# MODELING THE SELF-ASSEMBLY OF CAGE MOLECULES

## Raúl Pérez Soto

**ADVERTIMENT**. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

**ADVERTENCIA.** El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

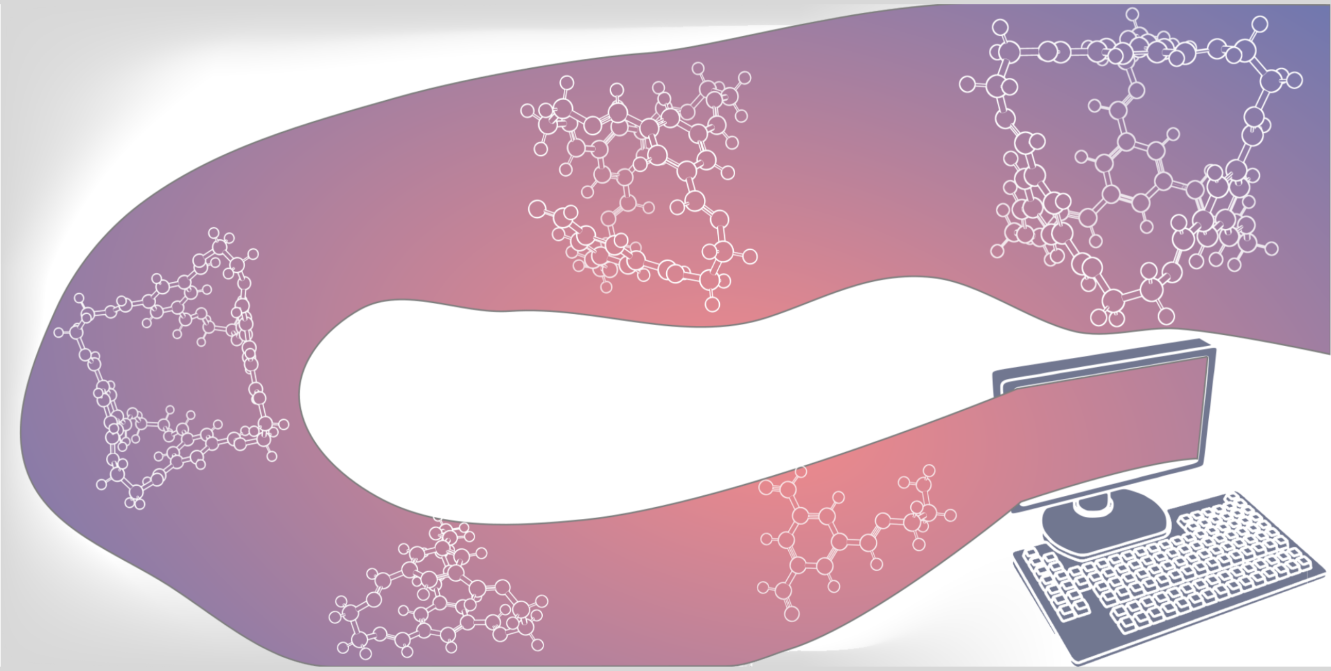
**WARNING**. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

UNIVERSITAT
ROVIRA i VIRGILI

# Modeling the Self-Assembly of Cage Molecules

---

RAÚL PÉREZ SOTO



**DOCTORAL THESIS**
**2022**

*Raúl Pérez Soto*

# Modeling the Self-Assembly of Cage Molecules

DOCTORAL THESIS
Supervised by:
Prof. Feliu Maseras Cuní & Dr. Maria Besora Bonet

UNIVERSITAT ROVIRA i VIRGILI
Tarragona
2022

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

Universitat Rovira i Virgili
Dept. Química Física i Inorgànica
Carrer Marcel·lí Domingo, 1
43007, Tarragona (Spain)

Prof. Feliu Maseras Cuní, Group Leader at the Institute of Chemical Research of Catalonia, and Dr. Maria Besora Bonet, Lecturer at the Universitat Rovira i Virgili.

WE STATE that the present study, entitled "Modeling the self-assembly of cage molecules" presented by Raúl Pérez Soto for the award of the degree of Doctor, has been carried out under our supervision at the Institute of Chemical Research of Catalonia and that it fulfills all the requirements to be eligible for the International Doctor distinction.

Tarragona, June 22nd, 2022

| Doctoral Thesis Supervisor | Doctoral Thesis Supervisor |
|---|---|
| Prof. Feliu Maseras Cuní | Dr. Maria Besora Bonet |

A mi familia.

# Acknowledgments

I would like to thank my parents, Carmen Soto Palomo and Paco Pérez Rodríguez, my sister Irene Pérez Soto and all of my family. Without their support, the doctoral studies that pulled me apart from them would have been impossible.

I have to also mention my previous mentors, Emilio Martínez-Nuñez, Francisco Maldonado-Hódar and Maciej Haranczyk who not only taught me how to be a good researcher but also made me feel comfortable in their groups. I can not forget to mention my previous and great research colleagues Ismael Serrano and Marco Bernabei.

My thesis supervisors, Feliu Maseras and Maria Besora. It was thanks to their careful guidance and insightful discussions that I was able to overcome the intrinsic difficulties of working in a field in which I had close to a null background.

I would like to thank all the wonderful people of the computational groups at ICIQ. Without you being almost like my family, when I was so far apart from my own family, I would have given up on this.

Finally, I would also like to thank Montse Diéguez and Rosa Caballol that have helped me so much with the administrative work as well as Mar Reguero from whom I learned so much about teaching.

*Becas de Formación de Profesorado Universitario*

*FPU18/01138*

ICIQ
Institut
Català
d'Investigació
Química

GOBIERNO
DE ESPAÑA

MINISTERIO
DE CIENCIA
E INNOVACIÓN

AGENCIA
ESTATAL DE
INVESTIGACIÓN

# List of publications

*The Challenge of Reproducing with Calculations Raw Experimental Kinetic Data for an Organic Reaction*

Pérez-Soto, R.; Besora, M.; Maseras, F. *Org. Lett.* **2020**, 22, 8, 2873-2877.

*Pyssian*

Pérez-Soto, R.; Besora, M., Maseras, F. **2021**
Zenodo. https://doi.org/10.5281/zenodo.5055860
Available at github.com/maserasgroup-repo/pyssian

# Table of Contents

**Abstract....................1**
**Introduction to Cage Molecules....................5**
- *1. Cage Molecules....................9*
- *2. Organic Cage molecules....................11*
  - *2.1 Classifications....................13*
    - *2.1.1 By synthetic approach....................13*
    - *2.1.2 By chemical nature....................13*
    - *2.1.3 By topology....................15*
    - *2.1.4 Others....................16*
- *3. Porosity and Applications....................17*
  - *3.1 Porosity....................17*
  - *3.2 Applications....................20*
  - *3.3 Related research....................21*

**Introduction to Kinetic Simulations....................23**
- *1. Anatomy of a Kinetic Simulation....................27*
- *2. Rate laws....................30*
- *3. Activation Energies....................34*
- *4. Reactor models....................36*
  - *4.1 Batch reactor....................36*
  - *4.2. Semi-continuous batch reactor....................38*
  - *4.3 Plug-Flow reactor....................40*
- *5. Practical Implementation....................42*

**General Objectives....................45**
**Chapter I. Study of the Imine Condensation....................49**
- *1. Motivation....................52*
- *2. Introduction....................53*
  - *2.1 Historical background....................53*
  - *2.2 Mechanistic Studies....................54*
  - *2.3 Model Reaction....................55*
  - *2.4 Computational Studies....................56*
- *3. Computational Methods....................58*
  - *Methods of Section 4. Mechanistic Study....................58*
  - *Methods of Section 5. Microkinetic Simulation....................60*
    - *Simulation details....................60*
    - *Nomenclature....................60*
    - *Modeling approach....................61*
  - *Methods of Section 6. Benchmark....................65*
    - *Methods of Section 6.1 Preliminary Benchmark....................65*
    - *Methods of Section 6.2 Implicit solvation and DLPNO-CCSD(T) 66*
    - *Methods of Section 6.3 Microkinetic Benchmark....................67*
  - *Methods of Section 7. Microkinetic Modeling....................67*

*4. Mechanistic Study*.........68
*5. Microkinetic Simulation*.........73
*6. Benchmark*.........74
  *6.1 Preliminary Benchmark*.........74
  *6.2 Implicit solvation and DLPNO-CCSD(T)*.........76
  *6.3 Microkinetic benchmark*.........77
*7. Microkinetic Modeling*.........80
  *7.1 First Hypothesis (H1)*.........80
  *7.2 Second Hypothesis (H2)*.........84
*8. Summary*.........88
**Chapter II. Exploratory Study of the CC1 Cage Synthesis..........91**
*1. Motivation*.........94
*2. Introduction*.........95
  *2.1 Synthesis and Design*.........96
    *2.1.1 CC1 Synthesis*.........100
  *2.2 Thermodynamic Calculations*.........102
  *2.3 Mechanistic Studies*.........105
    *2.3.1 Self-sorting Experiments*.........105
    *2.3.2 Kinetic Studies on Metallocages*.........106
    *2.3.3 Key Previous Studies*.........107
*3. Computational Methods*.........113
*4. Results*.........116
  *4.1 Cage Thermodynamics*.........116
  *4.2 Chain Growth*.........119
    *4.2.1 First Reaction*.........119
    *4.2.2 Amine Monomer Reactions,* $\mathbf{[m+n]^{i} \rightarrow [m+(n+1)]^{i+1}}$.........121
    *4.2.3 Aldehyde Monomer Reactions,* $\mathbf{[m+n]^{i} \rightarrow [(m+1)+n]^{i+1}}$.........123
    *4.2.4 Side Reactions, Diaza-Cope Rearrangement*.........125
    *4.2.5 Initial Chain Growth, Reactions* $\mathbf{[1+n]^{i} \rightarrow [2+m]^{j}}$.........126
  *4.3 Ring Closures*.........129
    *4.3.1* **[2+2]** *Ring*.........129
    *4.3.2* **[3+3]** *Ring*.........131
  *4.4 Cage closures*.........133
    *4.4.1* $\mathbf{[2+3]^{6}}$ *Cage*.........133
    *4.4.2* $\mathbf{[3+4]^{8}}$, $\mathbf{[4+5]^{10}}$ *and* $\mathbf{[4+6]^{10}}$ *incomplete cages*.........135
    *4.4.3* **[4+6]b** *Cage*.........137
    *4.4.4* **[4+6]a** *Cage*.........139
*5. Summary*.........141
**Chapter III. Automated Discovery of the Cage Self-Assembly Reaction Network.........145**
*1. Motivation*.........148
*2. Introduction*.........149
*3. Cage Breaking*.........152
  *3.1 Molecule Representations*.........152

*3.2 Compound and Edge Enumeration.........................................154*
*3.3 Duplicated Compounds Removal............................................156*
*3.3.1 Adducts..........................................................................158*
*3.4 Edge to Reaction Translation.................................................160*
*4. Geometry Generation...................................................................162*
*4.1 Generation...........................................................................162*
*4.2 MM Refinement....................................................................166*
*5. Energy Calculations.....................................................................168*
*5.1 Conformational Sampling.......................................................170*
*5.2 Optimizations and Frequency Calculations..............................170*
*5.3 Clustering.............................................................................170*
*5.4 DFT Single Point Calculations................................................172*
*6. Summary......................................................................................172*
**Chapter IV. Automated Study of the CC1 Cage Synthesis..........175**
*1. Motivation....................................................................................178*
*2. Computational Methods...............................................................179*
*2.1 Thermodynamics...................................................................179*
*MM Constrained optimization....................................................179*
*XTB Metadynamics...................................................................180*
*XTB Optimization and Frequency..............................................182*
*DFT Single Point Calculations...................................................183*
*2.2 Kinetic models.......................................................................184*
*Initial regime of the batch synthesis...........................................184*
*Final regime of the batch synthesis............................................185*
*Flow synthesis...........................................................................186*
*Reactions..................................................................................187*
*3. Thermodynamic Results..............................................................192*
*4. Model 1, RT.................................................................................197*
*5. Model 2, Reactor Effect...............................................................203*
*6. Model 3, Temperature effect........................................................208*
*7. Model 4, Energy per bond...........................................................212*
*8. Model 5, Relative Energies GFN2-xTB.........................................217*
*9. Model 6, Finding the difference between GFN2-xTB and DFT.....226*
*10. Summary....................................................................................233*
**General Conclusions...................................................................237**
**Annex: List of Intermediates of the Self-Assembly of CC1.........243**

# Abbreviations and Acronyms

| | |
|---|---|
| **CHDA** | 1,2-diaminocyclohexane |
| **EtDA** | Ethane-1,2-diamine |
| **TFB** | 1,3,5-triformylbenzene |
| **PhCHO** | Benzaldehyde |
| **NBut** | *n*-butylamine |
| **DCM** | Dichloromethane |
| **SCF** | Self Consistent Field |
| **DFT** | Density Functional Theory |
| **WFT** | Wavefunction Theory |
| **DFTB** | Density Functional Tight Binding |
| **MOF** | Metallic Organic Framework |
| **COF** | Covalent Organic Framework |
| **POM** | Polyoxometalate |
| **CC** | Covalent Cage |
| **OMIM** | Organic Molecule of Intrinsic Microporosity |
| **TCC** | Tubular Covalent Cage |
| **PB** | Porphyrine Box |
| **TS** | Transition State |
| **TST** | Transition State Theory |
| **SP** | Single Point |
| **IRC** | Intrinsic Reaction Coordinate |
| **qh** | quasi-harmonic |
| **TFA** | Trifluoroacetic acid |
| **THF** | Tetrahydrofuran |
| **RDS** | Rate Determining Step |

# Abstract

Porous molecular materials constituted by cage molecules are gathering the interest of the scientific community due to their promising applications, like chemical separations, gas adsorption or as agents to increase the porosity in composite materials.

In the recent years, the development of organic cage molecules has vastly improved. The synthesis of these compounds in one-pot reactions relying on the self-assembly of the constituting monomers is now a common practice. Some steps have been taken to understand the underlying mechanism of the molecule's self-assembly, mostly for metallocages, and to identify which parameters control the reaction outcome, but there is still much to uncover.

In this work we attempt to further expand the understanding on the self-assembly of cage molecules through computational modeling, more specifically focusing on imine cages, which are part of the family of organic cage molecules.

Initially, the reader will be introduced to the field of cage molecules with special emphasis on organic cages. Next, a fundamental tool in the upcoming chapters, kinetic simulations, that relate quantum chemistry calculations with concentrations, will be introduced. Afterwards, the reader will find four chapters containing the main results of this thesis.

In the first chapter, *Study of the Imine Condensation*, we focus on getting a detailed mechanistic insight on the reaction that governs the self-assembly of imine cages. After struggling with the appropriate computational methodology choice, we ended up achieving an acceptable match with experimental results, allowing us to set up the methodology for the study of the self-assembly.

Through the following chapter, *Exploratory Study of the CC1 Cage Synthesis*, we gather the current state of the art on the self-assembly of organic cages and, more specifically, imine cages. From the literature we select the CC1 imine cage as a model system and proceed to study key reactions of its self-assembly based on the knowledge of the previous chapter.

The last two chapters are dedicated to a more thorough exploration of the self-assembly and the impact of parameters like temperature, concentrations and the nature of the monomers. In the chapter *Automated Discovery of the Cage Self-Assembly Reaction Network* we present the algorithm that we designed to enumerate and generate all the intermediates, as well as to enumerate all the reactions. The chapter *Automated Study of the CC1 Cage Synthesis* contains the results of applying the algorithm, which allowed us to elucidate part of the mechanism governing the self-assembly and how it is affected by the different factors.

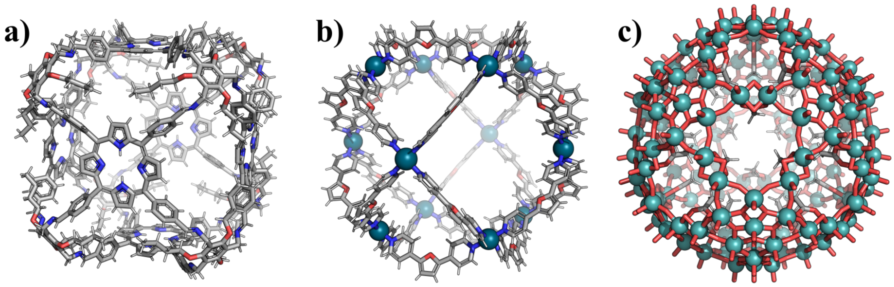
Finally, the general conclusions of each chapter are compiled to summarize the main results of the presented thesis.

# Introduction to Cage Molecules

## 1. Cage Molecules

During the last decades several promising porous materials have emerged, such as intrinsically porous molecular solids,[1] Metallic Organic Frameworks (MOFs) and Covalent Organic Frameworks (COFs). Porous molecular solids are composed of cage compounds. According to the IUPAC cage compounds are defined as *polycyclic compounds having the shape of a cage*.[2] Then, we can infer that the core characteristic of a cage compound is to be able to enclose another compound.[3]

If the space that would be occupied by the guest compound stays empty when no other compound is caged, it is termed as porous cage compound. The solid of a porous molecule is a very good candidate to form a porous solid, although the final porosity of the material will depend on the packing. Amongst the cage compounds of molecular entities we can differentiate three main groups, according to their composition.



$C_{456}H_{444}N_{48}O_{24}$
**Organic Cages**

$C_{336}H_{240}Pd_{12}O_{24}N_{48}$
**Metallocages**

$C_{60}H_{90}Mo_{132}O_{432}$
**Polyoxometalates**

*Figure 1: Example cage compounds. a) organic cage synthesized by Hong et al*[4b] *b) metallocage synthesized by Fujita and coworkers*[5b] *c) keplerate studied by Bo and coworkers.*[6c]

---

[1] Cooper, A. I. *ACS Cent. Sci.* **2017**, 3, 6, 544-553.
[2] Muller, P. *Pure Appl. Chem.* **1994**, 66, 1077.
[3] Acharyya, K.; Mukherjee, P. S. *Angew. Chem. Int. Ed.* **2019**, 58, 26, 8640-8653.

One of the groups is the organic cage molecules[4] (Figure 1a) that will be the main focus of this thesis. As their name indicates, these compounds are composed of main group elements without transition metals. Another group are the metallocages[5] (Figure 1b) usually constituted by organic molecules bridged by transition metals. Finally, we have the polyoxometalates[6] (or POMs, Figure 1c) which are mostly composed of transition metals and oxygens.

Metallocages and organic cages are traditionally closer to organic chemistry whereas POMs are traditionally closer to inorganic chemistry. As a consequence, the term molecular cage has been mostly used in literature to refer to metallocages and organic cages interchangeably, but POMs are rarely mentioned in molecular cages' literature.

Although the self-assembly processes of metallocages and organic cages may share similarities much is unknown yet. Due to the different nature of the bonds bridging the cages together we found reproducing the self-assembly of organic cages much more interesting and challenging. In the next section we will provide an overview on this type of compounds and in the last section of this chapter we will briefly comment on why these compounds have gained the attention of the community and our.

---

[4] a) Zhang, G.; Mastalerz, M. *Chem. Soc. Rev.* **2014**, 43, 1934-1947. b) Hong, S.; Rohman, R.; Jia, J.; Kim, Y.; Moon, D.; Kim, Y.; Ko, Y. H.; Lee, E.; Kim, K. *Angw. Chem. Int. Ed.* **2015**, 54, 13241. (CCDC 1405313)

[5] a) Pullen, S.; Tessarolo, J.; Clever, G. H. *Chem. Sci.* **2021**, 12, 7269-7293. b) Tominaga, M.; Suzuki, K.; Kawano, M.; Kusukawa, T.; Ozeki, T.; Sakamoto, S.; Yamaguchi, K.; Fujita, M. *Angew. Chem. Int. Ed.* **2004**, 43, 5621. (CCDC 238399)

[6] a) Vilà-Nadal, L.; Cronin, L. *Nat. Rev. Mater.* **2017**, 2, 17054. b) Gumerova, N.; Rompel, A. *Chem. Soc. Rev.* **2020**, 49, 7568-7601. c) Melgar, D.; Bandeira, N. A. G.; Avalos, J. B.; Bo, C. *Phys. Chem. Chem. Phys.* **2017**, 19, 5343-5350. (ioChem-BD https://www.iochem-bd.org//handle/10/261)

## 2. Organic Cage molecules

Drawing the line that separates the shape of a ring from the shape of a belt might be defined numerically for macroscopic objects. But, when going down to the molecular scale it is difficult to draw an accurate line between what is a cage and what is not (see Figure 2).

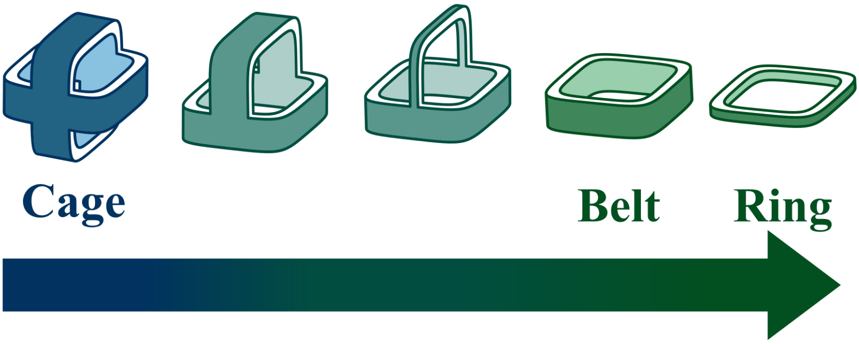


*Figure 2: Example geometries illustrating the transition between the shape of a cage and the shape of a ring.*

As a consequence, setting a starting point for cage molecules becomes difficult. Probably one the oldest influences of current organic cage compounds are the katapinand compounds reported by Park and Simmons in 1968[7] and the cryptands popularized by Dietrich, Lehn and Sauvage after 1969.[8] About a decade later a belt-shaped molecule, the cucurbituril[9] (of great importance in supramolecular chemistry), showing an intrinsically shape-persistent porosity was discovered. In the same decade two ancestors of the current organic cage molecules, the azaparacyclophane synthesized by Yukito,

---

[7] a) Simmons, H. E.; Park, C. H. *J. Am. Chem. Soc.* **1968**, 90, 2428-2429. b) Simmons, H. E.; Park, C. H. *J. Am. Chem. Soc.* **1968**, 90, 2431-2432.
[8] Dietrich, B; Lehn, J. M.; Sauvage, J. P. T*etrahedron Lett* **1969**, 10, 2889-2892.
[9] Freeman, W.; Mock, W.; Shih, N. *J. Am. Chem. Soc.* **1981**, 103, 7367-7368.

Jun-ichi and Takayuki[10] and an early ancestor of the family of imine cages synthetised by MacDowell and Nelson[11] were reported. Some years later, a diels-alder based-cage was synthesized by Ashton et al[12], setting an important precedent to the carbon-based cages.

The discovery of the Noria cage in 2006[13] can be seen as a turning point in the field as the development of organic cage molecules started to attract more attention which peaked in the 2009 paper published by Cooper and coworkers.[14] After this paper, imine cages started their reign over the other families of organic cage molecules. Probably one of the main consequences of this paper was that it showed that porous molecular materials were plausible competitors of other porous materials.

As a consequence a race started for obtaining the organic cage molecule with the largest stable intrinsic cavity, as well as to obtain molecular cage materials with surface areas comparable with other porous materials such as MOFs and COFs. Notable mentions along this race are the cages obtained by Warmuth and coworkers [15] and Mastalerz and coworkers [16].

---

[10] Yukito, M.; Jun-ichi, K.; Takayuki, H. *Chem. Lett.* **1987**, 16, 161-164.

[11] MacDowell, D.; Nelson, J. *Tetrahedron Lett.* **1988**, 29, 385-386.

[12] Ashton, P. R.; Girreser, U.; Giuffrida, D.; Kohnke, F. H.; Mathias, J. P.; Raymo, F. M.; Slawin, A. M. Z.; Stoddart, J. F.; Williams, D. J. *J. Am. Chem. Soc.* **1993**, 115, 5422-5429.

[13] Kudo, H.; Hayashi, R.; Mitani, K.; Yokozawa, T.; Kasuga, N. C.; Nishikubo, T. *Angew. Chem.* **2006**, 118, 8116-8120.

[14] Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I. *Nat. Mater.* **2009**, 8, 973-978.

[15] a) Liu, X.; Liu, Y.; Li, G.; Warmuth, R. *Angew. Chem. Int. Ed.* **2006**, 45, 901-904. b) Liu, X.; Warmuth, R. *J. Am. Chem. Soc.* **2006**, 128, 14120-14127.

[16] Zhang, G.; Presly, O.; White, F.; Oppel, I. M.; Mastalerz, M. *Angew. Chem. Int. Ed.* **2014**, 53, 1516-1520.

The race to increase the surface area gave a milestone in the field, the concept of porous liquid.[17] This concept promised a technological advantage of cage molecules over MOFs and COFs which drove more researchers into the field.

Finally, the last major milestone in the field, in our opinion, was achieved by Jelfs, Cooper and coworkers in 2018.[18] They coupled high throughput experimentation with high throughout computational predictions to synthesize a large amount of new imine cages, presenting a systematic procedure for the discovery of imine cages.

## 2.1 Classifications

### *2.1.1 By synthetic approach*

We can find different criteria to classify organic cage molecules. If we put the focus on the synthesis we find three big categories: synthesized through step-wise protocols, self-assembled and, templated or post-modified.

The step-wise protocols follow the same logic as lots of organic syntheses in chemistry. Self-assembled cages usually rely on one-pot reactions where the constitutional components of the cage end up forming the cage. These cages are the most prominent in the literature and they usually rely on reversible reactions. Finally, we refer to templated cages as cages resulting from modifying already synthesized cages.[19]

### *2.1.2 By chemical nature*

Related to the synthesis but not under the same criteria we can also group organic cages by their chemical nature. Research is mostly divided between three major families: imine cages, boronic ester cages and carbon-based cages.

---

[17] O'Reilly, N.; Giri, N.; James, S. L. *Chem. Eur. J.* **2007**, 13, 3020-3025.

[18] Greenaway, R. L.; Santolini, V.; Bennison, M. J.; Alston, B. M.; Pugh, C. J.; Little, M. A.; Miklitz, M.; Eden-Rump, E. G. B.; Clowes, R.; Shakil, A.; Cuthbertson, H. J.; Armstrong, H.; Briggs, M. E.; Jelfs, K. E.; Cooper, A. I. *Nat. Commun.* **2018**, 9, 2849.

[19] Wang, H.; Jin, Y; Sun, N.; Zhang, W. Jiang, J. *Chem. Soc. Rev.* **2021**, 50, 8874-8886.

Imine cages bridge their constitutional monomers through imine bonds. These are currently the most researched family of organic cages, and their synthesis usually relies on the self-assembly of two monomers, one with aldehyde groups and another with amine groups.

Boronic ester cages bridge their constitutional monomers through a cyclic boronic ester moiety. The interest on these cages has recently raised as examples of this family show promising qualities, such as large intrinsic pores and rigid backbones. Usually their synthesis also relies in the self-assembly of two monomers, one containing vicinial alcohol groups and another containing boronic acid groups. Also, combinations with imine chemistry have already been attempted.[20]

Carbon based cages are bridged through carbon based backbones. An example of this family of compounds can be the Noria cage[13] or the cage synthesized by Avellaneda et al. relying on alkyne metathesis. [21] This group of cages has a much wider variety of reactions and syntheses. Step-wise syntheses were initially used[12,22] and more recently self-assembly-based syntheses, relying on alkyne metathesis, have also been reported. [21,23]

---

[20] Hutin, M.; Bernardinelli, G.; Nitschke, J. R. *Chem. Eur. J.* **2008**, 14, 4585.

[21] Avellaneda, A.; Valente, P.; Burgun, A.; Evans, J. D.; Markwell-Heys, A. W.; Rankine, D.; Nielsen, D. J.; Hill, M. R.; Sumby, C. J.; Doonan, C. J. *Angew. Chem. Int. Ed.* **2013**, 52, 3746.

[22] a) Zhang, C.; Chen, C.-F. *J. Org. Chem.* **2007**, 72, 24, 9339-9341. b) Zhang, C.; Wang, Q.; Long, H.; Zhang, W. *J. Am. Chem. Soc.* **2011**, 133, 51, 20995-21001.

[23] a) Wang, Q.; Zhang, C.; Noll, B. C.; Long, H.; Jin, Y.; Zhang, W. *Angew. Chem. Int. Ed.* **2014**, 53, 10663. b) Lee, S.; Yang, A.; Moneypenny, T. P., II; Moore, J. S. *J. Am. Chem. Soc.* **2016**, 138, 7, 2182-2185.

### *2.1.3 By topology*

We can also classify the cages according to their topology. This classification becomes more prominent in cages synthesized from two monomers. As these organic cages constitute the largest set of studied organic cages, many authors started to refer to the organic cages by the number of monomers required to obtain the cage. With time became standard, specially for imine cages, the nomenclature [2+3], [4+6], [4+4] ... [n+m], where n and m correspond to the number of monomers of each type that constitute the cage (Figure 3) became standard.

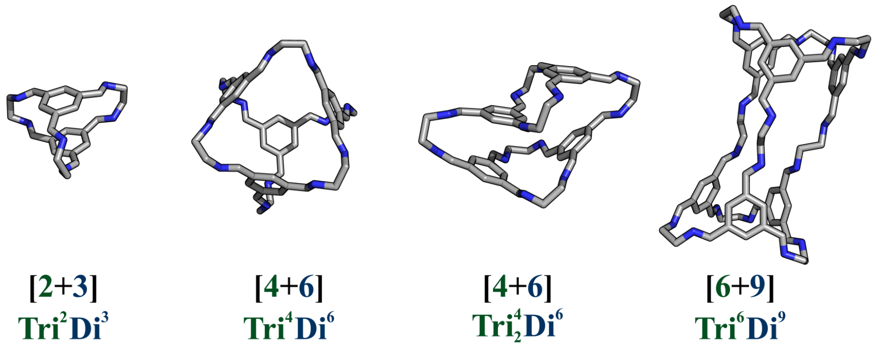


*Figure 3: Examples of possible imine cages topologically different coming from the same monomers.*

Years later, an interesting work by Jelfs and coworkers attempted to upgrade such nomenclature.[24] In this work they related the geometries with the nomenclature that was being proposed (lower row of Figure 3) while attempting to retain the simplicity of the already standarized nomenclature. The proposed nomenclature included the number of functional groups that each one of the monomers had as well as some information of their shape. In the examples showed in Figure 3 the monomers are ethane-1,2-diamine (Di, ditopic) and 1,3,5-triformylbenzene (Tri, Tritopic).

---

[24] Santolini, V.; Miklitz, M.; Berardo, E.; Jelfs, K. E. *Nanoscale* **2017**, 9, 5280-5298.

### *2.1.4 Others*

As a consequence of the growth of the field several other nomenclatures emerged. Tryptycene based cages can also be found under the abbreviation OMIM-1 (Organic Molecule of Intrinsic Microporosity 1), the imine cages developed by Cooper and coworkers initially followed the CC1 and TCC1 nomenclature (Covalent Cage 1 and Tubular Covalent Cage 1 respectively) and a similar nomenclature can be found in more recent cages such as the porphyrine boxes, PB-1 and PB-2 ( Porphyrine Box 1 and 2 respectively).[25] A curious example of how deep this nomenclature took roots is that the cage popularized under the name of CC3-R (usually referred directly as CC3) by Cooper and coworkers[14] which is one of the most studied cages, still retains such name. However, the first report of that compound was by Skowronek and Gawronski.[26]

---

[25] Hong, S.; Rohman, R.; Jia, J.; Kim, Y.; Moon, D.; Kim, Y.; Ko, Y. H.; Lee, E.; Kim, K. *Angew. Chem. Int. Ed.* **2015**, 54, 13241-13244.

[26] Skowronek, P.; Gawronski, J. *Org. Lett.* **2008**, 10, 21, 4755-4758.

