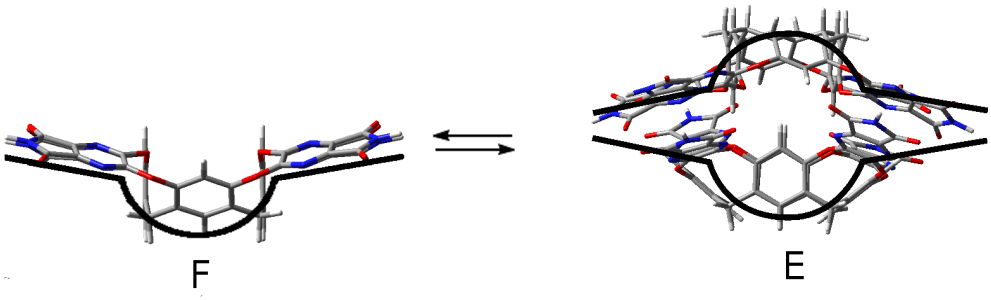



Figure 4.6: *Representation of the equilibrium taking place between two kite-like monomers.*

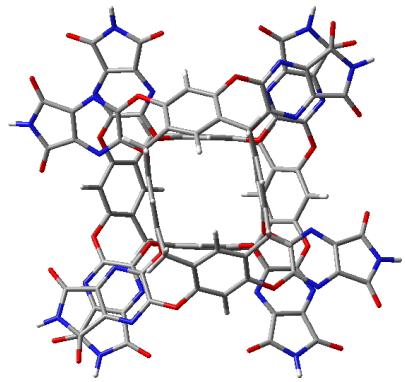


Figure 4.7: *Top-view of an aggregate of two monomers: the kite-kite complex. A face-to-face arrangement is favored while the monomers are rotated by almost 90 degrees.*

To obtain the relative stabilities of **I** and **E** for example, we have to take the difference of their relative energies in Table 4.1: $\Delta E_{sol}(\mathbf{E}) - \Delta E_{sol}(/\mathbf{I}_{-})$ for the enthalpy of solvation, and $\Delta G_{sol}(\mathbf{E}) - \Delta G_{sol}(\mathbf{I})$ for the free energy of solvation.

Intriguingly, **UFF** predicts **E** to be favored over **I** , when **F** is considered unstable. **UFF** considers non-bonded interactions to be strong enough to overcome an energetically unfavored structure, significant entropic losses as well as reduction of solvation energy in the formation of **E**.

For **MM3** and **AM1** the energy rises by respectively 56.9 and 51.7 kcal/mol. The enthalpy plays the major role in this phenomenon as the structural distortion rises the energy.

DFT functionals also give interesting results. Remember they all agree on **F** being moderately more stable than **I** , but for the closed kite-kite structure **E**, **B3LYP** (+23.8 kcal/mol), **PBE** (+19.4 kcal/mol), **M06**(+7.8

kcal/mol) and $\mathbf{M06}_{\mathrm{sol}}$(+10.4 kcal/mol) show an unstable aggregate. In contrast, **B97D** and $\mathbf{B97D}_{\mathrm{sol}}$ predict a very favorable aggregation process ($2\mathbf{I} \rightarrow \mathbf{E}$): respectively -25.4 and -22.8 kcal/mol.

It is clear that the methods we tested are in deep disagreement. We will come back to this issue in the discussion.

### 4.2.3 The Monomer as a Model of Host-Host and Host-Guest Interactions

Considering the interaction of a monomer with a single guest, experimental insights on related systems[144,145] report small aromatic compounds to bind the host in organic solvent. Also, the capsule has a strong ability to trap even impurities, *i.e.* solutes found in very low concentrations.[140] We therefore expect a stable interaction between the monomer and the guests.

We intend first to analyze the results provided by our calculations on the monomer in order to distinguish more clearly the numerous contributions to complexation: solvation, entropy and enthalpy. This will later on allow us to understand better the larger aggregates we will tackle.

Either of the guests **A** or **B** can enter the cavity of the monomer **I** to form respectively **K** and **L**. The vase-like monomer **I** can also dimerize to form **J**, see Figure 4.8.

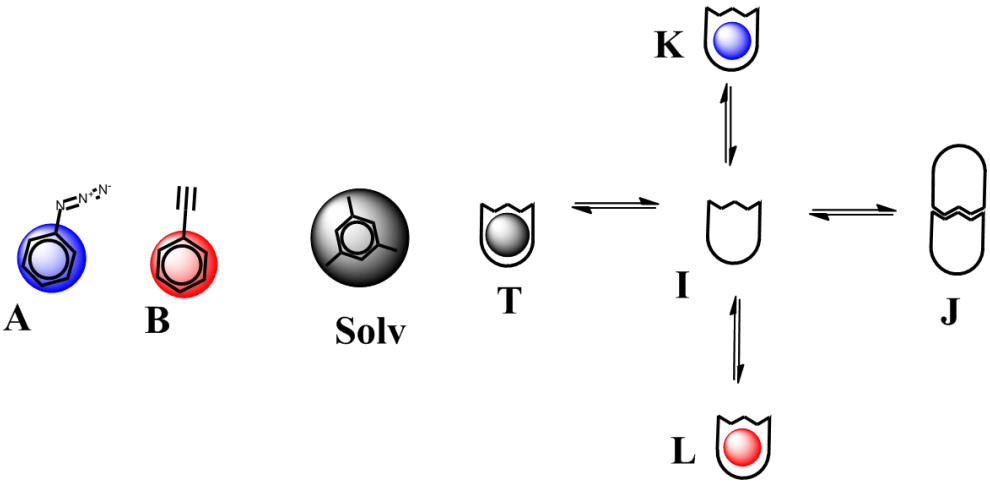



Figure 4.8: *Representation of different aggregates involving the vase-like monomer. The guests* **A** *and* **B** *are depicted by circles that can interact with the host cavity.*

The energetics are included in Table 4.1. It is striking to note that for **AM1, O-AM1, B3LYP** and **PBE**, enthalpic contributions do not favor

the inclusion of a single guest in the monomer ($\mathbf{I} + \mathbf{A} \rightarrow \mathbf{K}$ and $\mathbf{I} + \mathbf{B} \rightarrow \mathbf{L}$). For **AM1** and **O-AM1** the guest is simply ejected from the monomer during the optimization. The energies we propose are the results of a single point where the guest $\mathbf{A}$ or $\mathbf{B}$ would still be located inside the monomer $\mathbf{I}$. Therefore they are not fully optimized and values are approximate.

**UFF** and **O-UFF** favor complex $\mathbf{K}$ and $\mathbf{L}$ in a tight range between -10.5 and -13.4 kcal/mol.

According to the **B3LYP** and **PBE** methods, the inclusion of a guest is disfavored, due to the absence of enthalpic stabilization, as well as the loss of entropy and the cost of partial desolvation of both host and guests.

**M06** and $\mathbf{M06}_{\mathrm{sol}}$ predict a strong enthalpic interaction between $\mathbf{I}$ and the guests, between -7.5 and -13.8 kcal/mol. However free energy of solvation are ultimately in disfavor of the aggregation, with free energies ranging from 5.2 to 10.6 kcal/mol. **B97D** and $\mathbf{B97D}_{\mathrm{sol}}$ are the only DFT methods to find a favored total interaction ranging from 0.0 to -4.0 kcal/mol.

Two vase monomers can aggregate to form the hydrogen-bonded capsule ($2^*\mathbf{I} \rightarrow \mathbf{J}$). In enthalpic terms, this process is thermodynamically favored for all the methods except **UFF**. Also, no guests are present so **O-UFF** gives exactly the sames results.

Although the enthalpy of formation is favorable, adding the contribution of the entropy changes the sign of the interaction for **AM1** (+10.0 kcal/mol), and to a lesser extent, **B3LYP** (+0.3 kcal/mol).

**MM3** (-4.5 kcal/mol) and all DFT methods other than **B3LYP** (from -4.0 to -17.4 kcal/mol) favor the formation of capsule respectively to the vase form of the monomer.

### 4.2.4 Encapsulation of Guests

Rebek et al. studied the dynamic equilibrium that takes place with phenyl azide and phenyl acetylene as guests.[55] The study concluded that their relative affinities should be similar, with a few tenths of kcal/mol favoring an asymmetrically filled complex, with two reactants being inside. Furthermore, it has been reported that the product formed in the capsule remains trapped inside.

Before dealing with the dynamical character of the equilibrium observed with the guests, we propose in this section to compare the energies of several complexes of the capsule with zero, one or two guests inside, as depicted Figure 4.9.

The capsule may hold for example only one guest (compounds $\mathbf{M}$ and $\mathbf{N}$ ), or one solvent molecule ($\mathbf{U}$). Two solvent molecules form the complex $\mathbf{V}$. If two guests are inside, the complexes $\mathbf{O}$, $\mathbf{P}$ and $\mathbf{Q}$ can be formed. Finally,

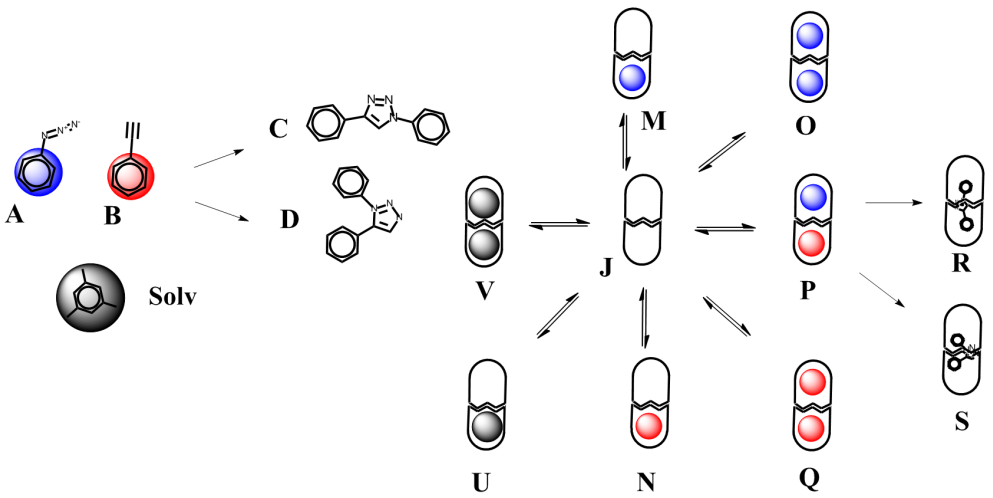



Figure 4.9: *Representation of different complexes involving a capsule.*

we address the stability of the unique product in the capsule (complex **R**). The results are shown in Tables 4.2 and 4.3.

The performance of the different methods for describing the inclusion of guests and solvent molecules into the capsule can be put in three categories. First, **UFF, MM3, O-UFF, O-MM3** as well as **B97D** and $\mathbf{B97D}_{\mathrm{sol}}$ find a host-guest interaction strong enough to overcome discouraging entropy changes that destabilize the complexes as the number of molecules increase. With **UFF** and **O-UFF**, the stabilizations are always very strong for the first guest and relatively mild for the second one. For instance with **UFF**, **M** is stabilized by -24.1 kcal/mol over **J**, but **O** lies only 4 kcal/mol below **M**. With **MM3**, **O-MM3**, **B97D** and **B97D**, the interaction is not as strong, but more balanced as the first and then the second guests are placed in the capsule. For example with **B97D**, **M** (one guest **A**) lies 5.3 kcal/mol below **J** (no guest), while **O** (two guests **A**) lies 5.7 kcal/mol below **M**.

**M06** and $\mathbf{M06}_{\mathrm{sol}}$ find a strong enthalpic stabilization for all guests but predict an increase in the free energy during this processes, due to large losses of entropy. For instance, the enthalpic stabilization of **N** is evaluated at -10.0 kcal/mol, but the free energy change is unfavorable: +1.3 kcal/mol.

**AM1**, **O-AM1**, **B3LYP** and **PBE** do not predict any advantageous enthalpic interaction upon inclusion of guests. The situation gets even worse between the first and the second guest. For example with **O-AM1** and **PBE**, **M** lies respectively 11.3 and 13.5 kcal/mol above **J**, while **O** lies respectively 21.7 and 20.0 kcal/mol above **M**.

Inclusion of one and two explicit mesitylene molecules are more difficult

Table 4.2: *Stabilities of a selected set complexes involving the full capsule, computed with molecular mechanics, semi-empirical and hybrid methods. All values are given in kcal/mol.*

| | **UFF** | | **MM3** | | **AM1** | |
|---|---|---|---|---|---|---|
| | ΔEsol | ΔGsol | ΔEsol | ΔGsol | ΔEsol | ΔGsol |
| **J** | 0.0 | **0.0** | 0.0 | **0.0** | 0.0 | **0.0** |
| **M** | -41.4 | **-24.1** | -19.2 | **-8.0** | 4.0 | **11.7** |
| **O** | -55.4 | **-28.1** | -36.1 | **-9.9** | 11.7 | **33.4** |
| **N** | -35.7 | **-20.2** | -18.4 | **-5.7** | 3.4 | **11.2** |
| **Q** | -53.5 | **-25.3** | -34.0 | **-8.0** | 13.7 | **34.0** |
| **P** | -55.3 | **-27.6** | -36.7 | **-11.7** | 12.2 | **31.2** |
| **R** | -7.1[a] | **30.5**[a] | -23.2[a] | **7.6**[a] | -28.2 | **0.2** |
| **U** | -31.3 | **-15.3** | -9.6 | **4.4** | 8.5 | **22.6** |
| **V** | -40.0 | **-5.2** | -8.1 | **30.1** | 23.2 | **50.3** |
| | **O-UFF** | | **O-MM3** | | **O-AM1** | |
| | ΔEsol | ΔGsol | ΔEsol | ΔGsol | ΔEsol | ΔGsol |
| **J** | 0.0[b] | **0.0**[b] | 0.0[c] | **0.0**[c] | 0.0[d] | **0.0**[d] |
| **M** | -39.6 | **-22.4** | -16.5 | **-5.4** | 3.9 | **11.3** |
| **O** | -64.1 | **-33.7** | -34.1 | **-8.0** | 9.8 | **31.4** |
| **N** | -35.6 | **-20.4** | -18.4 | **-7.1** | 3.4 | **10.9** |
| **Q** | -51.1 | **-24.7** | -28.5 | **-3.4** | 14.1 | **33.0** |
| **P** | -54.9 | **-26.9** | -27.7 | **-1.9** | 11.3 | **31.5** |
| **R** | -119.8 | **-84.1** | -100.7 | **-68.2** | -60.4 | **-32.5** |
| **U** | -34.6 | **-12.6** | -9.2 | **1.2** | 8.9 | **22.3** |
| **V** | -53.1 | **-15.5** | 0.3 | **29.1** | 24.3 | **49.7** |

a: Inconsistency due to the chemical reaction resulting in atom type change.
b, c, d: only the supermolecule appears , so the number respectively refers to a pure MM/SE calculation.

to rationalize, as no real trend can be found, probably due to their particular bulkiness. The inclusion of one or two solvent molecules in the capsule, **U** and **V** respectively, is destabilizing according to all methods except **UFF** and **O-UFF**.

Within the scope of a given method, the general trend of guest inclusion is the following: when the interaction is predicted to be stabilizing with one guest, the system becomes even more stable with a second guest, and vice versa.

### 4.2.5 Formation of the Product in the Capsule

Finally, the formation of the only observed product $\mathbf{P}\_ \rightarrow \_\mathbf{R}\_$ results in an energy change that ranges from -51.8 kcal/mol for **B97D** to -71.8 kcal/mol for **PBE**. The three notorious exceptions are **AM1**, with -31.1 kcal/mol

Table 4.3: *Relative energies of a selected set of complexes involving the full capsule computed using DFT methods. All values are given in kcal/mol.*

| | **B3LYP** | | **PBE** | | **M06** | |
|---|---|---|---|---|---|---|
| | ΔEsol | ΔGsol | ΔEsol | ΔGsol | ΔEsol | ΔGsol |
| **J** | 0.0 | **0.0** | 0.0 | **0.0** | 0.0 | **0.0** |
| **M** | 7.5 | **17.7** | 3.4 | **13.5** | -13.5 | **5.4** |
| **O** | 21.8 | **43.4** | 10.6 | **33.5** | -32.8 | **7.2** |
| **N** | 9.5 | **19.0** | 5.1 | **16.3** | -8.9 | **8.9** |
| **Q** | 20.8 | **42.2** | 9.2 | **32.7** | -22.3 | **14.8** |
| **P** | 19.9 | **41.5** | 11.2 | **33.5** | -25.3 | **10.7** |
| **R** | -46.7 | **-17.0** | -67.8 | **-38.3** | -87.7 | **-42.4** |
| **U** | 14.6 | **25.9** | 11.1 | **21.5** | -3.0 | **19.5** |
| **V** | 27.1 | **52.3** | 26.5 | **55.0** | -12.3 | **24.3** |
| | **M06$_{sol}$** | | **B97D** | | **B97D$_{sol}$** | |
| | ΔEsol | ΔGsol | ΔEsol | ΔGsol | ΔEsol | ΔGsol |
| **J** | 0.0 | **0.0** | 0.0 | **0.0** | 0.0 | **0.0** |
| **M** | -15.1 | **1.3** | -19.7 | **-5.3** | -20.6 | **-3.1** |
| **O** | -33.5 | **3.1** | -38.8 | **-11.0** | -39.8 | **-8.3** |
| **N** | -10.0 | **6.6** | -17.2 | **-2.7** | -18.0 | **-2.4** |
| **Q** | -22.4 | **13.1** | -31.7 | **-3.8** | -32.8 | **-2.6** |
| **P** | -27.2 | **10.3** | -34.7 | **-7.1** | -35.9 | **-6.0** |
| **R** | -89.1 | **-46.6** | -92.6 | **-58.8** | -100.6 | **-66.8** |
| **U** | -4.3 | **16.0** | -11.0 | **3.3** | -12.2 | **3.5** |
| **V** | -13.1 | **21.4** | -27.5 | **6.9** | -29.8 | **0.5** |

and the molecular mechanics methods: **UFF** and **MM3**, which are unable to model a chemical reaction with bond formation, as discussed earlier in this work.

### 4.2.6 Basis Set Superposition Error

Basis set superposition errors are incorporated in the results presented Tables 4.1 to 4.3. The results on BSSE are summarized in Table 4.4.

Large BSSE errors are collected for all the assemblies of the set we tested. The smaller the assembly, the smaller the BSSE value and vice-versa. It appears that we can divide the results in two groups: (1) **B3LYP** and **PBE** with slightly smaller BSSE and (2) **M06, M06$_{sol}$, B97D, B97D$_{sol}$** with larger BSSE.

Table 4.4: *Basis Set Superposition Errors computed for a set of species. Values in kcal/mol.*

| Compound | **B3LYP** | **PBE** | **M06** | **M06$_{sol}$** | **B97D** | **B97D$_{sol}$** |
|---|---|---|---|---|---|---|
| **E** | 14.9 | 16.2 | 23.5 | 22.0 | 25.2 | 23.9 |
| **K** | 6.2 | 7.9 | 11.4 | 10.8 | 11.9 | 11.0 |
| **L** | 6.0 | 6.7 | 11.3 | 8.6 | 9.6 | 9.1 |
| **J** | 16.2 | 14.3 | 14.4 | 14.3 | 14.9 | 14.5 |
| **M** | 24.5 | 22.3 | 26.3 | 26.1 | 27.3 | 26.9 |
| **O** | 35.5 | 32.6 | 39.5 | 39.5 | 40.3 | 40.0 |
| **N** | 24.5 | 22.4 | 25.7 | 25.3 | 27.1 | 26.6 |
| **Q** | 33.4 | 29.9 | 38.0 | 37.8 | 39.6 | 38.8 |
| **P** | 34.7 | 32.3 | 37.5 | 37.3 | 40.8 | 40.2 |
| **R** | 38.0 | 34.5 | 40.9 | 42.6 | 40.5 | 32.4 |
| **U** | 24.4 | 22.1 | 27.6 | 27.1 | 28.2 | 27.7 |
| **V** | 28.8 | 30.1 | 36.0 | 35.4 | 40.8 | 37.7 |

### 4.2.7 Summary of the Results Obtained by Different Methods and Overview of the Behavior of the Capsule

We reviewed the treatment offered by different families of methods: MM, semi-empirical, ONIOM, DFT and DFT-D. We gathered the different relations established in the previous sections in Figure 4.4. The behavior of the capsule and its guests in solution emerges as rather a complex scheme. It should be noted that this network is simplified and other relations may exist.

Given that all the results are difficult to visualize at once, we propose to select among them a specific path: **E**, **F**, **I** , **J**, **N**, **P** and finally **R**. This paths compiles all the critical points that should be computationally modeled adequately, see Figure 4.10. The capabilities of each methods are discussed in the following section.

## 4.3 Discussion on the Methods

We gather in this section remarks relative to each method or category of methods. As in the previous chapter, they have their advantages and drawbacks that the computational chemist has to take into consideration. Inevitably some trends will be similar, so we will go quickly on the general remarks and focus on the specificities of this system.

We point out that although solvation energies are reported with force fields and ONIOM methods, the corrections we proposed for the solvation effects are performed using DFT after converging the geometry, as explained in Chapter 2.