## 3. Porosity and Applications

The most important and attractive characteristics of organic cage molecules are the two characteristics that define them: having an accessible intrinsic pore and their molecular nature. First, we will briefly comment the peculiarities of how the intrinsic pore of the cages affects the materials constituted by these cage molecules and then we will provide an overview of the research focused on the application of these compounds.

### 3.1 Porosity

The intrinsic pore of cage molecules makes them very good candidates to form nanoporous solids and/or mesoporous solids. The solids formed from these compounds will be molecular solids, contrary to MOFs and COFs. Molecular solids, in general, are known to not have specially high melting points, which joined to the fact that porous organic cage solids can be dissolved in organic solvents, makes them attractive from an engineering point of view. Two main reasons account to this attractiveness. The first one is that the chemical synthesis and material shaping process can be decoupled. The second one is that the material can be casted into molds with very low energetic consumption compared to, for example, metals.

In these solids of organic cage molecules we can encounter two types of porosity: intrinsic and extrinsic.[27] The intrinsic porosity (highlighted in green in Figure 4) is the most intuitive one, which comes from the inter-connection of the intrinsic pores of the cages, leading to a porous material. The extrinsic porosity (highlighted in blue in Figure 4) instead comes from the inter-connection of the interstitial pores formed as a consequence of the packing.

---

[27] Hasell, T.; Cooper, A. I. *Nat. Rev. Mater.* **2016**, 1, 16053.

As a result, tuning the porosity of these materials can be carried out from various perspectives. One option is to design cages with large intrinsic pores, leading to solids where most of the porosity comes from the intrinsic pores and making the formation of porous amorphous crystals more likely.[28] Another option is to control the crystal polymorphs through the crystallization process, leading to a different connectivity of the intrinsic and/or extrinsic porosity thus obtaining materials with different porosity.

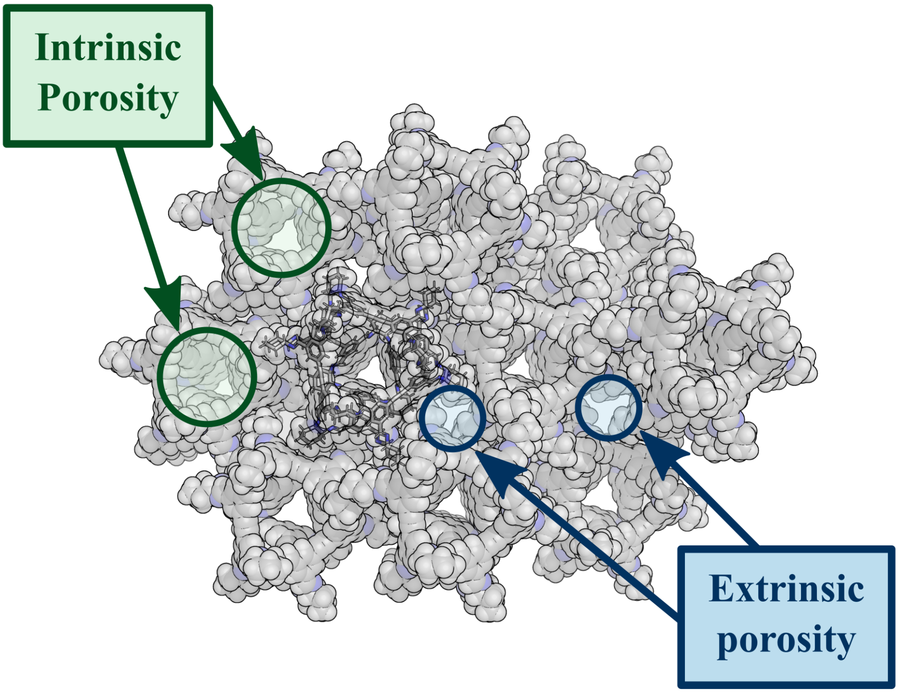


*Figure 4: Schematic slab of a molecular cage crystal highlighting the different types of porosity.*

---

[28] Song, Q.; Jiang, S.; Hasell, T.; Liu, M.; Sun, S.; Cheetham, A. K.; Sivaniah, E.; Cooper, A. I. *Adv. Mater.* **2016**, 28, 2629-2637.

Porosity in these compounds is not restricted to the solid state, they can also retain it in the liquid state. Any liquid shows some degree of transient porosity. We can think of this transient porosity as the equivalent to the extrinsic porosity of the solids, but in the liquid state. Due to the movement of the molecules the interstitial voids are not persistent in time. In 2007 O'Reilly, Giri and James[29] proposed the concept of porous liquid: a liquid also having non-transient porosity. The core idea behind porous liquids is the inclusion of motives in the liquid that have a persistent pore. Here, porous organic cages show promising applications as agents to include those persistent pores. Conceptually 4 types of porous liquids were proposed (see Figure 5).

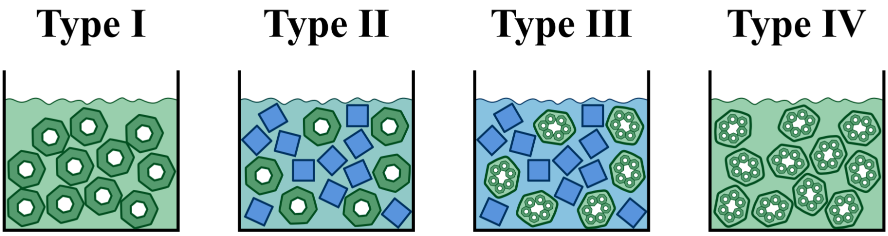


*Figure 5: Schemes illustrating the different types of porous liquids.*

Melting a porous organic cage solid, yields a type I porous liquid. Dissolving the solid instead (with a solvent whose size is too large to fit in the pore) of the cage yields a porous liquid of type II. Instead, dispersing the porous crystal in a liquid forms a porous liquid of type III. And, conceptually, we can think of the porous liquid of type IV as a liquid made entirely of porous crystals. This last type was formally conceived as the liquid or pseudo-liquid state of extended networks like MOFs and COFs.

---

[29] O'Reilly, N.; Giri, N.; James, S. L. *Chem. Eur. J.* **2007**, 13, 3020-3025.

### 3.2 Applications

The applications of organic cages are mostly focused on the applications of the porous material derived from them although we can find interesting computational[30] and experimental[31] work highlighting their ability to host other compounds.

The applications related to the porous materials derived from these compounds can be summarized into three concepts *storage, separation* and *additives*. As a consequence of their permanent porosity they can be used to capture contaminants such as $SO_2$[32] or $CO_2$ [33] and have shown similar capabilities as porous liquids[34].

As the shape of the cavity depends on the molecule, selective adsorption of compounds can be enhanced. This allows these compounds to be used for gas separation[35], as molecular sieves[36] and even in chiral chromatography[37]. Finally, under the application *additives* we refer to their usage to increase the microporosity of other materials (i.e. dispersing them to form porous liquids of type III) or directly their usage in composite materials.[38]

---

[30] Kravchenko, O.; Varava, A.; Pokorny, F. T.; Devaurs, D.; Kavraki, L. R.; Kragic, D. *J. Chem. Inf. Model.* **2020**, 60, 3, 1302-1316.

[31] Liu, Q.; McCaffrey, R.; Zhang, W. *Chem. Asian J.* **2018**, 13, 4, 362-372.

[32] Martínez-Ahumada, E.; He, D.; Berryman, V.; López-Olvera, A.; Hernandez, M.; Jancik, V.; Martis, V.; Vera, M. A.; Lima, E.; Parker, D. J.; Cooper, A. I.; Ibarra, I. A.; Liu, M. *Angew. Chem. Int. Ed.* **2021**, 60, 17556-17563.

[33] Krishnan, K.; Crawford, J. M.; Thallapally, P. K.; Carreon, M. A. *Ind. Eng. Chem. Res.* **2022**, ASAP Article. DOI: 10.1021/acs.iecr.2c00146

[34] Kai, A.; Egleston, B. D.; Tarzia, A.; Clowes, R.; Briggs, M. E., Jelfs, K. E.; Cooper, A. I., Greenaway, R. L. *Adv. Funct. Mater.* **2021**, 31, 2106116.

[35] a) Kewley, A.; Stephenson, A.; Chen, L.; Briggs, M. E.; Hasell, T.; Cooper, A. I. *Chem. Mater.* **2015**, 27, 3207-3210. b) Chen, L.; Reiss, P. S.; Chong, S. Y.; Holden, D.; Jelfs, K. E.; Hasell, T.; Little, M. A.; Kewley, A.; Briggs, M. E.; Stephenson, A.; Thomas, K. M.; Armstrong, J. A.; Bell, J.; Busto, J.; Noel, R.; Liu, J.; Strachan, D. M.; Thallapally, P. K.; Cooper, A. I. *Nat. Mater.* **2014**, 13, 954.

[36] He, A.; Jiang, Z.; Wu, Y.; Hussain, H.; Rawle, J.; Briggs, M. E.; Little, M. A.; Livingston, A. G.; Cooper, A. I. *Nat. Mater.* **2022**, 21, 463-470.

[37] Zhang, J.-H.; Xie, S.-M.; Wang, B.-J.; He, P.-G.; Yuan, L.-M. *J. Chromatogr. A* **2015**, 1426, 174-182.

### 3.3 Related research

A perspective of the literature related with understanding the self-assembly of these compounds will be covered in Chapter II. However, there is some promising research that did not fit the previous sections but it deserves some mention.

First, as previously mentioned, being able to melt the solids to later cast them is a very interesting property from an engineering point of view. Understanding how this process affects the final material was recently covered by Greenaway and coworkers showing a promising improvement of these compounds through their melting and quenching.[39]

From a computational point of view organic cages have been used to improve the state of the art of chemical descriptors[40], improve methodologies used for the high-throughput discovery and prediction of these materials[41] and finally to point towards a plausible application of porous organic cages as membranes for water desalination.[42]

---

[38] a) Bushell, A. F.; Budd, P. M.; Attfield, M. P.; Jones, J. T. A.; Hasell, T.; Cooper, A. I.; Bernardo, P.; Bazzarelli, F.; Clarizia, G.; Jansen, J. C. *Angew. Chem. int. ed.* **2013**, 52, 1253-1256. b) Jiang, S.; Chen, L.; Briggs, M. E.; Hasell, T.; Cooper, A. I. *Chem. Comm.* **2016**, 52, 6895-6898

[39] Brand, M. C.; Greenwell, F.; Clowes, R.; Egleston, B.; Kai, A.; Cooper, A. I.; Bennet, T. D.; Greenaway, R. L. *J. Mater. Chem. A* **2021**, 9, 19807-19816.

[40] Sturluson, A.; Huynh M. T.; York, A. H. P.; Simon, C. M. *ACS Cent. Sci.* **2018**, 4, 12, 1663-1676.

[41] a) Berardo, E.; Turcani, L.; Miklitz, M.; Jelfs, K. E. *Chemical Science* **2018**, 9, 8513-8527. b) Yuan, Q.; Szczypiński, F. T.; Jelfs, K. E. *Digital Discovery* **2022**, 1, 127-138.

[42] a) Kong, X.; Jiang, J. *Phys. Chem. Chem.* Phys. **2017**, 19, 18178-18185. b) Bernabei, M.; Soto, R. P.; García, I. G.; Haranczyk, M. *Mol. Syst. Des. Eng.* **2019**, 4, 912-920. c) Yuan, Y. D.; Dong, J.; Liu, J.; Zhao, D.; Wu, H.; Zhou, W.; Gan, H. X.; Tong, Y. W.; Jiang, J.; Zhao, D. *Nat. Commun.* **2020**, 11, 4927.

# Introduction to Kinetic Simulations

# 1. Anatomy of a Kinetic Simulation

The simulation of microkinetic models is becoming more widespread in the homogeneous computational chemistry and has been an usual tool in the heterogeneous computational chemistry for quite some time already.[43] As the usage of kinetic models is growing in importance we deemed necessary to include here a general overview on the anatomy of a kinetic model (Figure 6). The main goal of this brief section is to present the assumptions and limitations that are present when using kinetic models.

When we want to study a chemical event one of the first things that we need to account for is where is the event happening, is it a sealed flask? is it stirred? is it heated or cooled? By answering these questions we will be able to decide the boundaries of our system of study and how does it interact with the environment. An appropriate name for the target of our study and the identified boundaries, would be chemical system. It includes all species present in the system as well as all possible reactions. It is humanly impossible to consider all of the possible reactions and species present in any realistic chemical system. However, a wide amount of them will not be relevant to the event that we are interested in, as such we select a set of species and reactions that, based on our current knowledge, are most likely to be enough to reproduce the real chemical event. We will refer to this subset species, reactions and boundaries as chemical system model.

---

[43] a) Motagamwala, A. H.; Dumesic, J. A. *Chem. Rev.* **2021**, 121, 2, 1049-1076. b) Matera, S.; Schneider, W. F.; Heyden, A.; Savara, A. *ACS Catal.* **2019** c) Besora, M.; Maseras, F. *WIREs Comput. Mol. Sci.* **2018**, 8:e1372.

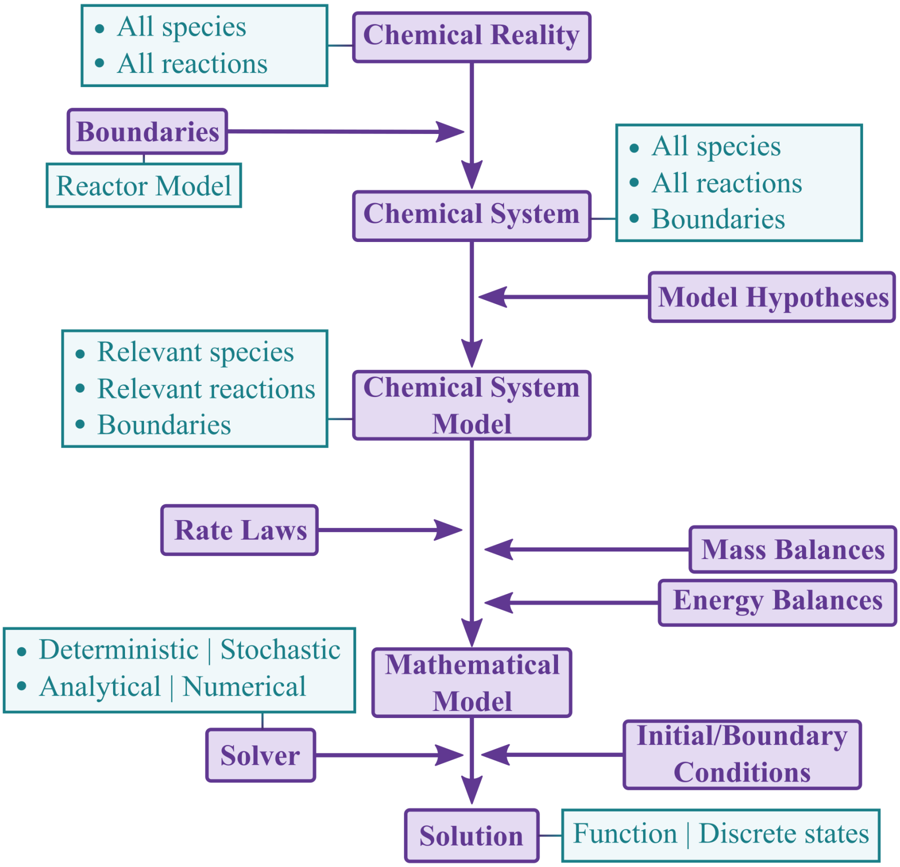

All species
All reactions
Chemical Reality
Boundaries
Reactor Model
All species
All reactions
Boundaries
Chemical System
Model Hypotheses
Relevant species
Relevant reactions
Boundaries
Chemical System Model
Rate Laws
Mass Balances
Energy Balances
Deterministic | Stochastic
Analytical | Numerical
Mathematical Model
Solver
Initial/Boundary Conditions
Solution
Function | Discrete states

*Figure 6: Scheme of the general assumptions and information involved in the simulation of a kinetic model.*

Now that we have our chemical system model we need to transform it into a mathematical model in order to simulate it computationally. For such transformation we need to know a mathematical expression that gives us information of how the concentration of each compound changes as a consequence of each reaction. This mathematical expression is commonly referred as rate law. Not only we need to know how the concentration of each compound varies as a consequence of each reaction but also as a consequence of the interaction with the boundaries. The overall mathematical formulation of this change is the mass balance. Similarly we might need to take into account the variation of energy of the system, when events like cooling or heating occur, which is formalized with the energy balance.

As the last step we need to actually perform the simulation of the mathematical model, which usually involves solving a system of equations (generally a system of differential equations). Solving such system is the task of the solver. The solver may be an algorithm, a collection of algorithms or even an human being. In order to run the simulation we need to provide the solver an initial state (usually the actual details of the reactor, concentrations, volume, influx of species...) these are the boundary or initial conditions. The solution of such mathematical model corresponds to either the mathematical formulation representing the evolution of the system or a discrete set of states that represent the evolution of the system.

Finally, it is important to highlight that the nomenclature of chemical system, chemical system model and boundaries is used here for clarification purposes. They are used to highlight the different hypotheses directly and indirectly considered when constructing a specific mathematical model. Depending on the context we might find that a kinetic model may refer to only the set of relevant species and relevant reactions or that it may directly refer to the mathematical model, which has the boundaries incorporated in its formulation. It would not be unusual also to find authors referring to the set of rate laws (which includes implicitly the set of relevant species and reactions with a specific mathematical formulation) as a kinetic model.

## 2. Rate laws

When describing a mechanism of a reaction, the reaction is subdivided into simpler and simpler steps through a divide and conquer approach. A reaction that corresponds to a step that cannot be further simplified is known as an elementary step. The rate law of an elementary step can be written as equation 1 and one approach to include the reversibility of an elementary step is to consider it as two independent irreversible elementary steps, the forward and reverse direction.

$$rate = k \prod_i C_i \qquad (1)$$

In equation 1, $i$ stands for all the reactants participating in the reaction, $C_i$ corresponds to the concentration of the compound $i$ and $k$ is known as a kinetic constant. A kinetic model composed entirely of elementary steps is usually referred to as microkinetic model.

Currently there are two main theories used for deriving the rate laws of a given elementary step, namely collision theory and transition state theory.[43a,44] As derived from its name, collision theory focuses on the collision of the molecules. Three main types of collisions can be considered, elastic collisions, inelastic collisions and reactive collisions (Figure 7).

In elastic collisions only a change in the trajectory of the colliding particles happens. In inelastic collisions not only we have a change of trajectory but also a change in the internal state of the particle. Reactive collisions involve a change of state and an internal rearrangement. Depending on how we model

---

[44] a) Chorkendorff, I.; Niemantsverdriet, J. W. *Concepts of Modern Catalysis and Kinetics* **2003**, Wiley b) Fernández-Ramos, A.; Miller, J. A.; Klippenstein, S.; Truhlar, D. G. *Chem. Rev.* **2006**, 106, 4518-4584

the molecules we can find different formulations of the kinetic constants with their advantages and drawbacks.[45]

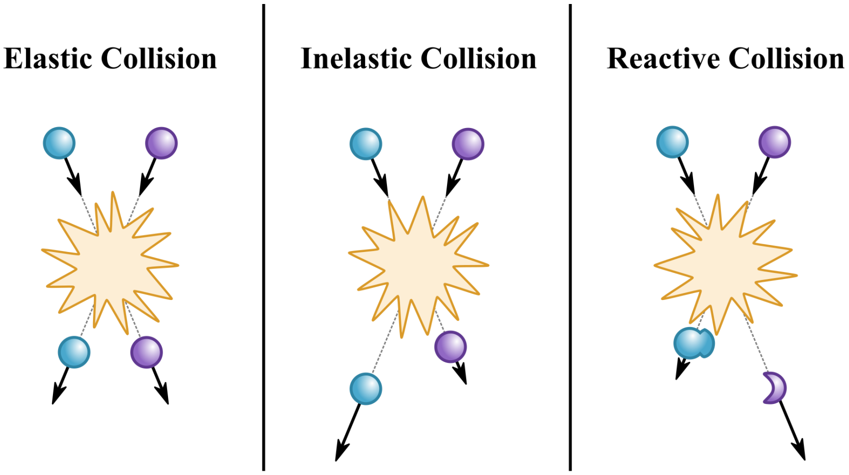


*Figure 7: Types of collisions considered in collision theory.*

Looking through the lens of collision theory helps understanding that the concentrations in the rate law of an elementary step are accounting for the probability of a collision between the reactants. It also becomes easier to understand why trimolecular reactions are not proposed as elementary steps in literature, since the probability of a simultaneous collision of three particles is very low.

Transition State Theory (TST) focuses instead in the concept of transition state. Before, the focus was on the state of the particles before and after the collision, but the core of TST revolves about that state lying in between. Amongst the assumptions of this theory there are two that are important to highlight (summarized in Figure 8). The first one is the hypothesis that we can treat that transition state (also referred as activated complex) as a state in equilibrium with the pre-reaction state (the reactants).

---

[45] a) LeRoy. R. L. *J. Phys. Chem.* **1969**, 73, 4338. b) Menzinger, M.; Wolfgang, R. *Angew. Chem. Int. Ed. Engl.* **1969**, 8, 438. c) Langevin, P. *Ann. Chem. Phys.* **1905**, 5, 245 d) Davidsson, J.; Nyman, G. *Chem. Phys.* **1988**, 125, 171-183.

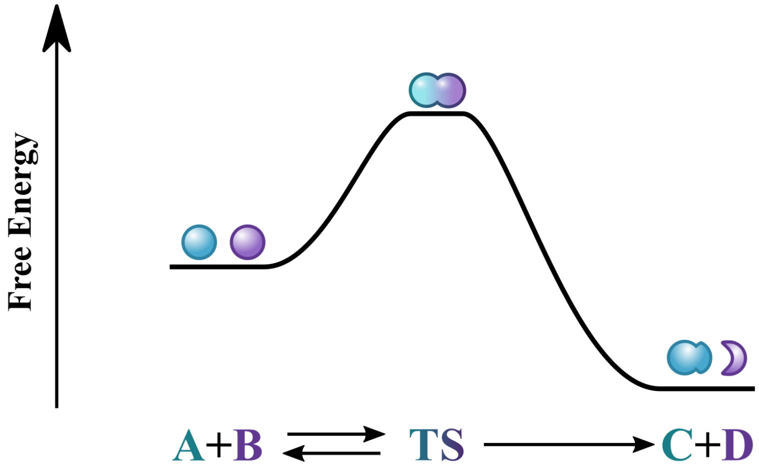


*Figure 8: General reaction happening under TST theory assumptions.*

As a consequence of this assumption, we can relate, using Boltzmann distributions, the relative population of the activated complex with the difference in free energies of both states (activation energy). The second hypothesis assumes that every particle that reaches the transition state progresses forward. With these two approximations the core concept underlying TST can be easily understood.

From TST theory emerges the widely used Eyring-Polanyi equation,[46] equation 2, that allows the mathematical formalization of the rate laws of the elemental steps based on their activation energies.

---

[46] a) Eyring, H. *J. Chem. Phys.* **1935**, 3, 107. b) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, 31, 875-894

$$k=\kappa\frac{k_B T}{h}e^{\frac{-\Delta G^{\ddagger}}{RT}} \qquad (2)$$

Where κ is the transmission factor (used to include the effect of tunneling), T is the temperature, $k_B$ is Boltzmann constant, h is Planck constant, R is the universal gas constant and $\Delta G^{\ddagger}$ is the activation energy, obtained as the difference in free energy between the transition state and the reactants. When no tunneling effect is included the value of κ is set to 1.

## 3. Activation Energies

Finding the barriers (activation energies) of reactions, comparing and discussing them is probably the most active area of research in computational quantum chemistry.[47] Rationalizing reaction mechanisms based on the energies of the different states during an overall reaction event has become customary and lots of information can be inferred from energy profiles, depicting the energies of each of the possible states of the chemical system model and the reactions connecting them (see Figure 8). The accurate calculation of these energies comes at the cost of computational power and time. For this task we find two large groups of methods, wavefunction theory (WFT) based methods and density function theory (DFT) based methods.[48]

Although these groups start from the same base theory and attempt to solve the same problem, reproducing the electronic structure of matter, the approach used to solve it is different. Whereas in WFT the individual character of the electron is emphasized in DFT that individuality is stripped off and the focus is instead the collective behavior of the electrons, the electronic density. Approximations to both, WFT and DFT, to reduce the computational cost in order to allow the calculation of larger systems led to the development of the so-called semi-empirical methods, which own part of their name to the parametrizations included in them. Coming from DFT we find the semi-empirical density functional tight binding (DFTB)[49] and its improvements, DFTB3,[50] GFN-xTB and GFN2-xTB.[51]

---

[47] a) Harvey, J. N.; Himo, F.; Maseras, F.; Perrin, L. *ACS Catal.* **2019**, 8, 6803-6813. b) Fey, N.; Lynam, J. M. *WIREs Comput. Mol. Sci.* **2021**, e1590. c) Lewis-Atwell, T.; Townsend, P. A.; Grayson, M. N. *WIREs Comput. Mol. Sci.* **2021**, e1593.

[48] a) Evangelista, F. *J. Chem. Phys.* **2018**, 149, 030901. b) Dasgupta, S.; Lambros, E.; Perdew, J. P.; Paesani, F. *Nat. Commun.* **2021**, 12, 6359. c) Atkins, P. W; Friedman, R. S. *Molecular Quantum Mechanics* **2011**, OUP Oxford. d) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* **1996**, Dover Publications Inc. e) Parr, R. G.; Yang, W. *Density-functional theory of atoms and molecules.* **1989**, Oxford University Press, New York, New York.

Going one step down on the accuracy ladder and one step up in speed, we find the force field methods. [52] These methods rely in Newtonian mechanics instead of quantum mechanics. Force field methods rely entirely on their parametrizations. They will reproduce experimental geometries and energy differences at a good speed-accuracy ratio if and only if they were parametrized for that specific system or the model has a great similarity with the systems used for the parametrization.

---

[49] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, Th.; Suhai, S.; Seifert, G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, 58, 7260-7268.

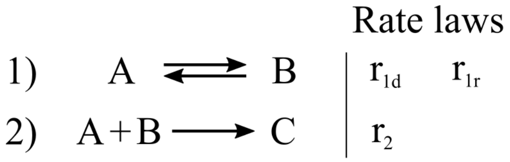
[50] a) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, 111, 10861-10873. b) Gauss, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, 7, 931-948.

[51] a) Grimme, S.; Bannwarth, C.; Shushkov, P. *J. Chem. Theory. Comput.* **2017**, 13, 1989-2009. b) Bannwarth, C.; Ehlert, S.; Grimme, S. *J. Chem. Theory Comput.***2019**, 15, 3, 1652-1671.

[52] Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. *Chem. Rev.* **2021**, 121,16,10142-10186.

## 4. Reactor models

The derivation of the mass and energy balances of the different basic reactor models can be found profusely in the literature.[53] Here we will only highlight the mathematical derivation under isothermal conditions of the mass balances of the three reactor models used in the thesis. To simplify the mathematical nomenclature we will consider the set of species and reactions in Scheme 1.



*Scheme 1: Reactions and variables (that represent their rate laws) used as examples for the balances*

### 4.1 Batch reactor

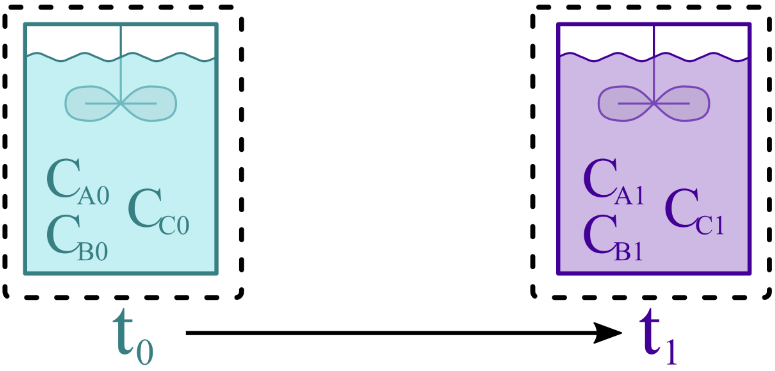


*Figure 9: Scheme of the batch reactor model*

---

[53] a) Smith, J. M.; Smith, J. M. *Chemical Engineering Kinetics* **1981**, McGraw-Hill. b) Brodkey, R. S.; Hershey, H. C. *Transport Phenomena: A Unified approach.* **1987**, McGraw-Hill.

In the batch reactor (Figure 9) we have a closed system with no interaction with the environment. We already indicated that the processes are assumed to be isothermal and as we will be assuming that there is no change in the volume we can write the mass balance of the compound A as equation 3.

$$\frac{\partial n_A}{\partial t} = -V r_{1d} + V r_{1r} - V r_2 \qquad (3)$$

Here we need to remember that the different rate laws ($r_{1d}$, $r_{1r}$, $r_2$) give the variation of the concentration with respect to time as a consequence of each reaction, which is the reason that each rate law is multiplied by the volume (V). $n_A$ in equation 3 corresponds to the number of molecules of A. As the volume is constant we can further simplify the expression to equation 4.

$$\frac{\partial C_A}{\partial t} = -r_{1d} + r_{1r} - r_2 \qquad (4)$$

Where $C_A$ corresponds to the concentration of A in the reactor. We then apply the same procedure to the other species, ending with the following system of differential equations (equations 5-7).

$$\frac{\partial C_A}{\partial t} = -r_{1d} + r_{1r} - r_2 \qquad (5)$$

$$\frac{\partial C_B}{\partial t} = +r_{1d} - r_{1r} - r_2 \qquad (6)$$

$$\frac{\partial C_C}{\partial t} = +r_2 \qquad (7)$$

The solution of this system in the time interval $[t_0, t_1]$ and with the initial conditions $C_{A0}$, $C_{B0}$ and $C_{C0}$ is the simulation of the reactor.

### 4.2. Semi-continuous batch reactor

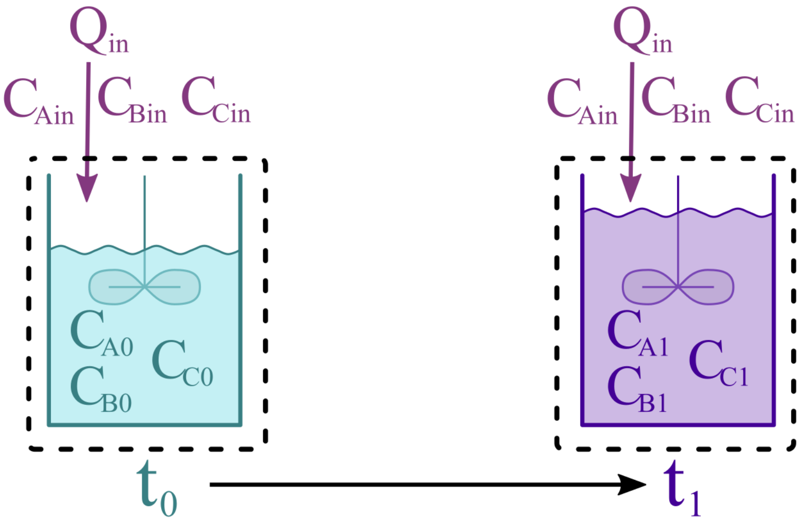


*Figure 10: Scheme of a semi-continuous batch reactor model*

The only difference between the semi-continuous batch reactor (see Figure 10) and the batch reactor (Figure 9) is that we have a volumetric flow ($Q_{in}$) that enters the reactor during the whole reaction process. As a consequence we cannot assume that the volume is constant. The mass balance for A cannot be as compact as equation 2 so we end up with equation 8 where $C_{Ain}$ is the concentration of the A compound in the inflow.

$$\frac{\partial n_A}{\partial t} = + Q_{\text{in}} C_{Ain} - V r_{1d} + V r_{1r} - V r_2 \tag{8}$$

As we cannot assume the volume constant and we need to describe how the volume varies in the system we need to include it in the system of differential equations (equation 12). To do so we will assume that the inflow has a constant volumetric rate and that the volumes are additive. As a consequence our final set of differential equations corresponds to equations 9-12.

$$\frac{\partial n_A}{\partial t}=Q_{\mathrm{in}}C_{Ain}+V\left(-r_{1d}+r_{1r}-r_2\right) \tag{9}$$

$$\frac{\partial n_B}{\partial t}=Q_{\mathrm{in}}C_{Bin}+V\left(+r_{1d}-r_{1r}-r_2\right) \tag{10}$$

$$\frac{\partial n_C}{\partial t}=Q_{\mathrm{in}}C_{Cin}+V\,r_2 \tag{11}$$

$$\frac{\partial V}{\partial t}=Q_{\mathrm{in}} \tag{12}$$

Solving the system in the time interval $[t_0,t_1]$ with the initial conditions $C_{A0}$, $C_{B0}$ and $C_{C0}$ and with the initial volume corresponds to the simulation of the reactor.

## 4.3 Plug-Flow reactor

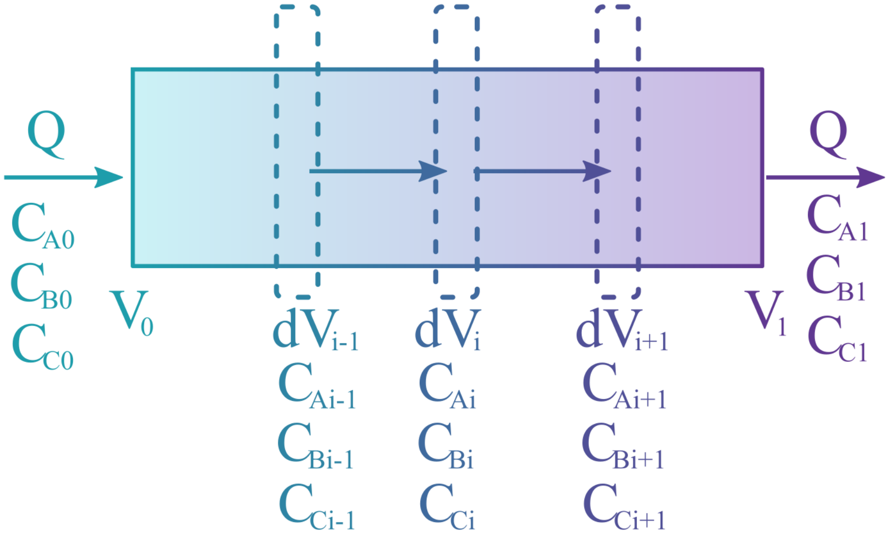


*Figure 11: Scheme of a plug-flow reactor model*

Contrary to the two previous models (batch and semi-continuous batch), where we were considering an homogeneous mixture, in the Plug-Flow reactor (Figure 11) we can only make that assumption on a differential volume slice. As with the semi-continuous batch reactor, we will assume that the volumetric flow is constant.

A general approximation used in the modeling of these type of reactors is to assume that the reactor is operating at the stationary state. At that state, the concentrations at each differential volume slice do not change with time. The mass balance for the concentration of A in a differential slab of the reactor can be seen in equation 13 which is equivalent to equation 14 which trivially reduces to equation 15.

$$\left.\frac{\partial n_A}{\partial t}\right]_i = Q\left(C_{Ai-1} - C_{Ai}\right) + dV\left(-r_{1d} + r_{1r} - r_2\right) \tag{13}$$

$$\left.\frac{\partial n_A}{\partial t}\right]_i = Q\left(C_A - \left(C_A + dC_A\right)\right) + dV\left(-r_{1d} + r_{1r} - r_2\right) \tag{14}$$

$$\left.\frac{\partial n_A}{\partial t}\right]_i = -Q\,dC_A + dV\left(-r_{1d} + r_{1r} - r_2\right) \tag{15}$$

As previously indicated, we are assuming a stationary state of the reactor which makes equation 15 equal to 0 and leading to the final differential equation, equation 16. Following suit with the other species we obtain the final system of differential equations (equations 16-18).

$$\frac{\partial C_A}{\partial V} = \frac{1}{Q}\left(-r_{1d} + r_{1r} - r_2\right) \tag{16}$$

$$\frac{\partial C_B}{\partial V} = \frac{1}{Q}\left(+r_{1d} - r_{1r} - r_2\right) \tag{17}$$

$$\frac{\partial C_C}{\partial V} = \frac{1}{Q}\left(+r_2\right) \tag{18}$$

Solving the system for the volume interval $[V_0, V_1]$ with the initial conditions $C_{A0}$, $C_{B0}$ and $C_{C0}$ corresponds to the simulation of the reactor.

## **5. Practical Implementation**

With the contents of the previous sections it would be natural to question how all of these different components were actually computed for the presented thesis.

We start with the computation of the activation energies with implementations of WFT and DFT. We used the Gaussian[54] and ORCA[55] packages for standard WFT and DFT calculations. As most of these calculations were carried out in Gaussian we developed a python library, named pyssian,[56] coupled with some useful command-line programs to alleviate the human effort.

For semi-empirical energies and dynamics we used the GFN2-xTB semi-empirical method, and consequently the XTB[57] software. For force field calculations we used the OpenBabel[58] software, which also served as file format converting tool.

---

[54] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, revision D.01; Gaussian, Inc. Wallingford, CT, 2013.

[55] Neese, F. *Wiley Interdiscip. Rev: Comput. mol. Sci.* **2017**, 8, e1327.

[56] Pérez-Soto, R.; Besora, M.; Maseras, F.; pyssian v1.0.2, **2021**. maserasgroup-repo/pyssian: 1.0.2 (v1.0.2). Zenodo. https://doi.org/10.5281/zenodo.5055860

[57] Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, P.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. *WIREs Comput. Mol. Sci.* **2020**, 11, e01493.

[58] (a) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminform* **2011**, 3, 33. (b) Open Babel, version 3.1.0, compiled from source code obtained at https://github.com/openbabel/openbabel.

Regarding the setup and calculation of the kinetic and microkinetic models nowadays there is a wide variety of software either already used by the community such as COPASI[59] or recently developed such as Overreact.[60] But we required a more flexible framework, ideally written in python, that would allow us to decouple the simulation of the model and the generation of the model. As a consequence of this, we developed our own framework to generate and simulate the models.

---

[59] Hoops, S.; Sahle, S.; Gauges, R.; Lee, C.; Pahle, J.; Simus, N.; Singhal, M.; Xu, L.; Mendes, P.; Kummer, U. *Bioinformatics* **2006**, 22, 24, 3067-3074.

[60] Schneider, F. S. S.; Caramori, G. F. *J. Comput. Chem.* **2022**, 123.

# General Objectives

Understanding the self-assembly processes of cage molecules facilitates their design, which pushes forward the development of tailored porous materials for specific compound separations and/or storage.

From a broader perspective, the main objective of this thesis is to set a base of knowledge and a systematic procedure to approach this study from a computational perspective, without neglecting the experimental conditions and their effects. We will study the self-assembly of the CC1 imine cage under different synthetic conditions which will serve as model of the self-assembly of the cage molecules. For such task we have divided the work into different chapters that cover the following general objectives:

- Understand the mechanism of the reaction that drives the self-assembly (Chapter I)
- Test if we can reproduce experimental results of the reaction that drives the self-assembly through computational means (Chapter I).
- Find trends in the barriers of the different reactions (chain growth, ring closures, cage closures) involved in the self-assembly (Chapter II).
- Design an algorithm to automate the generation and enumeration of all possible intermediates as well as the enumeration of all the reactions involved in the self-assembly. (Chapter III)
- Extract the underlying mechanism of the self-assembly of CC1 (Chapter IV).
- Extract which intermediates play a major influence in the prediction of the outcome (Chapter IV).
- Understand how concentration and temperature can influence the self-assembly mechanism (Chapter IV).

# Chapter I. Study of the Imine Condensation

## 1. Motivation

In the self-assembly of the imine cages, the imine bond formation is relevant as it is the responsible for the creation of new carbon-nitrogen bonds. We start this chapter by gathering detailed knowledge on the mechanism of the imine bond formation. We first compile the knowledge on the imine bond formation from literature and look for a model system to study computationally. As a result we select the reaction of benzaldehyde with n-butylamine mainly due to the availability of data and the similarity with some imine cage monomers.

We then move to explore computationally the mechanism of the reaction and perform kinetic simulations to compare it with experiments. Due to the unexpected results we then proceed to explore the computational methodology running a benchmark study. Finally, as consequence of the non-conclusive outcome of the benchmark we attempt to bridge the gap between simulation and experiment from a microkinetic modeling perspective.

# 2. Introduction

## 2.1 Historical background

An imine is a functional group or organic compound containing the C=N moiety.[1] By analogy with the C=O moiety where aldehydes and ketones are included, amongst the imines we find, for instance aldimines and ketimines.

In 1864 a landmark discovery by Hugo Schiff[2] led to the development of a new family of imine compounds, later known as Schiff bases. As the term imine was not coined until some years later, attributed to Albert Ladenburg in 1883,[3] the term Schiff base took roots into the community. Schiff reported a condensation reaction between aldehydes and amines bearing this new type of compounds.

With the passing of the years these compounds earned the attention of the scientific community. Such attention rose to the point that by 1940 we can even find in an imine review published by Sprung the sentence *"So many aromatic Schiff bases are described in the literature that it is virtually impossible to list even a representative number of them."* [4].

We can see how the condensation of an aldehyde (or ketone) and primary amine had become the most common synthesis of imine compounds in a review published in 1963 but, knowledge on its mechanism was still scarce.[5]

---

[1] IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by McNaught, A. D.; Wilkinson, A. Blackwell Scientific Publication, Oxford (1997). Online version (**2019**-) created by Chalk, S. J. ISBN 0-9678550-9-8.

[2] Schiff, H. *Justus Liebigs Ann. Chem.* **1864**, 131, 118−119.

[3] Ladenburg, A. *Ber. Dtsch. Chem. Ges.* **1883**, 12, 1125-1694.

[4] Sprung, M. A. *Chem. Rev.* **1940**, 26, 3, 297-338.

[5] a) Campbell, K. N.; Sommers, A. H.; Campbell, B. K. *J. Am. Chem. Soc.* **1944**, 66, 1, 82-84. b) Campbell, K. N.; Helbing, C. H.; Florkowski, M. P.; Campbell, B. K. *J. Am. Chem. Soc.* **1948**, 70, 11, 3868-3870. c) Layer, R. W. *Chem Rev.* **1963**, 63, 5, 489-510.

During the following years the mechanistic picture got clearer[6] and a general consensus was achieved, establishing the reaction as a two step reaction with an hemiaminal formation as the first step and an hemiaminal dehydration as the second step (see Figure 12).

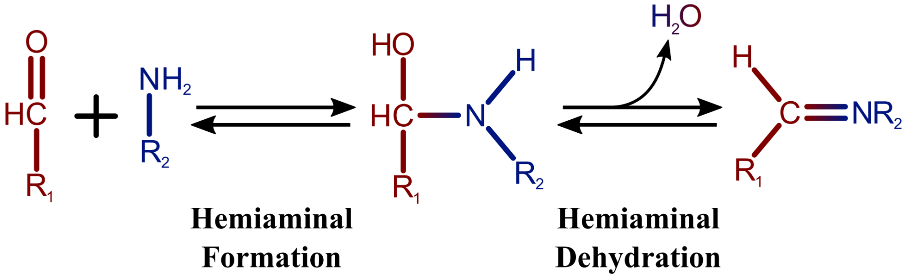


*Figure 12: Historically accepted mechanism of the Imine Condensation reaction.*

### 2.2 Mechanistic Studies

Probably, the most exhaustive experimental work on the imine condensation reaction is the series of studies carried out by Sayer and coworkers in the 1970s. In 1973 they published a study of the hemiaminal dehydration step[7] which they followed in 1974 with a study on the hemiaminal formation.[8] A few years later, in 1977, they improved the understanding of the hemiaminal dehydration step[9] and in 1980, through a study on the imine hydrolysis, they end up discussing the concertedness of the proton transfer at the dehydration.[10]

---

[6] a) Hine, J.; Menon, B. C.; Jensen, J. H.; Mulders, J. *J. Am. Chem. Soc.* **1966**, 88, 14, 3367-3373. b) Jencks, W. P. *Chem. Rev.* **1972**, 72, 6, 705-718. c) Hine, J.; Via, F. A. *J. Am. Chem. Soc.* **1972**, 94, 1, 190-194. d) Lopez-Nieves, M.; Wentzell, P. D.; Crouch, S. R. *Anal. Chem.* **1990**, 62, 3, 304-308. e) Urbansky, E. T. *J. Chem. Educ.* **2000**, 77, 12, 1644-1647.

[7] Sayer, J. M.; Peskin, M.; Jencks, W. P. *J. Am. Chem. Soc.* **1973**, 95, 13, 4277-4287.

[8] Sayer, J. M.; Pinsky, B; Schonbrunn, A.; Washtien, W. *J. Am. Chem. Soc.* **1974**, 96, 26, 7998-8009.

[9] Sayer, J. M.; Jencks, W. P. *J. Am. Chem. Soc.* **1977**, 99, 2, 464-474.

[10] Sayer, J. M.; Conlon, P. *J. Am. Chem. Soc.* **1980**, 102, 10, 3592-3600.

Amongst their results, they provide very interesting conclusions regarding the timings of the different events involved in the reactions. They found that in the hemiaminal formation the C-N bond formation is very likely coupled with a proton transfer to the oxygen atom. Also, that a stepwise mechanism involving a zwitterionic intermediate with the formation of the C-N bond prior to the protonation of the oxygen is improbable. Regarding the dehydration, they found that an initial proton abstraction from the N to form an anionic intermediate is very unlikely. They concluded that the formation of the C=N double bond is coupled to some extent with the C-O cleavage. And, with acid catalysis, they found that the proton transfer to the oxygen is also coupled with the C-O cleavage.

The thorough studies by Sayer and coworkers were carried out with water as solvent. However, the imine condensation reactions have been carried out in a a variety of different solvents, protic and aprotic, and experimental mechanistic studies in other solvents have been scarce, as highlighted in a 2015 review.[11] Ciaccia and Di Stefano, authors of the mentioned review, had a previous paper discussing the transimination mechanisms in aprotic media while also including experimental data on the condensation between the benzaldehyde and n-butylamine in dichloromethane.[12]

## 2.3 Model Reaction

It is from this work[12] that we select our model reaction (Figure 13b) to study the imine condensation, although it was not the main focus of the experimental article. This reaction between the benzaldehyde and the *n*-butylamine is an ideal model system for us due to two main factors.

---

[11] Ciaccia, M.; Di Stefano, S. *Org. Biomol. Chem.* **2015**, 13, 646-654.

[12] Ciaccia, M.; Cacciapaglia, R.; Mencarelli, P.; Mandolini, L.; di Stefano, S. *Chem. Sci.* **2013**, 4, 2253–2261.

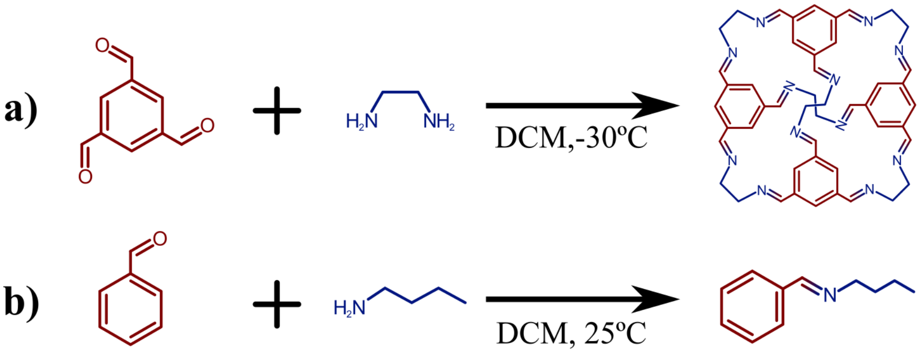


*Figure 13: Model reactions and solvent in the current thesis for a) Self-assembly of an imine cage, CC1, and b) Condensation between an aldehyde an amine.*

First, they followed the evolution of the reaction through NMR and reported the raw data (concentration *vs* time), which allows us to benchmark our computational results. Being able to compare our calculations with experiments helps us in the selection of an appropriate theoretical methodology (DFT functional and basis set mainly).

Second, the solvent (dichloromethane, DCM) and the reactants are similar to the monomers used in the synthesis of the CC1 imine cage (see Figure 13a) that we will study in the following chapters.

Because of this we will focus in this chapter on the computational exploration of the mechanism of the reaction between benzaldehyde and *n*-butylamine in dichloromethane.

## 2.4 Computational Studies

Moving to the computational studies on the mechanism of the imine condensation we find a remarkable change between the early amine to carbonyl addition studies of Yamataka et al. in 1986[13] and the most recent studies in which the the formation of the carbon-nitrogen bond is, just, another step of the more complex mechanism.[14]

Before DFT calculations came into play we can already find mechanistic studies where a proton shuttle mechanism is more favorable than the uncatalyzed mechanism.[15] Although such mechanistic hypothesis did not take root immediately[16] it soon grew in popularity.[17] Proof of it is the 2007 study on a related reaction by Patil and Sunoj where they carry out a very thorough investigation on the role of the proton shuttle.[18]

With the improvement of the computers and computational chemistry the size of the systems studied grew,[19] and the focus of the research shifted to other topics such as regioselectivity[20] or indirect applications such as reductive amination reactions.[14,21]

---

[13] Yamataka, H.; Nagase, S.; Ando, T.; Hanafusa, T. *J. Am. Chem. Soc.* **1986**, 108, 4, 601-606.

[14] Jameel, F; Stein, M. *J. Catal.* **2022**, 405, 24-34.

[15] a) Williams, I. H. *J. Am. Chem. Soc.* **1987**, 109, 21, 6299-6307. b) Hall, N. E.; Smith, B. J. *J. Phys. Chem. A* **1998**, 102, 25, 4930-4938.

[16] Pliego J. R.; Alcântara, A. F. de C.; Veloso, D. P.; Almeida, W. B. *J. Braz. Chem. Soc.* **1999**, 10, 381-388.

[17] a) Salvà, A.; Donoso, J.; Frau, J.; Muñoz, F. *J. Phys. Chem. A* **2003**, 107, 44, 9409-9414. b) Mascavage, L. M.; Sonnet, P. E.; Dalton, D. R. *J. Org. Chem.* **2006**, 71, 9, 3435-3443.

[18] Patil, M. P.; Sunoj, R. B. *J. Org. Chem.* **2007**, 72, 22, 8202-8215.

[19] a) Vilanova, B.; Gallardo, J. M.; Caldés, C.; Adrover, M.; Ortega-Castro, J.; Muñoz, F.; Donoso, J. *J. Phys. Chem. A* **2012**, 116, 8, 1897-1905. b) Casasnovas, R.; Adrover, M.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Muñoz, F. *J. Phys. Chem. B*, **2012**, 116,35,10665-10675. c) Solís-Calero, C.; Ortega-Castro, J.;Hernández-Laguna, A. ; Muñoz, F. *Theor. Chem. Acc.* **2012**, 131, 1263.

[20] Calow, A. D. J.; Carbó, J. J.; Cid, J.; Fernández, E.; Whiting, A. *J. Org. Chem.***2014**, 79, 11, 5163-5172.

[21] Boz, E.; Tüzün, N. Ş.; Stein, M. *RSC. Adv.* **2018**, 8, 64, 36662-36674.

## 3. Computational Methods

### Methods of Section 4. Mechanistic Study

| | |
|---|---|
| **Software** | Gaussian 09 revision D.01[22] |
| **Calculation type** | optimization and frequencies. *opt freq* |
| **Functional** | B3LYP[23] |
| **Basis Set** | cc-pVTZ for all atoms.[24] |
| **Empirical Dispersion** | GD3.[25] *empiricaldispersion=gd3* |
| **Symmetry** | disabled. *nosymm* |
| **Solvation** | Implicit. SMD for dichloromethane[26]. *scrf(smd,dichloromethane)* |
| **Integration grid** | 75 radial shells and 302 angular points per shell.[27] Gaussian09 default. *int=(Grid=finegrid)* |
| **Frequencies** | RRHO approximations. |
| **Free energy corrections** | Reference state corrections, liquid phase at 1M and 298.15K[28] (manually applied) |

---

[22] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, revision D.01*; Gaussian, Inc. Wallingford, CT, **2013**.

[23] a) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, 98, 11623−11627. b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, 37, 785−789. c) Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648−5652.

[24] Dunning, T. H. *J. Chem. Phys.* **1989**, 90, 1007.

[25] Grimme, S. *J. Comput. Chem.* **2006**, 27, 1787−1799.

[26] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, 113, 6378−6396.

[27] Lebedev, V. I. *Zh. Vychisl. Mat. Mat. Fiz.* **1976**, 16, 293-306.

[28] Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2011**, 115, 14556−14562.

| | |
|---|---|
| **Software** | Gaussian 09 revision D.01 |
| **Calculation type** | Forwards and Reverse IRC calculations. *irc=(calcfc,maxpoints=1000, Recalc=5,stepsize=4,lqa)* |
| **Functional** | B3LYP |
| **Basis Set** | cc-pVTZ for all atoms. |
| **Algorithm** | LQA |
| **Empirical Dispersion** | GD3. *empiricaldispersion=gd3* |
| **Symmetry** | disabled. *nosymm* |
| **Solvation** | Implicit. SMD for dichloromethane. *scrf(smd,dichloromethane)* |
| **Integration grid** | 75 radial shells and 302 angular points per shell. Gaussian09 default. *int=(Grid=finegrid)* |
| **Frequencies** | RRHO approximations. |

IRC calculations with this setup were only carried out for the Transition States with the water dimers. To ensure that the other Transition States are connected with the appropiate minima a manual displacement along the imaginary vibrational mode of 0.13 times the amplitude of the frequency was applied and the resulting geometries were optimized to minima.

## Methods of Section 5. Microkinetic Simulation

*Simulation details*

| | |
|---|---|
| **Boundaries** | Isothermal. (Constant temperature)<br>Liquid Phase. (Constant volume)<br>Batch reactor model (Closed System) |
| **Temperature** | 298.15K |
| **Kinetic constants** | Calculated through Eyring Equation.[29] |
| **Tunneling effect** | Not considered. |
| **Solver type** | Numerical |
| **Solver** | LSODA[30] (from ODEPACK, as used by scipy[31]) |
| **Tolerances** | Relative tolerance (rtol): $10^{-6}$<br>Absolute tolerance (atol): $10^{-11}$ |
| **Others** | Scipy defaults for the remaining parameters. |
| **Report time step** | 0.1s |
| **Simulation time** | 5000s |
| **Initial concentrations** | benzaldehyde: 0.060M<br>n-butylamine: 0.060M<br>water: 0.001M<br>other species: 0.000M |

*Nomenclature*

The following tables contain the set of reactions considered in all the sections of the current chapter. In these reactions **A** corresponds to benzaldehyde, **B** to *n*-butylamine, **W** to water, **H** to 1-butylamino-1-phenylmetanol and **I** to (E)-N-butyl-1-phenylmethanimine. The species following the convention **[i+j+k]** correspond to adducts of the corresponding **i**,**j**,**k** compounds, i.e. **[W+W]** would correspond to an adduct of two waters.

---

[29] a) Eyring, H. *J. Chem. Phys.* **1935**, 3, 107. b) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, 31, 875-894

[30] Hindmarsh, A. *IMACS Transactions on Scientific Computation* **1983**, 1, 55-64.

[31] a) Oliphant, T. E. *Comp. Sci. Eng.* **2007**, 9, 10-20. b) Millman, K. J.; Aivazis, M. *Comp. Sci. Eng.* **2011**, 13, 9-12.

*Modeling approach*

We detail here our modeling approach for the sake of clarity. Let us consider a general catalyst molecule **C** that catalyzes the hemiaminal dehydration, as example.

Overall reactions: $\mathrm{H + C \rightleftarrows I + W + C}$

The first main hypothesis considered is that every bond-forming and bond-breaking reaction which involves a calculated TS (hemiaminal formation and hemiaminal dehydration) is unimolecular. This means that: first, the molecules form an adduct; and second, that this adduct is the one that proceeds through the TS.

Overall reaction: $\mathrm{H + C \rightleftarrows I + W + C}$
Step 1: $\mathrm{H + C \rightleftarrows [H+C]}$
Step 2: $\mathrm{[H + C] \rightleftarrows [I + W + C]}$

The second main hypothesis considered is that all possible aggregations or dissociations are considered.

Overall reaction: $\mathrm{[I + W + C] \rightleftarrows I + W + C}$
Step 1a: $\mathrm{[I + W + C] \rightleftarrows [I + W] + C}$
Step 1b: $\mathrm{[I + W + C] \rightleftarrows [I + C] + W}$
Step 2a: $\mathrm{[I + W] \rightleftarrows I + W}$
Step 2b: $\mathrm{[I + C] \rightleftarrows I + C}$

The third main hypothesis is that every aggregation and dissociation steps are considered to be fast diffusion-controlled reversible elemental steps. Where by “fast” we mean that we can model them as reactions with an “estimated” barrier of 2.0 kcal mol$^{-1}$ in the exergonic direction of the reaction. This estimated barrier comes from previous studies on difussion processes.[32] The full set of elemental steps considered for the microkinetic models is provided in the following tables:

---

[32] a) Besora, M.; Maseras, F. *WIREs Comput. Mol. Sci.* **2018**, 8:e1372. b) Goehry, C.; Besora, M.; Maseras, F. *ACS Catal.* **2015**, 5, 4, 2445-2451.

**Reactant aggregations**

| **Reaction IDs Direct, Reverse** | **Reaction** | | |
|---|---|---|---|
| R000, R001 | A + B | ⇄ | [A+B] |
| R002, R003 | A + W | ⇄ | [A+W] |
| R004, R005 | B + W | ⇄ | [B+W] |
| R006, R007 | B + B | ⇄ | [B+B] |
| R008, R009 | W + W | ⇄ | [W+W] |
| R010, R011 | A + [B+W] | ⇄ | [A+B+W] |
| R012, R013 | A + [B+B] | ⇄ | [A+B+B] |
| R014, R015 | A + [W+W] | ⇄ | [A+W+W] |
| R016, R017 | B + [A+B] | ⇄ | [A+B+B] |
| R018, R019 | B + [A+W] | ⇄ | [A+B+W] |
| R020, R021 | B + [B+W] | ⇄ | [B+B+W] |
| R022, R023 | B + [W+W] | ⇄ | [B+W+W] |
| R024, R025 | W + [A+B] | ⇄ | [A+B+W] |
| R026, R027 | W + [A+W] | ⇄ | [A+W+W] |
| R028, R029 | W + [B+W] | ⇄ | [B+W+W] |
| R030, R031 | W + [B+B] | ⇄ | [B+B+W] |
| R032, R033 | W + [W+W] | ⇄ | [W+W+W] |
| R034, R035 | [A+B] + [B+W] | ⇄ | [A+B+B+W] |
| R036, R037 | [A+B] + [W+W] | ⇄ | [A+B+W+W] |
| R038, R039 | [A+W] + [B+W] | ⇄ | [A+B+W+W] |
| R040, R041 | [A+W] + [B+B] | ⇄ | [A+B+B+W] |

### Hemiaminal Formation

| Reaction IDs Direct, Reverse | Reaction | | |
|---|---|---|---|
| R042, R043 | [A+B] | ⇄ | H |
| R044, R045 | [A+B+W] | ⇄ | [H+W] |
| R046, R047 | [A+B+B] | ⇄ | [H+B] |
| R048, R049 | [A+B+B+W] | ⇄ | [H+B+W] |
| R050, R051 | [A+B+W+W] | ⇄ | [H+W+W] |

### Hemiaminal Diffusion

| Reaction IDs Direct, Reverse | Reaction | | |
|---|---|---|---|
| R052, R053 | [H+W] | ⇄ | H + W |
| R054, R055 | [H+B] | ⇄ | H + B |
| R056, R057 | [H+B+W] | ⇄ | H + [B+W] |
| R058, R059 | [H+B+W] | ⇄ | [H+B] + W |
| R060, R061 | [H+B+W] | ⇄ | [H+W] + B |
| R062, R063 | [H+W+W] | ⇄ | H + [W+W] |
| R064, R065 | [H+W+W] | ⇄ | [H+W] + W |

### Hemiaminal Dehydration

| Reaction IDs Direct, Reverse | Reaction | | |
|---|---|---|---|
| R066, R067 | H | ⇄ | [I+W] |
| R068, R069 | [H+W] | ⇄ | [I+W+W] |
| R070, R071 | [H+B] | ⇄ | [I+B+W] |
| R072, R073 | [H+B+W] | ⇄ | [I+B+W+W] |
| R074, R075 | [H+W+W] | ⇄ | [I+W+W+W] |

### Product Diffusion

| Reaction IDs Direct, Reverse | Reaction | | |
|---|---|---|---|
| R076, R077 | [I+W] | ⇄ | I + W |
| R078, R079 | [I+W+W] | ⇄ | I + [W+W] |
| R080, R081 | [I+W+W] | ⇄ | [I+W] + W |
| R082, R083 | [I+B+W] | ⇄ | I + [B+W] |
| R084, R085 | [I+B+W] | ⇄ | [I+B] + W |
| R086, R087 | [I+B+W] | ⇄ | [I+W] + B |
| R088, R089 | [I+B+W+W] | ⇄ | I + [B+W+W] |
| R090, R091 | [I+B+W+W] | ⇄ | [I+B] + [W+W] |
| R092, R093 | [I+B+W+W] | ⇄ | [I+W] + [B+W] |
| R094, R095 | [I+B+W+W] | ⇄ | [I+W+W] + B |
| R096, R097 | [I+B+W+W] | ⇄ | [I+B+W] + W |
| R098, R099 | [I+W+W+W] | ⇄ | I + [W+W+W] |
| R100, R101 | [I+W+W+W] | ⇄ | [I+W] + [W+W] |
| R102, R103 | [I+W+W+W] | ⇄ | [I+W+W] + W |

## Methods of Section 6. Benchmark

### *Methods of Section 6.1 Preliminary Benchmark*

**Details of the DFT calculations**

| | |
|---|---|
| **Software** | Gaussian 09 revision D.01 |
| **Calculation Type** | Single point, optimization and frequencies. *opt freq* |
| **Input Geometry** | Geometries optimized with B3LYP-D3/6-31+g(d,p) |
| **Functionals** | B3LYP, BP86,[33] PBEPBE[34] (commonly referred as PBE), PBE1PBE[35] (commonly referred as PBE0), M06,[36] M06-2X,[36] ωB97X-D [37] |
| **Basis Sets** | 6-31+g(d), 6-31+g(d,p) or 6-311++g(d,p) [38] |
| **Empirical Dispersion** | GD3 included for all functionals explicitly, except for ωB97X-D. *empiricaldispersion=gd3* |
| **Symmetry** | Disabled. *nosymm* |
| **Solvation** | Implicit. SMD for dichloromethane. *scrf(smd,dichloromethane)* |
| **Integration grid** | 75 radial shells and 302 angular points per shell. Gaussian09 default. *int=(Grid=finegrid)* |
| **Frequencies** | RRHO approximations. Gaussian09 default. |
| **Free energy corrections** | Reference state corrections, liquid phase at 1M and 298.15K (manually applied) |

---

[33] a) Becke, A. D. *Phys. Rev. A* **1988**, 38, 3098. b) Perdew, J. P. *Phys. Rev. B* **1986**, 33, 8822.

[34] a) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77, 3865-3868. b) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, 78, 1396.

[35] a) Ernzerhof, M.; Scuseria G. E. *J. Chem. Phys.* **1999**, 110, 5029. b) Adamo, C. *J. Chem. Phys* **1999**, 110, 6158.

[36] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 120, 215-241.

[37] Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, 10, 6615-6620.