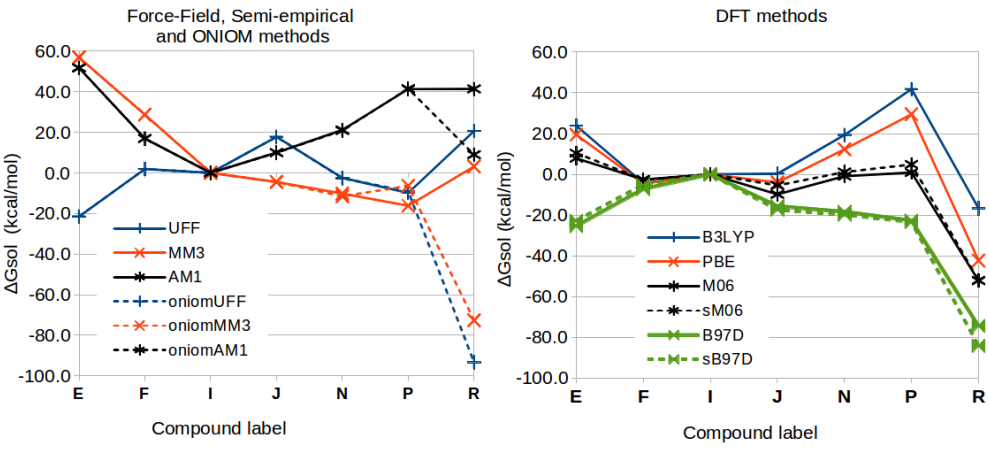



Figure 4.10: *Summary of results: relative stabilities calculated for a selected path:* $\mathbf{E} \rightarrow \mathbf{F} \rightarrow \mathbf{I} \rightarrow \mathbf{J} \rightarrow \mathbf{N} \rightarrow \mathbf{P} \rightarrow \mathbf{R}$. *Transition states are not included.*

### 4.3.1 Force-Field Methods

#### UFF

The structures obtained with **UFF** should not be trusted "as is". For instance, this method fails at describing the formation of the capsule since it does not properly describe hydrogen bonding.

The vase-kite equilibrium $\mathbf{I} \rightarrow \mathbf{F}$ reported in experiments on similar systems gave roughly acceptable results compared with DFT. It is important to describe this equilibrium correctly since it displays structural deformation characteristic of the mechanism expected for guest-exchange, see Figure 4.3.

However, even a better agreement than the one observed would not be sufficient by itself. Indeed, we only addressed here the question of the stability of complexes, but not the question of mechanisms involved. Transition states are necessarily more strained and MM methods are less reliable further from their "best" zones of the PES, where the force field has been built. MM methods used here cannot handle chemical reactions, but even conformational transition states inevitably involve strained structures, which are less likely to be correctly modeled than the surrounding minima.

On the positive side stands the kite-kite dimerization $2\mathbf{F} \rightarrow \mathbf{E}$ is in surprisingly good agreement with DFT. Probably cancellation of errors plays a significant role here. Finally, the inclusion of reactive guests in the capsule follows a good trend, but binding energies are overestimated.

#### MM3

The approach and parameters of this force field seem relatively successful. **MM3** has built-in capabilities to treat hydrogen bonds with an angle-dependent treatment. The formation of the capsule depends to a large extent on hydrogen bonds and leads to decent energies compared to DFT methods, although its stability appears underestimated.

Apart from reactivity, the vase - kite equilibrium appears to be a major shortage of the **MM3** method.

### 4.3.2 Semi-empirical Method: AM1

Approximate geometries can be obtained as a first guess and reused for later re-optimization, with more accurate methods. Nevertheless, we strongly discourage its use for the study of the energetics of van der Waals host-guest complexes for following main reasons: (i) Positive free energy for the formation of the capsule. (ii) Incapacity of the method to reproduce accurately non-bonded interactions which hampers the correct location of the minima on the potential energy surface and results for example in the ejection of guests from the hosts (**K** and **L**). (iii) Failure to deal with distorted structures *i.e.* **F** and **V**. (iv) Imprecise reaction energies (**P** $\rightarrow$ **R**).

Looking at Figure 4.10, it appears highly unsafe to rely on **AM1**, as the trend given is in complete disagreement with expected results. Given the numerous drawbacks of this semi-empirical method, we did not investigate the possibility of adding empirical dispersion corrections, although it has been done elsewhere.[146,147] This approach could however be an easy way of producing acceptable geometries much faster than with DFT.

### 4.3.3 ONIOM Methods

ONIOM calculations can be carried out quickly for supramolecular systems. Since the high layer will demand most the computational cost, 28 atoms are fairly acceptable.

#### ONIOM(B3LYP:UFF)

Although agreement with DFT-D is sketchy, **O-UFF** finds a stabilization for all guests of the capsule. Nevertheless, we cannot advice its use till the end of a computational project, since this scheme reproduces all the inaccuracies of **UFF**, except for the description of the chemical reaction. In particular, the formation of the box is still unfavored.

#### ONIOM(B3LYP:MM3)

The results are qualitatively similar with the **MM3** method. The energies of distorted structures such as the vase-kite equilibrium are particularly untrustworthy. As expected, the only significant improvement is the better description of the click reaction.

#### ONIOM(B3LYP:AM1)

**O-AM1** model failed as **AM1** in many critical aspects, and we discourage its use for host-guests systems, particularly because of the defective or absent description of non-bonded interactions: van der Waals and hydrogen bonds. It appears that empirical dispersion and hydrogen bonds corrections (-DH schemes[148]) would not be sufficient, since the energy of distorted structures are still amply erroneous.

### 4.3.4 DFT Methods

As observed in last chapter, all DFT methods lead to similar geometries, but it appears again mandatory to take into account medium and long range interactions to obtain accurate energies for host-guest complexes.

According to our results the popular **B3LYP** as well as **PBE** can reproduce reactivity as well as distortions (vase-kite equilibrium).

However, as for CB6, **B3LYP** and **PBE** fail to describe properly encapsulation processes, while it is experimentally known to be fast in solution,[32] implying at least a good thermodynamic interaction, not mentioning yet kinetics for the formation of complexes. In short, we cannot recommend any standard functional for the study of van der Waals complexes.

**M06** functional from Zhao and Truhlar belongs to the meta-hybrid GGA class. It contains a large number of parameters which are semi-empirically fit on various data, in order to minimize errors, notably on non-covalent interactions.

On supramolecular systems those errors may stem from the differences on the chemical structures between the training set and the system studied. **M06** seems to somewhat underestimate the strength of dispersion. Interaction energies are negative but entropy reverses the pattern in all cases involving guest inclusion, which indicates an insufficient strength of non-covalent binding. These errors are likely amplified because of the large surfaces of contact that supramolecular structures display.

We note that some concerns have been raised by Houk et al.[149] related to the quality of the integration grid as a source of errors with **M06**. Test

calculations proposed Table 4.5 show that free energies of complexation, corrected by solvent effects (ΔGsol) remains positive, although they are lower than with the standard, smaller grid. In this context and considering the significant extra-cost of more precise integration grids, we did not extend those tests to the whole set.

Table 4.5: *Refined values for the stability of several complexes obtained with* **M06**, *when using a more precise integration grid. Values include also BSSE and are presented in kcal/mol.*

| | **M06 (smaller grid)** | | **M06 (larger grid)** | |
|---|---|---|---|---|
| | ΔEsol | ΔGsol | ΔEsol | ΔGsol |
| **I** | 0.0 | **0.0** | 0.0 | **0.0** |
| **K** | -12.1 | **5.5** | -11.8 | **1.2** |
| **L** | -9.6 | **6.9** | -10.0 | **4.5** |
| **J** | -30.8 | **-9.8** | -30.6 | **-12.6** |

**B97D** is a semi-empirically corrected GGA-type functional. The strategy used here relies on the correction of the existing B97 functional.[96] It is explicitly parametrized by including atom-pairwise dispersion corrections, that work as a simple pair-wise force field.

All DFT methods consider the kite conformer more stable than the vase conformer. However, **B97D** and $\mathbf{B97D}_{\mathrm{sol}}$ are the only ones to allow the formation of the flat kite-kite aggregate **E**. Also, a proper stabilization of the guest is produced, as well as capsule formation, distorted structures and reactivity.

### 4.3.5 Solvent Optimization and Solvent Effects

The solvent optimizations ($\mathbf{B97D}_{\mathrm{sol}}$ and $\mathbf{M06}_{\mathrm{sol}}$) result in limited influence in both of systems studied. Structural changes are not significant, compared to their gas phase analogs, respectively **B97D** and **M06**. Optimizing in solution has an extra-cost that is not negligible.

An implicit solvation model is used, but we underline here that the supramolecular capsule solvent (mesitylene) has been purposely chosen not to fit inside. The deformed and destabilized structures **U** and **V** obtained in our calculations confirms the validity of the implicit model.

### 4.3.6 Basis Set Superposition Error

The BSSE we obtain is very large in all cases. Errors show mostly dependence on the size of the overlapping area of the different fragments, but not on the

functional used. For the study of smaller systems, one should consider larger basis sets, that would reduce the BSSE.

### 4.3.7 Conclusion on the Methods

We reviewed the performance of various families of methods. A least a hybrid scheme is mandatory for the reaction step. In this respect, although force fields such as **MM3** are promising, they do not seem reliable, particularly regarding conformation changes. Semi-empirical methods appear to be a very poor choice. Finally conventional DFT fails at describing van der Waals complexes.

**B97D** has been identified as the best method capable of ensuring accurate and exploitable results. We draw the same conclusion as for the last chapter and will use **B97D** for the rest of the study.

## 4.4 Dynamical Behavior of The Capsule

We continue the study started from the screening of methods, focusing on the results obtained with **B97D**, which will be further elaborated. As in the previous chapter, we want to gather information on various systems related to the capsule, as well as the relations that bind them. We will subsequently use the thermodynamic data of the complexes to produce a meaningful kinetic model that reproduces the experimentally observed behavior and the interplay of the species in solution.

In this section, we will intend to apprehend the dynamical character of the supramolecular capsule. As shown Figure 4.4, the interplay of the dimeric capsule and its guests is rather complex. A double competition takes place: the guests compete to enter the host until the reaction takes place in a capsule. In parallel, the capsule itself is in competition with a stable aggregate that does not leave room for guests.

### 4.4.1 Theoretical Study: Computing Mechanisms from Experimental Data

Some experimental insights and data gathered in the literature helps us to understand the behavior of the system and guides our investigation. We propose some of the most relevant ones here.[140,142,150]

First of all, the behavior of the capsule depends on the solvent and the guests used. For example, in protic solvent with large guests, a complete dissociation is necessary to exchange guests.

It has been observed that the capsule does not form in absence of suitable guests. Mesitylene has been chosen not to fit inside the capsule, in order to leave room for the Huisgen reaction and to facilitate NMR experiments at the same time. According to experiments, the mechanism of guest exchange does not require the complete dislocation of the capsule. A kinetic study is available and provides the rates of guest exchange. This study has been undertaken with p-xylene as the encapsulated guest and benzene as the replacing guest.

A possible interpretation of experimental insights is the following: without guests, the capsule does not form and collapses in an aggregate. On the other hand, when proper guests are present the capsule may form.

### 4.4.2 Gate Opening

To help us understanding the formation of the capsule, we start this section by exploring the interconversions of vase $\mathbf{I}$ and kite $\mathbf{F}$ monomers. To do so, we look for the relative stabilities and energy barriers of intermediates configurations.

A vase conformer $\mathbf{I}$ has four gates in a closed position, that is, the arrangement of each eight-membered rings make the gates point towards an axial direction. In a kite conformer, the arrangement of each eight-membered ring makes the gates point toward an equatorial direction. An interconversion between a vase $\mathbf{I}$ and a kite conformer $\mathbf{F}$ requires four steps, one step for each gate to open. A representation of the opening/closing of the fourth gate is shown Figure 4.11. The results we obtain are presented Table 4.6. In this table, the subscripts refer to intermediary conformations. For instance if one gate is open, we use $\mathbf{I}_{1\mathrm{op}}$, and if two gates are open we use $\mathbf{I}_{2\mathrm{op}}$, etc.

The energies of the first three TS are low and relatively similar. As we saw earlier, the kite form $\mathbf{F}$ of the monomer is more stable by -7.2 kcal/mol. The forward ($\mathbf{I} \rightarrow \mathbf{F}$) transition energy is evaluated at 7.1 kcal/mol, while the backwards transition energy is 14.3 kcal/mol for a monomer.

Simultaneous change of conformation on two or four sites of the monomer did not reveal to be an acceptable alternative. We obtained forward activation barriers of respectively 11.4 kcal/mol and 29.6 kcal/mol.

We remind the reader that the formation of supramolecular dimeric aggregates $\mathbf{J}$ and $\mathbf{E}$ displays the following stabilities. The capsule $\mathbf{J}$ is favored by -15.6 kcal /mol relative to separated vase monomers, while the kite-kite complex $\mathbf{E}$ displays a more consequent stabilization : -25.4 kcal/mol.

With these results, we are able to explain why the capsule does not form in the absence of guests. The vase-kite equilibrium is directed towards the kite conformation $\mathbf{F}$. Also, the superior stability of the kite-kite complex $\mathbf{E}$

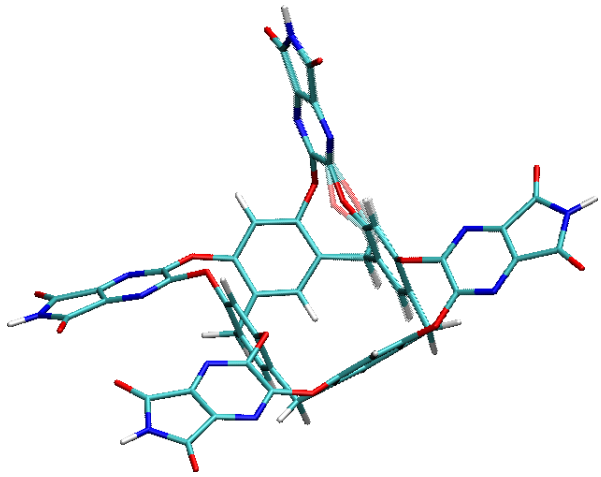


Figure 4.11: *Representation of the imaginary frequency associated to the opening/closing of the fourth gate:* $\mathbf{I}_{TS4}$. *On the structure of the transition state are superimposed the reactant and product immediately preceding/following the transition state.*

Table 4.6: *Energetics for the vase to kite interconversion mechanism* $\mathbf{I} \rightarrow \mathbf{F}$, *and of the formation of the kite - kite aggregate* **E**, *in kcal/mol. The formation of the capsule* **J** *is for comparison*

| complex | $\Delta E_{sol}$ | $\Delta G_{sol}$ |
|---|---|---|
| **I** | 0.0 | **0.0** |
| $\mathbf{I}_{TS1}$ | 7.7 | **7.1** |
| $\mathbf{I}_{1op}$ | 3.2 | **-0.4** |
| $\mathbf{I}_{TS2}$ | 8.8 | **6.4** |
| $\mathbf{I}_{2op}$ | 2.3 | **-0.6** |
| $\mathbf{I}_{TS3}$ | 10.4 | **5.2** |
| $\mathbf{I}_{3op}$ | 2.0 | **-3.8** |
| $\mathbf{I}_{TS4}$ | 7.7 | **2.3** |
| **F** | -0.9 | **-7.2** |
| **E** | -39.8 | **-25.4** |
| **J** | -34.6 | **-15.6** |

would actually prevent the formation of the capsule. It seems that one of the role of the guests is to stabilize the capsule in order to be able to compete with the flat kite-kite aggregate **E**.

### 4.4.3 Basic Aggregates

Table 4.7 shows the energies of guests **A**, **B** and **Solv** forming a complex with vase (**I** ) or kite (**F**) monomer. **A** and **B** can be trapped in the wrong orientation, that is the phenyl group pointing outside the bottom of the cavity. We call this wrong orientation "upside-down".

Table 4.7: *Energetics of a vase monomer as well as kite interacting with a guest. Energies are given in kcal/mol relative to separate species.*

| complex | $\Delta E_{sol}$ | $\Delta G_{sol}$ | complex | $\Delta E_{sol}$ | $\Delta G_{sol}$ |
|---|---|---|---|---|---|
| **I** | 0.0 | **0.0** | **F** | -0.9 | **-7.2** |
| **K** | -15.7 | **-2.9** | **G** | -8.2 | **-3.8** |
| **L** | -14.0 | **0.0** | **H** | -8.0 | **-3.4** |
| $\mathbf{K}_{\text{guest upside down}}$ | -14.2 | **-2.1** | | | |
| $\mathbf{L}_{\text{guest upside down}}$ | -13.7 | **-0.6** | | | |
| **T** | -20.1 | **-3.9** | | | |

The vase monomer (**I**) interacts strongly with every guest : from -13.7 kcal/mol for an upside-down phenylazide ($\mathbf{K}_{\text{guest upside down}}$) to -20.1 kcal/mol for mesitylene (**T**). However degrees of freedom are lost across all the complexes, explaining the weaker thermodynamic stabilities, see Figure 4.12.

We saw that a kite conformer is thermodynamically more stable than the vase conformer by 7.2 kcal/mol. However the interaction between a the kite conformer each of the guests is significantly reduced. This results in **G** and **H**, which are respectively 3.5 kcal/mol and 3.9 kcal/mol above separated species. This means they are unable to bind, although an interaction exists.

### 4.4.4 Encapsulation Energies

It has been reported that, depending on the type of guests, this capsule can simultaneously enclose one,[151] two[152,153] or three[154–156] guests in its core. The space inside the capsule has an elongated shape, thus each guest reactive groups can point towards the center or the extremities. We have two guests, each of which can have two possible orientations. So this represents 14 possible combinations of guests and orientations, including **M** **N**, **O**, **P** and **Q**. We also present the energies resulting from the encapsulation of one **U** or two **V** solvent molecules.

In Table 4.8 the stabilities of these structures are presented respect to the empty capsule.

In terms of free energy, a single guest **A** or **B** stabilizes the complex by respectively -5.4 and -2.7 kcal/mol, which is a bit more than what we observed

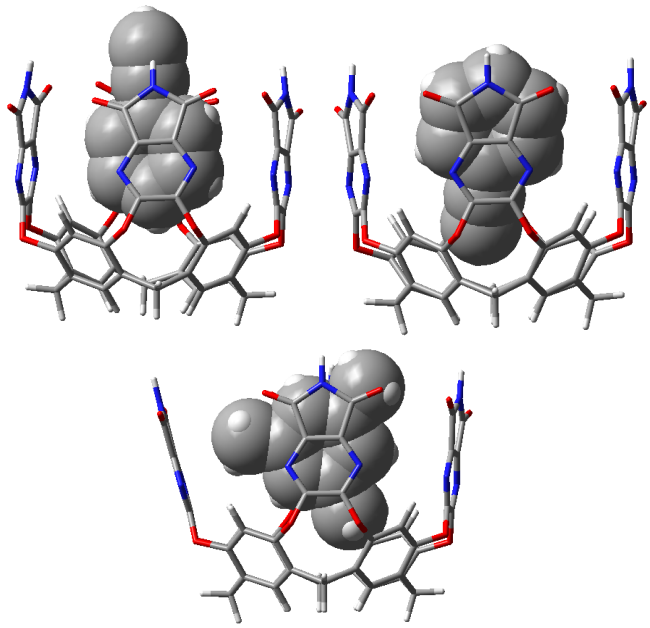


Figure 4.12: *Optimized structures of the monomer with guest* **B** *in the proper orientation (left:* **L***), with* **B** *upside down (middle:* $\mathbf{L}_{\text{guest upside down}}$*) or a mesitylene solvent molecule (right). Hosts: stick model. Guests: Space-filling model.*

in the case of the monomer (which was between 0.0 and -2.9 kcal/mol), apparently for enthalpic reasons. The shape of the capsule may be more suitable to the reactants **A** and **B** than the shape of the monomer. When we consider the encapsulation of two identical guests, the system almost doubles the stabilization respect to the stabilization occurring for the encapsulation of a single guest **A** or **B** . The most stable complex is **O**, which displays a comfortable stabilization of -11.0 kcal/mol, while **Q** displays a modest -3.8 kcal/mol. Guests **A** and **B** can be combined inside the capsule (**P**). The thermodynamic stabilization relative to the empty capsule is -7.1 kcal/mol.

A single guest **A** or **B** "upside down" in the capsule (respectively noted $\mathbf{M}_{\text{guest upside down}}$ and$\mathbf{N}_{\text{guest upside down}}$), is 5.0 and 1.3 kcal/mol above the "standard" orientation.

Finally, a mesitylene molecule also interacts strongly with the capsule. Nevertheless, the formation of this host-guest complex is disfavored by 3.3 kcal/mol, for entropic reasons. When two molecules of solvent are used a consequent enthalpic stabilization of -27.5 kcal/mol is found, however the final thermodynamic value is positive, 6.9 kcal/mol. This numbers show a

Table 4.8: *Energetics of the capsule and of several host-guest complexes, relative to separate species, in kcal/mol. Upside down guests are made explicit by the use of subscripts.*

| | ΔEsol | ΔGsol |
|---|---|---|
| **J** | 0.0 | **0.0** |
| **M** | -19.9 | **-5.4** |
| **N** | -17.2 | **-2.7** |
| $\mathbf{M}_{\text{guest upside down}}$ | -14.2 | **-0.4** |
| $\mathbf{N}_{\text{guest upside down}}$ | -13.1 | **-1.4** |
| **O** | -38.8 | **-11.0** |
| **Q** | -31.7 | **-3.8** |
| **P** | -34.7 | **-7.1** |
| $\mathbf{O}_{\text{one guest upside down}}$ | -27.1 | **0.9** |
| $\mathbf{O}_{\text{both guests upside down}}$ | -30.3 | **-1.9** |
| $\mathbf{Q}_{\text{one guest upside down}}$ | -28.6 | **-3.3** |
| $\mathbf{Q}_{\text{both guests upside down}}$ | -24.4 | **2.4** |
| $\mathbf{P}_{\text{guest A upside down}}$ | -28.4 | **-3.7** |
| $\mathbf{P}_{\text{guest B upside down}}$ | -29.5 | **-5.1** |
| $\mathbf{P}_{\text{both guests upside down}}$ | -24.7 | **1.2** |
| **U** | -11.0 | **3.3** |
| **V** | -27.5 | **6.9** |

good agreement with the experimental observation of the absence of encapsulation of the solvent.