[38] a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, 54, 724. b) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, 28, 213-222.

**Details of the DLPNO-CCSD(T) calculations**

| | |
|---|---|
| **Software** | ORCA4.0[39] |
| **Calculation Type** | Single point |
| **Input Geometry** | Geometries optimized with B3LYP-D3/6-31+g(d,p) |
| **Method** | DLPNO-CCSD(T)[40] |
| **Basis Sets** | cc-pVDZ, cc-pVTZ, cc-pVQZ or cc-pV5Z for all atoms. |
| **Auxiliary basis functions** | autoaux[41] option to select the matching auxiliary basis functions. |
| **PNO** | Normal PNO Defaults. |
| **Solvation** | Implicit, SMD for dichloromethane |
| **SCF convergence** | Tight convergence criteria. *TightSCF* |

*Methods of Section 6.2 Implicit solvation and DLPNO-CCSD(T)*

**Details of the vacuum calculations**

| | |
|---|---|
| **Software DLPNO-CCSD(T)** | ORCA4.0 |
| **Software B3LYP-D3** | Gaussian 09 revision D.01 |
| **Calculation Type** | Single point |
| **Input Geometry** | Geometries optimized with B3LYP-D3/cc-pVTZ |
| **Basis Set** | cc-pVTZ for all atoms. |
| **Details** | Same setup as the corresponding Preliminary benchmark calculations but without solvent. |

---

[39] Neese, F. *Wiley Interdiscip. Rev: Comput. mol. Sci.* **2017**, 8, e1327.
[40] Riplinger, C.; Neese, F. *J. Chem. Phys.* **2013**, 138, 034106.
[41] Stoychev, G. L.; Auer, A. A.; Neese, F. *J. Chem. Theory Comput.* **2017**, 13, 554-562.

*Methods of Section 6.3 Microkinetic Benchmark*

**Details of the DFT calculations**

| | |
|---|---|
| **Software** | Gaussian 09 revision D.01 |
| **Calculation Type** | Optimization and frequencies. *opt freq* |
| **Input Geometry** | Geometries optimized with B3LYP-D3/cc-pVTZ |
| **Functionals** | B3LYP, PBEPBE, M06-2X |
| **Basis Sets** | cc-pVTZ for all atoms. |
| **Empirical Dispersion** | GD3. *empiricaldispersion=gd3* |
| **Symmetry** | Disabled. *nosymm* |
| **Solvation** | Implicit. SMD for dichloromethane.<br>*scrf(smd,dichloromethane)* |
| **Integration grid** | 75 radial shells and 302 angular points per shell. Gaussian09 default.<br>*int=(grid=finegrid)* |
| **Frequencies** | RRHO approximations. Gaussian09 default. |
| **Free energy corrections** | Reference state corrections, liquid phase at 1M and 298.15K (manually applied) |
| **Quasi-harmonic corrections (only when specified)** | GoodVibes software used.[42]<br>Grimme style corrections.[43]<br>Frequency threshold 100cm$^{-1}$. |

Same setup for the microkinetic simulations as in Section 5 but with the obtained energies for each compound, adduct and transition state of the corresponding DFT functional.

**Methods of Section 7. Microkinetic Modeling**

Same microkinetic setup as Section 5 with the energies obtained using the setup of Section 6.3 corrected manually as indicated in Section 7 (page 80).

---

[42] Funes-Ardoiz, I.; Paton, R. S. *Goodvibes 2.0.2*, **2016.** DOI:10.5281/zenodo.595246
[43] Grimme, S. *Theory Chem. Eur. J.* **2012**, 18, 9955-9964.

## 4. Mechanistic Study

Following the early studies on the imine condensation, we started calculating the mechanism for the reaction without catalyst between the benzaldehyde (**A**) and the *n*-butylamine (**B**). As we can see in Figure 14 the pre-reaction adduct (**[A···B]** or **1**) is less stable than the dissociated reactants, and the hemiaminal formation barrier is of 36.7 kcal mol$^{-1}$. We then arrive to the hemiaminal intermediate (**5**) which is also less stable than the separate reactants. The reaction then proceeds through the transition state (**6**), 47.7 kcal mol$^{-1}$ over **0**, to arrive to the imine-water adduct (**9**) that ends up releasing the products (**10**). The overall reaction is exergonic, which is consistent with experimental observations, but the barriers are too high.

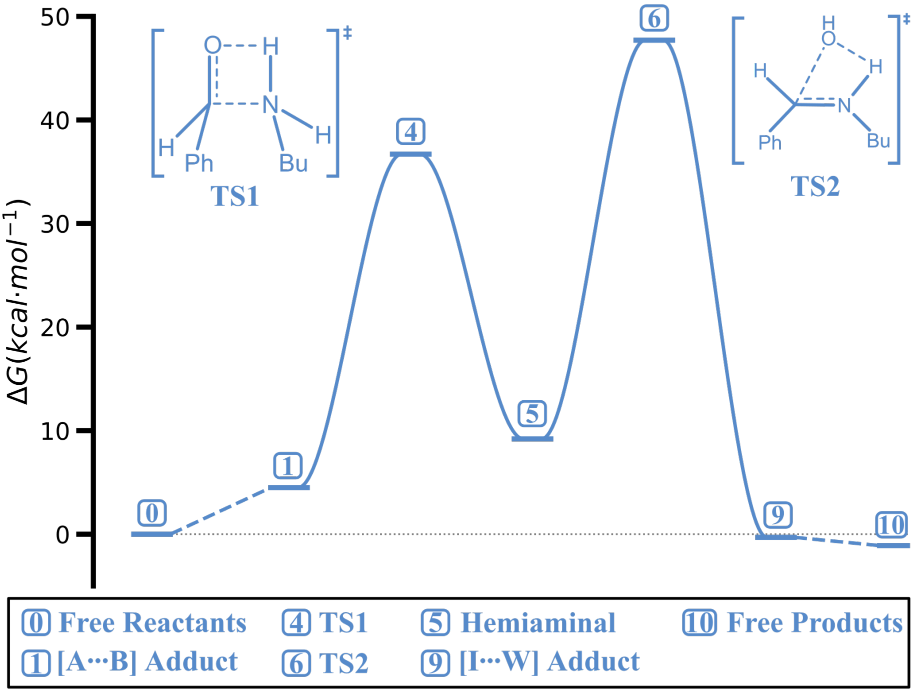


*Figure 14: Computed free energy profile of the imine condensation reaction without catalyst.*

This result was only mildly disappointing, as recent computational works were already ahead on the potential facilitating role of different proton shuttles. The best candidate molecules for the task are the amine (**B**) and the water (**W**) as both of them can be proton donors and proton acceptors. We started by including a single molecule as proton shuttle (see Figure 15). We see a significant decrease in the overall barrier of the reaction in both cases, the **water-catalyzed** and the **amine-catalyzed**. However, the obtained barriers are still too high to be consistent with a reaction happening at room temperature.

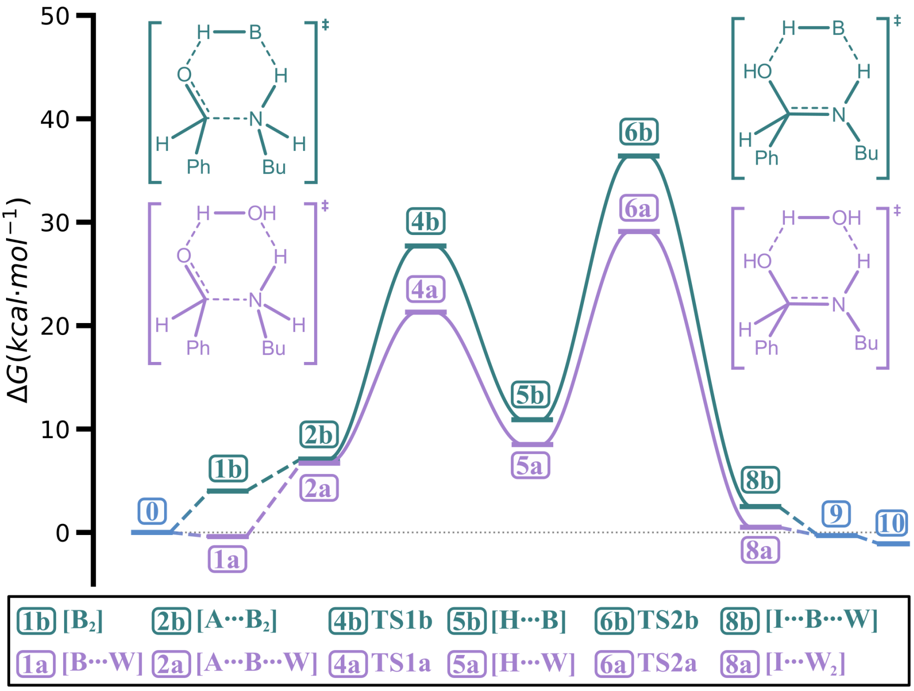


*Figure 15: Computed free energy profiles for the imine condensation using a single molecule as proton shuttle.*

As a single molecule did not yield low enough barriers we proceeded to calculate the mechanism using two larger shuttles: **the amine-water adduct**, and **the water dimer**.

We can see (Figure 16) that we still have the same qualitatively acceptable interpretation. The dehydration, **6c** and **6d**, is rate limiting. The hemiaminal intermediate, **5c** and **5d**, in both cases is unstable thus, its isolation would prove to be difficult experimentally. And, the overall reaction is still exergonic. But now, both profiles have overall barriers that can be considered acceptable (below 25.0 kcal mol$^{-1}$) for a reaction that happens at room temperature.

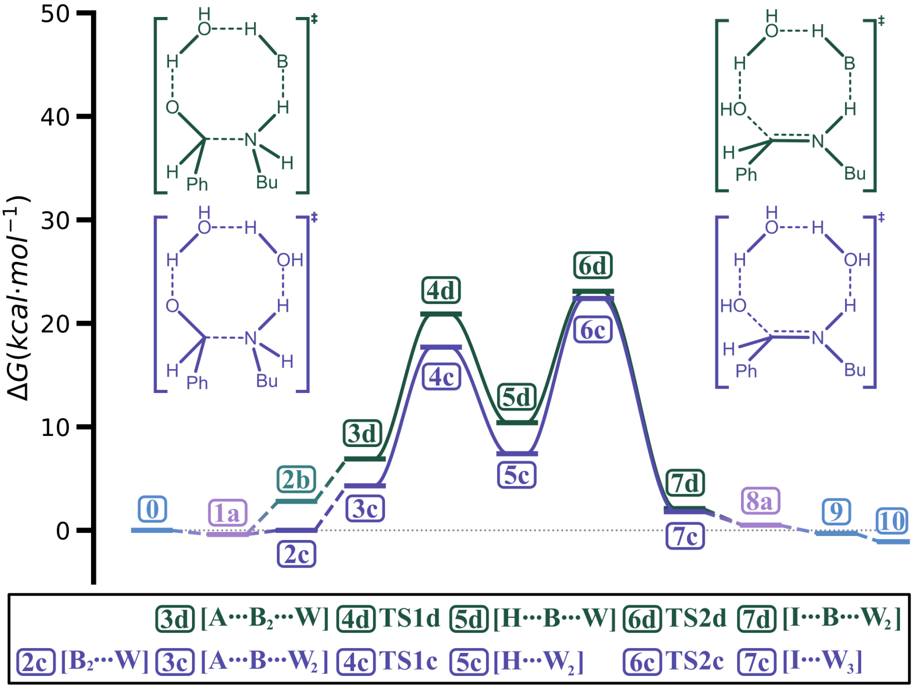


*Figure 16: Computed free energy profiles for the imine condensation using two molecules as proton shuttle.*

Upon close inspection of the IRC of the water dimer TSs, **TS1c** (**4c**) and **TS2c** (**6c**), we were pleasantly surprised to see how accurate were the experimental discussions in the 70s by Sayer and coworkers.[7,8,9,10] In Figure 17 we have the IRC calculations of **TS1c** and **TS2c**. As we can see in the energy profile, the **C-O** distance serves well as a reaction coordinate. Also we can see in the figure some arrows that serve as an approximate guide to the eye for the timing of the different chemical events relative to the **C-O** bond distance.

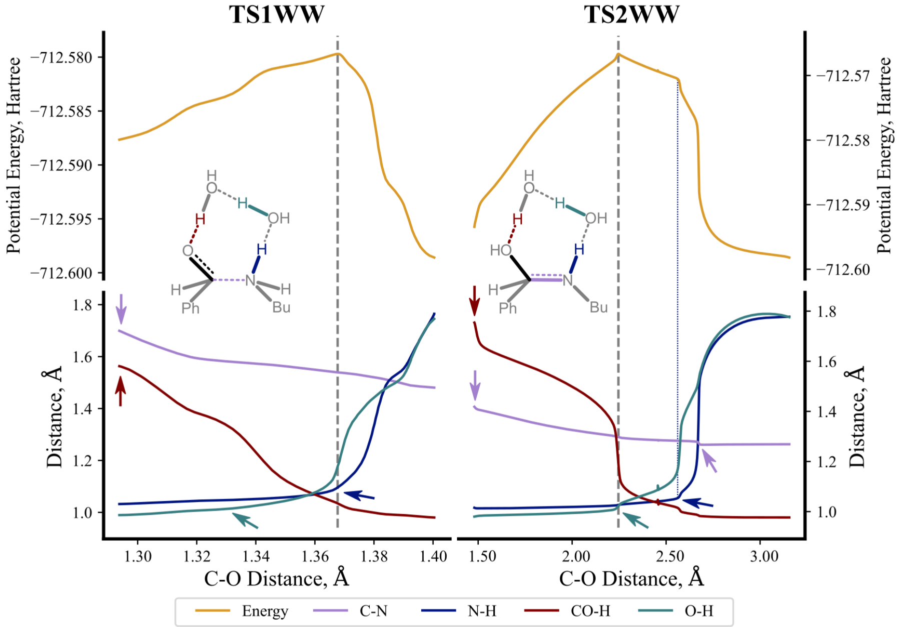


*Figure 17: IRC calculations of the water dimer transition states. C-O Distance at the Transition State marked by the dashed grey line. The arrows and the blue dashed lines are included as a guide to the eye.*

For the hemiaminal formation, **TS1c**, we can see that the **C-N** bond formation occurs coupled with the loss of the **C=O** double bond and the proton transfer to the same oxygen (**CO-H**). We can see that the relay mechanism of the proton shuttle activates afterward, being the the proton of the amine (**N-H**) the last one to transfer.

For the hemiaminal dehydration, **TS2c**, we can see how the C=N double bond formation is coupled with the **C-O** bond breaking. Although the proton transfer (**CO-H**) starts more or less simultaneously, we can see that the actual transfer takes place later. As a consequence of that proton transfer, the proton relay mechanism starts (**O-H**), and the last proton to be transferred is the proton from the nitrogen (**N-H**).

## 5. Microkinetic Simulation

Given the consistency of the model with the previous experimental and computational findings we decided that it was time to test if we could reproduce the kinetic experimental results of Ciaccia and Di Stefano[12] using microkinetic simulations. We calculated the initial concentrations of the reactants from the values provided in the original manuscript. Based on the specifications of Sigma-Aldrich's "100%" deuterated dichloromethane we decided to consider a water impurity of 1mM (ca. 1 order of magnitude smaller than the maximum concentration in the technical datasheet) in the solvent as initial water concentration.[44]

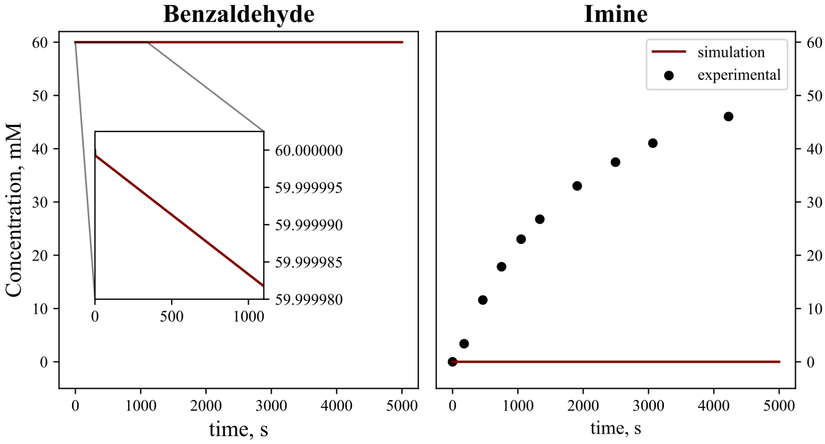


*Figure 18: Results of the initial microkinetic simulation.*

The initial results were underwhelming, depicted in Figure 18. On the left side we can see how, according to the simulation, the benzaldehyde is barely consumed. One the right side we can see how, as a consequence, the agreement with experiment (black dots) is non-existent.

---

[44] Sigma Aldrich webpage. Specification sheet, product 237027. https://www.sigmaaldrich.com/catalog/product/aldrich/237027 (accessed Dec 20, 2019).

## 6. Benchmark

### 6.1 Preliminary Benchmark

In order to improve the agreement between computational and experimental results, we decided to proceed with an evaluation of different computational descriptions for the calculation of the energy. In this reduced benchmark, instead of calculating all the species required for the microkinetic calculations we restricted the calculations to a reduced set, including the free reactants, the intermediate, the products, the uncatalyzed dehydration transition state and its equivalents in the process catalyzed by bimolecular proton shuttles (**1**, **5**, **10**, **6**, **6c** and **6d** in Figures 14-16).

We started by changing the size of the basis set (entries 1-4 in Table 1). Probably the most remarkable result is that entry 1 predicts an overall endergonic reaction. This sets a lower limit on the size of the basis set, requiring at least a basis set of the size of entry 2 for the energies. Nonetheless, the difference between the obtained geometries is negligible across entries 1-4.

Next, in entries 2 and 5-10, we have the results from different functionals. We can see that our initial idea of finding a clear trend between different functionals to guide our decision was thwarted. If we look at the TSs, we can categorize the results by lowest predicted barriers (5, 6), highest (7, 8, 10) and the remaining ones (2, 9). If instead we look at the stability of the hemiaminal intermediate we can group the results by least stable (2) most stable (9,10) and in between (5-8). The only remarkable trends related with the nature of the functionals is that both pure functionals seem to behave similarly (5,6).

Finally, we decided to use some CCSD(T) quality results (entries 11-14) in order to help us decide the most appropriate functional to use, and we were surprised by the outcome. If we compare the columns **TS2**, **H** or **I+W** of entries 2 and 14 we can see that the differences of energy are not too large, however the main problem comes when we include the proton shuttle catalyzed transition states (**TS2c** and **TS2d**) were the predicted barriers of DLPNO-CCSD(T) are *ca.* 10 kcal mol$^{-1}$ higher than the ones that we used for the microkinetic simulation of the previous section.

| | **Functional/ method** | **Basis Set** | **TS2 (6)** | **TS2c (6c)** | **TS2d (6d)** | **H (5)** | **I + W (10)** |
|---|---|---|---|---|---|---|---|
| 1 | B3LYP | 6-31+g(d) | 43.1 | 24.4 | 28.2 | 10.1 | 0.8 |
| 2 | B3LYP | 6-31+g(d,p) | 41.1 | 22.1 | 24.3 | 8.8 | -2.1 |
| 3 | B3LYP | 6-311++g(d,p) | 41.2 | 23.3 | 26.4 | 9.7 | -2.3 |
| 4 | B3LYP | cc-pVTZ | 47.7 | 22.4 | 23.1 | 9.2 | -1.1 |
| 2 | B3LYP | 6-31+g(d,p) | 41.1 | 22.1 | 24.3 | 8.8 | -2.1 |
| 5 | BP86 | 6-31+g(d,p) | 36.6 | 14.8 | 18.0 | 6.6 | -2.7 |
| 6 | PBEPBE | 6-31+g(d,p) | 37.0 | 15.9 | 18.6 | 6.6 | -2.8 |
| 7 | M06 | 6-31+g(d,p) | 44.8 | 28.1 | 31.2 | 5.5 | -2.0 |
| 8 | ωB97X-D | 6-31+g(d,p) | 43.5 | 26.6 | 29.5 | 5.3 | -1.9 |
| 9 | PBE1PBE | 6-31+g(d,p) | 42.8 | 21.3 | 23.6 | 3.0 | -3.0 |
| 10 | M06-2X | 6-31+g(d,p) | 44.5 | 25.4 | 27.3 | 2.6 | -3.6 |
| 11 | DLPNO-CCSD(T) | cc-pVDZ | 72.6 | 37.5 | 44.2 | 7.3 | -1.0 |
| 12 | DLPNO-CCSD(T) | cc-pVTZ | 56.8 | 33.8 | 38.0 | 6.1 | -3.5 |
| 13 | DLPNO-CCSD(T) | cc-pVQZ | 50.3 | 33.5 | 36.7 | 6.0 | -4.2 |
| 14 | DLPNO-CCSD(T) | cc-pV5Z | 46.5 | 33.7 | 36.2 | 6.0 | -4.6 |

*Table 1: Preliminary benchmark results. Values for entries 1-10 are relative free energies optimized with the specified functional and basis set. Entries 11-14 come from DLPNO-CCSD(T) single point corrections to the geometries optimized obtained from entry 2. All energies in kcal mol$^{-1}$.*

### 6.2 Implicit solvation and DLPNO-CCSD(T)

As we were unable to find the most accurate way of computing DLPNO-CCSD(T) energies in solution we included the SMD solvation as ORCA4.0 did allow it. Although some results do match the DFT computed results (TS2, H, I+W) our immediate hypothesis about the origin of such large differences was that it was due to the implicit solvation. We then proceeded to test such hypothesis. Using the geometries optimized with B3LYP/cc-pVTZ (and implicit solvation) we calculated the SP energies in vacuum and the SP DLPNO-CCSD(T) energies in vacuum and with SMD solvation.

| | **B3LYP** | | | **DLPNO-CCSD(T)** | | |
|---|---|---|---|---|---|---|
| | **$\Delta E_{SMD}$** | **$\Delta E_{vac}$** | **$\Delta\Delta E$** | **$\Delta E_{SMD}$** | **$\Delta E_{vac}$** | **$\Delta\Delta E$** |
| **TS2** | 37.6 | 43.0 | 5.5 | 45.2 | 52.2 | 7.1 |
| **TS2c** | -6.0 | -10.5 | -4.6 | 5.0 | 0.3 | -4.7 |
| **TS2d** | -2.1 | -5.2 | -3.1 | 8.9 | 5.2 | -3.7 |
| **H** | -3.9 | -5.7 | -1.8 | -6.8 | -8.9 | -2.1 |
| **I + W** | -1.0 | 0.2 | 1.2 | -3.2 | -2.7 | 0.5 |

*Table 2: Effect of solvation in the potential energy. $\Delta\Delta E = \Delta E_{vac}-\Delta E_{SMD}$ Energies in kcal mol$^{-1}$.*

If we compare the columns corresponding to $\Delta E_{vac}$ (Table 2) we can see that, consistently, the values of DLPNO-CCSD(T) for the transition states are *ca.* 10 kcal mol$^{-1}$ higher than the B3LYP values in vacuum. We observe the same trend for the results with SMD (which we have already mentioned).

As we can see when we look at the ΔΔE columns the solvation effects are significant, which is to be expected due to changes in polarity. But we can see that the effect of the implicit solvation is of the same magnitude in both cases and thus, cannot compensate the change in potential energy.

Therefore, DLPNO-CCSD(T) barrier heights taken at face value render the calculated mechanism much slower, which is not compatible with experiment.

### 6.3 Microkinetic benchmark

DLPNO-CCSD(T) gas phase potential energies are usually accepted as accurate reference values within the computational chemistry field. But literature on solvent phase DLPNO-CCSD(T) is much more scarce and, at least for this system, it raises more questions than it answers.

For this reason, we decided to resort to calculate all the species needed to set up the microkinetic models and run the simulations using the energies obtained from different functionals from the preliminary benchmark.

From Table 1 we selected PBEPBE and M06-2X (entries 6 and 10). We selected PBEPBE due to the low barriers as well as having a very similar behavior with the other pure functional. On the other hand we selected M06-2X due to its wide usage in computational organic chemistry as well as being one of the functionals that predict the hemiaminal more stable than others. We run the microkinetic simulations with the new profiles (see left column of Figure 19) without achieving a significantly better agreement with experiments (solid lines in Figure 20). As we could expect from the profiles, PBEPBE produced the largest amount of imine in the 5000s of simulation followed by M06-2X, but in both cases the barriers are too high for the timescale of the experiments. With the aim of improving the energies, we applied quasi-harmonic corrections to the results, obtaining the profiles in the right column of Figure 19), and re-run the kinetics. Nonetheless, no better agreement was achieved (dashed lines in Figure 20).

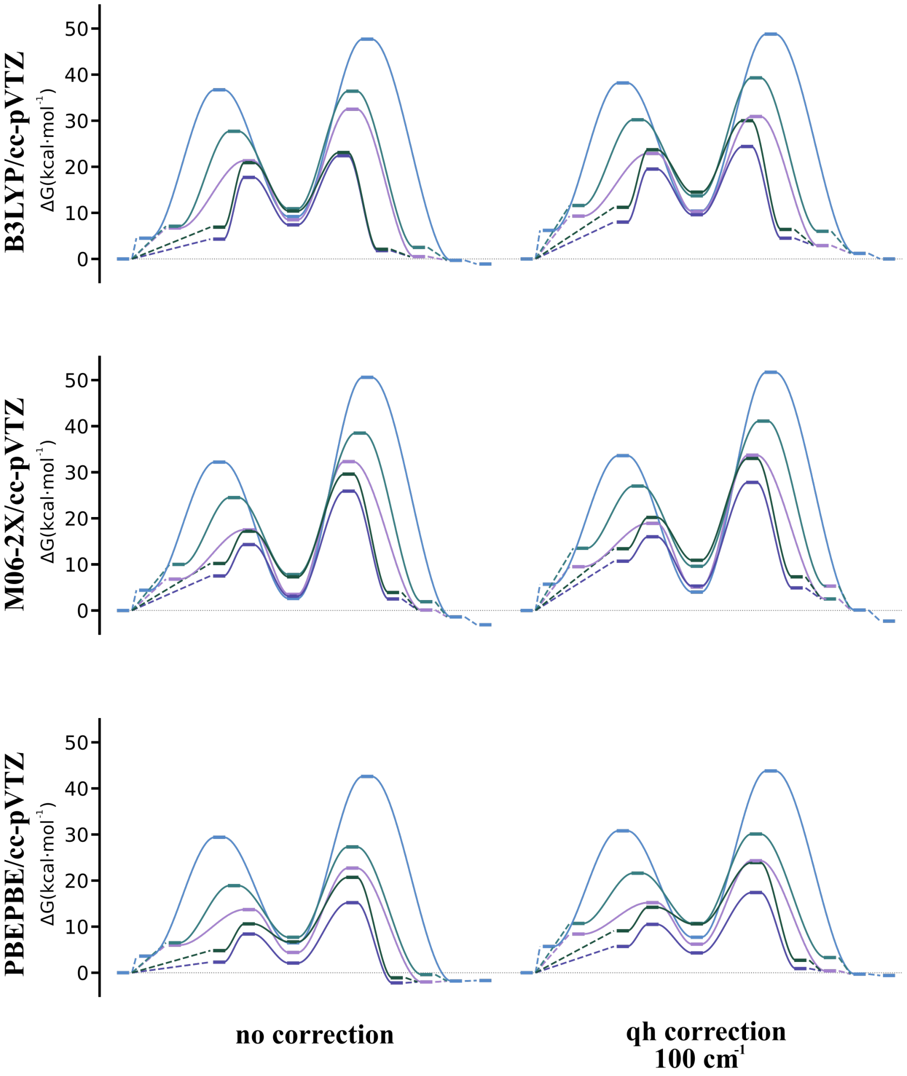


*Figure 19: Simplified representation of the free energy profiles used for the microkinetic simulations with the three selected functionals and their quasi-harmonic corrected versions.*

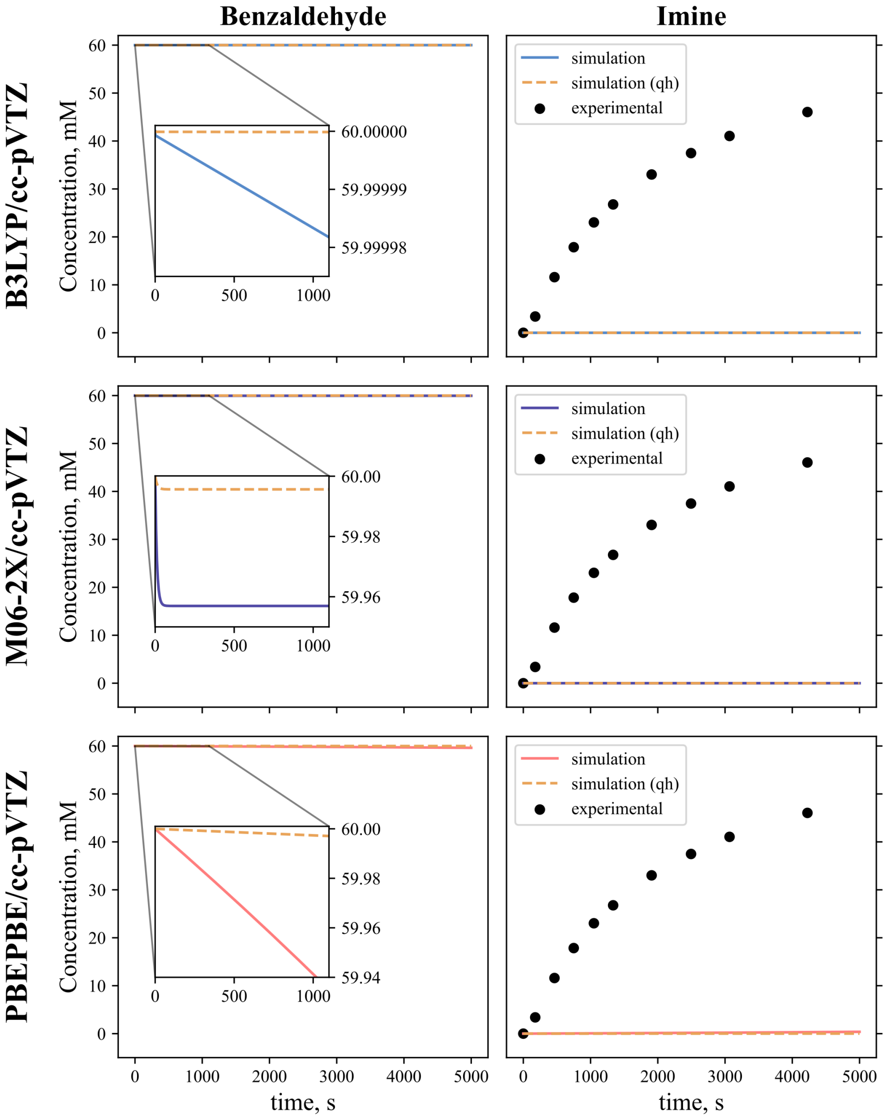


*Figure 20: Results of the microkinetic simulation benchmark. Dashed lines correspond to the simulation with the quasi-harmonic corrected free energies.*

## 7. Microkinetic Modeling

Qualitatively, computational results agree with experimental observations. Quantitatively, they do not.

Wondering on how to bridge this breach we found some examples in computational homogeneous catalysis where with the adjustment of the DFT results, the authors were able to close the gap between calculation and experiment.[45]

Based on this idea we decided to adjust our DFT profiles for the kinetics, and to decide the adjustment we formulated two hypotheses (H1 and H2) and tested them by using the microkinetic simulations.

### 7.1 First Hypothesis (H1)

One of the main fears of any computational chemist studying mechanisms is to have overlooked **the conformation** of a transition state that would make the model match the experiments. Inspired by this fear we came up with the first hypothesis for our adjustments (**H1**): What if we have overestimated all the TSs by X kcal mol$^{-1}$ ?

What would happen if the overall barrier was 0.1 kcal mol$^{-1}$ lower? and 1.0 kcal mol$^{-1}$ lower? and 2.0 kcal mol$^{-1}$? what if there is a completely different mechanism to bridge these two intermediates through a TS 8.0 kcal mol$^{-1}$ lower in energy? What if there is some sort of tunnel effect that lowers the effective barrier by 3.0 kcal mol$^{-1}$? All of these questions, directly related with **H1**, can be answered faster by shifting our focus to the microkinetic models instead of the DFT calculations. In Figure 21 we illustrate the practical effect of this hypothesis for the uncatalyzed reaction.

---

[45] a) Artús Suàrez, L.; Culakova, Z.; Balcells, D.; Bernskoetter, W. H.; Eisenstein, O.; Goldberg, K. I.; Hazari, N.; Tilset, M.; Nova, A. *ACS Catal.* **2018**, 8, 9, 8751−8762. b) Yu, Y.; Zhu, Y.; Bhagat, M. N.; Raghuraman, A.; Hirsekorn, K. F.; Notestein, J. M.; Nguyen, S. T.; Broadbelt, L. J. *ACS Catal.* **2018**, 8, 12, 11119−11133. c) Jaraíz, M.; Rubio, J. E.; Enríquez, L.; Pinacho, R.; López-Pérez, J. L.; Lesarri, A. *ACS Catal.* **2019**, 9, 6, 4804−4809.

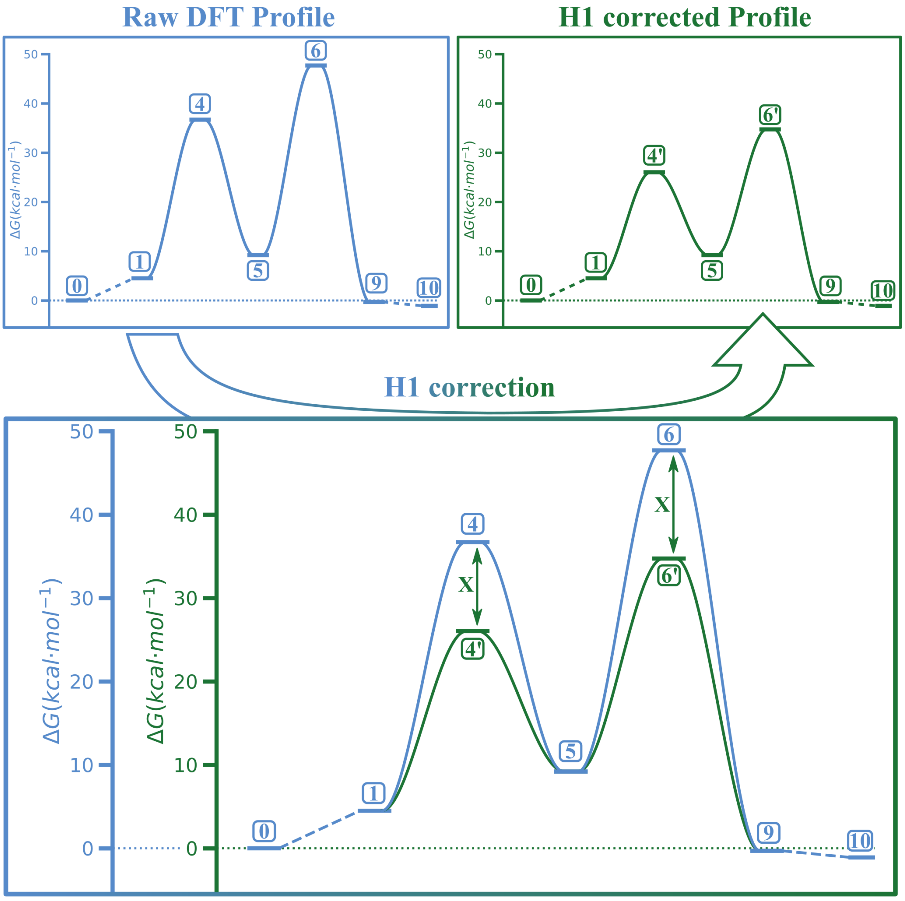


*Figure 21: Scheme illustrating the effect of the hypothesis H1 over the uncatalyzed profile calculated with B3LYP/cc-pVTZ.*

As we increase the correction X, we lower the overall barrier without changing the rate determining step. A good match between the experimental results and the simulation would be a great motivation to invest more efforts into calculations. And, at the very least, we might be able to calm our fears a bit.

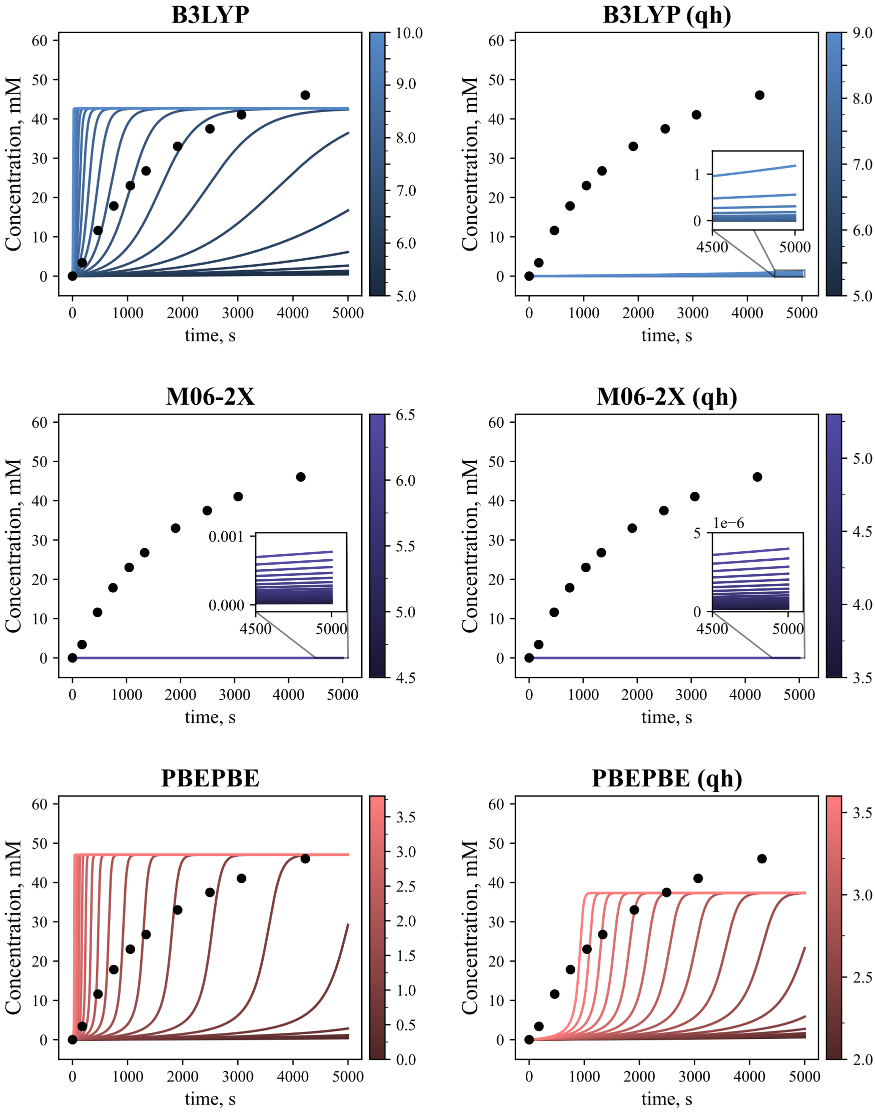


*Figure 22: Simulated total concentration of Imine vs time with the H1 based correction (coded by color) applied to different functionals. Corrections in kcal mol$^{-1}$.*

We started with the free energy profile of B3LYP and, as can be seen in Figure 22, at least a correction of 5.0 kcal mol$^{-1}$ was required for the imine concentration to start raising in the 5000s of simulation. We can see that the equilibrium state is clearly wrong as well as the shape of the curve. We get the best match between 7.0 and 8.0 kcal mol$^{-1}$ of correction to the transition states but it is clear that any corrected profile with this approach is going to be far from the real one.

We then moved to testing the same hypothesis for the energies obtained using M06-2X and PBEPBE. In the case of M06-2X exceeding the 6.5 and 5.3 kcal mol$^{-1}$ corrections would lead to lower the energy of some transition states below the minima connected to it. We can see that with this correction, the agreement with experiments is not much better than when no correction was applied (Figure 20). The results of PBEPBE, improve a bit the agreement with the equilibrium state but at the cost of reducing the match between the experimental curve and the simulated curves.

Following the simulations on the kinetic benchmark we proceeded to run the **H1**-corrected simulations with the quasi-harmonic corrected free energies of B3LYP, M06-2X and PBEPBE (right column of Figure 22). Surprisingly, in all the cases the match worsens. A possible explanation would be that a big contribution to the relative differences in energy of these species is due to the differences between the low laying frequencies. It might be that, due to the non-covalent interactions present in most of these species the quasi-harmonic corrections (that target the low laying frequencies) might end up making the free energies less realistic.

### 7.2 Second Hypothesis (H2)

Contrary to **H1**, the idea of our second hypothesis came from completely different sources. If we look at the raw values of energies (which we should not try to interpret directly) obtained from different quantum chemistry methods, we may wonder why those values shift more or less. Shifting those energies by a constant value is the end result of the correction of the free energies to change the reference state from 1atm to 1M. Also, some entropic corrections end up behaving (effectively) as a constant value added to all calculations.[46]

From these different sources we came up with the questions: what if there is a systematic error in the calculations? what if it behaves like a constant value? These were the origin of our second hypothesis (**H2**): What if we have a systematic underestimation of Y kcal $mol^{-1}$ per calculation?

A practical answer to that question is “we just need to correct each calculation by adding value” or in other words to bias the DFT results. If we think about a unimolecular reaction it is right to think of this hypothesis **H2** as a completely innocuous correction as it has no effect. However, the effects of this correction become more evident when a multimolecular reaction is considered.

In Figure 23 we can see the effect of this correction in the profile computed for the reaction catalyzed by a single water molecule. In the reference state (**0**) we have three different calculations corresponding to benzaldehyde, **A**, *n*-butylamine, **B**, and water, **W**. Because of that, its energy is increased (**0’**) three times **Y**, an arbitrary constant value. If we now look at the relative energy of the amine-water adduct, **[B···W]** or **1a**, to match the reference state it has to contain two different calculations, the free aldehyde **A** and **[B···W]**. Because of this, its energy is increased two times (**1a'**) instead of three. We proceed similarly with **[A···B···W]**, **2a**, that contains all the species in the same calculation and increase its energy only once.

---

[46] Besora, M.; Vidossich, P.; Lledós, A.; Ujaque, G.; Maseras, F. *J. Phys. Chem. A* **2018**, 122, 1392−1399.

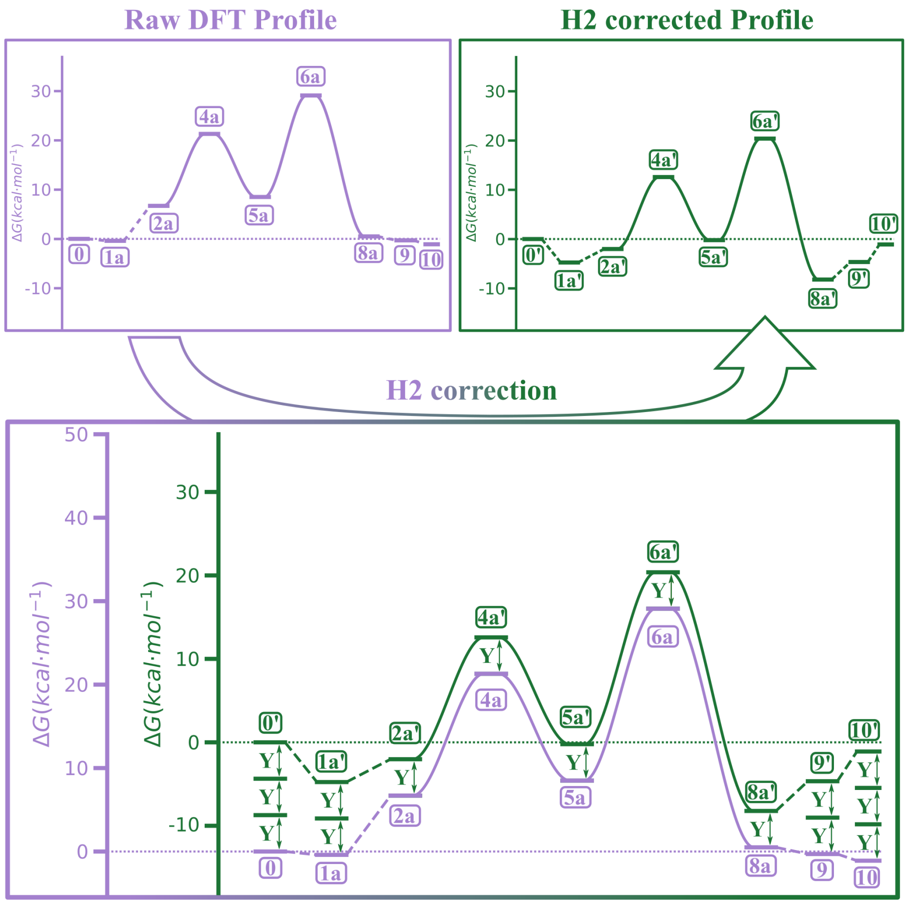


*Figure 23: Scheme illustrating the effect of the hypothesis H2 over the profile calculated with B3LYP/cc-pVTZ with a single water as proton shuttle.*

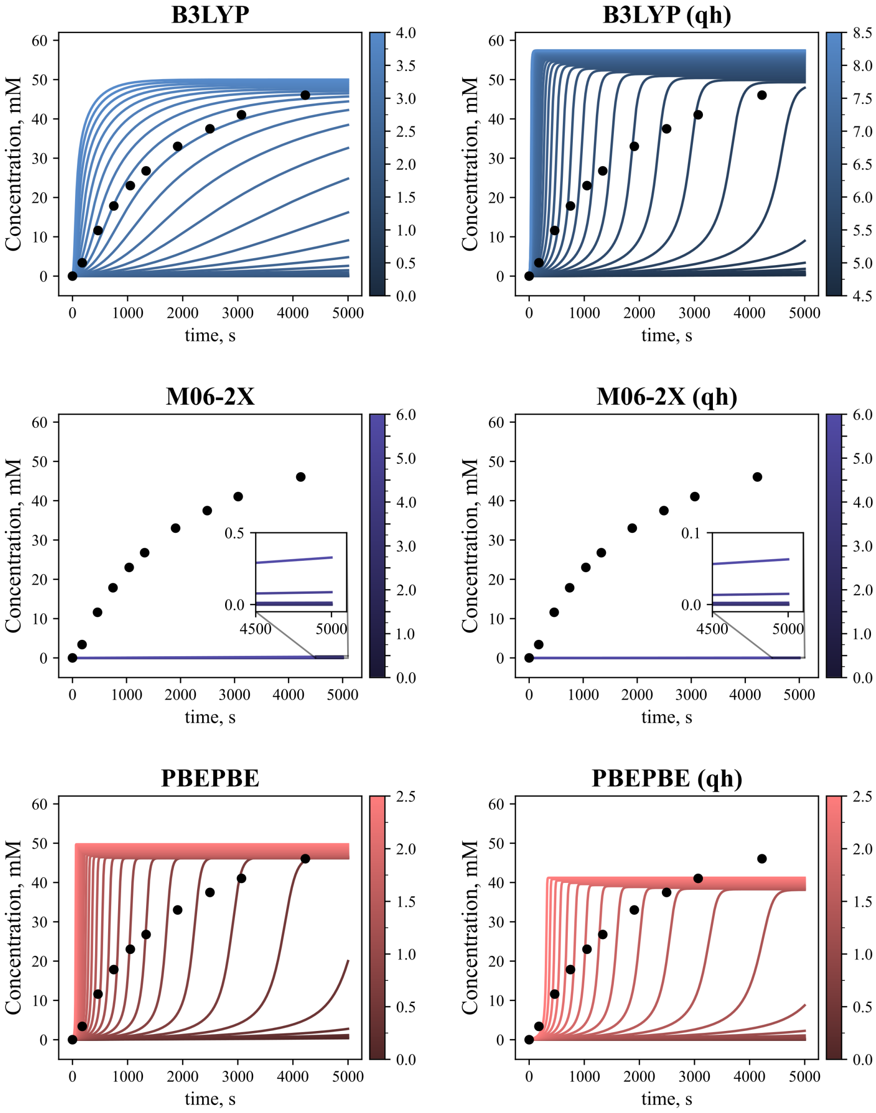


*Figure 24: Simulated total concentration of Imine vs time with the H2 correction (color) applied to different functionals. Corrections in kcal mol$^{-1}$.*

When then proceed to re-calculate the energies so that the new reference state (**0'**) corresponded to the zero in energy. Now we see that the difference in energy between **0’** and **1a’** is exactly the same one as **0** and **1** minus the correction, **Y**, making **1a’** more stable. However, for steps connecting the minima and the transition state (*i.e.* **2a’** and **4a’**) we see that there is no change since both shifted the same amount. The main effect of this correction ends up being a stabilization of the adduct species, such as **[B···W]**, relative to its constituents, **B** and **W**.

We then proceeded to run the microkinetic simulations using the correction derived from **H2.** If we compare the results from the PBEPBE or M06-2X calculations with the **H2** correction (Figure 24) we can see that they are very similar to the results from **H1** (Figure 22). In both cases we see that this correction makes the reactions faster. For M06-2X and its quasi-harmonic corrected results we can observe it in the y-axis of the inset plot and for the results of PBEPBE we can see it in the extent of the correction (looking at the color bars).

The main surprise is the results of B3LYP and its quasi-harmonic corrected version. The best match between the experimental and the simulated results is with the 3.2 kcal mol$^{-1}$ correction and we can see how the simulated lines have a major resemblance with the shape of the experimental results. Its quasi-harmonic counterpart (where **H1** had almost no effect), on the other hand, predicts a really fast reaction, showing the same trend as the PBEPBE results.

Finally we see that **H2** cannot be generalized to all functionals, as it is neither constant across functionals nor does it behave like a constant value for all the functionals selected. Nonetheless, we now have a free energy profile of the reaction that is more realistic than the one we had before and that also matches well with the previous experimental and computational studies.

## 8. Summary

In summary, we have reviewed the experimental and computational knowledge on the mechanism of imine condensation. We have studied in detail the mechanism of the condensation between the benzaldehyde and the *n*-butylamine.

We have found a mechanism similar to those previously reported by other computational studies that used different levels of theory. The mechanism found using a water dimer as a proton shuttle has a good qualitative match with previous experimental results in water and, coupled with the water-amine catalyzed mechanism, provides a suitable reaction pathway for the reaction to happen in dichloromethane. Independently of the catalyst and theory method, we have seen that the hemiaminal dehydration step is the rate determining step.

We have shown that the pure theoretical energies are not necessarily enough to be able to reproduce raw kinetic data and we have been able to bridge that gap with the careful use of microkinetic simulations coupled with a systematic bias of the DFT-computed energies. We have shown that such approach is not independent of the DFT functional and that not all functionals can be simply biased to achieve agreement with experiments.

Finally, we have highlighted the importance of the availability of raw kinetic data, concentration *vs* time, as well as the importance of reproducing its shape instead of just the equilibrium state.

# Chapter II. Exploratory Study of the CC1 Cage Synthesis

## 1. Motivation

In order to understand the self-assembly mechanism of imine cages we focus on the self-assembly mechanism of the CC1 imine cage. A thorough mechanistic investigation is manually not feasible due to the large amount of possible intermediates, conformations and transition states. Thus, the main aim of this chapter is to get some basic knowledge of the system that will allow us to run DFT-based kinetic simulations in Chapter IV without the computational cost of calculating every single transition state in the reaction network.

Because of this we are going to focus on how the imine condensation barrier changes as the CC1 synthesis progresses. Not only that but we want to find if we can categorize different types of reactions within the mechanism and either assign a specific barrier for each type or find a way to estimate the barrier for that reaction category. For these tasks we will calculate the transition states of key reactions from the pool of reactions involved in the self-assembly.

As we have observed in the previous chapter an accurate calculation of the barrier for the reaction that is compatible with experimental concentration *vs* time data is, at least, challenging. Thus, it seems reasonable to approach this study in a relative manner by comparing the results with the first monomer condensation barriers. Such relative approach will allow us to see if our results are compatible with the current knowledge of the self-assembly of imine cages.

## 2. Introduction

Understanding on the factors controlling the self-assembly of imine cages is vital for the rational design of these compounds and the materials derived from them. Gaining this knowledge is a necessary requisite for a possible future large scale production and applications. The precursor selection, the stability of the products, the reaction and the synthetic strategy can play a role. Most of current research focuses on understanding the effects of each variable but, due to the complex nature of self-assembly process, a full mechanistic picture is necessary but still far from reach.

We will start by reviewing the synthetic strategies and conditions as well as how they affect the reaction, with special attention to the specific procedure for the synthesis of CC1 imine cage. Next, we will review how these systems have been characterized computationally. And, lastly, we will review the studies that have shed some light into the mechanism.

## 2.1 Synthesis and Design

Imine cages[1] as well as imine macrocycles[2] are generally synthesized through one-pot reactions where an aldehyde monomer and an amine monomer are mixed together. Not only that but it is relatively usual to see them synthesized at room temperature. Depending on the compound desired the reaction times may range from hours to weeks and, overall, we can see a widespread usage of DCM as solvent. Finally, sometimes TFA or $Sc^{3+}$ are added to catalyze the imine condensation reactions.

Nowadays, a typical synthesis of an imine cage involves the dropwise addition of the solution of the aldehyde monomer to the solution of amine monomer in a small excess. In the case of imine macrocycles, direct mixture of two solutions containing both monomers in a 1:1 ratio is the preferred procedure.

It is to note that a simple synthesis does not necessarily correlate with simple product separation and characterization. As an example of this, in 2008 Warmuth and Xu synthetized an homochiral **[8+12]** cage starting from an enantiomerically pure amine.[3] However, similar attempts using instead a racemic mixture of the amine are scarce, being the recent study of Mastalerz and coworkers in 2021 one of the latest examples.[4] In this study they were able to synthesize the two homo enantiomers and a meso enantiomer of a **[8+12]** chiral cage through self-sorting.

---

[1] Briggs, M. E.; Cooper, A. I. *Chem. Mater.* **2017**, 29, 149-157.

[2] Borisova, N.; Reshetova, M. D.; Ustynyuk, Y. A. *Chem. Rev.* **2007**, 107, 1, 46-79.

[3] Xu, D.; Warmuth, R. *J. Am. Chem. Soc.* **2008**, 130, 24, 7520-7521.

[4] Philippe, W.; Rominger, F.; Zhang, W.-S.; Gross, J. H.; Elbert, S. M.; Schröder, R. R.; Mastalerz, M. *Angew. Chem. Int. Ed.* **2021**, 60, 16, 8896-8904.

The monomer selection and its effect on the product stoichiometry (or products) is probably where most of the literature focuses.[5] A good example of this is the study published in 2018 by Cooper and coworkers.[6] In this study they self-assembled a wide variety of **[3+2]**, **[6+4]** and **[4+4]** imine cages under the same synthetic conditions (Figure 25).

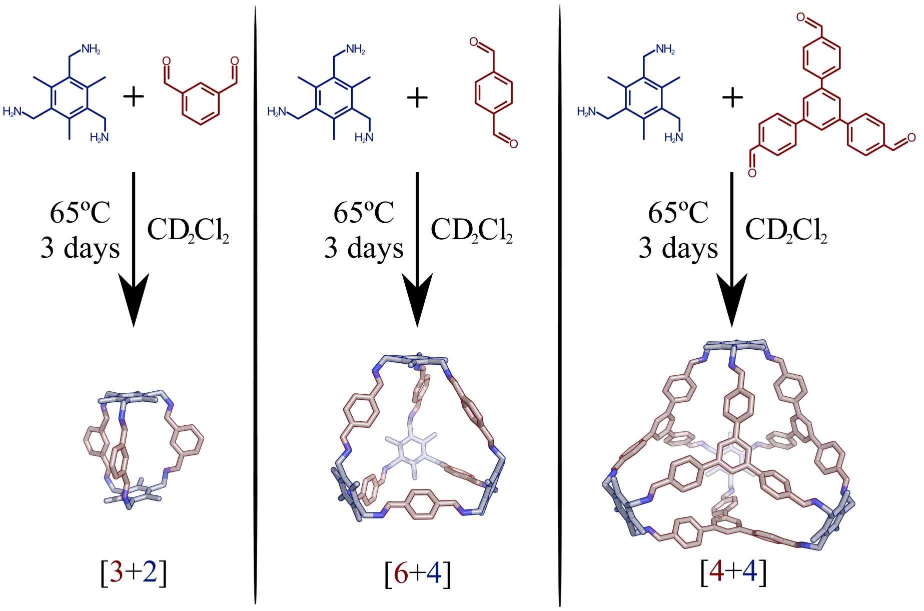


*Figure 25: Examples extracted from reference 6 of the outcome of different monomer combinations under the same conditions.*

---

[5] a) Kwit, M.; Grajewski, J.; Skowronek, P.; Zgorzelak, M.; Gawroński, J. *Chem. Rec.* **2019**, 19, 213-237. b) Tarzia, A.; Jelfs, K. E. *Chem. Commun.* **2022**, 58, 3717-3730.

[6] Greenaway, R. L.; Santolini, V.; Bennison, M. J.; Alston, B. M.; Pugh, C. J.; Little, M. A.; Eden-Rump, E. G. B.; Clowes, R.; Shakil, A.; Cuthbertson, H. J.; Armstrong, H.; Briggs, M. E.; Jelfs, K. E.; Cooper , A. I. *Nat. Commun.* **2018**, 9, 2849.

Some studies have highlighted interesting odd-even effects in the reaction outcomes.[7] These effects consist in an alternation on some property (i.e. cage stoichiometry, conformation) of the synthesized cage depending on the number of carbons in the alkyl chain of one of the monomers (see Figure 26a and 26b). It is also remarkable the studies on the self-sorting of chiral cages (See an example in Figure 26c).[4,8]

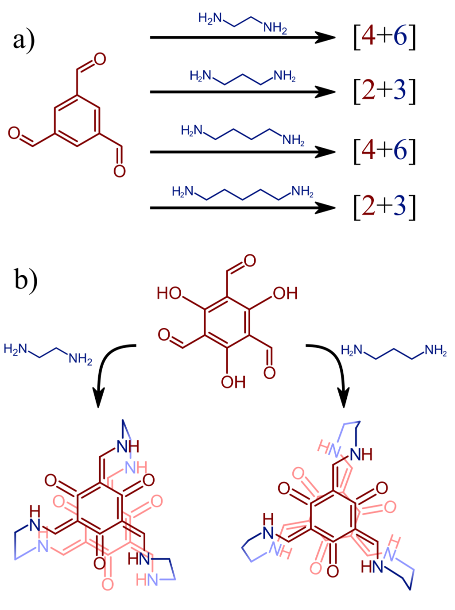


*Figure 26: Odd-even effect and chiral self-sorting examples. a) reaction of TFB with amine linkers of different length (ref 7a), b) reactions of 1,3,5-triformylphloroglucinol with two different amine linkers leading to two different conformations (ref 7b), c) Chiral self-sorted cubes synthesized by Philippe et al. (ref 4).*

---

[7] a) Jelfs, K. E.; Eden, E. G. B.; Culshaw, J. L.; Shakespeare, S.; Pyzer-Knapp, E. O.; Thompson, H. P. G.; Bacsa, J.; Day, G. M.; Adams, D. J.; Cooper, A. I. *J. Am. Chem. Soc.* **2013**, 135, 9307-9310. b) Bera, S.; Basu, A.; Tothadi, S.; Garai, B.; Banerjee, S.; Vanka, K.; Banerjee, R. *Angew. Chem. Int. Ed.* **2017**, 56, 2123-2126. c) Su, K.; Wang, W.; Du, S.; Ji, C.; Zhou, M.; Yuan, D. *J. Am. Chem. Soc.* **2020**, 142, 42, 18060-18072.

[8] a) Beaudoin, D.; Rominger, F.; Mastalerz, M. *Angew. Chem. Int. Ed.* **2017**, 56, 5, 1244-1248. b) Slater, A. G.; Little, M. A.; Briggs, M. E.; Jelfs, K. E.; Cooper, A. I. *Mol. Syst. Des. Eng.* **2018**, 3, 223-227.

Finally, it has been highlighted in different manners that the solvent plays an important role in all the process.[9] It can change the outcome of the reaction for the same reactants[10] (Figure 27a) or it can have an effect in the distribution of the different species obtained from the dynamic scrambling of the imine cages[11] (Figure 27b) or even it may drive a change in stoichiometry of a cage (Figure 27c).[12]

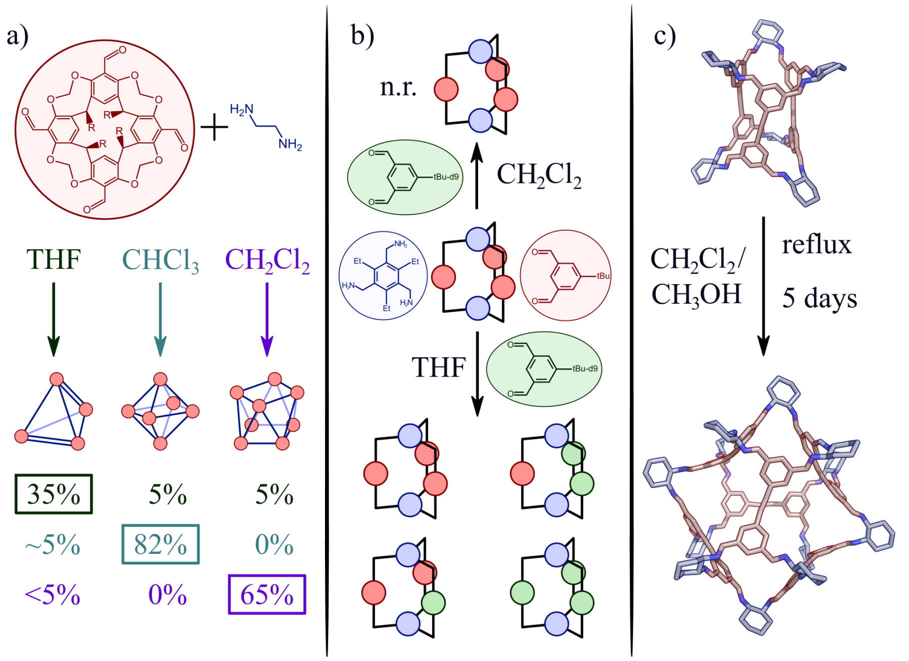


*Figure 27: Examples of the influence of the solvent on imine cages. a) Yields of each cage when synthesized in different solvents (ref 10). b) Effects of solvent on the scrambling of a cage (ref 11). c) Change in stoichiometry of a cage due to the solvent (ref 12).*

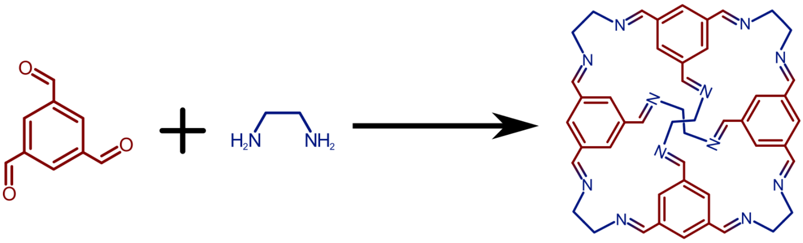
---

[9] Zhang, G.; Mastalerz, M. *Chem. Soc. Rev.* **2014**, 43, 1934-1947.
[10] Liu, X.; Warmuth, R. *J. Am. Chem. Soc.* **2006**, 128, 14120-14127.
[11] Schick, T. H. G.; Rominger, F.; Mastalerz, M. *J. Org. Chem.* **2020**, 85, 21, 13757-13771.
[12] Pugh, C. J.; Santolini, V.; Greenaway, R. L.; Little, M. A.; Briggs, M. E.; Jelfs, K. E.; Cooper, A. I. *Cryst. Growth Des.* **2018**, 18, 5, 2759-2764.