The results we present show that most of the combinations of guests are favorable in thermodynamic terms, compared to an empty capsule. The most stable configurations are formed by $\mathbf{O}$, followed by $\mathbf{P}$, $\mathbf{P}_{\text{guest B upside down}}$. All those three complexes are capable of internal hydrogen bonding between the guests, while $\mathbf{Q}$ is unable of doing so and is less stable. Looking at the trends, we can see the better compatibility of the benzyl moiety for the bottom of the capsule.

In the reference experimental paper by Rebek,[55] the best results where obtained with concentrations of $\mathbf{A}$, $\mathbf{B}$ , and $\mathbf{I}$ respectively of 25mM, 50 mM, 5 mM. Impurities from the solvent such as benzene and p-xylene, are reported to be *c.a* 7mM.

The concentration of $\mathbf{B}$ is the double of the concentration of $\mathbf{A}$. The same strategy was used in the previous chapter, as one guest fits better than the other. One should understand the stabilities of $\mathbf{O}$, $\mathbf{P}$ and $\mathbf{Q}$ as a competition described in Equations 4.1 and 4.2. $\mathrm{K}$ is an equilibrium constant, experimentally evaluated at 9± 3.

$$\mathbf{O} + \mathbf{Q} \overset{K}{\rightleftharpoons} 2 * \mathbf{P} \tag{4.1}$$

$$K_D = \frac{[\mathbf{P}]^2}{[\mathbf{O}] * [\mathbf{Q}]} \tag{4.2}$$

If we sum the stabilities of the complexes on the left of equation 4.1 ($\mathbf{O}$ and $\mathbf{Q}$), and compared them to the (weighted) stability of $\mathbf{P}$, we get an idea of why this equilibrium is taking place.

Table 4.9: *Weighted stabilities of the three complexes in dynamic equilibrium:* $\mathbf{O} + \mathbf{Q} \overset{\mathrm{K}}{\rightleftharpoons} 2 * \mathbf{P}$. *The empty capsule represents the reference. Values in kcal/mol.*

| | ΔGsol |
|---|---|
| **J** | **0.0** |
| 1* **O** + 1***Q** | **-14.8** |
| 2* **P** | **-14.1** |

As shown Table 4.9, it seems that the method **B97D** does not favor the hetero-dimer (by only 0.7 kcal/mol) but accounts very well for the equilibrium observed, as the authors find the "unsymmetrical complex" is "the most abundant species" and suggest a difference of "the preference for the homo-dimer is only a few tens of kcal/mol".[55] The precision of the method may be responsible for the wrong ordering of energies but is very reasonable considering the substantial size of the system. Furthermore, the equilibrium is "pushed" towards $\mathbf{Q}$ and $\mathbf{P}$, as larger concentrations of guest $\mathbf{B}$ than guest $\mathbf{A}$ are used.

### 4.4.5 The Click Reaction in Solution and in the Capsule

The azide $\mathbf{A}$ and the alkyne $\mathbf{B}$ may react in the solvent or inside the capsule. Figure 4.13 shows the two possible products in solvent. In Table 4.10 we present the energetics for the mechanism of the cycloaddition.

In solution, separated species are more stable than the adducts. We get a free energy of activation of 27.1 kcal/mol for Route 1 and 25.0 kcal/mol for Route 2. In the capsule, as explained before trapped reactants in close contact form stable aggregates. The transition states are 18.7 and 34.2 kcal/mol higher than the adducts for respectively Route 1 (1,4 product) and Route 2 (1,5 product).

The capsule has a strong binding power which brings guests in close contact and in the proper orientation. The entropic cost of the cycloaddition reaction is reduced rather than the enthalpic cost. On Route 1 we can compare the potential energies of activation. They are evaluated to be 16.5

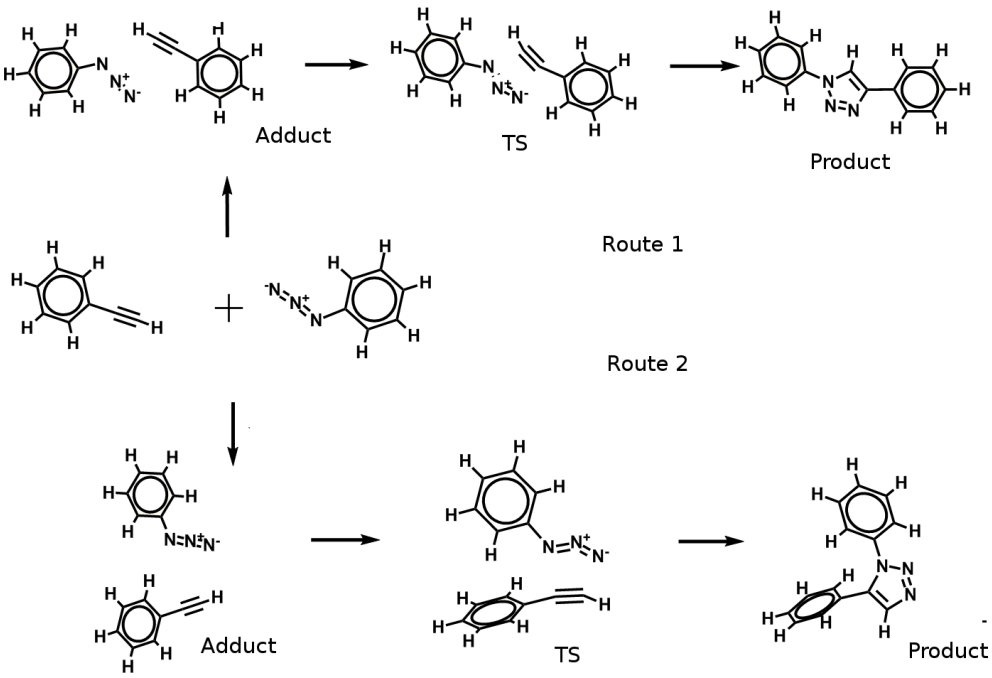



Figure 4.13: *Representation of the two possible routes for the Huisgen cycloaddition in solvent.*

Table 4.10: *Energies for the cycloaddition reaction path, whether in the solvent (left) or in the capsule (right). Energies in kcal/mol.*

| In solution | | ΔEsol | ΔGsol | In the capsule | | ΔEsol | ΔGsol |
|---|---|---|---|---|---|---|---|
| **A+B** | | 0.0 | **0.0** | **J+A+B** | | 0.0 | **0.0** |
| **Adduct** | | -2.1 | **4.5** | **P** | | -34.7 | **-7.1** |
| **TS** | Route1 | 16.5 | **27.1** | **TS** | Route1 | -18.0 | **11.7** |
| **Product** | | -60.4 | **-44.3** | **R** | | -88.8 | **-55.0** |
| **Adduct** | | -2.7 | **6.0** | **P** | | -34.7 | **-7.1** |
| **TS** | Route2 | 13.6 | **25.0** | **TS** | Route 2 | -1.2 | **27.1** |
| **Product** | | -58.7 | **-41.8** | **S** | | -71.8 | **-39.4** |

kcal/mol in solvent and 16.7 kcal/mol in the nanoreactor. Meanwhile, the entropic cost and other thermal corrections are cut down from 10.6 kcal/mol to 2.0 kcal/mol.

These results account very well from a qualitative point of view for the acceleration experimentally observed as well as for the absolute regioselectivity. According to these numbers, the acceleration from 25.0 to 18.7 kcal/mol should be much higher than experimentally observed . Other factors discussed below hinder the reaction in the host.

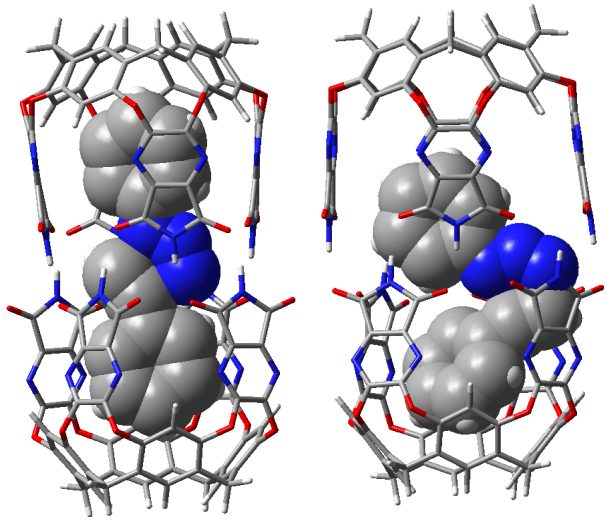


Figure 4.14: *Optimized structures of the two possible transition states for the cycloaddition taking place in the capsule. Left: transition state for* $\mathbf{P} \rightarrow \mathbf{R}$. *Right: transition state for* $\mathbf{P} \rightarrow \mathbf{S}$. *Hosts: stick model. Guests: Space-filling model.*

Figure 4.14 shows the encapsulated transition states for Route 1 (left) and for Route 2 (right). We can explain easily the factors at the origin of the absolute regioselectivity. (i) In Route 2, the aromatic moiety of the phenylazide gets away from the bottom of the capsule. This results in weakened middle range favorable interactions between the guest and the bottom of the capsule. (ii) We also remark for this product a steric clash between the guest and the gates of the capsule. To minimize this unfavorable steric repulsion, the capsule has to distort.

### 4.4.6 Capsule Dislocation or Gate Opening ?

According to experimental results, the trapped guests are in dynamic equilibrium: a guest can be replaced by another within "a few minutes".[32] How do the guests enter and leave the capsule? Either the capsule totally dislocates, or the flexibility of the host plays a central role, as suggested in the literature.[56,142] In this section, both possibilities are computationally explored and compared.

#### Full dissociation

We could imagine the following dissociation: $\mathbf{O} \rightarrow 2^{*}\mathbf{K}$. The capsule is disaggregated in this example, but the guests in the solution cannot enter the monomers, since the space is still occupied. To exchange the guests we have to empty one of the monomers. For this reason, the energies we selected here always show at least an empty monomer, for example: $\mathbf{O} \rightarrow \mathbf{K}+\mathbf{I}+\mathbf{A}$. All possible conformations of dis-aggregation for the most relevant aggregates are presented in Table 4.11 together with their relative energies.

Table 4.11: *Energy required to dislocate a set of complexes involving the capsule. Values in Kcal per mol.*

| | ΔEsol | ΔGsol |
|---|---|---|
| $\mathbf{J} \rightarrow 2^{*}\mathbf{I}$ | 34.6 | **15.6** |
| $\mathbf{O} \rightarrow \mathbf{K}+\mathbf{I}+\mathbf{A}$ | 57.7 | **23.7** |
| $\mathbf{Q} \rightarrow \mathbf{L}+\mathbf{I}+\mathbf{B}$ | 52.3 | **19.5** |
| $\mathbf{P} \rightarrow \mathbf{K}+\mathbf{I}+\mathbf{B}$ | 53.5 | **19.8** |
| $\mathbf{P} \rightarrow \mathbf{K}+\mathbf{I}+\mathbf{A}$ | 55.3 | **22.7** |
| $\mathbf{U} \rightarrow \mathbf{I}+\mathbf{T}$ | 29.0 | **7.6** |
| $\mathbf{V} \rightarrow \mathbf{I}+\mathbf{T}+\mathbf{Solv}$ | 45.5 | **4.1** |

The dislocation of the empty capsule is given as a point of comparison. The first thing we notice is the extremely high potential energies of dislocation, from 29.0 kcal/mol for $\mathbf{U}$ to 57.7 kcal/mol for $\mathbf{O}$. Free energies are fairly lower, as the number of fragment increases, entropy can be released. They range from 4.1 kcal/mol for $\mathbf{V}$ to 23.7 kcal/mol for $\mathbf{O}$.

The dislocation energies for the aggregates made of combinations of reactive guests are found in the tight range 19.5 to 23.7 kcal/mol. When one or two solvent molecules are present, the capsule is already destabilized and is therefore easier to break apart.

#### Opening Gates in Absence of Guests

A pair of ether links binds each of the gates to the calixarene scaffolds. Their $\mathrm{sp}^3$ hybridization allows the conformational changes that are necessary to open each of the gates, (see also Figure 4.5 and 4.11). Figure 4.15 details this mechanism in three steps, on an empty capsule for more clarity.

Once an empty capsule opens, the gate leaves a large opening behind, shown Figure 4.16.

Table 4.12 presents the values for the gate-opening process. We propose here more details on the different components of the energy, since this example is particularly interesting. The free energy barrier to overcome is

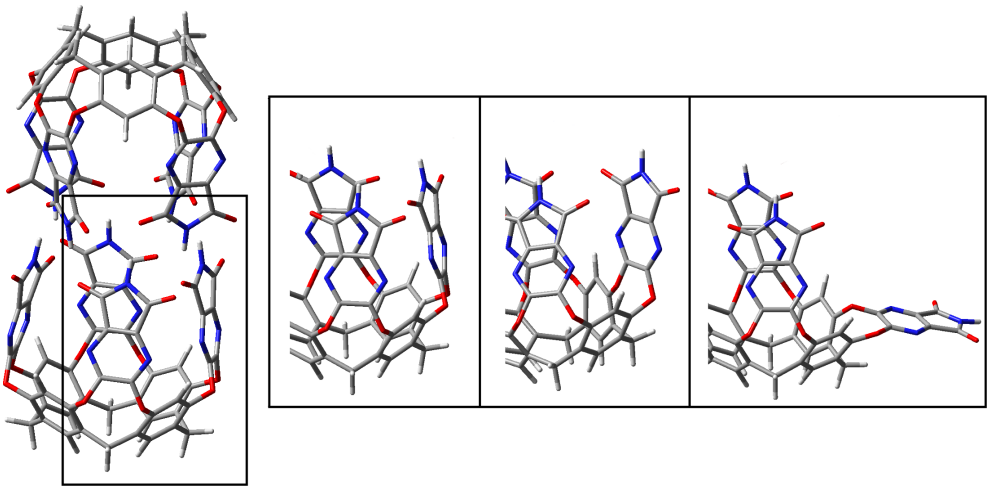


Figure 4.15: *Mechanism of a gate opening for an empty capsule. Left: region of interest on the capsule. Right: Zoom on this region in three frames accounting for the mechanism. First: closed conformation. Second: transition state. Third: Open conformation.*

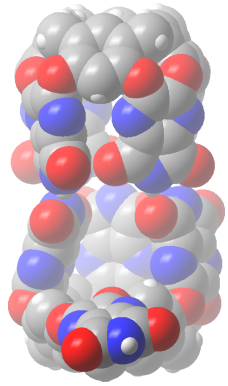


Figure 4.16: *Space-filling model showing the space after opening a gate of the capsule*

Table 4.12: *Detailed energetic profile for the first gate opening. Energies relative to the empty, closed capsule. Values are given in kcal/mol*

| mechanism | ΔH | -TΔS | Δsolv | total=$\Delta G_{sol}$ |
|---|---|---|---|---|
| **J** | 0.0 | 0.0 | 0.0 | **0.0** |
| $\mathbf{J}_{TS1}$ | 14.3 | -0.1 | 0.2 | **14.4** |
| $\mathbf{J}_{1\ gate\ open}$ | 13.6 | -4.4 | -3.3 | **5.9** |

relatively low, with 14.4 kcal/mol. In this process, entropy and solvent effects are negligible.

Once the barrier is passed, the electronic energy remains substantially the same (from 14.4 kcal/mol to 13.6 kcal/mol). Meanwhile, some entropy is liberated through increased conformational freedom and solvent contributions become more favorable with the increase of solvation surface. Those components stabilize partly the open gate.

### 4.4.7 Guest Exchange Mechanism

We reported above a test case where opening one gate of an empty box was an easy process. The journey of a molecule from the solvent to the core of the capsule is more complicated. Here we aim at drawing a coherent description of the mechanism of guest inclusion and exchange.

#### Capture of Guests

We mentioned earlier that it is unlikely to find an empty capsule in solution. Nevertheless, we want to understand how guest exchange is achieved. We present here the energies of the complexes related to this process together with the transition state energies required to open one gate. This will give us information on guest exchange, assuming an "empty site mechanism".

We present the results in Table 4.13, which is divided in three sections. The first section concerns the energies obtained as a guest enters. In the second section and third sections the different combinations of complexes related to the second inclusion of guests are presented. The table should be read from top to bottom. For example on the left side, we start from an empty, closed capsule and we first enclose $\mathbf{A}$ (first section). In the same left column, once $\mathbf{A}$ is enclosed, we have two possibilities that are: trapping $\mathbf{A}$ (second section) or trapping $\mathbf{B}$ in the remaining half of the capsule (third section). A star (*) indicates the gate to open is not located on the monomer containing the guest.

Figure 4.17 displays this process with the corresponding labels for clarity. If we only trap $\mathbf{A}$ sequentially two times, we have to look on the left of Table 4.13, in the first and second section. We present two times values for $\mathbf{P}$, since both sides of the asymmetric capsule can open and have different activation energies.

The results show that the energetics of guest inclusion are reasonable, the highest barriers being evaluated at respectively 15.3 and 15.9 kcal/mol. After the first inclusion, the barriers to open gates are lower, since the guest constrains the capsule and more entropy can be released as a gate opens.

Table 4.13: *Proposed intermediates and transition states for two consecutive inclusions of guests* **A** *and* **B**. *Values in kcal/mol.*

| | first guest to capture: **A** | | | first guests to capture: **B** | |
|---|---|---|---|---|---|
| | ΔEsol | ΔGsol | | ΔEsol | ΔGsol |
| **J** | 0.0 | **0.0** | **J** | 0.0 | **0.0** |
| $\mathbf{J}_{TS}$ | 15.5 | **14.4** | $\mathbf{J}_{TS}$ | 15.5 | **14.4** |
| $\mathbf{J}_{1open}$ | 10.8 | **5.9** | $\mathbf{J}_{1open}$ | 10.8 | **5.9** |
| $\mathbf{M}_{1open}$ | -8.8 | **0.7** | $\mathbf{N}_{1open}$ | -6.2 | **3.3** |
| $\mathbf{M}_{TS}$ | -3.7 | **6.8** | $\mathbf{N}_{TS}$ | -1.5 | **9.3** |
| **M** | -19.9 | **-5.4** | **N** | -17.2 | **-2.7** |
| $\mathbf{M}_{TS}$* | -2.8 | **9.9** | $\mathbf{N}_{TS}$* | -0.6 | **13.1** |
| $\mathbf{M}_{1open}$* | -9.7 | **-1.1** | $\mathbf{N}_{1open}$* | -6.4 | **3.6** |
| | second guest to capture: **A** | | | second guest to capture: **A** | |
| | ΔEsol | ΔGsol | | ΔEsol | ΔGsol |
| $\mathbf{O}_{1open}$ | -28.2 | **-4.1** | $\mathbf{P}_{1open}$ | -25.6 | **-1.1** |
| $\mathbf{O}_{TS}$ | -23.5 | **1.9** | $\mathbf{P}_{TS}$ | -19.5 | **5.5** |
| **O** | -38.8 | **-11.0** | **P** | -34.7 | **-7.1** |
| | second guests to capture: **B** | | | second guest to capture: **B** | |
| | ΔEsol | ΔGsol | | ΔEsol | ΔGsol |
| $\mathbf{P}_{1open}$ | -28.1 | **-4.1** | $\mathbf{Q}_{1open}$ | -24.8 | **-0.7** |
| $\mathbf{P}_{TS}$ | -23.4 | **1.7** | $\mathbf{Q}_{TS}$ | -20.2 | **3.3** |
| **P** | -34.7 | **-7.1** | **Q** | -31.7 | **-3.8** |

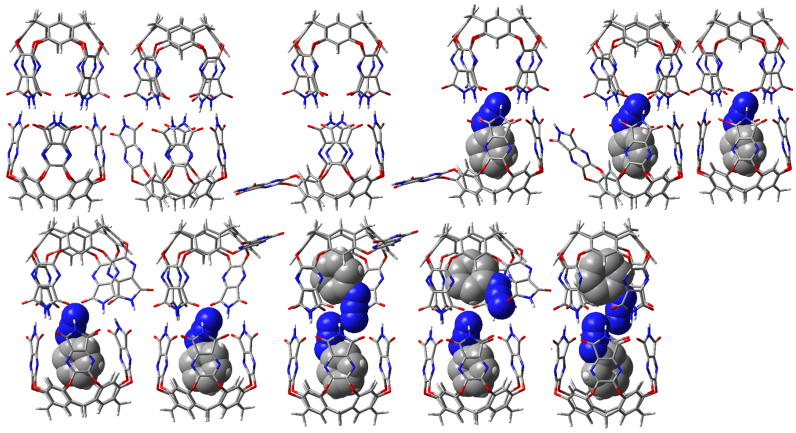


Figure 4.17: *Proposed structures associated to the consecutive inclusion of two guests* **A**. *In reading order:* **J**, $\mathbf{J}_{TS}$, $\mathbf{J}_{1open}$, $\mathbf{M}_{1open}$, $\mathbf{M}_{TS}$, **M**, $\mathbf{M}_{TS}$*, $\mathbf{M}_{1open}$*, $\mathbf{O}_{1open}$, $\mathbf{O}_{TS}$ *and* **O**. *Host: stick model. Guests: space-filling model*

An empty capsule is unlikely to exist in solution as all the complexes are more stable, by -3.8, -7.1 and -11.0 kcal/mol for respectively **Q**, **P** and **O**.