### *2.1.1 CC1 Synthesis*

At this point, it is important to describe the synthetic procedure of the CC1 cage to know which effects might play a role in the synthesis of the target of our computational study.

Since the first report of this compound[13] its synthesis has always relied in the same overall reaction (Figure 28). In 2011, an optimized synthetic protocol in batch was reported by Cooper and coworkers[14] and later, the same group published a flow synthesis for the same compound, reducing the reaction times while maintaining the high yield. [15]



TFB EtDA CC1

*Figure 28: Overall reaction for the synthesis of the CC1 imine cage.*

In Figure 29 the synthetic procedure of the optimized batch synthesis is schematically presented. It is important to note that although this synthesis is commonly referred as "batch synthesis" it behaves as a semi-continuous reactor during the first 48h since the aldehyde is being added. Below the scheme of the batch synthesis we can see a representation of the flow synthesis proposed by the same group. It is important to note the different temperatures used in both syntheses.

---

[13] Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I. *Nat. Mater.* **2009**, 8, 973-978.

[14] Lydon, D. P.; Campbell, N. L.; Adams, D. J.; Cooper, A. I. *Synth. Comm.* **2011**, 41, 14, 2146-2151

[15] Briggs, M. E.; Slater, A. G.; Lunt, N.; Jiang, S.; Little, M. A.; Greenaway, R. L.; Hasell, T.; Battilocchio, C.; Ley, S. V.; Cooper, A. I. *Chem. Commun.* **2015**, 51, 17390-17393.

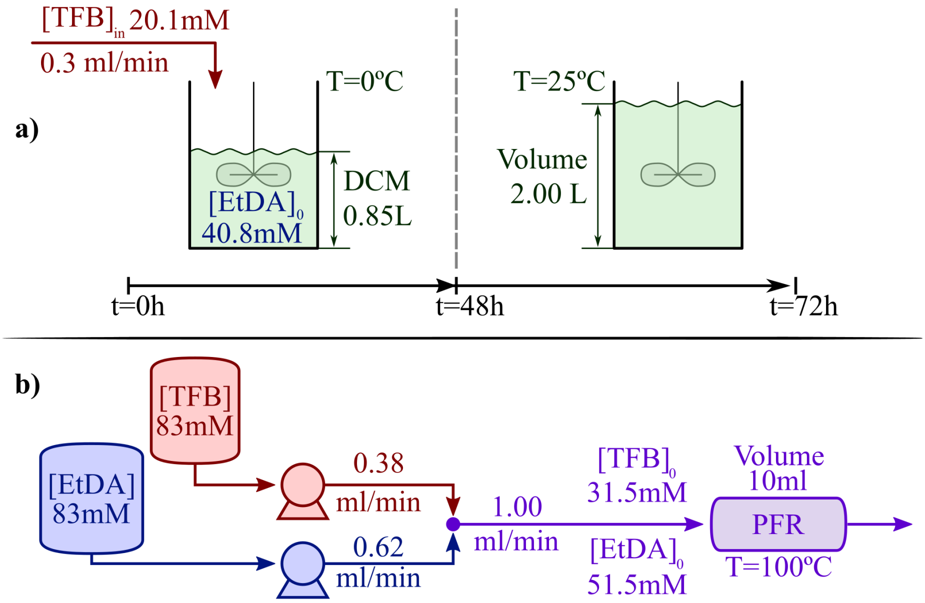

a)
[TFB]in 20.1mM
0.3 ml/min
T=0ºC
[EtDA]0
40.8mM
DCM
0.85L
T=25ºC
Volume
2.00 L
t=0h
t=48h
t=72h
b)
[TFB]
83mM
[EtDA]
83mM
0.38
ml/min
0.62
ml/min
1.00
ml/min
[TFB]0
31.5mM
[EtDA]0
51.5mM
Volume
10ml
PFR
T=100ºC

*Figure 29: Schemes of the a) batch and b) flow synthesis of CC1.*

Regarding the chemicals, in both cases the solvent employed is DCM (obtained from Fisher), and the TFB (Manchester Organics, UK) and the EtDA (Sigma-Aldrich) were purchased from the same sources. Finally, they reported that all chemicals were used as received in both cases.

## 2.2 Thermodynamic Calculations

The study of imine macrocycles from a computational perspective has been approached in a very different way than the study of imine cages. In imine macrocycles the main focus of discussion of computational studies is the obtained geometric characteristics of the monomers and products.[5a,16] It is not unusual to find studies relying on semi-empirical methods or force field-based models. On the other hand, imine cage literature tends to focus the discussion on free energies. We can find that even in geometry focused papers such as the 2018 work of Jelfs and coworkers,[17] energy is a key point in the discussion. Although we can find force field based models for cages it is becoming more and more common to find investigations using semi-empirical or DFT based models.

Computational models have been used to understand crystal polymorphism in cases such as the CC1 cage[18] where the difference in energies between the different conformers of the CC1 cage was used to understand the different phases found in the CC1 crystals. In some other cases the computational models have been used to rationalize the outcome of scrambling experiments. An example of this application is the 2010 work of Mukherjee and coworkers[19] where the results of the dynamic re-assembly of a **[3+2]** cage were rationalized. A more recent example of this is the 2020 work of Greenaway et al.[20] where a similar re-assembly upon a **[3+2]** cage was explained with the help of computational results.

---

[16] Nour, H. F.; Lopez-Periago, A. M.; Kuhnert, N. *Rapid Commun. Mass Spectrom.* **2012**, 26, 9, 1070-1080.

[17] Santolini, V.; Miklitz, M.; Berardo, E.; Jelfs, K. E. *Nanoscale* **2017**, 9, 5280-5298.

[18] Jelfs, K. E.; Schiffmann, F.; Jones, J. T. A.; Slater, B.; Cora, F.; Cooper, A. I. *Phys. Chem. Chem. Phys.* **2011**, 13, 20081-20085.

[19] Acharyya, K.; Mukherjee, S.; Mukherjee, P. S. *J. Am. Chem. Soc.* **2010**, 135, 2, 554-557.

[20] Greenaway, R. L.; Santolini, V.; Szczypiński, F. T.; Bennison, M. J.; Little, M. A.; Marsh, A.; Jelfs, K. E.; Cooper, A. I. *Chem. Eur. J.* **2020**, 26, 17, 3718-3722.

As previously commented, the perspective used in the design of the cages has been mostly focused on monomer selection. In this sense, the usage of high-throughput calculations coupled with experiments is raising in importance[21] and a very nice example is the 2018 paper of Cooper and coworkers previously highlighted.[6]

In this work the authors raise a very interesting point about how to compare the relative energies of the different cages. Up to that date, (and even afterwards since it is not completely adopted by the community) the free energy of cages is usually reported relative to the free reactants following equation 1.

$$\Delta G^{[n+m]^i} = G_{DFT}^{[n+m]^i} + iG_{DFT}^{H_2O} - \left(n G_{DFT}^{TFB} + m G_{DFT}^{EtDA}\right) \tag{1}$$

Where we have extended the **[n+m]** nomenclature where **n** and **m** stand for the number of each one of the monomers required to form the intermediate. The superindex **i** stands for the number of imine bonds formed (which equals the number of water molecules formed). To maintain a consistent nomenclature across this thesis we always follow the **aldehyde** first and **amine** second convention. For example, $\mathbf{[1+2]^2}$ would correspond to a compound formed by **1 aldehyde**, **2 amine** and having a total of **2 imine bonds**. $G_{DFT}$ stands for the free energy obtained from the DFT calculation.

Comparing the relative energy of two cages (or intermediates) of differing size can be misleading at first sight. The reference state for a $\mathbf{[4+6]^{12}}$ cage is going to be 2 times the reference state of a $\mathbf{[2+3]^6}$ cage. This however becomes trickier when comparing other stoichiometries, and becomes a problem when comparing intermediates. Thus, Cooper and coworkers promote the usage of the of relative energy per bond, (see equation 2).

---

[21] Evans, J. D.; Jelfs, K. E.; Day, G. M.; Doonan, C. J. *Chem. Soc. Rev.* **2017**, 46, 3286-3301.

$$\Delta G_{bond}^{[n+m]^i} = \frac{\Delta G_r^{[n+m]^i}}{i} \qquad (2)$$

Both measures of relative stability give information on how the system behaves. The relative free energy gives a wider representation of the free energy profile of the system, whereas the relative free energy per bond provides a better estimation of stability of a certain compound.

## 2.3 Mechanistic Studies

### *2.3.1 Self-sorting Experiments*

In the literature we can find experiments with the aim of understanding the preferences between monomers to self-assemble. An approach is to start with a mixture of monomers and see how the system evolves and characterize it (Figure 30, left side). One way to study this is to use a racemic mixture of one of the monomers instead of the enantiomerically pure monomer like the recent studies by the Mastalerz group.[4,8a] Another way is to start from an asymmetric monomer[13,16] or even a mixture of monomers[22]. Although full understanding of these studies is difficult, they shed some light into the preferential synthetic pathways of the early stages of the reaction.

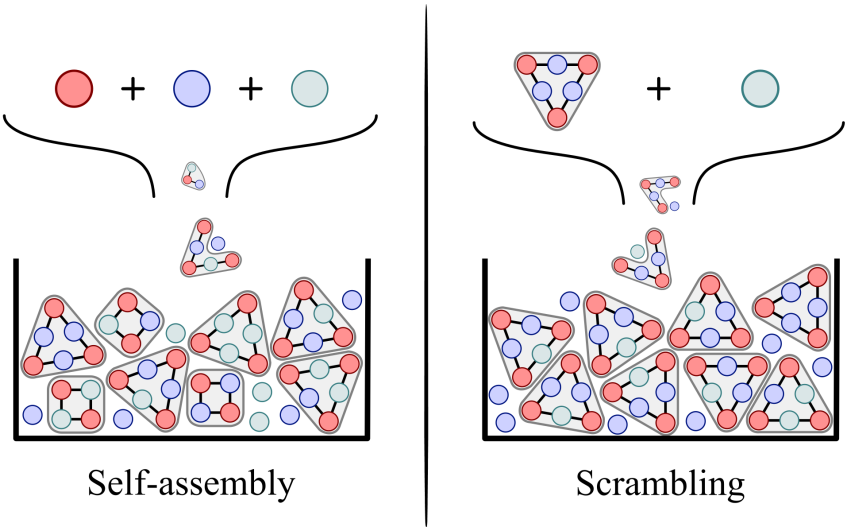


*Figure 30: Scheme of the forward and the scrambling self-sorting experiments*

A second, opposite, approach consists on exposing a synthesized cage either to another cage or to another monomer (Figure 30, right side).[11,22] These studies provide valuable information on the last steps of the cage formation and support the reversibility of the imine bond formation. However, plainly assuming that these imine bonds will always hydrolize easily would be an error as proved by the work of 2012 by Hasell et al. where they demonstrated that a crystal of the CC3 cage was able to last for 4h in boiling water.[23]

---

[22] Jiang, S.; Jones, J. T. A.; Hasell, T.; Blythe, C. E.; Adams, D. J.; Trewin, A.; Cooper, A. I. *Nat. Commun.* **2011**, 2, 207.

### *2.3.2 Kinetic Studies on Metallocages*

Computational mechanistic studies on imine cages are generally focused on thermodynamics. However we can find some relevant studies on the self-assembly processes of metallocages. We can highlight mainly two schools here. On one hand we have the Molecular dynamics based studies, and on the other hand we find the kinetic simulation based studies of the Sato group.

Under the molecular dynamics based studies we find the study of the self-assembly of a Pd $M_6L_8$ cage and a Pd $M_{12}L_{24}$ cage.[24] A few years after these studies we see the equivalent to the self-sorting experiments in MD simulations, first for a smaller asymmetric cage[25] and later for the large dynamic exchange of a $M_{12}L_{24}$ cage[26]. Finally it is remarkable from a computational modeling perspective the self-assembly and encapsulation studies carried out by Jiang et al. in 2017.[27]

The more recent kinetic based studies, by the group of Sato, use a stochastic approach to simulate the self-assembly of smaller cages while successfully reproducing experimentally observed concentration *vs* time profiles.[28,29] Their method is based on the assignment of some experimental kinetic constants to the different types of reactions that they identify. One of main outcomes of these studies is that they locate the rate determining step (RDS) at the cage closure. Such claim has not been issued in imine cage self-assembly where it is usually assumed that if a Cage presents a dynamic chemistry behavior (reversibility of imine bond formation) then most of the intermediates in its synthesis should behave similarly.

---

[23] Hasell, T.; Schmidtmann, M.; Stone, C. A.; Smith, M. W.; Cooper, A. I. *Chem. Commun.* **2012**, 48, 4689-4691.

[24] a) Yoneya, M.; Yamaguchi, T.; Sato, S.; Fujita, M. *J. Am. Chem. Soc.* **2012**, 134, 35, 14401-14407. b) Yoneya, M.; Tsuzuki, S.; Yamaguchi, Y.; Sato, S.; Fujita, M. *ACS Nano* **2014**, 8, 2, 1290-1296.

[25] Mishra, S. S.; Kompella, S. V. K.; Krishnaswamy, S.; Balasubramanian, S.; Chand, D. K. *J. Phys. Chem. Lett.* **2017**, 8, 9, 2082-2086.

[26] Tachi, Y.; Sato, S.; Yoneya, M.; Fujita, M.; Okamoto, Y. *Inorg. Chem.* **2020**, 59, 17, 12884-12894.

[27] Jiang, T.; Zhang, H.; Cui, Z.; Tan, T. *J. Phys. Chem. Lett.* **2017**, 8, 9, 2082-2086.

[28] Matsumura, Y.; Hiraoka, S.; Sato, H. *Phys. Chem. Chem. Phys.* **2017**, 19, 20338-20342.

[29] Takahashi, S.; Sasaki, Y.; Hiraoka, S.; Sato, H. *Phys. Chem. Chem. Phys.* **2017**, 21, 6341-6347.

### *2.3.3 Key Previous Studies*

Finally we move to a selection of studies (mainly of experimental nature) that provide deep insights into the possible mechanism of the self-assembly. In these studies they were able to couple mass spectrometry and liquid chromatography to identify intermediates of these reactions.

First, we start with a study on imine macrocycles by Kuhnert and coworkers in 2012 (summarized in Figure 31).[16] Among the compounds included in the study, two monomer combinations are of special relevance. In both cases they maintain the amine monomer (1,2-diaminecyclohexane) but change the aldehyde monomer (*p*-benzenedialdehyde and *m*-benzenedialdehyde). First they characterize the intermediates of both self-assembly processes in a 1:1 ratio at 0.1 M concentrations. They find that the product of the *para* aldehyde is a clean **[3+3]**$^{6}$ imine macrocycle however a mixture of **[2+2]**$^{4}$ and **[3+3]**$^{6}$ macrocycles is found for the *meta* aldehyde. Not only that but upon characterization of the main intermediates we can see different trends between the *meta* and the *para* monomers. For the *para* the **[2+1]**$^{2}$ intermediate prevails over the **[1+2]**$^{2}$ intermediate however, we see the opposite tendency for the *meta* monomer where the **[1+2]**$^{2}$ is favored.

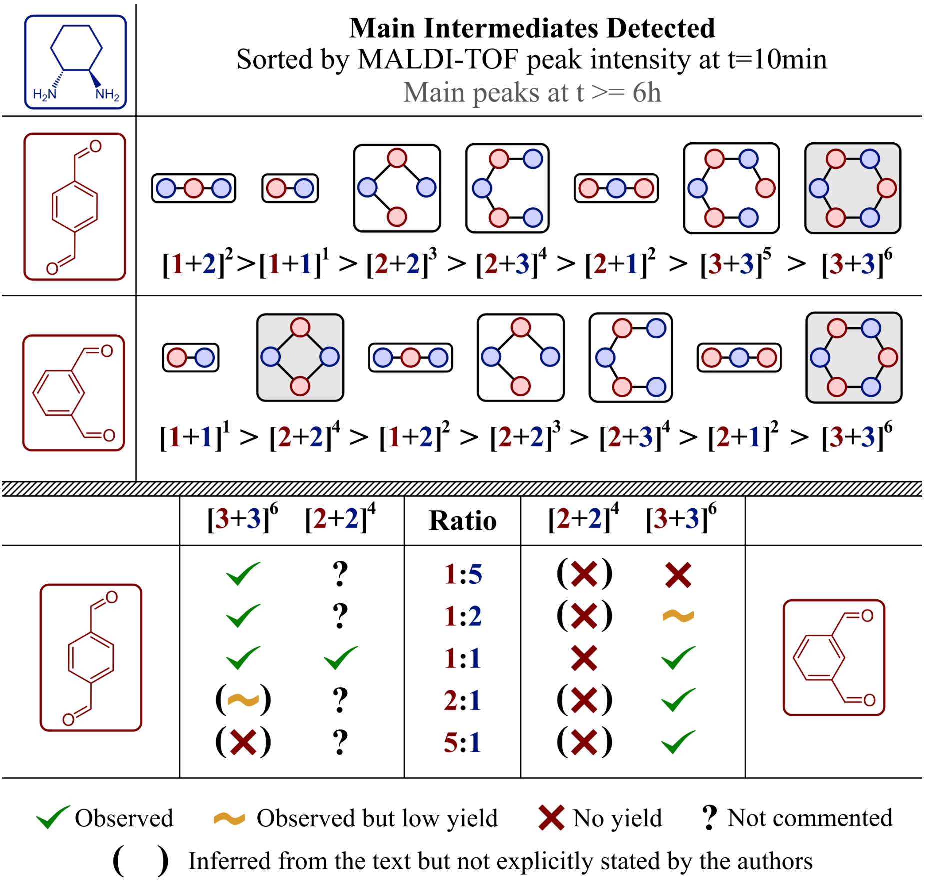

$H_2N$
$NH_2$
**Main Intermediates Detected**
Sorted by MALDI-TOF peak intensity at t=10min
Main peaks at t >= 6h
$[1+2]^2 > [1+1]^1 > [2+2]^3 > [2+3]^4 > [2+1]^2 > [3+3]^5 > [3+3]^6$
$[1+1]^1 > [2+2]^4 > [1+2]^2 > [2+2]^3 > [2+3]^4 > [2+1]^2 > [3+3]^6$
O
O
$[3+3]^6$ $[2+2]^4$ Ratio $[2+2]^4$ $[3+3]^6$
✓ ? 1:5 (✗) ✗
✓ ? 1:2 (✗) ~
✓ ✓ 1:1 ✗ ✓
(~) ? 2:1 (✗) ✓
(✗) ? 5:1 (✗) ✓
✓ Observed
~ Observed but low yield
✗ No yield
? Not commented
( ) Inferred from the text but not explicitly stated by the authors

*Figure 31: Summary of the results by Kuhnert and coworkers (ref 16). Main intermediates detected, products and effects of the change in the monomer ratios.*

Putting aside the cause, the consequence seems obvious: the *para*-substituted monomer favors the chain growth of the molecule through the addition of aldehydes whereas in the *meta*-substituted prefers to grow through amine additions. Further support for this is seen at the outcomes at differing monomer ratios. Where the *para*-monomer reduces the yield of the cage with excess of the amine whereas the *meta*-monomer reduces the yield with excess of aldehyde. Finally, it is to note that in the case of the *meta*-substituted aldehyde the excess of amine favors the formation of the **[3+3]**$^{6}$ cycle.

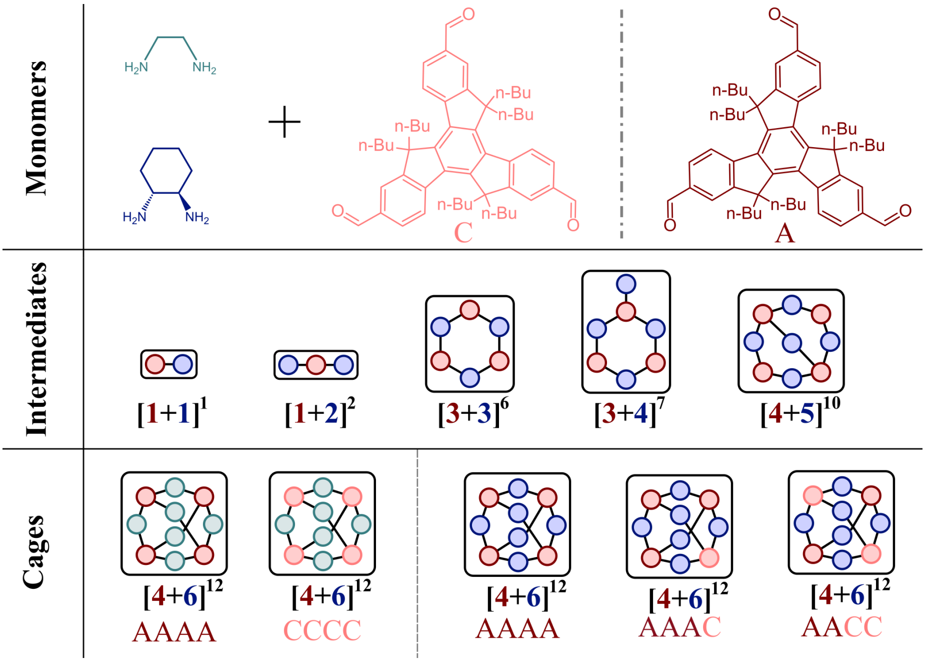

Monomers
Intermediates
Cages
H2N
NH2
H2N
NH2
+
O
n-Bu
n-Bu
n-Bu
n-Bu
n-Bu
n-Bu
O
O
C
O
n-Bu
n-Bu
n-Bu
n-Bu
n-Bu
n-Bu
O
O
A
[1+1]1
[1+2]2
[3+3]6
[3+4]7
[4+5]10
[4+6]12
AAAA
[4+6]12
CCCC
[4+6]12
AAAA
[4+6]12
AAAC
[4+6]12
AACC

*Figure 32: Summary of the work by Cao and coworkers (ref 30). A and C are the two faces of the same compound which end up exposed to the outer side of the final cage.*

Next, we move to the results from Cao and coworkers (summarized in Figure 32). In 2016 they study the self-assembly kinetics of a large **[4+6]**$^{12}$ chiral cage.[30] Although this imine cage uses a significantly different aldehyde monomer compared with the CC1 cage it highlights a key feature of these cages: the anti-clockwise chirality. Later, in 2017, a study on the kinetics of the racemization of the cage[31] which sheds some light into the latest steps of the self-assembly the cage closure/opening.

---

[30] Wang, X.; Wang, Y.; Yang, H.; Fang, H.; Chen, R.; Sun, Y.; Zheng, N.; Tan, K.; Lu, X.; Tian, Z.; Cao, X. *Nat. Commun.* **2016**, 7, 12469.

[31] Wang, Y.; Fang, H.; Zhang, W.; Zhuang, Y.; Tian, Z.; Cao, X. *Chem. Commun.* **2017**, 53, 8956-8959.

In 2016 a study by Wang et al.[30] identified the main species when using EtDA with a chiral trialdehyde, finding two enantiomerically pure cages. However, in the case of CHDA a mixture is observed. It is for these cages that they attempted to identify the different intermediates. The main species detected were the **[1+1]**[1] and **[1+2]**[2] acyclic intermediates, then the **[3+3]**[6] and the **[3+4]**[7] intermediates, both containing a **[3+3]** macrocycle. Finally they identified the presence of a **[4+5]**[10] intermediate which is difficult to decouple from the doubly substituted cage. Finally, they also characterized the obtained homo and mono-substituted cages. The thermodynamic products of the EtDA cage are both homo-cages, however each enantiomer of the CHDA leads to a single homo-cage. It seems reasonable to think that the cause of this is the freedom of the EtDA to rotate the N-C-C-N dihedral.

In 2017, a study of high relevance was reported by the same research group.[31] The authors were able to prove that the racemization proceeds significantly faster than the self-assembly. This means that a full decomposition of the cage into its monomers is very unlikely. Unfortunately, finding the exact pathway for the racemization is difficult as the authors discuss in the study, mainly due to the inability to differentiate experimentally the various pathways that may contribute to the racemization.

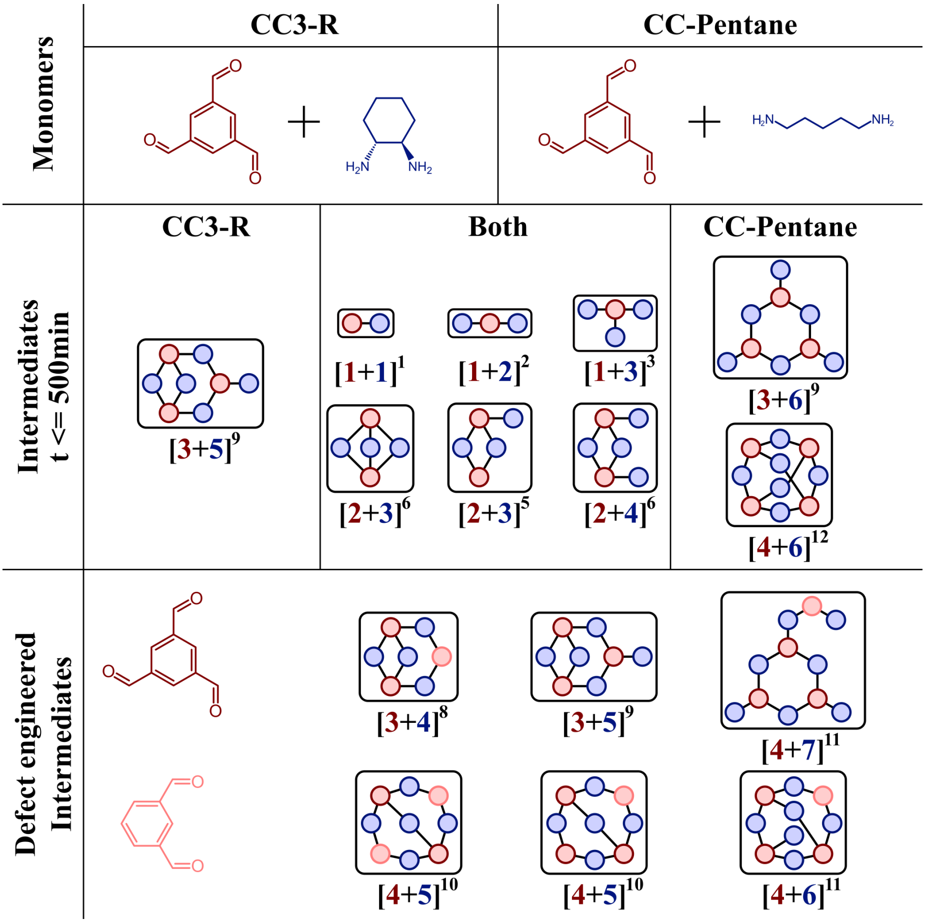

Monomers
CC3-R
CC-Pentane
$H_2N$
$NH_2$
Intermediates t <= 500min
CC3-R
Both
CC-Pentane
[3+5]9
[1+1]1
[1+2]2
[1+3]3
[2+3]6
[2+3]5
[2+4]6
[3+6]9
[4+6]12
Defect engineered Intermediates
[3+4]8
[3+5]9
[4+7]11
[4+5]10
[4+5]10
[4+6]11

*Figure 33: Summary of the intermediates detected by Zhu et al., reference 32.*

Last, but not least, we have the study of Zhu et al. in 2018.[32] This study is carried out for two imine cages, the CC3-R imine cage and the CC-pentane. Not only this is the most relevant study for us due to compound similarity but also because, up to our knowledge, it is the first time that such a wide amount of intermediates have been characterized.

Starting with the results of the CC3-R imine cage, the acyclic **[1+1]**$^1$, **[1+2]**$^2$ and **[1+3]**$^3$ intermediates, the **[2+3]**$^5$ and **[2+4]**$^6$ macrocycles of **[2+2]** rings, the **[2+3]**$^{6}$ cage and a **[3+5]**$^{9}$ intermediate are detected in the first 500 mins of reaction. By the time of the first measurement the **[1+1]**$^{1}$ has already been converted to either **[1+2]**$^{2}$ or **[1+3]**$^{3}$, which matches well with preference for an amine directed growth of the *meta*-substituted dialdehyde observed by Kuhnert and coworkers. However, contrary to their observations, Zhu et al. observe the formation of a **[2+2]** macrocycle. From the evolution over time of the species we can infer that the **[2+3]**$^{6}$ cage is more stable than **[2+3]**$^{5}$. Otherwise it would be difficult to observe accumulation of the **[2+3]**$^{6}$ in the early stages of the reaction. Finally, the detection of the **[3+5]**$^{9}$ intermediate which has both a **[2+2]** and a **[3+3]** ring in its structure seems to point that the **[2+2]** ring closure happens as soon as possible but later than the **[2+2]** ring opening to allow the formation of the CC3-R cage.

---

[32] Zhu, G.; Liu, Y.; Flores, L.; Lee, Z. R.; Jones, C. W.; Dixon, D. A.; Sholl, D. S.; Lively, R. P. *Chem. Mater.* **2018**, 30, 1, 262-272.

Moving to the results of the CC-pentane cage, they observe the **[3+6]**$^{9}$ intermediate, that does not have a **[2+2]** ring, instead of the **[3+5]**$^{9}$ intermediate. And, if we look at the free energy profiles obtained by Zhu et al. we see that the CC3-R profile is steeper than the profile of the CC-pentane.

Lastly it is also worth mentioning the detection of other possible intermediates for the CC3-R cage through the usage of a mixture of TFB and *m*-benzenedialdehyde and changing their ratios. With this procedure they were able to identify as plausible intermediates the **[3+4]**$^{8}$ (**[3+5]**$^{9}$ without a terminal amine) the **[4+6]**$^{11}$ open cage and the **[4+5]**$^{10}$ intermediate with 2 **[3+3]** rings (where for both intermediates they find two different versions depending on the amount of dialdehyde in the structure).

## 3. Computational Methods

| | |
|---|---|
| **Software** | Gaussian 09 revision D.01[33] |
| **Calculation type** | optimizations. *opt* |
| **Functional** | B3LYP[34] |
| **Basis Set** | 6-31+g(d) for all atoms.[35] |
| **Empirical Dispersion** | GD3.[36] *empiricaldispersion=gd3* |
| **Symmetry** | disabled. *nosymm* |
| **Solvation** | Implicit. SMD for dichloromethane[37]. *scrf(smd,dichloromethane)* |
| **Transition state optimizations** | *opt=(ts,calcfc,noeigentest,maxstep=4)* |
| **Difficult SCF convergence** | *scf=xqc, guess=always* |
| **Integration grid** | 75 radial shells and 302 angular points per shell.[38] Gaussian09 default. *int=(Grid=finegrid)* |

---

[33] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, revision D.01*; Gaussian, Inc. Wallingford, CT, **2013**.

[34] a) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, 98, 11623−11627. b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, 37, 785−789. c) Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648−5652.

[35] Dunning, T. H. *J. Chem. Phys.* **1989**, 90, 1007.

[36] Grimme, S. *J. Comput. Chem.* **2006**, 27, 1787−1799.

[37] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, 113, 6378−6396.