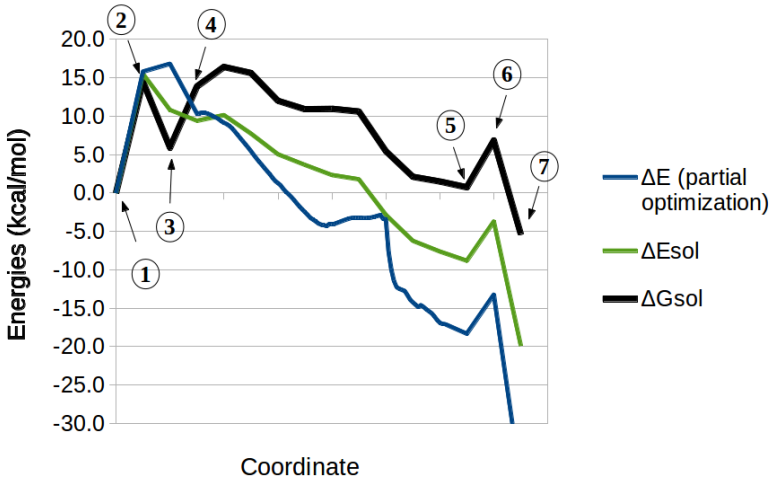



Figure 4.18: *Relaxed scan for the inclusion of a phenylazide. Points 1 to 7 have been added as a complement, see text.*

Therefore, the complex majoritarily found in solution should be $\mathbf{O}$, as we saw earlier. What we learn from this new data is that the transition energies for guest inclusion and guest exchange are very reasonable.

To check that no higher point than the ones presented lies above the conformations presented, we perform a scan between $(\mathbf{J}_{1\text{open}})$ and $(\mathbf{M}_{1\text{open}})$. Guest $\mathbf{A}$ gradually approaches the opening of the gate and enters the core of the capsule. A step size of 0.1 Å was chosen and the coordinate has been chosen as a bond distance between the phenyl moiety of $\mathbf{A}$ and an $sp^3$ carbon at the bottom of the capsule. This coordinate was selected among several other to avoid unwanted distortions and therefore minimize the "noise" of the scan. We present this scan Figure 4.18 on which we added several points including transition states from Table 4.13 to obtain a full profile of azide inclusion on an empty capsule: $(\mathbf{J}+\ \mathbf{A}\rightarrow\ \mathbf{M})$. In Figure 4.19, we present some structures calculated along the scan.

Points 1, 2, 3 and 4 correspond respectively to an empty capsule $(\mathbf{J})$, the TS for opening a gate $(\mathbf{J}_{\text{TS}})$, the energy for a gate open $(\mathbf{J}_{1\text{open}})$, followed by the beginning of the scan. Step 5, 6 and 7 are added after the scan and respectively represent $(\mathbf{M}_{1\text{open}})$, $(\mathbf{M}_{\text{TS}})$ and $(\mathbf{M})$. The blue curve corresponds to a first "rough" scan, with a step size of 0.1 Å and only 10 cycles of optimization allowed to save computing time.

We select 10 out of 93 scan points (Figure 4.19) for a more refined treatment displayed in green and black, where full (but still constrained) optimization is enabled, with the usual corrections of this work.

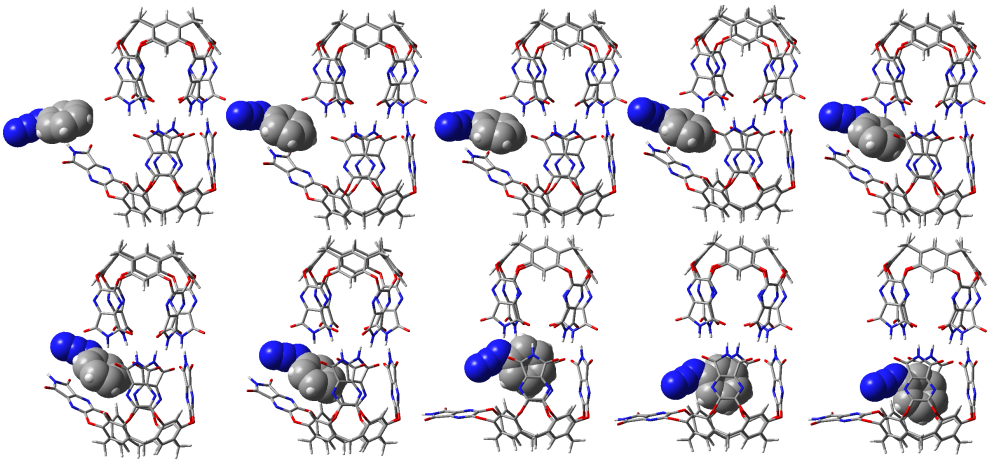


Figure 4.19: *Mechanism for the inclusion of the azide in the empty capsule:* $\mathbf{J} + \mathbf{A} \rightarrow \mathbf{M}$.

It appears that most of the energetic cost is borne by the opening of one gate. As the new guest approaches the near environment of the capsule, the energy rises, primarily for entropic reasons. Further on, the energy only decreases until the guest reaches its optimal location inside the host. Finally, closing the gate appears easy and stabilizes the complex.

These observations are very similar with guest $\mathbf{B}$ , and we assume here that the profile will be similar with a second guest, since this process only concerns a half of the capsule an leaves enough space on the other side.

Overall, the only energetic barriers to overcome corresponds to the opening of the gate and the approach of the guest. As a gate is open, the aperture of the host does not seem to hinder the progress of the guest inside the cavity.

The capsule is dynamic system: guests may exchange quickly based on criteria such as size, compatibility with the capsule and between the guests. How does a guest replace another? Rebek and coworkers suggested that a mechanism demanding an empty site is unlikely, and dislocation is not necessary either.[142] We already investigated those two possibilities by studying the general behavior of the capsule and the stability of various complexes. Indeed, the reverse mechanism of guest inclusion is guest release. A released guest leaves room for a new guest in the capsule. Also, the dislocation of the capsule has been studied as a possible mechanism. We continue these investigations in the following sections.

**Exchanging Gates With Two Open Gates**

A possible mechanism proposed by the authors involves simultaneous opening of two gates, so an outgoing guest departs at the same time an incoming guest arrives, see Figure 4.3. Experimental evidence supports a mechanism of guest exchange involving a single monomer. In other words, the guests do not travel from one side of the capsule to the other during guest exchange. We will therefore focus our investigations based on this idea.

The mechanism we propose here is composed of three main processes: gate opening, guest exchange and gate closure. The starting point of the mechanism is a fully formed capsule with two guests inside: **O**, **P** and **Q**. The energies computed for this mechanism are reported Figure 4.14. Figure 4.20 shows three intermediary conformations: : **Ex_6**, **Ex_8** and **Ex_10**.

The sequence of steps is the following:
First, we open two gates:

| | |
|---|---|
| **Ex_1** | The capsule is closed, with two guests inside. (e.g.: **O**). |
| **Ex_2TS** | TS for first gate opening. |
| **Ex_3** | A gate is open. |
| **Ex_4TS** | TS for second gate opening. |
| **Ex_5** | Two gates are open. |

Then, an incoming guest displaces the outgoing one:

| | |
|---|---|
| **Ex_6** | The incoming guest interacts with an open gate. |
| **Ex_7TS** | TS for the incoming gate closing. |
| **Ex_8** | The incoming gate is closed, the functional group of the incoming guest is still partly outside. |
| **Ex_9TS** | TS: the functional group of the incoming guest squeezes in the capsule. |
| **Ex_10** | The incoming guest enters and displaces the outgoing guest outside the core of the capsule. |
| **Ex_11** | The outgoing guest gets out. |

Finally, the capsule is closed again:

| | |
|---|---|
| **Exch12TS** | TS: outgoing gate closing. |
| **Exch13** | The box is closed, two guests are in. |

A significant part of the energy cost relies in the sequential opening of the two gates (**Ex_1** to **Ex_5**) and requires 10.5 to 15.1 kcal/mol. The aggregation step **Ex_5** $\rightarrow$ **Ex_6** is rather straightforward: it costs between 1.8 and 4.7 kcal/mol. The full inclusion of the incoming guest in the core of

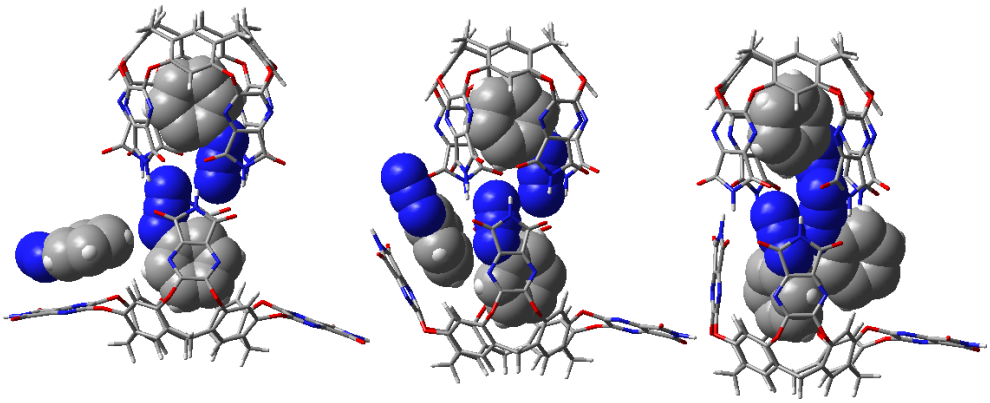


Figure 4.20: *Selected steps for the guest exchange with two open gates, example with guest* **A**. *In reading order:* **Ex_6**, **Ex_8** *and* **Ex_10**. *Host: stick model. Guests: space-filling model*

Table 4.14: *Energetics of a proposed mechanism for guest exchange involving a capsule with two open gates . Values in kcal/mol.*

| replacement | **O → O** ΔGsol | **Q → Q** ΔGsol | **O → P** ΔGsol | **Q → P** ΔGsol |
|---|---|---|---|---|
| **Ex_1** | 0.0 | 0.0 | 0.0 | 0.0 |
| **Ex_2TS** | 12.9 | 7.1 | 12.9 | 7.1 |
| **Ex_3** | 6.8 | 3.1 | 6.8 | 3.1 |
| **Ex_4TS** | 21.1 | 17.2 | 21.1 | 17.2 |
| **Ex_5** | 15.1 | 10.5 | 15.1 | 10.5 |
| **Ex_6** | 17.4 | 15.1 | 18.6 | 12.3 |
| **Ex_7TS** | 26.2 | 23.1 | 27.1 | 21.4 |
| **Ex_8** | 22.2 | 18.5 | 24.2 | 18.9 |
| **Ex_9TS** | 30.7 | 28.4 | 34.4 | 26.0 |
| **Ex_10** | 19.4 | 9.3 | 14.9 | 6.1 |
| **Ex_11** | 6.8 | 3.1 | 6.9 | 2.7 |
| **Ex_12TS** | 12.9 | 7.1 | 12.6 | 9.3 |
| **Ex_13** | 0.0 | 0.0 | 3.9 | -3.2 |

the capsule going through **Ex_9TS** is the limiting step, lying 26.0 to 34.4 kcal/mol above the closed capsule with encapsulated guests.

This description is not exhaustive: there might be several conformers of the different guests for each step, and we have no guarantee to have picked the minimum energy pathway. We have nevertheless an upper-limit and the energies of the various steps do not appear to depend so much on the guests.

The energies of this mechanism through the opening of two gates appears however not acceptable to provide a actual description of a dynamic exchange

of guests. While only opening two gates appears reasonable, destabilizing components pile up: entropic cost for the incoming guest, displacement of a outgoing guest and the squeezing the incoming one, the results plead against the idea of a fast and accessible mechanism.

#### Alternative Mechanism: A Single Open Gate

Are there other mechanisms operating? A yet unexplored idea consists in aggregating three guests in the capsule core, just for one moment, the time of the exchange to take place.

To support this idea, the literature comes handy with an helpful experiment. Rebek et al. studied the rotation motion of [2,2]-paracyclophane.[157,158] This guest is very bulky and rigid but still accommodated in the capsule.

The study reveals that despite its large size, the paracyclophane molecule is still able to rotate inside the host at a speed depending on the other co-guest. Probably the rotation is more favored in the central region of the capsule, which is more flexible. A small co-guest will then favor the access to this region.

We explore the possibility of having three encapsulated guests as a metastable intermediate for guest exchange, with a new series of calculations. The proposed mechanism involves four main processes: Gate opening, entry of the new guest, exit of the previous guest and gate closure. Results are reported in Table 4.15, and Figure 4.21 displays three intermediates: **Alt_4**, **Alt_8** and **Alt_12**.

The sequence of steps is the following:
First, we open one gate:

| | |
|---|---|
| **Alt_1** | The capsule is closed, with two guests inside. |
| **Alt_2TS** | TS for incoming gate opening. |
| **Alt_3** | A gate is open. |

An incoming guest enters the capsule:

| | |
|---|---|
| **Alt_4** | Incoming new guest interacts with the open gate. |
| **Alt_5TS** | TS: the incoming gate closes. |
| **Alt_6** | The incoming gate is closed, the functional group of the incoming guest is still outside. |
| **Alt_7TS** | TS: the functional group of the incoming guest squeezes in the capsule. |
| **Alt_8** | The incoming guest enters. Three guests are is the capsule. |

The outgoing guest leaves the capsule:

**Alt_9TS** TS: the functional group of the outgoing guest squeezes out of the capsule.
**Alt_10** The functional group of the outgoing guest is outside.
**Alt_11TS** TS: the outgoing gate opens.
**Alt_12** The outgoing gate is open, with the outgoing guest still interacting with it.
**Alt_13** The outgoing guest gets out.

Finally, the capsule is closed again:
**Alt14TS** TS outgoing gate closes
**Alt15** Box is closed, two guests are in.

This alternative mechanism involves a state in which three guests are simultaneously encapsulated. This metastable state, although high in energy (**Alt_8**: between 13.8 and 18.6 kcal/mol), presents an advantage: only one gate is required to open at the time. This strategy leads to a lower energy pathway than the previous "two gates" mechanism.

Table 4.15: *Energetics of an alternative mechanism for guest exchange. This mechanism only require one gate to be open at the time. Values in kcal/mol.*

| replacement | **O → O** ΔGsol | **Q → Q** ΔGsol | **O → P** ΔGsol | **Q → P** ΔGsol |
|---|---|---|---|---|
| **Alt_1** | 0.0 | 0.0 | 0.0 | 0.0 |
| **Alt_2TS** | 12.9 | 7.1 | 12.9 | 7.1 |
| **Alt_3** | 6.8 | 3.1 | 6.8 | 3.1 |
| **Alt_4** | 11.7 | 11.9 | 16.5 | 8.7 |
| **Alt_5TS** | 20.8 | 21.8 | 22.3 | 18.6 |
| **Alt_6** | 16.8 | 20.6 | 17.9 | 17.1 |
| **Alt_7TS** | 17.8 | 20.6 | 23.4 | 23.6 |
| **Alt_8** | 18.1 | 13.8 | 18.6 | 14.3 |
| **Alt_9TS** | 25.2 | 21.7 | 27.0 | 17.8 |
| **Alt_10** | 18.9 | 14.4 | 20.7 | 11.7 |
| **Alt_11TS** | 22.4 | 21.3 | 27.0 | 19.4 |
| **Alt_12** | 18.3 | 11.8 | 15.3 | 6.0 |
| **Alt_13** | 6.8 | 3.1 | 6.9 | 2.7 |
| **Alt_14TS** | 12.9 | 7.1 | 12.6 | 9.3 |
| **Alt_15** | 0.0 | 0.0 | 3.9 | -3.2 |

The complex with three encapsulated guest (**Alt_8**) lies between 13.8 kcal/mol and 18.6 kcal/mol above their respective two-guest complexes. Also, steps **Alt_6**, **Alt_7TS**, **Alt_9TS** and **Alt_10** where the functional group of the leaving/entering guest is squeezed out of the capsule and the

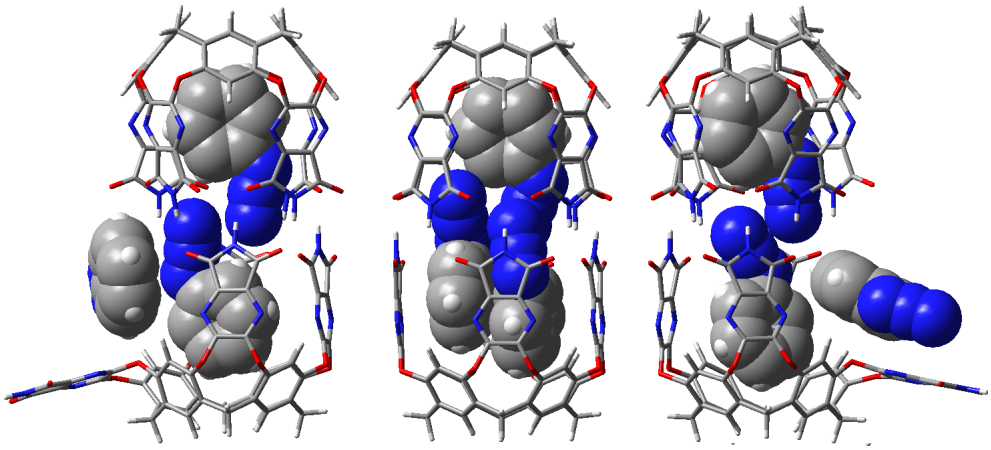


Figure 4.21: *Selected steps for the guest exchange with a single open gate, illustrated with guest* **A**. *In reading order:* **Alt_4**, **Alt_8** *and* **Alt_12**. *Host: stick model. Guests: space-filling model*

ring is inside, may not be entirely necessary. We presume that the energies of the gate-closing **Alt_5TS** and **Alt_11TS** would remain similar if the guests squeezed inside the capsule directly.

As we observed earlier, there are differences in the energies, depending which guest is involved. Here again, we have computed only one conformer for each point of the path, and we could imagine slight improvements in terms of stabilities of intermediates and transition states. However, the trend is generally similar for the all series of calculations performed. Additionally, those calculations are extremely costly, with 242 atoms involved.

The barriers with this mechanism range from 21.8 to 27.0 kcal/mol, so this "one-gate" mechanism should also be discarded, as was the "two-gates" mechanism.

#### Conclusion on the Guest Exchange Mechanisms

We investigated four possibilities for guest exchange. In Figure 4.22, we propose a summary involving only guest **A**, for more clarity.

The "empty site" mechanism involves a vacant site in the capsule, and is basically a backwards and forward guest releasing-guest trapping. Its cost is evaluated in the range 7.1 - 12.9 kcal/mol. The "dislocation" mechanism implicates the dislocation of the box into disconnected monomers, one of which should be empty. The cost of this mechanism ranges from 19.5 to 23.7 kcal/mol. The "two gates" mechanism requires two gates to be simultaneously open, in such a way to let the incoming guest enter as the

outgoing guest leaves the core of the capsule from 26.0 to 34.4 kcal/mol are required for this mechanism to take place. The "one gate" mechanism intends to reproduce the "two gates" mechanism, with the particularity of opening a single gate. The "one gate" mechanism has an intermediate that displays three simultaneously encapsulated guests and requires from 21.8 to 27.0 kcal/mol, depending of the guests involved.

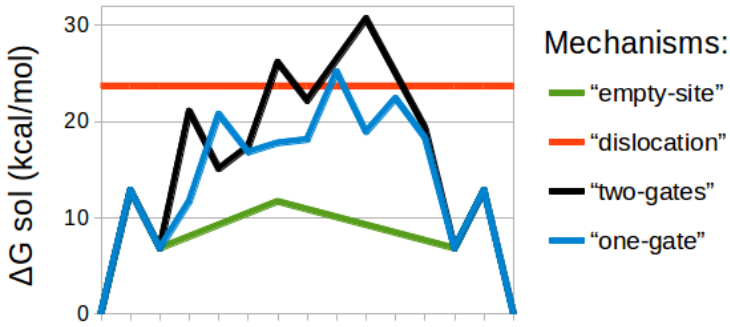



Figure 4.22: *Comparison of the four mechanisms proposed, where guests* **A** *is replaced by another guest* **A**. *Values in kcal/mol.*

Overall, simply offering a vacant site seems to be by far the less costly solution. A single open gate with one empty site comes at little cost, confirming "fast" guest exchange. Also the deformation remains minimal while the process, being dissociative, is entropically favored.

### 4.4.8 (Dis-)Aggregation Processes

When compounds approach each other or move away, the treatment theoretically applied to calculate free energies is doubtful, as we discussed in Section 3.5.2 in last chapter. Two independent calculations of arbitrary compounds are systematically inconsistent entropically with a single calculation of the two same compounds placed at very large distances.

To ensure that only entropy and not potential energy is problematic, we perform several scans to estimate the profile of an aggregation process (such as $\mathbf{I}+\mathbf{A}\rightarrow\mathbf{K}$) , and look at the components forming free energies. The scans do not show any potential energy barrier. The same stands for other guests and the formation of the capsule. In this context, we will follow the same reasoning as in the previous chapter to obtain pseudo-activation energies and diffusion-limited rate constants.

Using again equation 3.1 and applying it on the systems of interest in this chapter, all the diffusion-limited processes give a pseudo-activation energy in

the range 3.0-3.6 kcal/mol, depending on the size and molecular weight of the systems.

### 4.4.9 Mechanism of the Formation of the Capsule

The formation of the supramolecular capsule was reported to be induced by the presence of suitable guests.[157] In other words, it will not form in their absence.

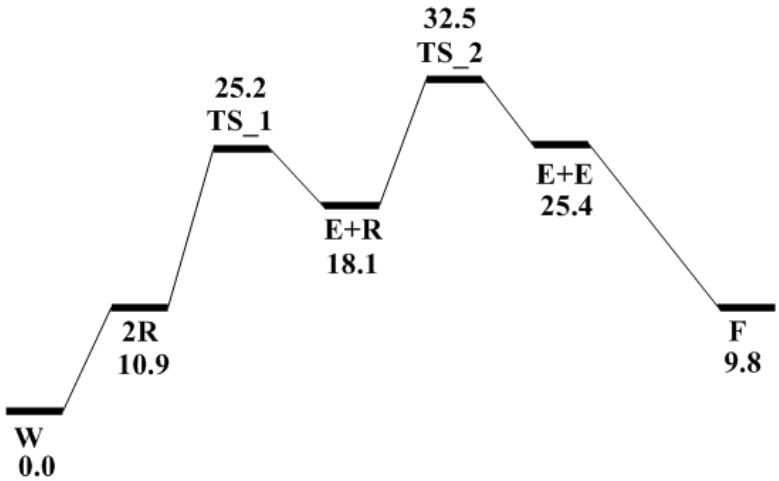



Figure 4.23: *Energetic span of the formation of the capsule. Values in kcal/mol.*

We represent in Figure 4.23 the elementary steps of the formation of the capsule. Each of the individual steps $\mathbf{F} \rightarrow \mathbf{I}$ bears an activation energy of 14.4 kcal/mol. As both monomers have to change their conformation, the mechanism is sequential and goes through two transition states TS_1 and TS_2. With this mechanism, the global activation energy required for the formation of the capsule is 32.5 kcal/mol. This value is higher than the uncatalyzed reaction activation energies: 27.2 and 25.0 kcal/mol, to form respectively $\mathbf{C}$ and $\mathbf{D}$.

Regardless of the mechanism, $\mathbf{J}$ is less stable than $\mathbf{E}$. This explains why guests are required to stabilize the capsule, as including solvent always results in higher-energy complexes, while including guests stabilizes $\mathbf{J}$. Nevertheless, a barrier of 32.5 kcal/mol appears quite high. We therefore have to look for another, less demanding mechanism for this process.

We searched for a new, less energetic mechanism for the formation of the capsule and found that there exists a path that does not require the formation

of the monomers **F** and **I** . Instead, the kite-kite dimer **E** would gradually rearrange by closing each of its gates one by one. This new mechanism involves a key intermediate **CapI_11** , see Figure 4.24 that presents the features of a pre-formed capsule. This intermediate displays two pairs of face-to-face gates, held together by hydrogen bonds

This is a particularly difficult case for a computational study, due to the size of the systems involved. We are considering a large dimer of 200 atoms and the possibilities of rearrangements are numerous. We present in Figure 4.24 the lowest energy path we could find so far.

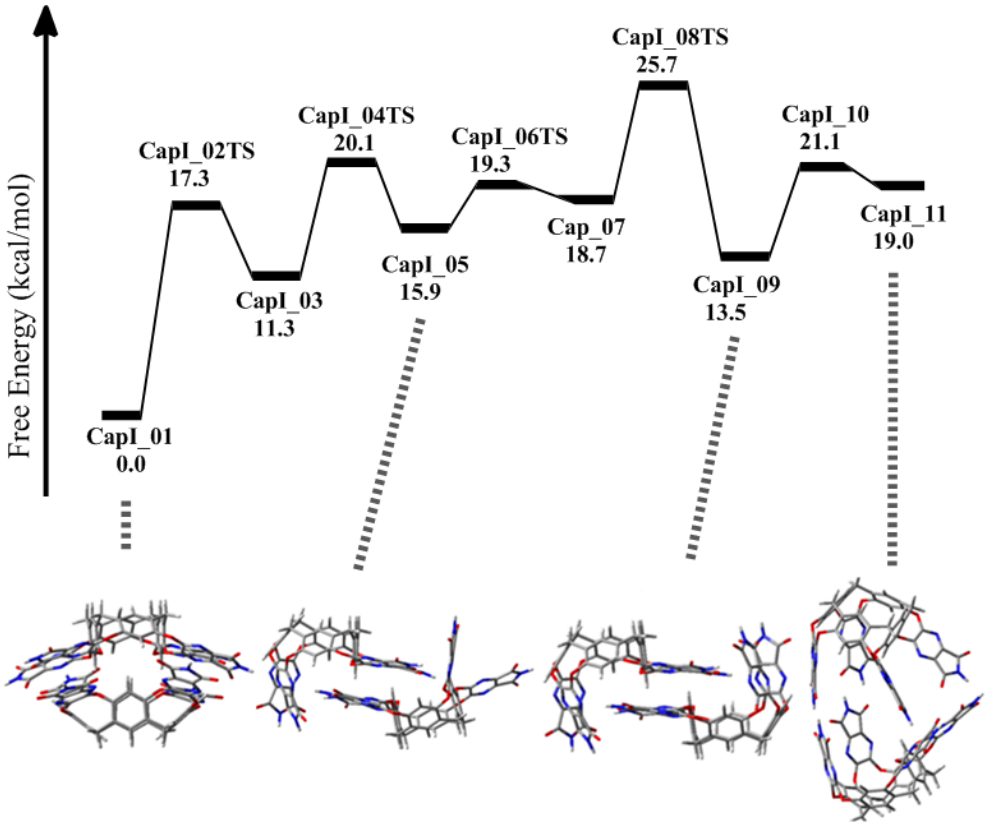



Figure 4.24: *Energetic span of the first part of the formation of the capsule. On the bottom of the figure selected structures are represented along the path, they are linked to the corresponding energy by dotted lines. Free energies in kcal/mol.*

The results show a reasonable mechanism that we describe here in 7 steps, not counting the transition states:

**CapI_01** The Kite-kite dimer **E**
**CapI_03** The configuration of one gate inverts.
**CapI_05** The configuration of a second gate inverts.
**CapI_07** The configuration of a third gate inverts.
**CapI_09** Four gates are now closed, two on each monomer. This is a rather stable intermediate.
**CapI_10** The monomers separate from each other.
**CapI_11** The monomer come back together, in a box-like configuration, while each monomer bears two open gates and two closed gates.

The successful strategy here consists in keeping or compensating the favorable interactions between the monomers, while changing their conformations, step by step. When keeping van der Waals interactions on a large surface was not possible, we tried to form the hydrogen bonds to prepare the shape of the capsule. Finally, as the capsule gets in shape, a space appears available in **CapI_11**. At this point, two options are possible: either we keep forming the capsule, either we introduce guests in the core of the pre-formed capsule directly.

#### Second Part: Without Guests

The first possibility is to close the capsule without the help of the guests. We propose in Figure 4.25 the energetics and mechanism of the formation of the capsule.

We found a reaction path in 5 steps:
**CapIIa_01** The shape of the capsule is preformed (it is the last step of previous the mechanism: **CapI_11**).
**CapIIa_03** One of the four remaining gates closes.
**CapIIa_05** One of the three remaining gates closes. Only two gates are left open, which face each other to maximize energy gain from hydrogen bonds.
**CapIIa_07** One of the two remaining gates closes.
**CapIIa_09** The last gate closes. The empty capsule is now fully formed.

This mechanism goes through four transition states, one for each gate to close. The highest TS is **CapIIa_04TS**, which displays an energy of 35.0 kcal/mol. The fully formed capsule **CapIIa_09** (also called **J**) stands 9.8 kcal/mol above the closed host (noted **CapI_01** or **E**).

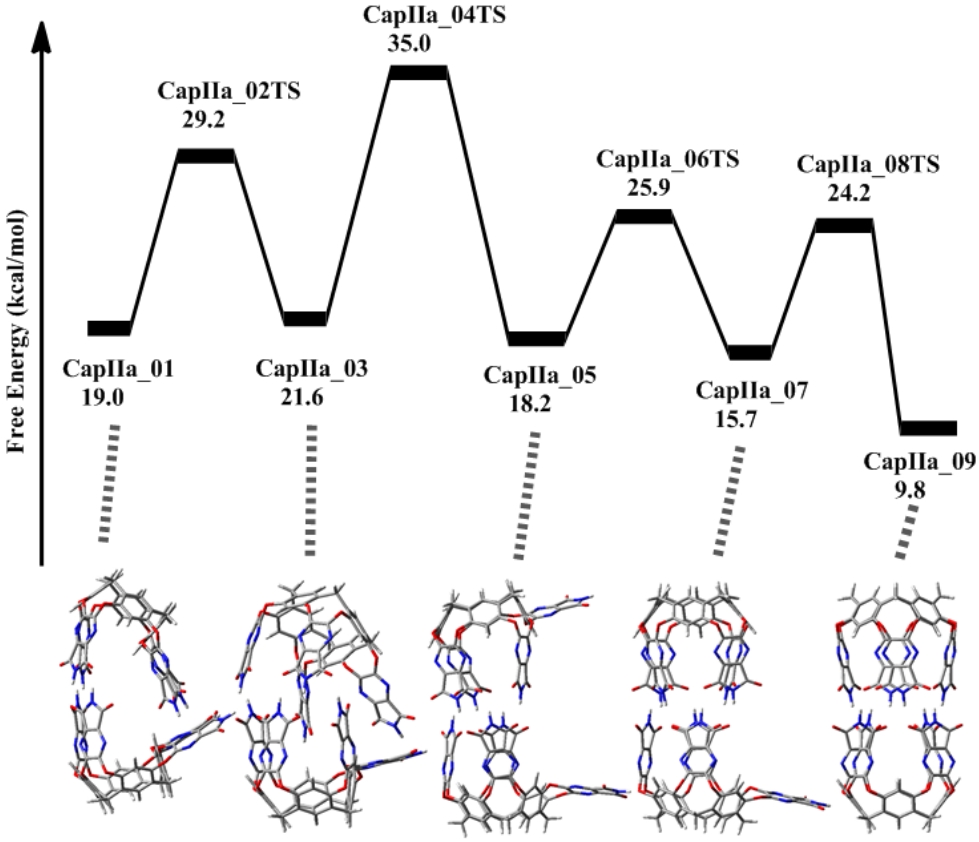



Figure 4.25: *Energetic span of the second part of the formation of the capsule, when no guests are participating in the mechanism. On the bottom of the figure structures of the minima of this mechanism are represented along the path, they are linked to the corresponding energy by dotted lines. Free energies in kcal/mol.*

With this mechanism, the high activation energy encourages us to pursue our investigations. 35.0 kcal/mol is even higher than the mechanism that may take place when monomers are separated, as detailed in section 4.4.9.

**Second Part: With the best guest: A**

Here we intend to explore the role of the guests in the formation of the capsule. We will consider the encapsulation of two guests **A** (phenylazide), as they where found to have a good affinity for entering **J**. As stated above, the first part of the mechanism is common to the empty capsule, so the starting point (**CapIIb_01**) also corresponds to (**CapI_11**). We present

this mechanism in Figure 4.26. The deformed structure displays a large space inside that can readily accommodate two guests **A**, as it displays large openings (steps involving **CapIIb_02** and **CapIIb_03**).
The mechanism we propose takes place in 7 steps, and includes the entrance of the guests in the pre-formed capsule:

**CapIIb_01** The shape of the capsule is preformed (it is the last step of the first part of the mechanism: **CapI_11**).
**CapIIb_02** One guest **A** enters the core of the preformed capsule.
**CapIIb_03** A second guest **A** enters the core of the preformed capsule.
**CapIIb_05** One of the four remaining gates closes.
**CapIIb_07** One of the tree remaining gates closes. The two gates remaining open face each other.
**CapIIb_09** One of the two remaining gates closes.
**CapIIb_11** The last gate closes. The empty capsule is now fully formed. This step corresponds to **O**, *i.e.* the fully formed capsule with two guests **A** inside.

With a maximum transition energy of 27.8 kcal/mol, this mechanism is less energetic than the two others: the mechanism **E** $\rightarrow$ **F** $\rightarrow$ **I** $\rightarrow$ **J** costs 32.5 kcal/mol, while the mechanism of pre-formation of the capsule without guests costs 35.0 kcal/mol.

This new findings seem to account very well for the guest-induced formation of the capsule. As this mechanism is more accessible than the previous ones, it is also more relevant. The presence of guests in the core of the pre-formed capsule facilitate the closing of the gates and therefore the formation of the host. Meanwhile, the host-guest system (**CapIIb_11**, or **O**) is stabilized, making more difficult the backwards mechanism to reform **E** (also noted **CapI_01**).

### 4.4.10 Energy of the Huisgen Reaction in Solvent

We reported that the free energy of activation for the formation of **C** and **D** are respectively of 27.1 and 25.0 kcal/mol. However, the rate constant for the formation of **C** and **D** is reported to be 4.3*$10^{-9}$$M^{-1}s^{-1}$ in mesitylene. This is a second order rate constant which corresponds to a free energy of activation of 28.8 kcal/mol. The functional B97D may have underestimated the real activation energy.

Houk et al. compared B3LYP and SCS-MP2[159] and found some differences in the activation energies of strained Huisgen cycloadditions. Also, Grimme reported that SCS-MP2 outperforms MP2 and standard DFT func-

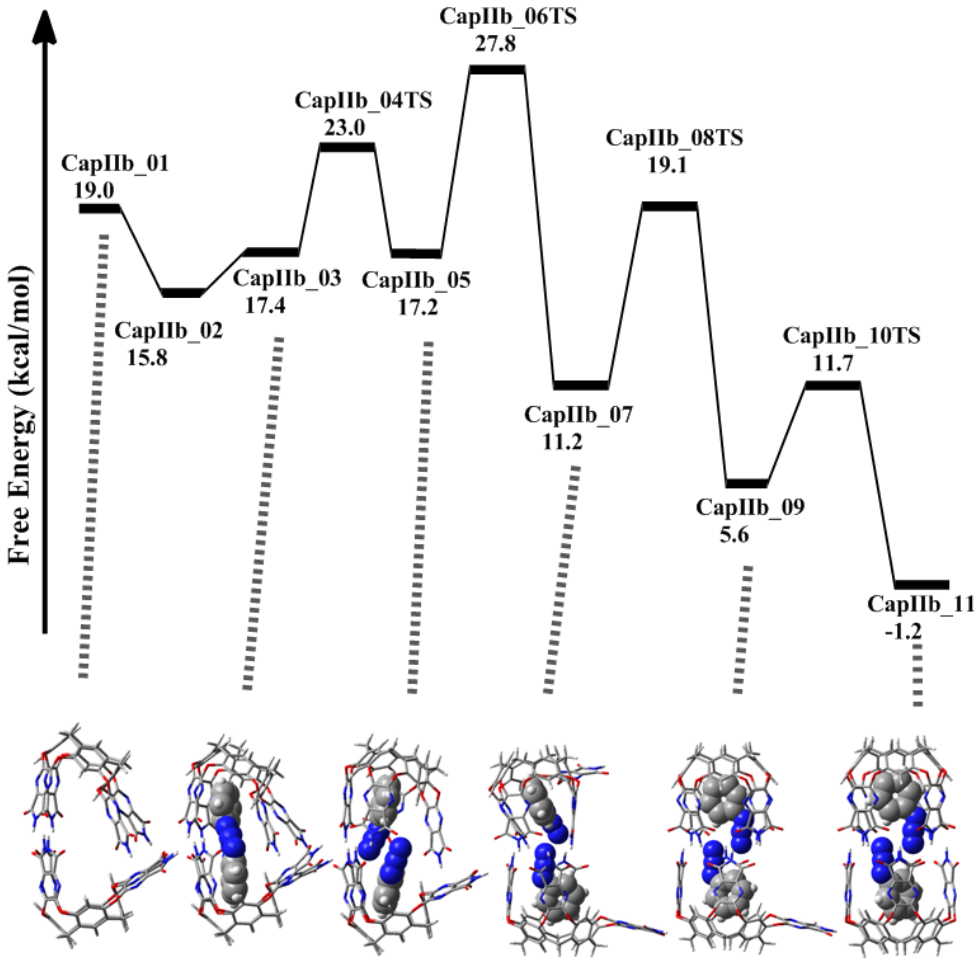



Figure 4.26: *Energetic span of the formation of the capsule. On the bottom of the figure selected structures are represented along the path, they are linked to the corresponding energy by dotted lines. Free energies in kcal/mol.*

tionals B3LYP in the calculation of activation energies of 1,3 dipolar cycloaddition reactions.[160]

To improve the quality of the results, we explored again the two pathways for the formation of **C** and **D** with a large SCS-MP2, with a basis set of 6-311G*. The results we obtained showed a free activation energy of respectively 32.5 kcal/mol and 31.7 kcal/mol for respectively **C** and **D**. These barriers are significantly higher, but both products have more similar activation energies.

Switching to SCS-MP2/6-311G*, we have found a difference of 5.4 kcal/mol for $\mathbf{C}$ and 6.8 kcal/mol for $\mathbf{D}$. We cannot use the SCS-MP2 method with a basis set of 6-311G* on our large host-guest system as it would be impossible to treat it computationally in a reasonable time. As an approximation, the activation free energies for the formation of both encapsulated products can then be considered to be affected by the same amount resulting in free energy barriers of 24.1 and 40.9 kcal /mol for the activation energies of respectively $\mathbf{R}$ and $\mathbf{S}$.We decided to use these larger and more accurate free energies for the cycloadditions (formation of $\mathbf{C}$, $\mathbf{D}$, $\mathbf{R}$ and $\mathbf{S}$) in our kinetic model.

## 4.5 Building a Comprehensive Kinetic Model

Throughout this chapter, we saw that the behavior of the capsule is by far too complex to be tackled with an energetic span model, which consists in calculating the free energy difference between the most stable intermediate and the highest transition state. This value is the effective activation energy of the global process.

The energetic span model does not take into account the weight of the concentrations of different chemical species involved and it can be difficult to arrange ramifications and subtleties of a complex reaction scheme. Having in hand the valuable data that represent free energies, one can access reaction rates, in a quite straightforward manner, using Eyring's equation.

In this section, we will describe a kinetic model that accounts for the very diversified behavior of the host and its guests in solution, in a similar fashion to what has been done in the last chapter. The labels used in this model are based on Figure 4.4. The model takes into account the background cycloaddition reaction happening in solution. It also includes the changes of conformation of the host leading to the formation of the capsule through several paths. The kinetic model also handles the processes of guest inclusion and exchange. All reversible reactions are implemented as well as the possibility of dislocation of the capsule. To summarize, all the relevant complexes and the transition states that relate them enter a kinetic model that is conceptually similar to the one we built in the previous chapter.

We set a value of 25 mM for $\mathbf{A}$, 50 mM for $\mathbf{B}$ and 5 mM for the supramolecular catalyst $\mathbf{E}$, as well as 6.67 M for the solvent, all taken from the experimental conditions. Please note that logarithmic scales have been used to plot the data that spans many orders of magnitude in concentration and time. The results obtained are summarized in Figure 4.27. We do not present all the species but we select the most relevant ones for clarity.

The background Huisgen reaction takes place at a constant rate and the

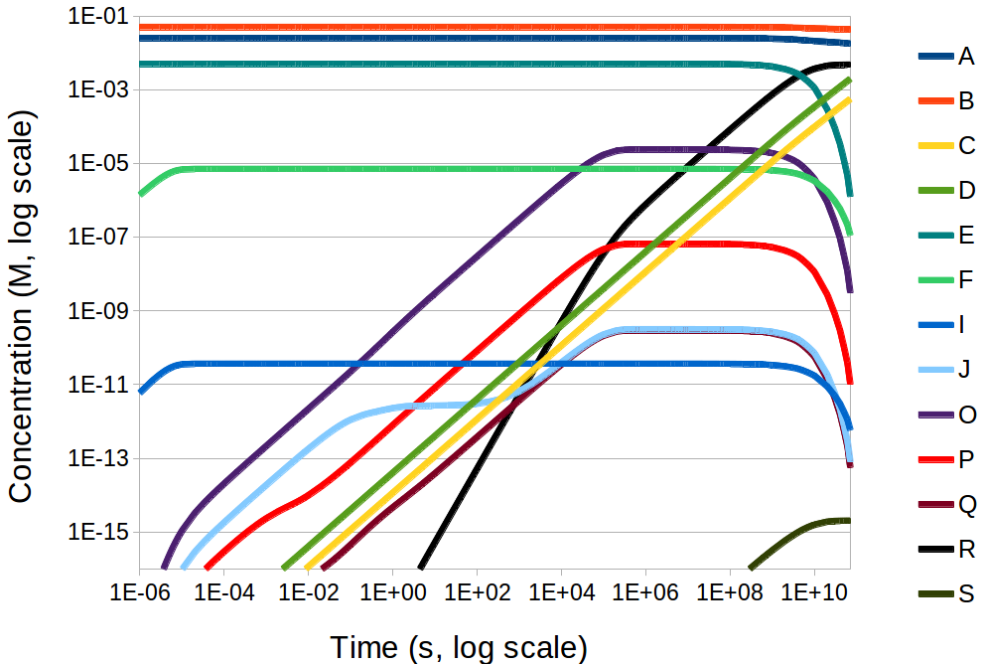



Figure 4.27: *Evolution of the concentration of the most relevant complexes related to the capsule and its guests.*

production of **C** and **D** only slows down as the reactants start to be visibly consumed, from *c.a.* $10^8$ seconds, or about 3 years.