[38] Lebedev, V. I. *Zh. Vychisl. Mat. Mat. Fiz.* **1976**, 16, 293-306.

| | |
|---|---|
| **Software** | Gaussian 09 revision D.01 |
| **Calculation Type** | frequencies. *freq* |
| **Input Geometry** | Geometries optimized with B3LYP-D3/6-31+g(d) |
| **Functionals** | B3LYP |
| **Basis Sets** | 6-31+g(d) for all atoms. |
| **Empirical Dispersion** | GD3. *empiricaldispersion=gd3* |
| **Symmetry** | Disabled. *nosymm* |
| **Solvation** | Implicit. SMD for dichloromethane.<br>*scrf(smd,dichloromethane)* |
| **Difficult SCF convergence** | *scf=xqc, guess=always* |
| **Integration grid** | 75 radial shells and 302 angular points per shell. Gaussian09 default.<br>*int=(Grid=finegrid)* |
| **Frequencies** | RRHO approximations. Gaussian09 default. |
| **Software used for corrections** | GoodVibes software used.[39] |
| **Reference state corrections** | Liquid phase at 1M<br>298.15K |
| **Quasi-harmonic corrections** | Grimme style corrections.[40]<br>Frequency threshold 50cm$^{-1}$. |

---

[39] Funes-Ardoiz, I.; Paton, R. S. *Goodvibes 2.0.2*, **2016.** DOI:10.5281/zenodo.595246
[40] Grimme, S. *Theory Chem. Eur. J.* **2012**, 18, 9955-9964.

| | |
|---|---|
| **Software** | Gaussian 09 revision D.01 |
| **Calculation Type** | Single point calculations |
| **Input Geometry** | Geometries optimized with B3LYP-D3/6-31+g(d) |
| **Functionals** | B3LYP |
| **Basis Sets** | 6-311+g(d,p) for all atoms. |
| **Empirical Dispersion** | GD3. *empiricaldispersion=gd3* |
| **Symmetry** | Disabled. *nosymm* |
| **Solvation** | Implicit. SMD for dichloromethane. *scrf(smd,dichloromethane)* |
| **Difficult SCF convergence** | *scf=xqc, guess=always* |
| **Integration grid** | 75 radial shells and 302 angular points per shell. Gaussian09 default. *int=(Grid=finegrid)* |

# 4. Results

In this section we will start presenting the thermodynamics of different cages and then we will move to the study of the reactions involved in the initial chain growth. Amongst these we will find the first monomer condensation as well as the different initial reactions. We will move afterwards to reactions leading to the formation of macrocycles, ring closures. Finally we will study some of the closures of the cages whose thermodynamics were presented at the beginning.

## 4.1 Cage Thermodynamics

Following previous computational results on imine cages we started by computing different cage stoichiometries to compare their stability. We considered the two possible **[4+6]**$^{12}$ cages, following the nomenclature of Santolini et al.[17] the $Tri^{4} Di^{6}$ cage (that we will refer as **[4+6]a** for simplicity) and the $Tri^{4}_{2} Di^{6}$ cage ( from now referred as **[4+6]b** ). In Figure 34 we can see the different cages that we calculated as well as their relative energies.

On the left axis we have the free energy relative to the initial monomers (please note that the energies are negative). According to this value the most stable cage would be the **[6+9]**$^{18}$ cage[41]. But as it was previously commented, it is more appropriate to compare the free energies per bond, which are in the right axis. According to this scale we can see that the **[4+6]a** cage, CC1, is the most stable. Nonetheless we can see that the relative energy grows larger in absolute value as we increase the size of the cage which depicts an overall exergonic free energy profile. We need to highlight that we have also added in this figure the **[3+4]**$^{8}$ “cage” which is the second most stable per bond cage of the cages shown. This incomplete cage is an intermediate found by Zhu et al.[32] as a defect engineered CC3-like cage and its **[3+5]**$^{9}$ version was found as an intermediate in the synthesis of the CC3-R. One of the key features of this compound is that it contains a **[2+2]** and a **[3+3]** macrocycle, which coupled with its low energy per bond, make it a good candidate for connecting the **[2+3]**$^{6}$ cage (and the **[4+6]b**) with the **[4+6]a**.

---

[41] Due to problems in the convergence of the scf we calculated this cage in Gaussian 16 with the same options as the rest of calculations but using the keyword “g09defaults”.

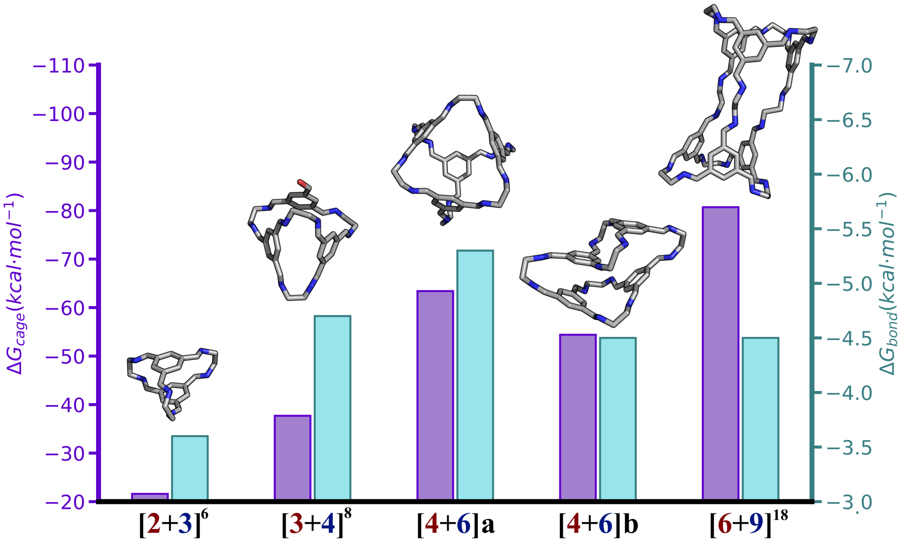


*Figure 34: Relative free energies and relative free energies per bond for the different cages calculated for the TFB and EtDA monomers*

Second, we can see that most of the cages depicted are prone to be relatively flexible, which means that we will have to tackle conformations to the best of our capabilities. In fact, only the geometry of the CC1 was extracted from a crystal structure[42] and then optimized. Up to date we have no record of any of the other cages depicted being characterized experimentally for this set of monomers, thus we generated the geometries and optimized them.

---

[42] Tozawa, T.; Jones, J. T. A.; Swamy, S. I.; Jiang, S.; Adams, D. J.; Shakespeare, S.; Clowes, R.; Bradshaw, D.; Hasell, T.; Chong, S. Y.; Tang, C.; Thompson, S.; Parker, J.; Trewin, A.; Bacsa, J.; Slawin, A. M. Z.; Steiner, A.; Cooper, A. I., CCDC 720848: Experimental Crystal Structure Determination, 2009. DOI: 10.5517/ccs6357

Finally, from a technical perspective, we observed that in SP calculations it is common to find problems in the convergence of the energy at some of the steps during the optimizations for these types of compounds, specially as they grow larger. In most of the cases we found that increasing the number of steps for the scf algorithm and/or enforcing a guess at each step of the optimization was able overcome the problem.

## 4.2 Chain Growth

In this section we will investigate what we have termed as chain growth reactions. Under this term we include the bimolecular reactions that lead to bimolecular products. That is, intermediate or monomer plus a second intermediate or monomer to produce water and the grown intermediate.

### *4.2.1 First Reaction*

We calculated the reaction profile for the reaction with the monomers, using the information from Chapter I. Contrary to the previous chapter, where the presence of water could be scarce, in the synthesis of the cage up to 12 water molecules per cage are formed. Thus we restricted the characterization to the water dimer pathway. We can see the results plotted in teal in Figure 35 and compare them with the results of the previous chapter in purple.

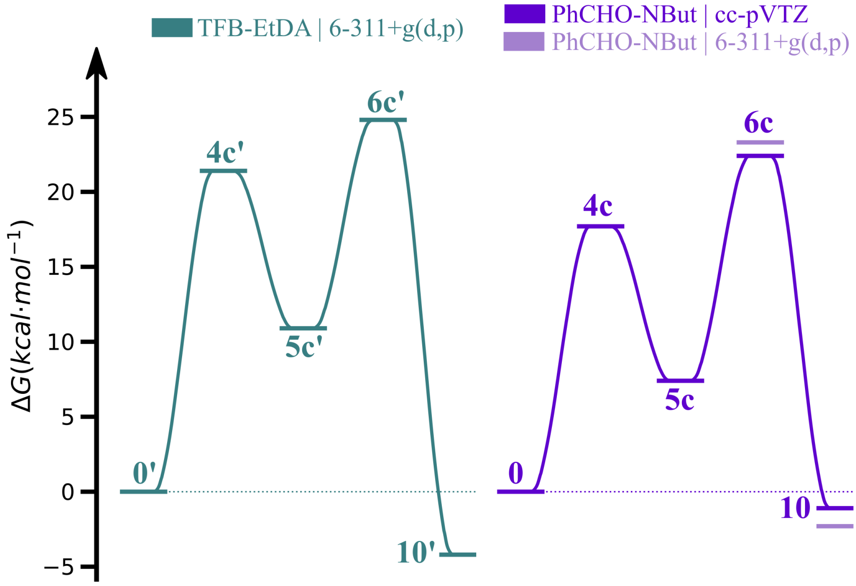


*Figure 35: Free energy profiles of the Imine condensation reaction. In* ***teal*** *the TFB and EtDA profile. In* ***purple*** *the benzaldehyde (PhCHO) and n-butylamine (NBut) profile. Numbers correspond to labels used in Chapter I.*

We can see that the profile follows the same shape with higher barriers for both TSs, hemiaminal formation (**4c’**) and dehydration (**6c’**) both catalyzed by a water dimer acting as a proton shuttle, and a less stable hemiaminal intermediate (**5c’**). We believe that the main cause for that, a part from electronic effects, is the difference in the computational methodology used. Such idea is confirmed when we plot in **light purple** the dehydration TS and the products of the previous chapter calculated with the same basis set.

Overall the key results that we can extract from here are that the mechanism does not change and that the dehydration TS still marks the rate determining step. Thus, we can simplify our exploration by only characterizing such TS instead of the whole profile per each reaction that we study.

### *4.2.2 Amine Monomer Reactions,* **[m+n]$^{i}$→ [m+(n+1)]$^{i+1}$**

Next we move to study the reactions that increase the size of the intermediates by incorporating a new amine monomer. We can see the sequential condensation of a single aldehyde monomer (TFB) with three amine monomers in Figure 36.

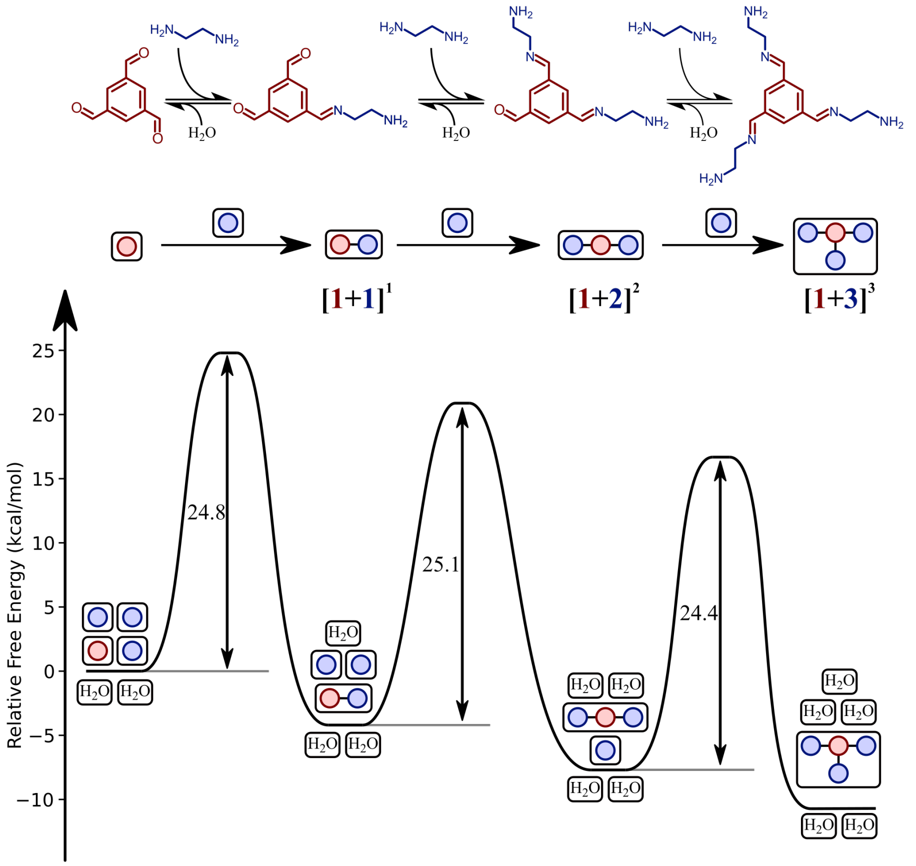


*Figure 36: First reactions that add an amine monomer. Their barriers and their representations.*

Under the reactions we can see the representations that we have been already using in the chapter as well as the free energy profile obtained with a visual depiction of the reference state for the each relative free energy of each intermediate. From the representation we can already guess that there is not a big difference when introducing a new amine monomer. This becomes even more obvious if we look at the results in Table 3 the energetic demand is similar. Not only when we “saturate” a single aldehyde monomer but when we add an amine to a “saturated” amine (**[2+1]**$^{2}$) or when we add it to a macrocyclic species (**[4+5]**$^{10}$).

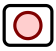
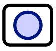
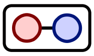
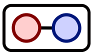
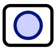
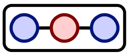
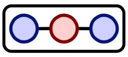
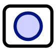
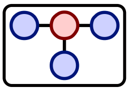
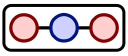
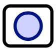
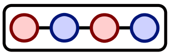
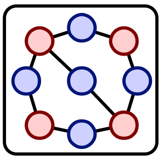
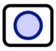
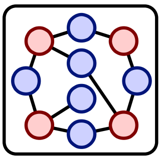
| Reactants | | | → | Product | $\Delta G_d^\ddagger$ | $\Delta G_r^\ddagger$ | $\Delta\Delta G_d^\ddagger$ | $\Delta\Delta G_r^\ddagger$ |
|---|---|---|---|---|---|---|---|---|
|  | + |  | → |  | 24.8 | 29.0 | 0.0 | 0.0 |
|  | + |  | → |  | 25.1 | 28.7 | 0.3 | -0.3 |
|  | + |  | → |  | 24.5 | 27.4 | -0.3 | -1.6 |
|  | + |  | → |  | 24.1 | 25.1 | -0.7 | -3.9 |
|  | + |  | → |  | 25.5 | 27.1 | 0.7 | -1.9 |
| **Std. deviation** | | | | | 0.5 | 1.4 | | |

*Table 3: Amine Saturation reactions. Energies in kcal mol*$^{-1}$*.* $\Delta\Delta G^\ddagger$ *is the difference between the barrier and the barrier of the condensation between the two monomers. The standard deviation does not include the first entry.*

Throughout the chapter, we will see several tables similar to Table 3. It displays the direct ($\Delta G_d^\ddagger$) and reverse barrier ($\Delta G_r^\ddagger$) of the reaction, allowing us to detect more likely outcomes of competing reactions as well as the main direction of the reaction. It also shows how the barriers differ from the first monomer condensation which is going to facilitate the detection of stable biases of the different types of reactions that we consider. As this type of table is going to be common, we will shorten the caption unless the table contents differ significantly.

Going back to the results, it is interesting to note that there is less difference in the direct barriers than in the reverse barriers. We are characterizing the hemiaminal dehydration step, whose TS is geometrically more similar to the imine. However, it seems that it would be better to predict the energy of the TS using the the direct barrier instead of the reverse. If we look at the relative differences with the first reaction we can conclude that adding an amine monomer behaves like the first monomer condensation, not significantly faster nor slower.

### *4.2.3 Aldehyde Monomer Reactions,* **$[m+n]^{i} \rightarrow [(m+1)+n]^{i+1}$**

Next we move to the competing reaction, the reaction with an aldehyde monomer. This type of reaction has a higher impact on the molecular weight of the resulting product. We started incrementally adding reactions to study. Table 4 summarizes these reactions. Rows 2-4 involve reactions going from **$[1+n]^{i}$** to **$[2+n]^{i+1}$** species. Rows 5-8 reactions going from **$[2+n]^{i}$** to **$[3+n]^{i+1}$** species and the final row corresponds to the formation of a branched **$[4+n]^{i+1}$** species.

Overall we can see that when increasing the length of the chain, we have a more stable behavior, although this time coming from the reverse barrier. Contrary to the amine monomers, it seems better to predict the reverse barrier for these reactions from the product. A simple explanation that is diluted in the representation of the species through graphs is that conformational complexity that we introduce with the addition of the aldehyde is higher, and thus the TS is more similar to products than to reactants.

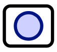
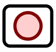
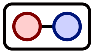
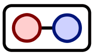
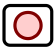
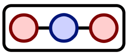
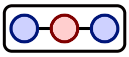
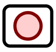
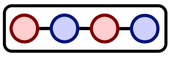
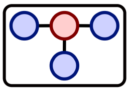
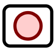
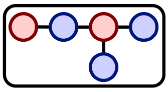
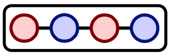
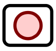
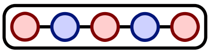
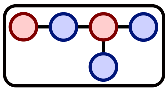
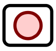
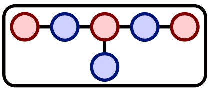
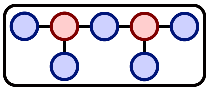
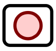
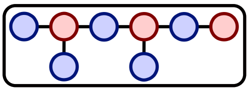
| **Reactants** | | **→** | **Product** | $\Delta G_d^‡$ | $\Delta G_r^‡$ | $\Delta\Delta G_d^‡$ | $\Delta\Delta G_r^‡$ |
|---|---|---|---|---|---|---|---|
|  | +  | → |  | 24.8 | 29.0 | 0.0 | 0.0 |
|  | +  | → |  | 27.7 | 29.5 | 2.9 | 0.5 |
|  | +  | → |  | 25.6 | 27.8 | 0.8 | -1.2 |
|  | +  | → |  | 24.7 | 28.1 | -0.1 | -0.9 |
|  | +  | → |  | 29.0 | 27.5 | 4.2 | -1.5 |
|  | +  | → |  | 28.0 | 27.1 | 3.2 | -1.8 |
|  | +  | → |  | 29.2 | 27.1 | 4.4 | -1.8 |
| **Std. deviation** | | | | 1.8 | 0.8 | | |

*Table 4: Aldehyde Saturation reactions.*

Comparing the direct barriers of tables 3 and 4 we see that there is preference for the addition of amine monomers over the addition of aldehyde monomers. This result couples very well with the literature results. We commented how in the study of Kuhnert and coworkers[16] they found that the meta-substituted aldehyde monomer had a preference for the **[1+2]**$^2$ intermediate instead of the **[2+1]**$^2$ and how such preference played a significant role when varying the ratios of monomers. Also no intermediate with a terminal aldehyde was detected in the work of Zhu et al.[32]

### *4.2.4 Side Reactions, Diaza-Cope Rearrangement*

In the same work of Zhu et al. a test calculation of the Diaza-Cope rearrangement[43] was reported for the **[2+1]$^2$** intermediate. The authors explored if such reaction is a relevant side reaction for these types of cages. We proceeded to calculate the product of such reaction and, for completeness, the TS (see Figure 37). We found an endergonic reaction of 12.7 kcal mol$^{-1}$ with a direct barrier of 30.7 kcal mol$^{-1}$. The difference in stability coupled with barrier led us to believe that for this specific system is not going to play a significant role and thus, we can skip its calculation for other species.

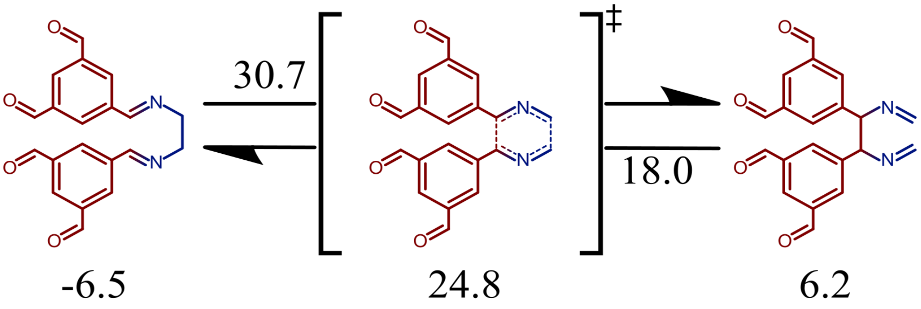


*Figure 37: Diaza-Cope Rearrangement reaction. Direct (above) and reverse (below) energy barriers in kcal mol$^{-1}$.*

The results are also compatible with the results found by Zhu et al. where they obtain a difference in energy of 9.3 kcal mol$^{-1}$ favoring the **[2+1]$^2$** intermediate. Also, older studies such as the 2009 study of Kim et al.[44] got similar conclusions. With these results the use of the reverse reaction to synthesize the monomers for the cages by Giri et al.[45] and Melaugh et al.[46] becomes even easier to understand.

---

[43] Enders, D.; Knopp, M.; Schiffers, R. *Tetrahedron:Asymmetry*, **1996**, 7, 1847.

[44] Kim, H.; Staikova, M.; Lough, A. J.; Chin, J. *Org. Lett.* **2009**, 11, 1, 157-160.

[45] Giri, N.; Davidson, C. E.; Melaugh, G.; Del Pópolo, M. G.; Jones, J. T.; Hasell, T.; Cooper, A. I.; Horton, P. N.; Hursthouse, M. B.; James, S. L. *Chem. Sci.* **2012**, 3, 2153-2157.

[46] Melaugh, G.; Giri, N.; Davidson, C. E.; James, S. L.; Del Pópolo, M. G. *Phys. Chem. Chem. Phys.* **2014**, 16, 9422-9431.

### *4.2.5 Initial Chain Growth, Reactions* $[1+n]^i \rightarrow [2+m]^j$

After the reactions with the monomers, we also need to consider that as we generate new intermediates, those are also candidates for reacting with themselves or other species. Thus we calculated the different reactions connecting $[1+n]^i$ species with $[2+m]^j$ species. The results are in Tables 5, 6 and 7.

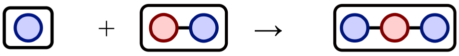
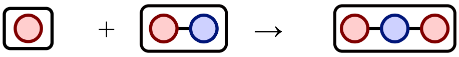
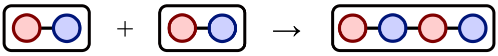
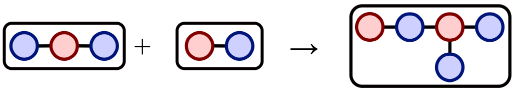
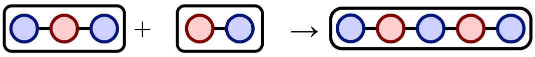
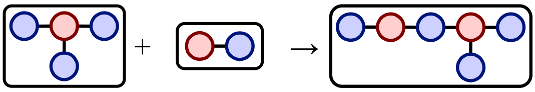
| | $\Delta G_d^\ddagger$ | $\Delta G_r^\ddagger$ | $\Delta\Delta G_d^\ddagger$ | $\Delta\Delta G_r^\ddagger$ |
|---|---|---|---|---|
|  | 25.1 | 28.7 | 0.3 | -0.3 |
|  | 27.7 | 29.5 | 2.9 | 0.5 |
|  | 28.6 | 30.2 | 3.8 | 1.2 |
|  | 26.0 | 28.1 | 1.2 | -0.8 |
|  | 25.7 | 28.3 | 0.9 | -0.7 |
|  | 24.0 | 27.1 | -0.8 | -1.9 |

*Table 5:* $[1+n]^i \rightarrow [2+(n+1)]^{i+2}$ *reactions.*

Overall, we can see from Tables 6 and 7 that the reactions with those species seem to be easier to predict than the reactions of Table 5. Of special interest in this table are the entries of the reaction with the $[1+2]^2$ where we can see that the barrier is similar, independently of which functional group of the $[1+1]^1$ is reacting, the amine (row 4) or an aldehyde (row 5). Another interesting result is that the dimerization of $[1+1]^1$ has the highest barrier. If we had to choose a strategy for predicting the barrier it might be better to use the reverse barriers.

Next, moving to Table 6, it seems that a bias in the barriers is easier to identify. Probably the best approach in this case is to predict the energy of the transition state from the backwards reaction. If we look at the dimerization reaction (row 5), it seems to be a more relevant kinetic competitor compared to the dimerization of the **[1+1]**$^1$ intermediate (row 3 Table 5). Nonetheless, the addition of an amine (row 1) is preferred over dimerization as well as the reaction with the product of such addition (row 6). That both early dimerizations (row 3 Table 5 and row 5 Table 6) are not favored, coupled with the dropwise addition of the aldehyde reactant in the optimized batch synthesis might point towards an irrelevant role of dimerizations in the overall mechanism but such claim would be very difficult to prove.

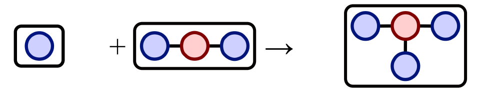
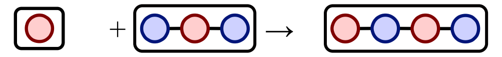
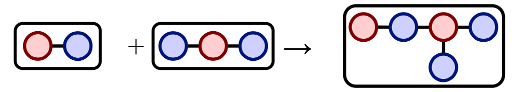
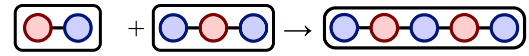
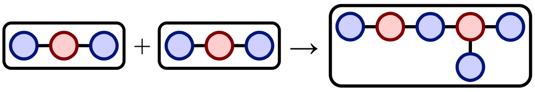
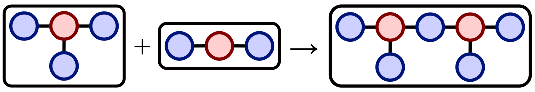
| | $\Delta G_d^{\ddagger}$ | $\Delta G_r^{\ddagger}$ | $\Delta\Delta G_d^{\ddagger}$ | $\Delta\Delta G_r^{\ddagger}$ |
|---|---|---|---|---|
|  | 24.5 | 27.4 | -0.3 | -1.6 |
|  | 25.6 | 27.8 | 0.8 | -1.2 |
|  | 26.0 | 28.1 | 1.2 | -0.8 |
|  | 25.7 | 28.3 | 0.9 | -0.7 |
|  | 25.3 | 27.8 | 0.5 | -1.2 |
|  | 23.7 | 27.9 | -1.0 | -1.1 |

*Table 6: [**1**+**n**]$^{i}$ → [**2**+(**n**+2)]$^{i+3}$ reactions.*

Finally, arriving to the results of Table 7 we find the behavior of a "saturated aldehyde". If we consider the intermediate as a triamine monomer we see the same behavior as with the amine monomer additions. Estimating the energy of the TS seems easier from the direct barrier and there is very little difference that it adds to an already substituted aldehyde or not.

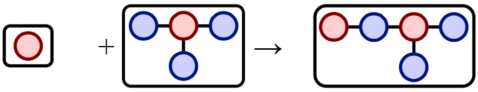
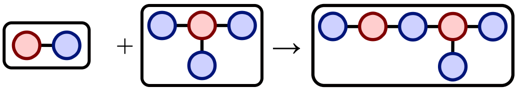
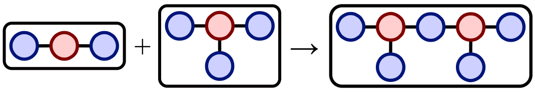
| | $\Delta G_d^\ddagger$ | $\Delta G_r^\ddagger$ | $\Delta\Delta G_d^\ddagger$ | $\Delta\Delta G_r^\ddagger$ |
|---|---|---|---|---|
|  | 24.7 | 28.1 | -0.1 | -0.9 |
|  | 24.0 | 27.1 | -0.8 | -1.9 |
|  | 23.7 | 27.9 | -1.0 | -1.1 |

*Table 7:* ***[1+n]***$^{i}$ *→* ***[2+(n+3)]***$^{i+4}$ *reactions.*

Summarizing, in the early stages of the reaction there is preference to react with the amine monomer so the reaction is likely to proceed through the pathway **[1+1]**$^1$→**[1+2]**$^2$→**[1+3]**$^3$ leading to the accumulation of the **[1+3]**$^3$ intermediate which is reactive enough to join with whichever of the other species present to give a **[2+(n+3)]**$^i$ species. Going back to the work of Zhu et al.[32] they arrived to the same conclusion from an experimental perspective.

## 4.3 Ring Closures

Now we move to the Ring Closures. Under this name we consider the intramolecular reactions that have a non-macrocyclic reactant and a macrocyclic product.

### *4.3.1 [2+2] Ring*

The smallest macrocycle that we can find along the CC1 self assembly is the **[2+2]**$^4$ macrocycle. This macrocycle may adopt two main conformations depicted in Figure 38. As we can see from the figure, the A conformation is favored over the B conformation. Such difference in stability is likely due to the different interactions between the aromatic groups as well as their interaction with the solvent.

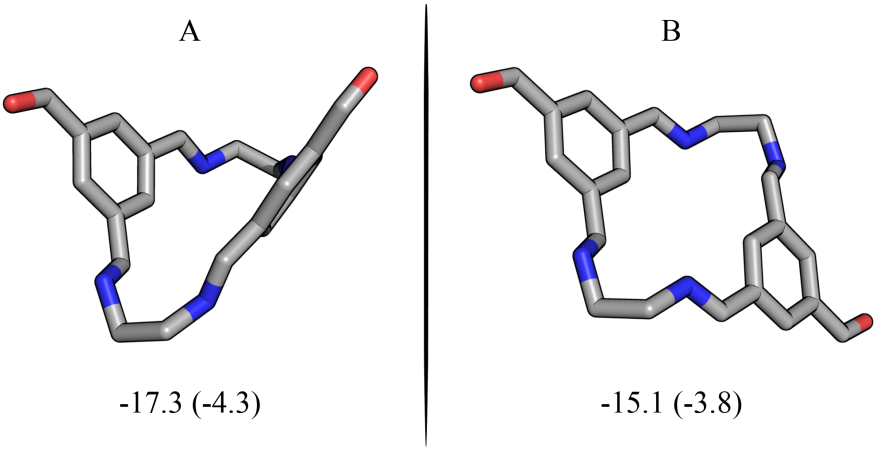


*Figure 38: Main conformations of the **[2+2]**$^4$ macrocycle with their free energies of formation (and free energy of formation per bond) in kcal mol$^{-1.}$*

We then proceed to calculate all the possible **[2+2]** ring closures from the different **[2+n]**$^i$ species (Table 8). The results obtained are interesting as ring closures are favored over any of the chain growth reactions studied and reported above. We observe small differences on the different ring closures overall in both directions. From the first two entries we can see that predicting the energy of the transition state from the reactant is probably the safest strategy. The next two rows are two different closures that arrive to the same product with similar energy barriers.

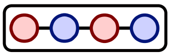
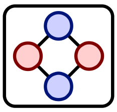
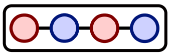
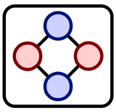
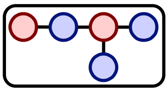
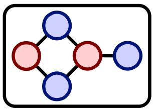
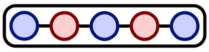
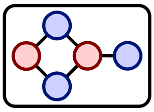
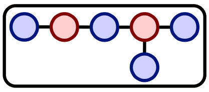
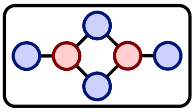
| **Reactants** | → | **Product** | $\Delta G_d^\ddagger$ | $\Delta G_r^\ddagger$ | $\Delta\Delta G_d^\ddagger$ | $\Delta\Delta G_r^\ddagger$ |
|---|---|---|---|---|---|---|
|  | → |  A | 21.1 | 28.5 | -3.7 | -0.5 |
|  | → |  B | 21.6 | 26.9 | -3.1 | -2.1 |
|  | → |  | 21.4 | 29.2 | -3.4 | 0.2 |
|  | → |  | 22.1 | 29.5 | -2.7 | 0.5 |
|  | → |  | 22.0 | 27.7 | -2.7 | -1.2 |
| **Std. deviation** | | | 0.4 | 1.0 | | |

*Table 8: **[2+2]** Ring closing reactions.*

Barriers for **[2+2]** ring closures (Table 8) are on the range of 21-22 kcal mol$^{-1}$ while favorable chain growth reactions have barriers on the range of 24-26 kcal mol$^{-1}$ (Tables 5-7). With these differences in energy the main conclusion would be that as soon as a **[2+2]** ring closure can happen, it happens. Again, it agrees with the findings of Zhu et al.[32] as no intermediate with two aldehydes without a **[2+2]** ring was reported. Such claim seems to go against the results of Kuhnert and coworkers[16] as they do observe these intermediates. A plausible cause could be the 1:1 ratio of the monomers as well as the impossibility to close a cage with ditopic monomers. In the study of Zhu et al. they do find the **[2+3]**$^{6}$ cage and its immediately related intermediates (**[2+3]**$^{5}$,

**[2+4]**$^{6}$). Nonetheless, they find a more intense signal of the **[2+2]**$^{4}$ than the **[2+1]**$^{2}$, **[2+2]**$^{3}$ or the **[2+3]**$^{4}$ which agrees with a relatively rapid ring closure.