In the first microseconds, the concentration of **F** takes off until reaching at 2.0*$10^{-5}$ seconds an equilibrium around 7.1*$10^{-6}$ M. Subsequently, **I** may be formed and also reaches an equilibrium at a low concentration of *c.a.* 3.7*$10^{-11}$ M around 2*$10^{-5}$ seconds. Being more stable, but not easily accessible, the empty capsule (**J**) concentration overcomes the concentration of **I** in 1*$10^4$ seconds (less than 3 hours). It remains however at very modest levels since it can quickly encapsulate the reactive guests **A** and **B** .

While the empty capsule **J** forms, the concentration of the capsule filled with two phenylazide guests (**O**) takes off much more rapidly, to reach a maximum at around 3*$10^5$ seconds (more than 3 days), with a maximal concentration of *c.a.* 2.3 *$10^{-5}$M. At 8*$10^3$ seconds (about 2 hours), the concentration of **R** overcomes the both the concentrations of **C** and **D**.

We successfully implemented a kinetic model that reproduces a network of reactions. The processes to pass from one point to the other on the network are controlled by a rate constant, that is itself derived from free energies obtained by DFT calculations. A kinetic program handles the interplay of the numerous processes taking place simultaneously.

### 4.5.1 Comparing Experimental and Theoretical Data

In this section, we compare the results of the theoretical kinetic model with the experimental kinetic results. For ease of understanding, we propose in Table 4.16 the comparison of the different values proposed by the experimentalists and by our calculations.

Table 4.16: *Comparison of experimentally and theoretical kinetic data on the supramolecular capsule*

| | Experimental data | | | Theoretical data | | | |
|---|---|---|---|---|---|---|---|
| reaction | kinetic constant | → | ΔG | kinetic constant | ← | ΔG | reaction order |
| $\mathbf{A} + \mathbf{B} \rightarrow \mathbf{C}$* | $4.3*10^{-9}$ | → | 28.8 | $8.7*10^{-12}$ | ← | 32.5 | 2 |
| $\mathbf{A} + \mathbf{B} \rightarrow \mathbf{D}$* | $4.3*10^{-9}$ | → | 28.8 | $3.3*10^{-11}$ | ← | 31.7 | 2 |
| $\mathbf{P} \rightarrow \mathbf{R}$ (over 6 days) | $1.3*10^{-9}$ | → | 29.5 | $6.6*10^{-13}$** | → | 34.0 | 0 |

*: No precision was given on regioselectivity.

**: Estimated from the results of the kinetic model, see text. (respectively 25mM and 50mM), see text.

The background reaction $\mathbf{A} + \mathbf{B} \rightarrow \mathbf{C} + \mathbf{D}$ is experimentally very slow[161] and is reported to have a formation rate constant of $4.3*10^{-9}M^{-1}s^{-1}$. Our model predicts a rate constant of $8.7*10^{-12}M^{-1}s^{-1}$ and $3.4*10^{-11}M^{-1}s^{-1}$ for respectively $\mathbf{C}$ and $\mathbf{D}$, two figures slower than the rate experimentally reported.

Concerning the reaction in the capsule, which is found to be completely regioselective on both experimental and theoretical sides, the values are discussed in what follows. The rate constant for the 1,4 product ($\mathbf{R}$) formation is experimentally measured over 6 days and determined to be $1.3*10^{-9}Ms^{-1}$. Note the units used, denoting data interpreted as a global zero-order reaction. It is therefore interpreted as not depending on the concentrations of the reactants, or of the capsule. Given that our kinetic model generates a concentration of $\mathbf{R}$ of $3.4*10^{-7}M$ in 6 days, it would corresponds to a zero-order rate constant of $6.6*10^{-13}Ms^{-1}$. It appears that we have a disagreement with the experiments. Arguably, the kinetic data should be interpreted as first-order respect to the concentration of the capsule. It is true that the rate does not directly depend on the concentration of the reactive guests in solution, but the reaction is limited by the amount of formed capsules, which are found in very small amounts. Although the agreement is not perfect, our study allows us to explain some of the experimental incongruities.

The authors of the experimental study[55] comment that as the volume of the capsule is about 450 $Å^3$, both reactants are therefore found at a

concentration of 3.7M. They calculate a pseudo zero-order rate constant, derived from the background reaction:

$$4.3 * 10^{-9} M^{-1} s^{-1} * 3.7 M^2 = 5.9 * 10^{-8} M s^{-1}$$

This value is calculated to compare concentrations effects to the actual rate. It is slightly larger than the value experimentally obtained (1.3*$10^{-9}$Ms$^{-1}$). The authors suppose that the orientation of the reactants is not ideal within the capsule. According to our model, the problem is not here but relies in the formation of the capsule, as the activation energy required to form $\mathbf{O}$ starting from $\mathbf{E}$ is 27.8 kcal/mol. As this value is very close to the activation energy required to form $\mathbf{R}$, the major product, starting from $\mathbf{O}$, the most stable complex (28.0 kcal/mol), this explains the ambiguity.

Finally they propose a zero-order rate constant for the background reaction, in order to compare the efficiency of both processes. They calculate:

$$4.3 * 10^{-9} M^{-1} s^{-1} * 50 mM * 25 mM = 5.4 * 10^{-12} M s^{-1}$$

The authors conclude that the reaction is 240 times faster inside the capsule than outside.

$$\frac{1.3 * 10^{-9} M s^{-1}}{5.4 * 10^{-12} M s^{-1}} = 241$$

With our data the pseudo-zero-order rate constant for the background reaction $\mathbf{A}+\mathbf{B} \rightarrow \mathbf{C}$ is:

$$87 * 10^{-12} M^{-1} s^{-1} * 50 mM * 25 mM = 1.1 * 10^{-14} M s^{-1}$$

Therefore, we calculate (using their method), that the reaction is accelerated 60 times.

$$\frac{6.6 * 10^{-13} / M s^{-1}}{1.1 * 10^{-14} M s^{-1}} = 60$$

This agreement is good considering the formulas used and the great complexity of the scheme presented. Combining the knowledge obtain by both experimental and theoretical sides, one could certainly perform experimentally a more refined kinetic study.

## 4.6 Conclusions

We assessed in the first part of this chapter the performance of various computational methods regarding the description of a large host-guest complex. Molecular mechanics present the advantage of velocity in terms of computer

time, but lacks of reliability, particularly because of the conformational flexibility of the system. A semi-empirical method is also fast, but extremely sketchy and should be avoided for van der Waals complexes.

ONIOM models present the advantage of correcting the difficulties encountered by molecular mechanics on the bond formation/rupture, while presenting the advantage of a modest computational cost. However, we did not find its results sufficiently accurate to undertake a full kinetic study. We will explore in the next chapter possibilities of improvement regarding this method.

DFT methods known as a reliable electronic structure method, do not appear sufficient when they are not corrected to account for dispersion effects, considering the size of the structure studied, since the surfaces of non-bonded contact are very large and require special treatment.

For this reason, DFT-D method B97D was found successful and reproduces accurately the experimental results, although at a large computational cost. We then used this method to calculate the free energies of various complexes as well as the activation free energies required to pass from one to another.

We found using a kinetic model that a stable aggregate prevents to a large extent the formation of the capsule, while a symmetrically loaded capsule with two phenyl azide units inside is only slightly more stable. The interconversion energy being quite large, the formation of the capsule is the rate-determining step rather than the Huisgen cycloaddition reaction.

We use the Eyring equation to obtain rate constants that are subsequently used in this kinetic model. The results obtained allow us to describe in great detail the processes that are taking place in solution. Notably, it makes possible to bridge an extremely wide range of time and concentration scales. We could reproduce the behavior of the capsule and its guests. We found that the limiting step was the formation of the capsule, which can easily trap and subsequently exchange guests. With this kinetic model, one could easily see the effect of varying the concentration of a given compound. Also, given that some reactions depend on the concentration of the guests only, or the concentrations of the hosts and the guests, or the hosts only, one cannot only rely on the use of an energetic span model, which does not allow this and cannot deal with the complexity of the mechanisms encountered.

To improve the performance of the capsule we conclude that a possible course of action would be to covalently connect two of their neighboring gates, located on two different monomers, in order stabilize the capsule, while still allowing guest exchange from the other gates. Another possibility would be to stabilize the host-guest complex, in order to drive further the

equilibrium between e.g $\mathbf{E}$ and $\mathbf{O}$ on the side of $\mathbf{O}$ or even better on the side of $\mathbf{P}$. We remind the reader that $\mathbf{O}$ is only -1.2 kcal/mol more stable than $\mathbf{E}$, and -3.9 kcal/mol more stable than $\mathbf{P}$. Three ways of achieving this may be imagined. Either (i) modifying the host to bind better to the reactants, (ii) modifying the guests to bind more strongly to their host or (iii) modifying the compatibility of the guests.

Of course such modifications may completely prevent turnover, which should be avoided but is probably the most difficult aspect of rational design of supramolecular catalysts. Results from the previous chapter suggest that limiting the space available in the host may help in this respect. A given host with two binding sites stabilizes very much either of the two reacting guests, but preferably individually. When together, the stabilization would be still sufficient to allow binding, but the destabilization created by their simultaneous presence would help the product formed to evacuate the cavity to allow for a new cycle to take place.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Chapter 5

# Force-Field Tuning as a Possible Solution

## 5.1 Introduction

The description of two systems in Chapters 3 and 4 with various computational methods highlighted that hybrid methods such as ONIOM are promising, particularly when the system is rigid, as it is the case with the cucurbit[6]uril.

The rigidity of the system increases error-cancellation in the MM part. Although convenient, it is not desirable. Ideally, the computational chemist is looking for accuracy and velocity, within a meaningful model and does not want to rely on such error-cancellation.

Unless stated otherwise, in this chapter we will only discuss potential energies: $\mathrm{E_{gas}}$. When we compare molecular mechanics results to the ones obtained by DFT, the latter refers to $\mathrm{E_{gas}}$ after eventual BSSE correction. The numerical values proposed in this chapter may therefore vary from the values of former chapters, since they do not include corrections accounting for entropy and solvation effects.

## 5.2 Choosing a Force Field

We saw in Chapter 2 that the potential energy produced by a force field depends on parameters such as force constants and equilibrium values. The formulas may vary from a force field to another, particularly for non-bonded interactions.

To make a choice, the approach usually consists of checking the historical background of different force fields, that explains what it has been designed

for. Outside this scope, it may be completely inefficient, since parametrization is made on a specific set of systems.

Within Gaussian, only three force fields are available: Amber,[162] Dreiding[163] and UFF.[118] Amber is more focused on nucleic acids and biochemical systems, while Dreiding is more generic and focuses on the prediction of structures. Finally, the universal force field UFF covers the entire periodic table.

We find that Dreiding and Amber are not adapted to our purpose and we saw that the results provided by UFF are largely questionable. To extend the possibilities, W. M. C. Sameera developed in our lab an interface that allows use of other force fields through the program Tinker. Briefly, the interface is a set of scripts that allows Tinker to be easily used with Gaussian, generally (but not necessarily) within a QM/MM model.

## 5.3 Identifying the problem

We used in Chapters 3 and 4 the Force Field MM3, which is often considered as a reference, and to which the performance of other force fields should be compared. In this section, we want to understand what components of the force field generate the error observed, and hope to find which parameters are erroneous.

MM3 revealed to be promising with CB6, and to a lesser extent with the capsule. The problem we identified with Rebek's supermolecule was the ability of the monomers to distort, in other words, the description of his conformational space. While dispersion and hydrogen bonds were evaluated to be acceptable on the system, the description vase-kite equilibrium was largely erroneous. The response of this force field to large distortions is here in question.

In Table 5.1 we recall some of the results obtained with $\mathbf{B97D}$ on the capsule, which we selected as the best method from the benchmark. We then add $\mathbf{UFF}$ and $\mathbf{MM3}$. As we saw already, $\mathbf{MM3}$ performs better than $\mathbf{UFF}$, except for the vase-kite equilibrium, for which a large error is returned.

We cannot provide a detailed breakdown of the potential energy within DFT. However we note that apart from the vase-kite equilibrium, $\mathbf{MM3}$ gas-phase enthalpies are in good agreement. Correcting a force field can then be tricky as improving an aspect of it may deteriorate others. The function types (bond stretching, Van der Waals etc. ) are different with $\mathbf{UFF}$ and $\mathbf{MM3}$, but we can group them in meaningful families and compare them. This breakdown is shown Table 5.2 for $\mathbf{E} \rightarrow \mathbf{R}$.

Geometries and enthalpies obtained with $\mathbf{UFF}$ and $\mathbf{MM3}$ are different,

Table 5.1: *Gas-phase enthalpies of a set of compounds for* **B97D**, **UFF** *and* **MM3**. *Values are in kcal/mol. For* **B97D**, *BSSE corrections are included where applicable.*

| | **B97D** | **UFF** | **MM3** |
|---|---|---|---|
| E → R | 6.5 | 7.3 | 38.5 |
| E → G | -18.3 | -29.2 | -19.9 |
| E → J | -17.3 | -28.3 | -20.1 |
| E → F | -41.5 | -7.6 | -43.9 |
| F → H | -25.3 | -35.6 | -21.8 |
| F → I | -46.9 | -63.7 | -41.1 |
| F → K | -24.1 | -34.7 | -23.5 |
| F → L | -44.0 | -60.6 | -39.9 |

and one can expect a different distribution of the components of the energy as well as error-cancellations. Still, it appears clearly that most of the final energy difference comes from the torsion terms, which is conceptually consistent with an inaccurate description of a conformational equilibrium in **MM3**. Indeed, the vase-like conformer and the kite-like conformer mainly differ by angle and dihedral conformational distortions.

Table 5.2: *Gas-phase energy differences sorted by class for the vase-kite equilibrium, for* **B97D UFF** *and* **MM3**, *in kcal/mol.*

| | **B97D** | **UFF** | **MM3** |
|---|---|---|---|
| Coulomb | | 0.0 | 0.0 |
| Dipole-Dipole | | 0.0 | 2.3 |
| Van der Waals | | 23.8 | 9.5 |
| Bond Stretching | | 7.6 | 1.8 |
| Angle Bending | | 32.9 | 19.0 |
| Torsion | | -57.2 | 5.3 |
| Out-of-Plane Bending | | 0.2 | 0.5 |
| total | 6.5 | 7.3 | 38.5 |

## 5.4 Missing Parameters

At this point, it is worth mentioning that to describe our system, we needed some parameters that are not defined for the MM3 force field. These missing parameters have been added to the MM3 force field we use. We adapted them from other parameters of the standard MM3 parameter set.

For example, at the bottom of the monomer lies 8 ether links between aromatic rings. In organic molecules a "C-O-C" pattern, where carbon atoms

are included in a carbonyl group is quite common, and is defined in the force field. However, if the carbon atoms are regular $sp^2$ alkenes, this is also a "C-O-C", but it appears to be missing. We used the parameters of the first to create the second.

While this kind of choice is questionable, it allows us to define this parameter. A similar procedure is then observed for missing bonds, angles, torsions and other parameters. As combinatorial explosion makes very difficult the task of universal parametrization of all possible combinations of e.g. atom types, angle types and dihedral types, this procedure is common practice. A single parameter missing would prevent an MM calculation from running.

## 5.5 Largest Structural Changes

Considering that the resorcinarene monomer contains 100 atoms, 121 bonds, 195 angles, 304 torsion angles, 4638 van der Waals pairs (and so on), we cannot easily analyze the data by reading the files. Luckily, the use of a parameter is repeated several times by the force field, which simplifies the analysis. Also, we can rely on scripts or simple programs to automatize such procedure.

First, we want to reduce the area to investigate. Coulomb interactions are absent, since we did not assign any explicit charge. We do not want to touch van der Waals parameters, since they appear to be functioning well in the benchmark. Concerning bond stretching and out of plane bending the values obtain should not arise from large error-cancellation, since they are reasonably low in both force fields. Rather, we want to look at the largest structural changes in angles and torsional terms for the vase-kite conformation change.

We wrote a FORTRAN program that allows use to quickly get the information we need. It lists the largest conformational changes, using a threshold that can be applied either on the energy difference provided by a parameter, either on the structural difference. In other words, we can either ask the program to list the largest angle distortion, or the angles that correspond to the largest energy change. We remind the reader that the total energy is the sum of many individual contribution such as the energy of an angle. A angle, e.g. C-C-C that would have a lowest energy a 120 degrees could result in a large change of energy by changing to 140 degrees. However if in a first optimized conformer, it is 130 degrees and in another one it is 110 degrees, the change is still 20 degrees, but the energy may not change as much. Similarly, some parameters may have very low constant, so a large conformational change may result in a low energy change. We want to have

all the cards in hand to improve the force field, this is why we use the two approaches just mentioned.

If we pick the angles that change by more than 2 degrees, we recover 25 out of a total of 195 terms that account together for 16.3 kcal/mol out of 19.0 kcal/mol, which represents more than 86% of the angle energy class. We extract and locate the central atom of those terms. Without surprise, they all belong to the lower part of the monomer, as illustrated Figure 5.1.

About half (10.8 kcal/mol) of the energy change in the angle class originates from the ether links. Specifically, it concerns only four out of the eight C-O-C terms, not the numerous C-C-O terms. Individually it represents between around 2-3 kcal/mol. The main changes in the angle class are therefore well identified.

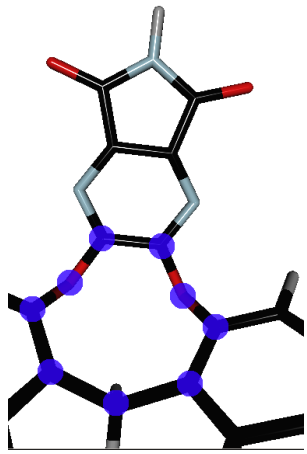


Figure 5.1: *Representation of a substructure of the monomer. The center of the largest angle changes are highlighted in blue (see text).*

## 5.6 Focusing Around the Ether Links

If we pick the torsion angles that change by more than 2 degrees, we recover 172 out of 304 terms. This criterion does not appear to be selective enough. At 5 degrees, we still get 118 terms. They account for 86% of the energy changes in the torsion class. We look at the details and find that all torsional angles involving the oxygen of the ether, (in second or third position, e.g C-O-C-C) return an energy of exactly 0.0 kcal/mol.

Specifically, those torsions corresponds to C($sp^2$)-C($sp^2$)-O(ether,$sp^3$)-C($sp^2$). The attentive reader will notice that *the oxygen of the ether, in second or third position* should in principle affects N($sp^2$)-C($sp^2$)-O(ether,$sp^3$)-

C($sp^2$) dihedrals. The latter was a missing parameter in MM3 which had been copied from the former. From now on, we will refer to them as C-C-O-C and N-C-O-C dihedrals for convenience.

The C-C-O-C parameter in question is a line in the parameter file, which looks like:

**torsion 2 2 6 2 0.000 0.0 1 0.000 180.0 2 0.000 0.0 3**

Let us see what this line means in detail. The torsional contribution of a single torsional angle to the MM3 force field is given by the equation 5.1.

$$\begin{aligned} \mathrm{E}_{\mathrm{dihedral}} = \mathrm{V}_1/2 * (1 + \cos(\psi - \psi_1)) + \mathrm{V}_2/2 * (1 - \cos(2 * (\psi - \psi_2))) \\ + \mathrm{V}_3/2 * (1 + \cos(3 * (\psi - \psi_3)) \end{aligned} \tag{5.1}$$

**torsion** stands for the type of parameter. **2 2 6 2** identifies the dihedral through atom types.

**0.000 0.0 1** says that the first parameter has a constant $\mathrm{V}_1$=0.000 and a equilibrium value $\psi_1$=0.0. The same hold for the rest of the line, with $\mathrm{V}_2$, $\psi_2$, $\mathrm{V}_3$ and $\psi_3$.

To be clear, the parameter line is provided in the parameter file, but will always return an energy of 0.0 kcal/mol regardless of the value of the dihedral, since $\mathrm{V}_1$, $\mathrm{V}_2$ and $\mathrm{V}_3$ all have a zero value. In practice, the parameter is present to allow the calculation running, but its description is missing. This behavior has to be fixed. Among the 118 terms found to correspond to a change of more than 5 degrees, 32 fall in the category C-C-O-C or N-C-O-C. Most of those dihedral correspond to large changes in the value of the dihedral. To discard other possibilities, we check that no other type of dihedrals have "zero" parameters.

We conclude that the error observed in MM3 for the vase-kite equilibrium is very likely to stem from this missing parameter.

In Figure 5.2, we propose to visualize the individual contributions of first, second and third-order terms to the torsional energy, when $\mathrm{V}_1$=$\mathrm{V}_2$=$\mathrm{V}_3$=1. The formula is designed to allow many energy profiles to be modeled, such as the well-known butane C-C rotation diagram. In general, at least one of the terms $\mathrm{V}_1$, $\mathrm{V}_2$ or $\mathrm{V}_3$ is set to 0, depending on the hybridization of the atoms involved.

### 5.6.1 The C-C-O-C parameter

To fix the C-C-O-C (and C-C-O-N) parameter, we use the following procedure:

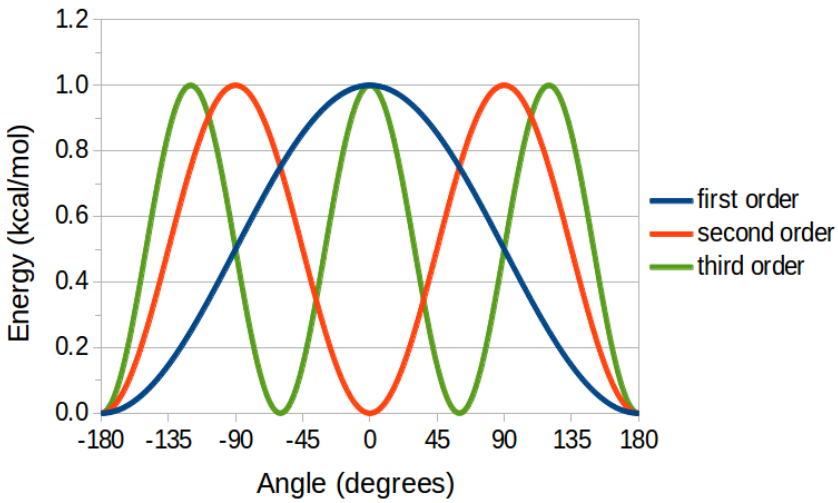



Figure 5.2: *Left: Contribution of the first, second and third order terms to the torsional energy in MM3, with V1=V2=V3=1. Values in kcal/mol.*

(i) Get a full QM- profile of the dihedral using a detailed scan. (ii) Use the set of structures produced by the scan to get an MM profile of the dihedral, where the C-C-O-C dihedral parameters are zero. (iii) Take the difference to extract the profile of the dihedral alone. (iv) Fit the parameters V1, V2 and V3 to this curve. (v) Use the new parameters to try to reproduce the QM profile. The results are proposed figure 5.3.

We should start simplifying the system to the maximum. A diphenyl ether would be a good choice, but we decide to use vinyl ether instead, to get rid of more steric problems. We perform a relax scan around one of the two C-C-O-C dihedrals, with a DFT functional (B97D, blue curve). Doing so, we freeze the other C-C-O-C dihedral to take away its contribution, as the geometries will be reused "as is" to build the C-C-O-C-dihedral-free MM profile (red, Figure 5.3). In green is plotted the difference, which we manually fit with the help of V1,V2 and V3 parameters. The result is the black line. Finally, the purple line shows the total MM energy, once the dihedral parameters are corrected. The agreement seems satisfactory, although not perfect.

The parameters we get are V1=-3.0, V2=6.5 and V3=-1.5. Many arguments go against including them directly in the force field parameters. First, a high peak is observed with MM3, in the absence of the dihedral parameter. If we trust the van der Waals parameters, this could mean that an other parameter such as an -angle bending responds too much for instance. Second, the fitted V1, V2 and V3 are quite far from others is the same category, say, organic *-C-O-* with same hybridization. The latter are on average at

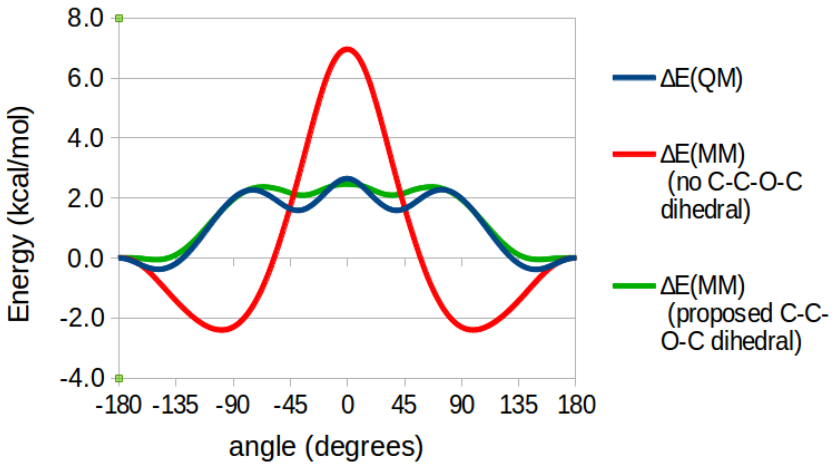



Figure 5.3: *Results obtained of a scan over the C-C-O-C dihedral in a simple test molecule, by QM and MM, before and after correction. Values in kcal/mol.*

V1=0.9, V2=2.8 and V3=-0.8 and variations are small, while the signs do not change. Third, a rotation around a single bond should be quite easy. On the contrary, a high V2 parameter is characteristic of double bonds. Last, those "fitted" parameters do not give acceptable vase versus kite energies (-39.4 kcal/mol, when both structures are optimized again).

We present in Table 5.3 the results they give on our systems. The new torsional parameters favor by a long way the kite conformation, when the situation was reverse before.

Table 5.3: *Gas-phase energy differences sorted by class for the vase-kite equilibrium, for* **UFF** *and* **MM3**, *after the modifications just mentioned, in kcal/mol.*

| | **B97D** | **UFF** | **MM3** | MM3 modified | MM3 modified |
|---|---|---|---|---|---|
| V1/V2/V3 (CCOC) | | | 0.0/0.0/0.0 | -3.0/6.5/-1.5 | 0.0/2.3/0.0 |
| Coulomb | | 0.0 | 0.0 | 0.0 | 0.0 |
| Dipole-Dipole | | 0.0 | 2.3 | 4.6 | 3.8 |
| Van der Waals | | 23.8 | 9.5 | 14.4 | 11.9 |
| Bond Stretching | | 7.6 | 1.8 | 1.5 | 1.3 |
| Angle Bending | | 32.9 | 19.0 | 26.5 | 26.4 |
| Torsion | | -57.2 | 5.3 | -86.4 | -27.9 |
| Out-of-Plane Bending | | 0.2 | 0.5 | -0.1 | 0.0 |
| total | 6.5 | 7.3 | 38.5 | -39.4 | 15.6 |

We therefore decide to abandon temporarily the correction of dihedral. To

go on with the study, we copy the parameters from a line that corresponds to C($sp^2$)-C($sp^2$)-O(ether,$sp^3$)-C(carbonyl). This line contains: V1=0.0, V2=2.3 and V3=0.0, which appears more reasonable. The energy difference we get is now of 15.6 kcal/mol, with both structures optimized, see Table 5.3. The structure of kite appears now much closer to the structure we obtain with B97D. We decide therefore to keep this combination.

### 5.6.2 The C-O-C parameter

We would like to have closer energies as well. To do so, we get back on the issue of the angles. Surprisingly, no C($sp^2$)-O(ether,$sp^3$)-C($sp^2$) exist in the standard MM3 implementation. So far, we have been using the parameters from C(carbonyl)-O(ether,$sp^3$)-C(carbonyl). This choice seems reasonable, although carbonyls are not "generic" $sp^2$ carbons. The MM3 implementation to calculate the angular contribution to the energy is as follows:

$$\mathrm{E_{angle}} = (\mathrm{k_2}/2)(\Delta\phi)^2[1 + \mathrm{k_3}(\Delta\phi) + \mathrm{k_4}(\Delta\phi)^2 + \mathrm{k_5}(\Delta\phi)^3 + \mathrm{k_6}(\Delta\phi)^4] \tag{5.2}$$

$\Delta\phi$ is the deviation of the angle from the equilibrium value set by the force field. $\mathrm{k_3}$ to $\mathrm{k_6}$ are shared parameters across all angles in MM3, so we will not change them. $\mathrm{k_2}$ is the set to 0.77 in the parameter line that defines the C($sp^2$)-O($sp^3$)-C($sp^2$) angle, and 106.8 is the reference angle, from which the deviation is calculated:

| | | | | | |
|---|---|---|---|---|---|
| **angle** | **2** | **6** | **2** | **0.7700** | **106.8** |

We proceed to another scan and fitting with the same vinyl ether molecule, but on the C-O-C angle. Results are shown in Figure 5.4.

The objective being the blue line (QM). The red line is obtained when we put 0 in the force constant $\mathrm{k_2}$ (0.77 $\rightarrow$ 0.00). The current parameters give the green line (MM, with the C-O-C parameter set to 0.77). The situation is not terrible but can be improved. Using our program, we find that all C-O-C angles of vases and kites fall in the range 100-130 degrees. This may change slightly if we implement modifications. To perform the fitting, we want to focus on this region.

To obtain the black (fitted line), we used a $\mathrm{k_2}$ of 1.00, which corresponds to an angle harder to deform. We also displaced the reference angle from 106.8 to 111.3 degrees. The fit is satisfying, so we propose to use it to recalculate the energy of the vase-kite equilibrium. We summarize the results in Table 5.4.

Also the magnitude of the contributions of angle and dihedral bending

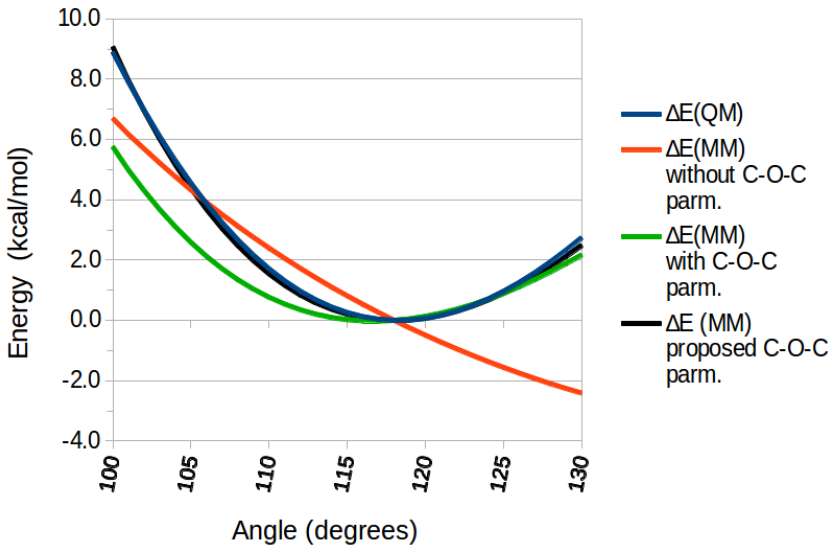



Figure 5.4: *Results of a scan preformed a two levels of theory on a vinyl ether molecule : QM and MM. For MM three curves are proposed, depending on the parameters used. in kcal/mol.*

has changed, the final energy obtain is still 7.3 kcal/mol above the expected value.

### 5.6.3 The C-C-O and N-C-O parameters

The next parameters that may be of relative importance are the C($sp^2$)-C($sp^2$)-O($sp^3$) and N($sp^2$)-C($sp^2$)-O($sp^3$) parameters. They are involved as before in the distortions taking place at the eight ether links. In the MM3 parameter file, the following lines are concerned:

| | | | | | |
|---|---|---|---|---|---|
| **angle** | **2** | **2** | **6** | **0.600** | **121.90** |
| **angle** | **6** | **2** | **37** | **1.200** | **120.00** |

We propose a similar set of QM and MM calculations to test the quality of the C($sp^2$)-C($sp^2$)-O($sp^3$) and N($sp^2$)-C($sp^2$)-O($sp^3$) parameters. To keep the tests as simple as possible and reduce noise from unwanted contributions, we use vinyl alcohol and formamide in its pure enol form to fit the respective angle parameters. The concerned angles are without surprise within 115-125 degrees, we will therefore focus on this region, see Figure 5.5.

We propose a force constant of 1.25 and 1.55, with an equilibrium angle of 119.2 and 120.7 for respectively the C-C-O and N-C-O parameters. We shall see the effect of this new modification in Table 5.5

Table 5.4: *Gas-phase energy differences sorted by class for the vase-kite equilibrium, for* **UFF** *and* **MM3**, *after the modifications just mentioned, in kcal/mol.*

| | **B97D** | **UFF** | MM3 modified |
|---|---|---|---|
| V1/V2/V3 (CCOC) | | | 0.00/2.30/0.00 |
| k / $\phi_0$ (COC) | | | 1.00/ 111.30 |
| Coulomb | | 0.0 | 0.0 |
| Dipole-Dipole | | 0.0 | 3.4 |
| Van der Waals | | 23.8 | 12.2 |
| Bond Stretching | | 7.6 | 1.5 |
| Angle Bending | | 32.9 | 23.7 |
| Torsion | | -57.2 | -27.1 |
| Out-of-Plane Bending | | 0.2 | 0.0 |
| total | 6.5 | 7.3 | 13.8 |

It turns out that the new parameters rise the relative energy of the kite conformer to 20.6 kcal/mol. This is not the result we expected. We could still check the parameters ruling the interactions around the $sp^3$ carbons at the bottom of the calixarene. However, looking at the breakdown of the energy, we do not see more space for improvement on the side of the angles. C($sp^2$)-C($sp^3$)-C($sp^2$) and C($sp^3$)-C($sp^2$)-C($sp^2$) do not appear to undertake major changes, in value or energy. Modifying parameters could result in an endless cascade of changes. We cannot predict if it will be the case, but went through the most obvious alterations. We therefore decide to stop our investigations here.

## 5.7 Conclusion

In this chapter, we presented a possible strategy for improving the quality of a partially defective Force Field. Unfortunately, we were unable to fix the MM3 parameters for our system within a reasonable amount of work.

One should stick to the idea of a minimal the amount of changes and each one of them should be carefully implemented. The missing parameters are generally adapted from other similar ones, so that the interplay of stretching, angles, torsions, van der Waals and other parameters such as out-of-plane bending still corresponds to the same "tone". For instance an angular distortion deforms at the same time surrounding torsional angles, puts pressure on van der Waals "spheres" and finally on covalent bonds.

The heavy setup required to run a Force Field calculation is nowadays only justified by the large size of molecular assemblies. Structural flexibility

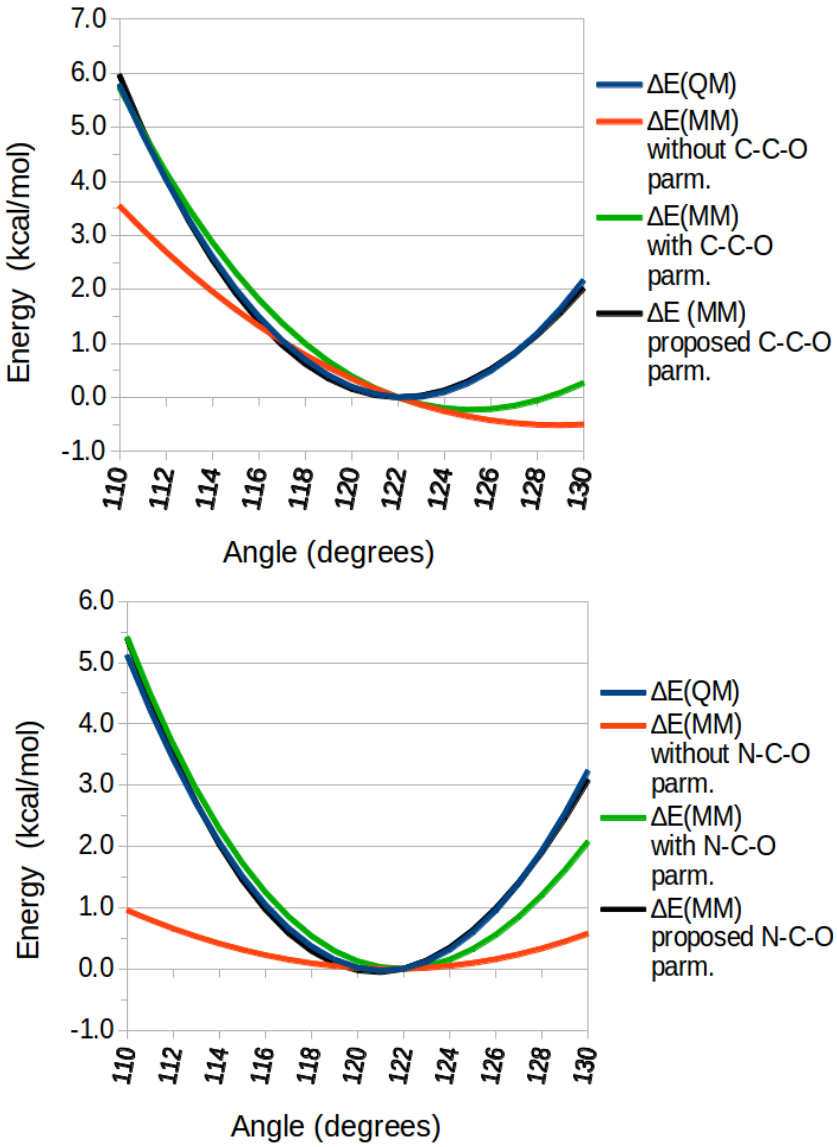



Figure 5.5: *Results of a scan performed on the C-C-O angle of a vinyl alcohol molecule (top) and the N-C-O a formamide molecule (bottom). Two levels of theory are used: QM and MM. For MM, three curves are proposed, depending on the parameters used. in kcal/mol.*

is a major problem for hybrid QM/MM schemes in supramolecular chemistry. Thus, computational chemists interested in this approach will focus on the choice, design or modification of a robust Force Field. In this respect,

Table 5.5: *Gas-phase energy differences sorted by class for the vase-kite equilibrium, for* **UFF** *and* **MM3**, *after the modifications for the C-C-O and N-C-O parameters, in kcal/mol.*

| | **B97D** | **UFF** | MM3 modified |
|---|---|---|---|
| V1/V2/V3 (CCOC) | | | 0.00/2.30/0.0 |
| k / $\phi_0$ (COC) | | | 1.00/ 111.30 |
| k / $\phi_0$ (CCO) | | | 1.25/ 119.20 |
| k / $\phi_0$ (NCO) | | | 1.55/ 120.70 |
| Coulomb | | 0.0 | 0.0 |
| Dipole-Dipole | | 0.0 | 3.6 |
| Van der Waals | | 23.8 | 12.7 |
| Bond Stretching | | 7.6 | 1.7 |
| Angle Bending | | 32.9 | 27.8 |
| Torsion | | -57.2 | -25.3 |
| Out-of-Plane Bending | | 0.2 | 0.0 |
| total | 6.5 | 7.3 | 20.6 |

accurate relative conformational energies represent a milestone. However, transition energies between the set of conformations must not be neglected, since they play a central role in dynamic models.

We think an hybrid approach is still very promising, given the results we present across this work, not forgetting the potential of calculations of much enhanced velocity.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Conclusion

The computational study of the mechanism of the Huisgen cycloaddition between an azide and an acetylene catalyzed by two different hosts, a cucurbit[6]uril and a resorcin[4]arene-base hydrogen-bonded capsule, has led as to the following conclusions concerning host-guest catalysis:

- A good description of the chemistry of these systems is provided by density functional methods with the inclusion of dispersion corrections (DFT-D), for instance B97D, complemented with the use of double-$\zeta$ plus polarization basis sets and the explicit introduction of basis set superposition error (BSSE) corrections. Other tested approaches were not so efficient. Widely used functionals without dispersion corrections (as B3LYP) provided an incorrect description, and M06, despite its large parametrization, was also insufficient.

  The semi-empirical method tested (AM1) provided a very poor description. Molecular mechanics (MM) were not able to describe bond breaking/formation, as expected, but the force fields we tested (UFF, MM3) also struggled in the description of the critical host-guest interactions. The bad performance of MM in describing host-guest interactions hindered also the performance of the ONIOM QM/MM approach, although this may be corrected through a careful tuning of the force field.

- The mechanistic description of these complex reaction systems requires the construction of a reaction network and its evaluation through a code for modeling reaction kinetics. The computed relative free energies can be converted into rate constants through the Eyring equation. Apparent rates taken from experiment can be used for steps under diffusion control.

  The use of this kinetic models is superior to that of the more commonly used energetic span model because it allows the introduction of concentration effects, which are important whenever reaction steps are not purely sequential.

- The operational mechanism of the cucurbit[6]uril-hosted Huisgen reaction is rather simple: the two substrates get into the host and they react. The reaction is complicated by the competition of possible guests for the available positions in the host, but the rigidity of the host simplifies significantly the computational study.

  The catalytic role of the cucurbit[6]uril host on the reaction is to bring the reactants together, rather than stabilizing specifically the transition state. The enthalpic interaction between the host and the transition state is similar to that with the pre-reaction adduct, and cannot thus explain the catalytic effect of the host.

- The catalytic effect of the resorcinarene-based capsule host on the Huisgen reaction is also mostly due on the entropic gains associated to bringing the two reactants together. The enthalpic interaction between the host and the transition state is similar to that with the pre-reaction adduct, and cannot thus explain the catalytic effect of the host. In this case, however, the reaction mechanism is much complicated by the need to form the self-assembled capsule from its components, and by the flexibility of the capsule itself.

- The most stable form for the two resorcin[4]arene components of the capsule is a dimeric kite-kite form where they are mostly planar, and where no space is available for the guests. The transformation from this kite-kite form to the vase-vase form that defines the capsule is induced by the presence of the guests, and happens to be the more energy demanding step of the whole catalytic cycle. The first steps consist of internal rearrangements within the dimer, with no initial separation into monomers.

- Although the presence of guests plays a key role in the resorcinarene-based capsule formation, the guest exchange steps take place through empty capsule intermediates. Exchange mechanisms by guest displacement (one guest pushes out another), or capsule sharing (three guests together inside the capsule) have higher energies.

- Computed values are in general in good agreement with available experimental values. Special care must be taken in choosing the precise parameters to compare, as interpretation of experimental kinetic data is not always straightforward.

- The problem of substrate inhibition remains a difficult one in the systems that have been studied. The product remains stuck in the host, and this blocks the catalytic cycle.

- Computational chemistry is valid tool for the study of host-guest catalysis and, in collaboration with experiment, may help in the development of more efficient systems.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

# Bibliography

[1] J.-M. Lehn, *Angew. Chem. Int. Ed.* **27**(1), 89 (1988).

[2] C. J. Walter, H. L. Anderson, J. K. M. Sanders, *J. Chem. Soc.,1 Chem. Commun.* p. 458 (1993).

[3] Thordarson Pall, Bijsterveld Edward J. A., Rowan Alan E., Nolte Roeland J. M., *Nature* **424**(6951), 915 (2003).

[4] M. Ghadiri, J. Granja, R. Milligan, D. Mcree, N. Khazanovich, *Nature* **366**(6453), 324 (1993).

[5] Adam D., et al., *Nature* **371**(6493), 141 (1994).

[6] D. H. Appella, L. A. Christianson, I. L. Karle, D. R. Powell, S. H. Gellman, *J. Am. Chem. Soc.* **118**(51), 13071 (1996).

[7] P. Terech, R. G. Weiss, *Chem. Rev.* **97**(8), 3133 (1997).

[8] M. M. Conn, J. Rebek, *Chem. Rev.* **97**(5), 1647 (1997).

[9] Fujita Makoto, et al., *Nature* **378**(6556), 469 (1995).

[10] D. Fiedler, R. G. Bergman, K. N. Raymond, *Angew. Chem. Int. Ed.* **43**(48), 6748 (2004).

[11] J. W. Steed, J. L. Atwood, *Supramolecular chemistry*. Wiley (2009).

[12] E. W. Spanagel, W. H. Carothers, *J. Am. Chem. Soc.* **57**(5), 929 (1935).

[13] C. J. Pedersen, *J. Am. Chem. Soc.* **89**(26), 7017 (1967).

[14] C. J. Pedersen, *J. Am. Chem. Soc.* **92**(2), 391 (1970).

[15] C. H. Park, H. E. Simmons, *J. Am. Chem. Soc.* **90**(9), 2431 (1968).

[16] B. Dietricj, J. Lehn, J. Sauvage, *Tetrahedron Lett.* (34), 2885 (1969).

[17] C. Dietrich-Buchecker, J. Sauvage, J. Kintzinger, *Tetrahedron Lett.* **24**(46), 5095 (1983).

[18] C. O. Dietrich-Buchecker, J. P. Sauvage, J. M. Kern, *J. Am. Chem. Soc.* **106**(10), 3043 (1984).

[19] I. T. Harrison, S. Harrison, *J. Am. Chem. Soc.* **89**(22), 5723 (1967).

[20] A. Zinke, E. Ziegler, *Chem. Ber.* **74**(11), 1729 (1941).

[21] B. T. Hayes, R. F. Hunter, *J. Chem. Technol. Biotechnol.* **8**(11), 743 (1958).

[22] R. Breslow, P. Campbell, *J. Am. Chem. Soc.* **91**(11), 3085 (1969).

[23] R. Behrend, E. Meyer, F. Rusche, *Liebigs Ann.* **339**, 1 (1905).

[24] W. Freeman, W. Mock, N. Y. Shih, *J. Am. Chem. Soc.* **103**(24), 7367 (1981).

[25] D. J. Cram, M. T. Blanda, K. Paek, C. B. Knobler, *J. Am. Chem. Soc.* **114**(20), 7765 (1992).

[26] L. Zang, Y. Che, J. S. Moore, *Acc. Chem. Res.* **41**(12), 1596 (2008).

[27] P. Jonkheijm, P. van der Schoot, A. P. H. J. Schenning, E. W. Meijer, *Science* **313**(5783), 80 (2006).

[28] T. Koblenz, J. Wassenaar, J. Reek, *Chem. Soc. Rev.* **37**(2), 247 (2008).

[29] M. Yoshizawa, J. K. Klosterman, M. Fujita, *Angew. Chem. Int. Ed.* **48**(19), 3418 (2009).

[30] R. Wyler, J. de Mendoza, J. Rebek, *Angew. Chem. Int. Ed.* **32**(12), 1699 (1993).

[31] R. S. Meissner, J. Rebek Jr, J. de Mendoza, *Science* **270**(5241), 1485 (1995).

[32] T. Heinz, D. Rudkevich, J. Rebek, *Nature* **394**(6695), 764 (1998).

[33] P. Ballester, G. Gil-Ramírez, *Proc. Natl. Acad. Sci.* **106**(26), 10455 (2009).

[34] M. Fujita, M. Tominaga, A. Hori, B. Therrien, *Acc. Chem. Res.* **38**(4), 369 (2005).

[35] D. L. Caulder, R. E. Powers, T. N. Parac, K. N. Raymond, *Angew. Chem. Int. Ed.* **37**(13-14), 1840 (1998).

[36] F. London, *Trans. Faraday Soc.* **33**, 8b (1937).

[37] A. S. Shetty, J. Zhang, J. S. Moore, *J. Am. Chem. Soc.* **118**(5), 1019 (1996).

[38] T. Yamamoto, et al., *J. Am. Chem. Soc.* **120**(9), 2047 (1998).

[39] G. B. McGaughey, M. Gagné, A. K. Rappé, *J. Biol. Chem.* **273**(25), 15458 (1998).

[40] D. Small, et al., *J. Am. Chem. Soc.* **126**(42), 13850 (2004).

[41] L. R. Rutledge, S. D. Wetmore, *Can. J. Chem.* **88**(8), 815 (2010).

[42] C. A. Hunter, J. K. M. Sanders, *J. Am. Chem. Soc.* **112**(14), 5525 (1990).

[43] M. J. Minch, *An Introduction to Hydrogen Bonding*, vol. 76. Oxford university press New York (1999).

[44] C. Schmuck, W. Wienand, *Angew. Chem. Int. Ed.* **40**(23), 4363 (2001).

[45] S. Anderson, H. L. Anderson, A. Bashall, M. McPartlin, J. K. M. Sanders, *Angew. Chem. Int. Ed.* **34**(10), 1096 (1995).

[46] M. Marty, Z. Clyde-Watson, L. J. Twyman, M. Nakash, J. K. M. Sanders, *Chem. Commun.* pp. 2265–2266 (1998).

[47] R. Cacciapaglia, S. Di Stefano, L. Mandolini, *Acc. Chem. Res.* **37**(2), 113 (2004).

[48] P. Ballester, A. Vidal-Ferran, *Introduction to Supramolecular Catalysis*, pp. 1–27. Wiley-VCH Verlag GmbH & Co. KGaA (2008).

[49] Meeuwissen Jurjen, Reek Joost N. H., *Nat. Chem.* **2**(8), 615 (2010).

[50] R. L. VanEtten, J. F. Sebastian, G. A. Clowes, M. L. Bender, *J. Am. Chem. Soc.* **89**(13), 3242 (1967).

[51] D. C. Rideout, R. Breslow, *J. Am. Chem. Soc.* **102**(26), 7816 (1980).

[52] G. L. Trainor, R. Breslow, *J. Am. Chem. Soc.* **103**(1), 154 (1981).

[53] L. G. Marinescu, M. Bols, *Angew. Chem. Int. Ed.* **45**(28), 4590 (2006).

[54] J. Kang, J. Rebek, *Nature* **385**(6611), 50 (1997).

[55] J. Chen, J. Rebek, *Org. Lett.* **4**(3), 327 (2002).

[56] X. Wang, K. N. Houk, *Org. Lett.* **1**(4), 591 (1999).

[57] C. D. Gutsche, J. A. Levine, *J. Am. Chem. Soc.* **104**(9), 2652 (1982).

[58] J. Moran, S. Karbach, D. Cram, *J. Am. Chem. Soc.* **104**(21), 5826 (1982).

[59] D. W. Johnson, K. N. Raymond, *Inorg. Chem.* **40**(20), 5157 (2001), PMID: 11559075.

[60] C. Hastings, M. Pluth, R. Bergman, K. Raymond, *J. Am. Chem. Soc.* **132**(20), 6938 (2010).

[61] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **40**(11), 2004 (2001).

[62] R. Huisgen, *Angew. Chem. Int. Ed.* **2**(11), 633 (1963).

[63] G. Wittig, A. Krebs, *Ber.* **94**(12), 3260 (1961).

[64] S. T. Abu-Orabi, M. A. Atfah, I. Jibril, F. M. Mari'i, A. A.-S. Ali, *J. Heterocyclic Chem.* **26**(5), 1461 (1989).

[65] S. S. van Berkel, et al., *ChemBioChem* **8**(13), 1504 (2007).

[66] D. H. Ess, G. O. Jones, K. N. Houk, *Org. Lett.* **10**(8), 1633 (2008).

[67] F. Schoenebeck, D. H. Ess, G. O. Jones, K. N. Houk, *J. Am. Chem. Soc.* **131**(23), 8121 (2009), PMID: 19459632.

[68] F. Himo, et al., *J. Am. Chem. Soc.* **127**(1), 210 (2005).

[69] C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **67**(9), 3057 (2002), PMID: 11975567.

[70] H. C. Kolb, K. Sharpless, *Drug Discovery Today* **8**(24), 1128 (2003).

[71] A. K. Feldman, B. Colasson, V. V. Fokin, *Org. Lett.* **6**(22), 3897 (2004), PMID: 15496058.

[72] E. Balducci, L. Bellucci, E. Petricci, M. Taddei, A. Tafi, *J. Org. Chem.* **74**(3), 1314 (2009).

[73] D. Sahu, B. Ganguly, *Tetrahedron Lett.* **54**(38), 5246 (2013).

[74] P. Carlqvist, F. Maseras, *Chem. Commun.* pp. 748–750 (2007).

[75] W. Mock, T. Irra, J. Wepsiec, M. Adhya, *J. Org. Chem.* **54**(22), 5302 (1989).

[76] L. Xu, W. Hua, S. Hua, J. Li, S. Li, *J. Org. Chem.* **78**(8), 3577 (2013).

[77] K. Jongmin, R. Julius, *Nature* **382**(6588), 239 (1996).

[78] D. Cantillo, et al., *Org. Biomol. Chem.* **9**, 7638 (2011).

[79] M. P. Frushicheva, S. Mukherjee, A. Warshel, *J. Phys. Chem. B* **116**(45), 13353 (2012).

[80] W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* **110**(6), 1657 (1988).

[81] A. Szabo, N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Dover Publications (2012).

[82] C. Møller, M. S. Plesset, *Phys. Rev.* **46**, 618 (1934).

[83] J. A. Pople, R. Seeger, R. Krishnan, *Int. J. Quantum Chem.* **12**(S11), 149 (1977).

[84] S. Boys, F. Bernardi, *Mol. Phys.* **19**(4), 553 (1970).

[85] S. Simon, M. Duran, J. Dannenberg, *J. Chem. Phys.* **105**, 11024 (1996).

[86] M. Dewar, E. Zoebisch, E. Healy, J. Stewart, *J. Am. Chem. Soc.* **107**(13), 3902 (1985).

[87] J. J. P. Stewart, *J. Comput. Chem.* **10**(2), 209 (1989).

[88] P. Hohenberg, W. Kohn, *Phys. Rev.* **136**, B864 (1964).

[89] W. Kohn, L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).

[90] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **58**(8), 1200 (1980).

[91] J. P. Perdew, Y. Wang, *Phys. Rev. B* **45**, 13244 (1992).

[92] J. P. Perdew, *Phys. Rev. B* **33**, 8822 (1986).

[93] A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).

[94] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).

[95] Y. Zhang, W. Yang, *Phys. Rev. Lett.* **80**, 890 (1998).

[96] A. D. Becke, *J. Chem. Phys.* **107**(20), 8554 (1997).

[97] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).

[98] N. Handy, A. J. Cohen, *Molecular Physics* **99**(5), 403 (2001).

[99] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* **91**, 146401 (2003).

[100] Y. Zhao, D. G. Truhlar, *J. Chem. Phys.* **125**(19), 194101 (2006).

[101] A. Becke, *J. Chem. Phys.* **98**(7), 5648 (1993).

[102] C. Adamo, V. Barone, *J. Chem. Phys.* **110**, 6158 (1999).

[103] V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, *J. Chem. Phys.* **119**(23), 12129 (2003).

[104] Y. Zhao, D. Truhlar, *Theor. Chem. Acc.* **120**(1), 215 (2008).

[105] Y. Zhao, B. J. Lynch, D. G. Truhlar, *J. Phys. Chem. A* **108**(21), 4786 (2004).

[106] S. Grimme, *J. Chem. Phys.* **124**(3), 034108 (2006).

[107] A. Tarnopolsky, A. Karton, R. Sertchook, D. Vuzman, J. M. L. Martin, *J. Phys. Chem. A* **112**(1), 3 (2008).

[108] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **123**(15), 154101 (2005).

[109] M. Krauss, D. Neumann, W. Stevens, *Chem. Phys. Lett.* **66**(1), 29 (1979).

[110] A. Warshel, M. Levitt, *J. Mol. Biol.* **103**(2), 227 (1976).

[111] M. Svensson, et al., *J. Phys. Chem.* **100**(50), 19357 (1996).

[112] S. Dapprich, I. Komáromi, K. Byun, K. Morokuma, M. J. Frisch, *Theochem* **461–462**(0), 1 (1999).

[113] S. Miertuš, E. Scrocco, J. Tomasi, *Chemical Physics* **55**(1), 117 (1981).

[114] J. Tomasi, B. Mennucci, R. Cammi, *Chemical Reviews* **105**(8), 2999 (2005).

[115] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **113**(18), 6378 (2009).

[116] S. Grimme, J. Antony, T. Schwabe, C. Muck-Lichtenfeld, *Org. Biomol. Chem.* **5**, 741 (2007).

[117] M. J. Frisch, et al., 'Gaussian 09 Revision D.01', Gaussian Inc. Wallingford CT 2009.

[118] A. Rappe, C. Casewit, K. Colwell, W. Goddard Iii, W. Skiff, *J. Am. Chem. Soc.* **114**(25), 10024 (1992).

[119] N. L. Allinger, Y. H. Yuh, J. H. Lii, *J. Am. Chem. Soc.* **111**(23), 8551 (1989).

[120] W. Sameera, F. Maseras, *Unpublished results* .

[121] J. W. Ponder, 'TINKER Molecular Modeling - Software Tools for Molecular Design - http://dasher.wustl.edu/ffe/' (Version 6.2 Feb © 1990-2013).

[122] M. Francl, et al., *J. Chem. Phys.* **77**, 3654 (1982).

[123] W. Sameera, F. Maseras, *Phys. Chem. Chem. Phys.* **13**, 10520 (2011).

[124] S. Grimme, *J. Comput. Chem.* **27**(15), 1787 (2006).

[125] C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **110**(32), 16066 (2006).

[126] H. Eyring, *J. Chem. Phys.* **3**, 107 (1935).

[127] W. Braun, J. T. Herron, D. K. Kahaner, *Int. J. Chem. Kinet.* **20**(1), 51 (1988).

[128] A. Monari, et al., *J. Chem. Theory Comput.* **3**(2), 477 (2007).

[129] D. Asturiol, M. Duran, P. Salvador, *J. Chem. Phys* **128**(14), (2008).

[130] D. Hugas, S. Simon, M. Duran, *Chem. Phys. Lett.* **386**(4–6), 373 (2004).

[131] S. Grimme, *J. Comput. Chem.* **25**(12), 1463 (2004).

[132] T. L. Hill, *J. Chem. Phys.* **16**(4), 399 (1948).

[133] J.-H. Lii, N. L. Allinger, *J. Comput. Chem.* **19**(9), 1001 (1998).

[134] Y.-M. Jeon, J. Kim, D. Whang, K. Kim, *J. Am. Chem. Soc.* **118**(40), 9790 (1996).

[135] M. Von Smoluchowski, *Zeitschrift für physikalische Chemie* **92**(2), 129 (1917).

[136] Barshop, Wrenn, Frieden, 'http://bililite.com/tenua/'.

[137] W. L. Mock, N. Y. Shih, *J. Org. Chem.* **51**(23), 4440 (1986).

[138] A. G. S. Hoegberg, *J. Am. Chem. Soc.* **102**(19), 6046 (1980).

[139] A. G. S. Hoegberg, *J. Org. Chem.* **45**(22), 4498 (1980).

[140] L. C. Palmer, J. Rebek Jr, *Org. Biomol. Chem.* **2**(21), 3051 (2004).

[141] J. Moran, et al., *J. Am. Chem. Soc.* **113**(15), 5707 (1991).

[142] S. Craig, S. Lin, J. Chen, J. Rebek, *J. Am. Chem. Soc.* **124**(30), 8780 (2002).

[143] J. A. Bryant, C. B. Knobler, D. J. Cram, *J. Am. Chem. Soc.* **112**(3), 1254 (1990).

[144] M. Vincenti, E. Dalcanale, P. Soncini, G. Guglielmetti, *J. Am. Chem. Soc.* **112**(1), 445 (1990).

[145] E. Dalcanale, P. Soncini, G. Bacchilega, F. Ugozzoli, *J. Chem. Soc., Chem. Commun.* pp. 500–502 (1989).

[146] M. Korth, M. Pitoňák, J. Řezáč, P. Hobza, *J. Chem. Theory Comput.* **6**(1), 344 (2010).

[147] J. Řezáč, J. Fanfrlík, D. Salahub, P. Hobza, *J. Chem. Theory Comput.* **5**(7), 1749 (2009).

[148] M. Korth, *J. Chem. Theory Comput.* **6**(12), 3808 (2010).

[149] S. E. Wheeler, K. N. Houk, *J. Chem. Theory Comput.* **6**(2), 395 (2010).

[150] T. Amaya, J. Rebek, *J. Am. Chem. Soc.* **126**(43), 14149 (2004).

[151] A. Scarso, L. Trembleau, J. Rebek, *Angew. Chem. Int. Ed.* **42**(44), 5499 (2003).

[152] A. Shivanyuk, A. Scarso, J. Rebek, *Chem. Commun.* (11), 1230 (2003).

[153] Y.-L. Zhao, K. Houk, D. Rechavi, A. Scarso, J. Rebek Jr, *J. Am. Chem. Soc.* **126**(37), 11428 (2004).

[154] O. Hayashida, A. Shivanyuk, J. Rebek, *Angew. Chem. Int. Ed.* **41**(18), 3423 (2002).

[155] M. Yamanaka, A. Shivanyuk, J. Rebek, *Proc. Natl. Acad. Sci.* **101**(9), 2669 (2004).

[156] D. Ajami, J. Rebek, *Angew. Chem. Int. Ed.* **120**(32), 6148 (2008).

[157] A. Scarso, H. Onagi, J. Rebek, *J. Am. Chem. Soc.* **126**(40), 12728 (2004).

[158] J. Rebek Jr, *Chem. Commun.* (27), 2777 (2007).

[159] S. Grimme, *J. Comput. Chem.* **24**(13), 1529 (2003).

[160] S. Grimme, et al., *J. Phys. Chem. A* **110**(8), 2583 (2006).

[161] W. Kirmse, L. Horner, H. Hoffmann, *Liebigs Ann.* **614**(1), 19 (1958).

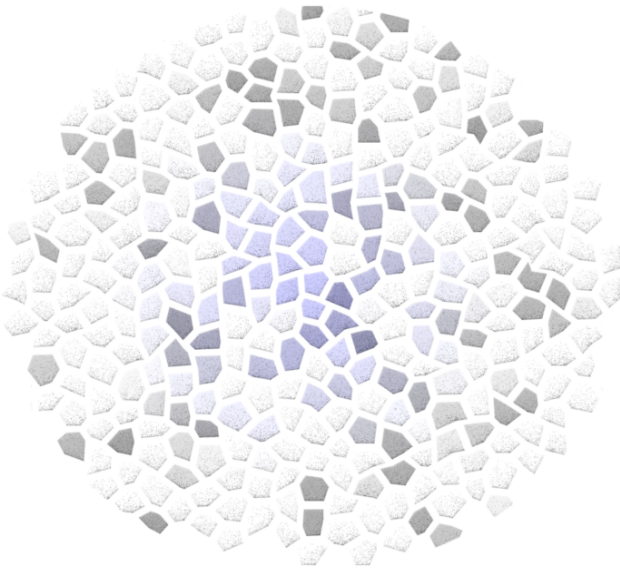
[162] W. D. Cornell, et al., *J. Am. Chem. Soc.* **117**(19), 5179 (1995).

[163] S. L. Mayo, B. D. Olafson, W. A. Goddard, *J. Phys. Chem. A* **94**(26), 8897 (1990).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON HOST-GUEST CATALYSIS.
Charles Goehry
Dipòsit Legal: T 1545-2014



UNIVERSITAT ROVIRA I VIRGILI