### *4.3.2 **[3+3]** Ring*

As we move to a larger macrocycle we realize how much the conformational complexity has increased. In Figure 39 we have the different conformations studied of the TS for the ring closure. The figure highlights the main conformational changes considered between the TS geometries.

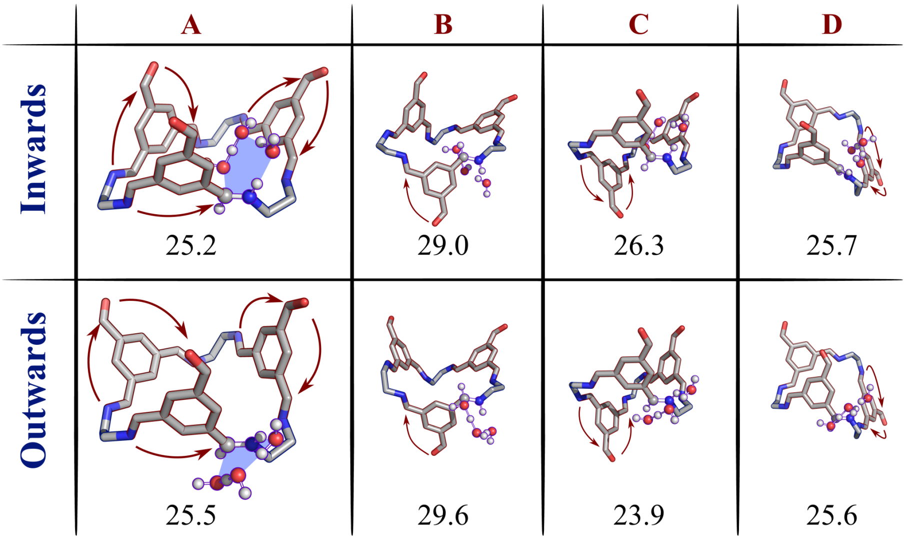


*Figure 39: Conformations of the TS for the **[3+3]** Ring closure of the **[3+3]**$^{5}$ intermediate along with the direct barriers. Inwards and Outwards refers to the orientation of the proton shuttle. B,C and D are variations of A by flipping an aldehyde moiety. The arrows show the clockwise or anticlockwise orientation or the aldehyde moieties.*

First, we have two possible orientations of the TS, inwards (top row) and outwards (bottom row). Second, we consider the orientations of the aromatic moieties with respect to each other and relative to the position of the bond being formed (A being the reference, B flipping the closest moiety, C flipping the furthest and D flipping the remaining one). Finally, we have to consider the clockwise/anticlockwise face of the aldehyde monomer (as a consequence of a flip, the face is reversed), which gives rise to 8 extra enantiomeric (these are redundant from an energetic point of view, and thus are omitted in the figure).

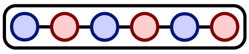
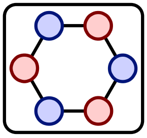
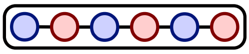
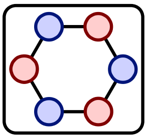
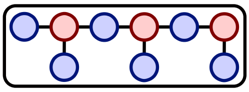
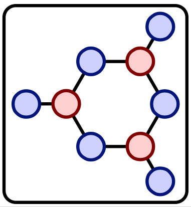
| | | | | $\Delta G_d^‡$ | $\Delta G_r^‡$ | $\Delta\Delta G_d^‡$ | $\Delta\Delta G_r^‡$ |
|---|---|---|---|---|---|---|---|
|  | → |  | OC | 23.9 | 33.3 | -0.9 | 4.3 |
|  | → |  | OA | 25.5 | 34.8 | 0.7 | 5.9 |
|  | → |  | OA | 26.8 | 34.9 | 2.0 | 6.0 |

*Table 9:* ***[3+3]*** *Ring closing reactions.*

Within the inwards and outwards we find small energy differences except in the case of the lowest TS (outwards C or OC, Figure 39). We believe that such difference is likely due to the strain of the IC (inwards C, Figure 39) conformation rather than a stabilizing effect of the OC. We also calculated the OA (outwards A, Figure 39) conformer of a different **[3+3]** closure (row 3 Table 9) which indicates a similar behavior for the **[3+3]** ring openings of the same conformation.

## 4.4 Cage closures

Finally, we arrive to the Cage Closures. Here we consider a cage closure as a intramolecular reaction with a macrocycle as reactant and a macrocycle as a product where at least 1 new cycle has been formed in the process.

### *4.4.1 **[2+3]**$^6$ Cage*

We start with the smallest cage, the **[2+3]**$^6$ cage. Guided by the results of the work of Jelfs and coworkers[7a] shown in Figure 26 (page 98) we started calculating different conformations of the cage as no crystal structure of this specific cage is reported to our knowledge. In Figure 40 we have depicted the top view of three different conformers of the cage **[2+3]**$^6$ as well as their relative energies.

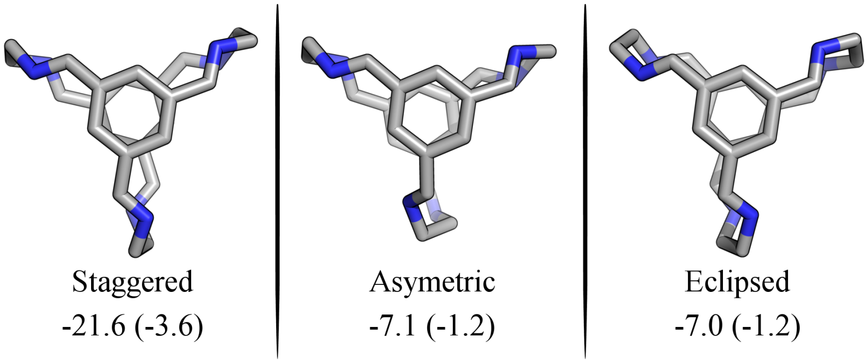


*Figure 40: Conformations of the **[2+3]** Cage values and the relative free energies in kcal mol$^{-1}$ and in parenthesis are the relative free energy per bond.*

A relatively unexpected result is that the “asymmetric” conformer is not so far from the eclipsed conformer energetically speaking. Nonetheless, there is no doubt about the most stable conformation, the staggered conformation. A likely explanation for it might come from the repulsion of the two aromatic rings and the distance between them as well as the strain of the alkyl chains.

Next, we moved to calculate the cage closure for the eclipsed (Eclip) conformation. Here we found a much higher barrier (first row of Table 10) than the barriers that we found during the ring closure reactions. This made us wonder if one of the deciding factors to obtain a **$[2+3]^6$** cage or a **$[4+6]^{12}$** cage that were obtained in the study of Jelfs and coworkers[7a] was this barrier instead of pure thermodynamics. We proceeded to calculate the same reaction for other cages made of a different amines (rows 2-6 of Table 10).

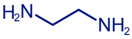
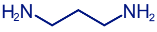
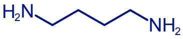
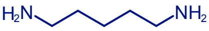
| **Amine** | **Conf** | **$\Delta G_d^\ddagger$** | **$\Delta G_r^\ddagger$** | **$\Delta G^{[2+3]^5}$** | **$\Delta G^{[2+3]^6}$** |
|---|---|---|---|---|---|
|  | Stagg | 27.5 | 27.2 | -21.8 (-4.4) | -21.6 (-3.6) |
|  | Eclip | 19.1 | 24.4 | -21.4 (-4.3) | -26.7 (-4.5) |
| | Stagg | 18.5 | 26.5 | -15.0 (-3.0) | -22.9 (-3.8) |
|  | Stagg | 26.2 | 25.3 | -13.3 (-2.7) | -12.4 (-2.1) |
|  | Stagg | 18.1 | 25.0 | -8.1 (-1.6) | -14.9 (-2.5) |
| | Eclip | 17.4 | 27.2 | -21.0 (-4.2) | -30.9 (-5.1) |

*Table 10: **[2+3]** Cage closing reactions for different diamine monomers. Energies in kcal mol$^{-1}$. Values in parenthesis are relative free energies per imine bond.*

One clear odd-even effect is the stability of conformation of the cage. The cages of the odd-numbered amines (amines with alkyl chains with odd number of carbons) are more stable in the eclipsed meanwhile the even-numbered amines prefer the staggered conformation which supports the findings of by Bera et al.[7b] and Su et al.[7c] where they observe the same trend.

If we look at the barriers the trend becomes even clearer. The odd-numbered have barriers lower than the barriers that we found for ring closures but the even-numbered have significantly higher barriers. Surprisingly this trend is diluted for the reverse barrier. Here we can find a third key point: the reactions of the odd-numbered amines are clearly exergonic whereas the reactions of the even-numbered are not.

Finally, following the same discussion that we had in the previous sections and given the results from the other cages it seems that for this cage closure estimating the energy of the TS from the reverse barrier is probably the best strategy.

### *4.4.2 **[3+4]⁸**, **[4+5]¹⁰** and **[4+6]¹⁰** incomplete cages*

The next cage in size that we consider is the "incomplete" **[3+4]**$^8$ Cage. Contrary to the Cage **[2+3]**$^6$ there are various possible reactions that are going to lead to this cage, we start from a **[3+3]** ring and then close a **[2+2]** ring or start with a **[2+2]** ring and close the **[3+3]** ring afterwards. (rows 1-3 of Table 11 ).[47]

The next cage in size that we can form is the **[4+5]**$^{10}$ "incomplete" cage. Again we have two different reactions leading to this one but in this case both start from a **[3+3]** ring and form a new **[3+3]** ring. (rows 4-5 of Table 11 )

Finally similar to both of these we have the **[4+6]**$^{10}$ incomplete cage closure. In this one we start from a **[4+4]** ring and then close a **[3+3]** ring (row 6 of Table 11 ).

Following the results of the **[2+3]**$^6$ cage we see again how guessing from the reverse reaction seems to be the safest bet. Interestingly we find significantly lower barriers (15-21 kcal mol$^{-1}$), this would lead to think that these cage closures happen faster than any of the previous reactions. If we look back to the experimental results of Zhu et al.[32] and Wang et al.[30] we see a good agreement with our results as these species (or a similar version with terminal amines) are amongst the ones experimentally found.

---

[47] After realizing that we had overlooked this reaction we decided to not calculate it due to the results obtained from the other closures. Nonetheless, we included it in the table to show all possible cage closures for the **[3+4]**$^8$

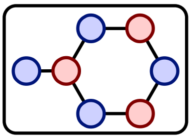
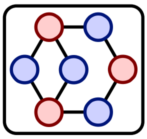
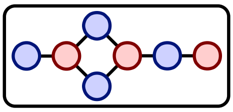
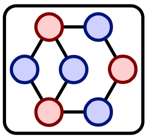
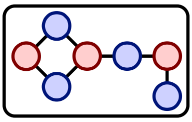
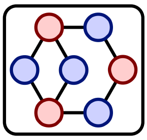
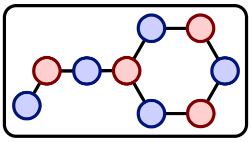
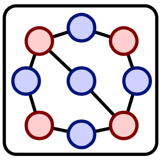
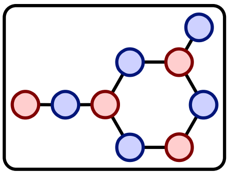
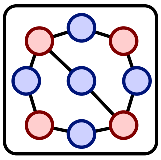
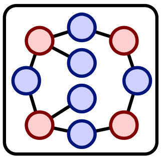
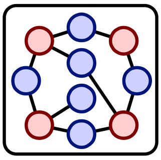
| Reactants | → | Product | $\Delta G_d^\ddagger$ | $\Delta G_r^\ddagger$ | $\Delta\Delta G_d^\ddagger$ | $\Delta\Delta G_r^\ddagger$ |
|---|---|---|---|---|---|---|
|  | → |  | 21.3 | 28.0 | -3.5 | -1.0 |
|  | → |  | 18.4 | 28.0 | -6.4 | -0.9 |
|  | → |  | – | – | – | – |
|  | → |  | 19.4 | 29.5 | -5.4 | 0.6 |
|  | → |  | 16.8 | 29.6 | -8.0 | 0.6 |
|  | → |  | 15.3 | 28.5 | -9.5 | -0.5 |

*Table 11: Incomplete cages' closing reactions.*

### *4.4.3 **[4+6]b** Cage*

Now we move to the alternative **[4+6]**$^{12}$ Cage (**[4+6]b**). Although this cage is not an incomplete cage it can be obtained from two different cage closures. A closure of a **[2+2]** ring or a closure of a **[4+4]** ring summarized in Table 12. Yet again we see that the strategy of predicting the barrier from the backwards reaction might be the best one. If we compare the direct barriers obtained we see again a significantly higher barrier compared to the previous ring closures, being the **[4+4]** ring closure the most likely to happen.

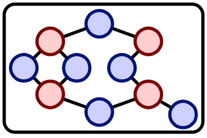
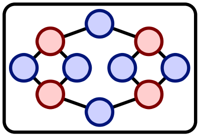
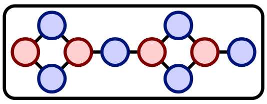
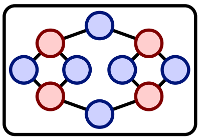
| | | | $\Delta G_d^{\ddagger}$ | $\Delta G_r^{\ddagger}$ | $\Delta\Delta G_d^{\ddagger}$ | $\Delta\Delta G_r^{\ddagger}$ |
|---|---|---|---|---|---|---|
|  | → |  | 26.0 | 27.8 | 1.2 | -1.2 |
|  | → |  | 24.5 | 27.3 | -0.3 | -1.6 |

*Table 12: **[4+6]b** Cage closing reactions.*

When we compare the backwards reaction, ring opening, with the other **[2+2]** ring openings in cages (first rows of Tables 11 and 12) we see similar values. We also find a similar situation when comparing the results with Table 8. And finally if we were to consider the **[4+4]** ring closure (row 2 of Table 12) as a **[2+2]** ring closure (it comes when you consider the full **[2+2]** ring as a "monomer") its behavior is not very different.

Going back to the results of Zhu et al. they have one very puzzling result. In the first hour they do not detect the **[4+6]**$^{12}$ cage for the CC3-R synthesis, which is known to give the **[4+6]a** (proved already by X-ray crystallography[42]). However, they do detect an **[4+6]**$^{12}$ intermediate in the first hour for the CC-pentane, where the crystal structure of a **[2+3]**$^{6}$ cage was reported by Jelfs and coworkers.[7a]

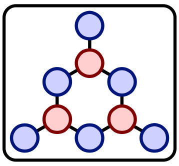
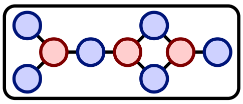
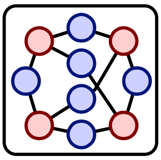
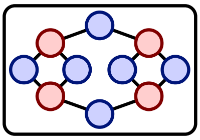
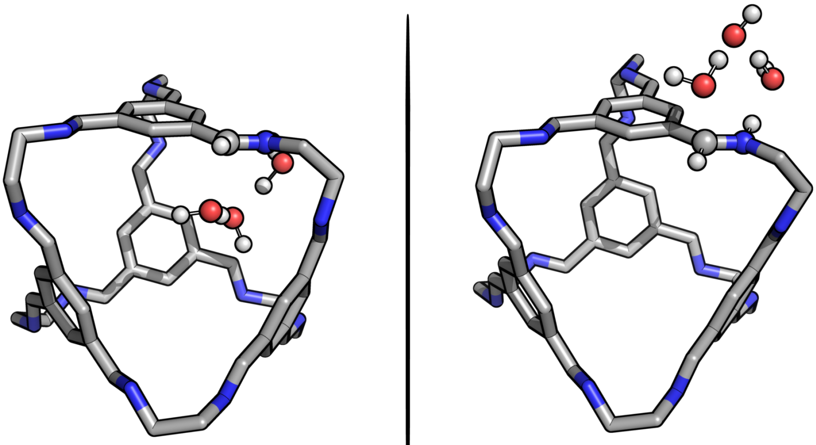
| | | |
|---|---|---|
|  | **[3+6]**$^{9}$ |  |
|  | **[4+6]**$^{12}$ |  |

*Table 13: Equivalent intermediates in mass spectrometry.*

A plausible explanation for this is a missasigment of the **[3+6]**$^{9}$ and the **[4+6]**$^{12}$ species as their work relied on mass spectrometry in which the compounds shown in Table 13 would be equivalent. This, coupled with the previous results that we found for dimerizations and the preference of the **[4+4]** ring closure for the **[4+6]b** cage would outline reasonable explanation to why these intermediates are detected for the CC-pentane but not for the CC3-R in the first hour.

### *4.4.4 **[4+6]a** Cage*



19.6 (31.5) 16.8 (28.7)

*Figure 41: Inside and outside conformations of the transition state. Direct (and reverse) barriers in kcal mol$^{-1}$.*

Finally, we move to the closure of the CC1 cage (**[4+6]a**). As we did with the **[3+3]** ring closure we consider the two possible orientations of the transition state (Figure 41). Here we find a stronger preference for the outward orientation compared with the **[3+3]** ring closures. In an attempt to mimic what we did in section 4.4.1 we calculated the TS of other **[4+6]a** cages coming from different monomers (Table 14).

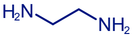
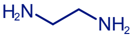
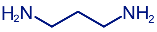
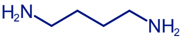
| **Amine** | **Conf** | **$\Delta G_d^‡$** | **$\Delta G_r^‡$** | **$\Delta G^{[4+6]^{12}}$(bond)** | **$\Delta G^{[2+3]^{6}}$(bond)** |
|---|---|---|---|---|---|
|  | inside | 19.6 | 31.5 | -64.2 (-5.3) | -21.6 (-3.6) |
|  | outside | 16.8 | 28.7 | -64.2 (-5.3) | -21.6 (-3.6) |
|  | outside | | 27.8 | -15.0 (-3.0) | -26.7 (-4.5) |
|  | outside | | 27.8 | -13.3 (-4.1) | -12.4 (-2.1) |

*Table 14: **[4+6]a** Cage closing reactions for different diamine monomers. Barriers in kcal mol$^{-1}$.*

We soon realized that the reverse barriers are similar which follows well the results from other cage closures. As no conformational study nor crystal structure was used for these “new” cages we decided to stop as the goal was already achieved. An important remark is that although the direct barrier is not the lowest we have found, it is significantly lower than any type of chain growth or ring closure reaction as well as the closure of the **[2+3]**$^{6}$ and the closures of the **[4+6]b**. Although the conformations calculated are not necessarily the lowest ones, they were constructed using the CC1 as template so the tendency observed in the energy per bond, which agrees with the odd-even effects observed by Jelfs and coworkers[7a], should be more or less consistent.

## 5. Summary

Within this chapter we have summarized the current knowledge on the mechanism of the self-assembly of imine cages. We have also highlighted the key experimental results and we have discussed how these findings cope with the results that we have obtained.

Regarding the chain growth reactions, we have checked that the Imine Condensation mechanism does not show major mechanistic changes compared with the mechanism studied in the previous chapter. We found that the reaction of an amine monomer behaves generally like the first condensation, and is clearly favored over the reaction with an aldehyde monomer. The Diaza-Cope has been discarded as a relevant side reaction and we have seen how the chain growth is more likely to happen between a large and a small fragment rather than through dimerization reactions. We also observed that the less free aldehydes a compound has, the more it behaves like an amine monomer.

Regarding the Ring closures and Cage closures we found that the **[2+2]** ring closure is favored over the chain growth reactions but the **[3+3]** ring closure is only slightly favored. We have seen how the **[4+6]a** and the incomplete cages closures are favored over the closure of the **[2+3]**$^{6}$ and **[4+6]b** cages. We also saw how the thermodynamics and the kinetics follow the same odd-even trends that are reported in the literature.

Finally we have been able to establish some approaches for estimating the energy of the TSs associated with the different types of reactions (summarized in Table 15). Overall we have seen that most reactions are better approached from the reverse barrier rather than the direct barrier with the exception of the amine monomer, the **[1+3]**$^{3}$ intermediate and the **[2+2]** ring closures.

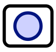
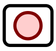
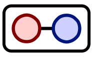
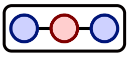
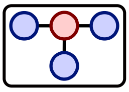
| **Reaction type** | **$\Delta\Delta G_d^{\ddagger}$ (stdev)** | **$\Delta\Delta G_r^{\ddagger}$(stdev)** |
|---|---|---|
| $[n+m]^{i}$ +  → $[n+(m+1)]^{i+1}$ | 0.0 (0.6) | |
| $[n+m]^{i}$ +  → $[(n+1)+m]^{i+1}$ | | -1.1 (0.8) |
| $[n+m]^{i}$ +  → $[(n+1)+(m+1)]^{i+2}$ | | -0.3 (1.0) |
| $[n+m]^{i}$ +  → $[(n+1)+(m+2)]^{i+3}$ | | -1.1 (0.3) |
| $[n+m]^{i}$ +  → $[(n+1)+(m+3)]^{i+4}$ | -0.6 (0.4) | |
| $[2+2]^{i} \rightarrow [2+2]^{i+1}$ | -3.1 (0.4) | |
| $[3+3]^{i} \rightarrow [3+3]^{i+1}$ | | 4.3 (–)* |
| $[4+4]^{i} \rightarrow 2\ [3+3]^{i+1}$ | | 0.2 (0.5) |
| [2+3] → ([2+2] closure)→ Cage [2+3] | | -1.8 (–)* |
| $[4+6]^{11}$ →([4+4] to 2[3+3])→ Cage [4+6]a | | -0.3 (–)* |
| $[4+6]^{11}$ → ([2+2] closure) → Cage [4+6]b | 1.2 (–)* | -1.2 (–)* |
| $[4+6]^{11}$ → ([4+4] closure) → Cage [4+6]b | -0.3 (–)* | -1.6 (–)* |
| $[3+4]^{7}$ → ([2+2] closure) → $[3+4]^{8}$ | | -1.0 (–)* |
| $[3+4]^{7}$ → ([3+3] closure) → $[3+4]^{8}$ | | -0.9 (–)* |

*Table 15: Reaction types and derived energy estimations. * Standard deviation calculation not applicable.*

# Chapter III. Automated Discovery of the Cage Self-Assembly Reaction Network

## 1. Motivation

In the previous chapters we studied in detail the imine condensation reaction and with that knowledge we simplified the study of some key reactions in the self-assembly process of the CC1 Imine Cage. Nonetheless, the study of such a dynamic process requires a wider view on the thermodynamic and kinetic picture of the cage generation process.

Looking back at the chain growth reactions in the previous chapter, enumerating all the pathways explored gave us a measure of the difficulty of such task. Manually enumerating all the intermediates and reactions is prone to human mistake and thus we need to resort to automation. This need becomes imperative when we also take into account the conformational complexity of these intermediates.

Thus, the aim of this chapter is to present to the reader an automated procedure for this purpose. This procedure aims to:

- Generate all the relevant intermediates.
- Explore their conformations.
- Calculate their free energies.
- Enumerate and store all the possible reactions interconnecting them.

Due to nature of the chapter we have decided to adapt its format. The reader will often find a full page figure on the right-hand page and the text discussing that figure on the left-hand page. We encourage to read the text while looking the figure. A final comment concerns the details of the energy calculations, which are given only as far as they are relevant to the procedure. The more technical description will not be supplied until Chapter IV (Computational Details section).

## 2. Introduction

The core concept of automation, to make machines perform tasks that at some point before were carried out by humans, is nowadays difficult to decouple from electronic devices such as computers.[1] Leveraging work to such devices so that we humans can invest our time in other tasks is a desirable aim in many aspects of our lives. Science, and in particular chemistry, is not an exception.

Automation runs deep in computational chemistry, as calculations themselves are automated procedures. Automation is also important in experimental chemistry, and we can see overall an increasing interest in the automation of tasks in many areas of chemistry in the last years.[2] We can highlight very interesting applications such as extraction of experimental procedures from literature,[3] assistance in the synthesis of compounds and their characterization, [4] or creation of robots to perform the experiments. [5] We can even go closer to the topic of this chapter and thesis to find cases where automation was applied to the construction of the geometries of porous organic cages[6].

As highlighted in the work of Tarzia and Jelfs,[6] the main focus in the field of the computational modeling of cages is mostly associated with the modeling of the cages themselves and not the intermediate species. In contrast, here we aim to develop an automated way of keeping track of all the reactions and compounds that can appear in the transition between the free monomers and the cage (Figure 42 shows an overview of the algorithm), as well as find the free energy landscape to ultimately perform kinetic simulations.

---

[1] Janssen, C. P.; Donker, S. F.; Brumby, D. P.; Kun, A. L. *Int. J. Hum. Comput. Stud.* **2019**, 131, 99-107.

[2] Sanderson, K. *Nature* **2019**, 568, 577-579.

[3] Mehr, S. H. M.; Craven, M; Leonov, A. I.; Keenan, G.; Cronin, L. *Science* **2020**, 370, 6512, 101-108

[4] (a) Trobe, M.; Burke, M. D. *Angew. Chem. Int. Ed.* **2018**, 57, 4192-4214. (b) Caramelli, D.; Granda, J. M.; Mehr, S. H. M.; Cambié, D.; Henson, A. B.; Cronin, L. *ACS Cent. Sci.* **2021**, 7, 11, 1821-1830.

[5] Burger, B.; Maffettone P. M.; Gusev, V. V.; Aitchison, C. M.; Bai, Y.; Wang, X.; Alston, B. M.; Li, B.; Clowes, R.; Rankin, N.; Harris, B.; Sprick, R. S.; Cooper, A. I. *Nature* **2020**, 583, 237-241.

[6] Tarzia, A.; Jelfs, K. E. *Chem. Commun.* **2022**, 58, 3717-3730.

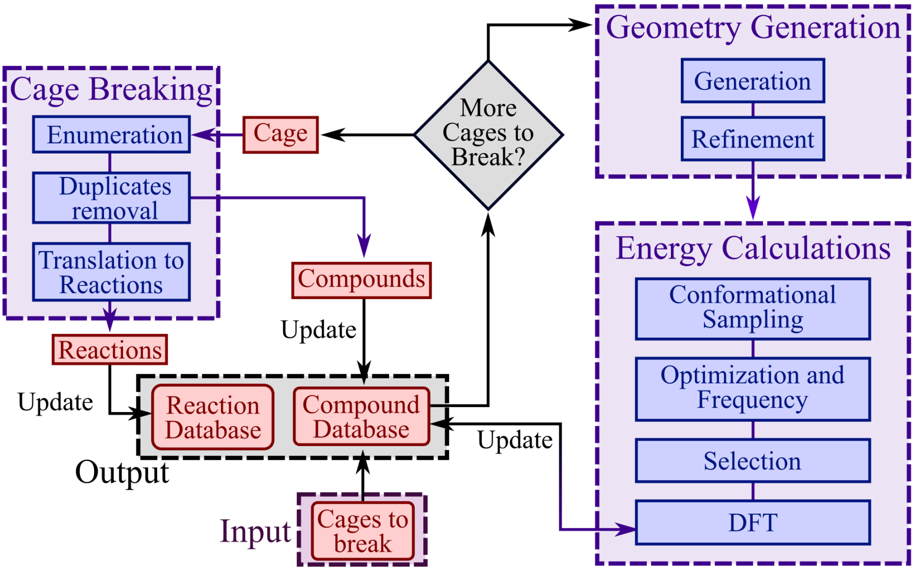

Geometry Generation
Generation
Refinement
Cage Breaking
Enumeration
Duplicates removal
Translation to Reactions
Cage
More Cages to Break?
Compounds
Update
Reactions
Update
Reaction Database
Compound Database
Update
Output
Input
Cages to break
Energy Calculations
Conformational Sampling
Optimization and Frequency
Selection
DFT

*Figure 42: General depiction of the algorithm for compound and reaction discovery.*

We decided to approach this starting from the cage and working our way backwards, “hydrolyzing each imine bond” systematically. We will refer to this as “Cage Breaking” for simplicity (Section 3. Cage Breaking). This approach has both a drawback and a big advantage.

The drawback is that breaking a specific cage is not going to generate all the possible monomer combinations. Only monomer combinations that are a motif in the cage will appear. For example, the CC1 cage does not contain any **[2+2]** macrocycle, thus intermediates such as the **[2+2]**$^4$ will not appear.

The advantage is that we can easily re-use previous results from the decomposition of other cages. If we want to consider all the intermediates derived from a different cage we can just break that new cage, find the new compounds that were not calculated and add the new reactions that were not previously included. This allows us to work in an incremental manner and can be used to partially overcome the mentioned drawback.

The following steps after we have identified the intermediate compounds are related with assigning a free energy value to each one of them. To do so we need to generate the geometry (Section 4. Geometry Generation) and then calculate the energy of each intermediate (Section 5. Energy Calculations).

Finally, it is important to indicate two important design choices. First, we have decided to restrict the intermediate discovery to only covalently bonded molecules. Second, we have decided to ignore the hemiaminal intermediates that would be present in each condensation reaction. These constraints were considered to reduce the complexity and quantity of the obtained compounds. Otherwise the amount of possible combinations becomes intractable with a dubious gain in chemical insight.

We give here some measure of the effect of such choices, we found a total of 85 different compounds from breaking 4 different cages and identified 400 reactions with the developed algorithm. To these 85 compounds we have to add the monomers and the water molecule, 88 compounds in total. Including the hemiaminals would lead to at least 400 new compounds, 1 per reaction. If we include the adducts the same reaction has to be calculated several times i.e. $A + B \rightarrow C$, $[A\cdots B] + B \rightarrow [C\cdots B]$, $[B\cdots A\cdots B] \rightarrow [C\cdots B]$ … which would dramatically increase the number of reactions and the number of intermediates.

# 3. Cage Breaking

## 3.1 Molecule Representations

It is important to remark here that we aim to a generic description of the cages, and thus there will be an independent labeling system for each of the starting cages, thus four different systems. The different ways in which we can represent those molecules becomes very relevant when digging into the details of the algorithm. A summary of these representations can be seen in Figure 43.

We start with the 3D representations of molecules, which is widely used in computational chemistry. The moieties coming from the aldehyde are in red and those coming from the amine are in blue. In the figure we can see two cage molecules and an intermediate obtainable from breaking either of them. Just below we have the Standard representation ($[\mathbf{n}+\mathbf{m}]^{i}$) used in previous chapters.

Next we go to the 2D or graph representations of the molecules. These represent each moiety by a node following the same color criteria (as the Standard representation) and an edge connecting the nodes corresponding to the imine bond. We can have alternative 2D representations for the same compound as we can see for the $[\mathbf{1}+\mathbf{1}]^{\mathbf{1}}$. These 2D representations depend on the cage and can be directly translated to the Binary representations. This representation comes as a result writing the existence (1) or absence (0) of an edge with respect to the 2D representation of the cage. The Decimal representation is the number in base 10 of the Binary representation.

The Text representation is a combination of the name arbitrarily assigned to the cage in which the compound was first identified and the Binary representation. Finally the Id is a unique number assigned upon discovery.

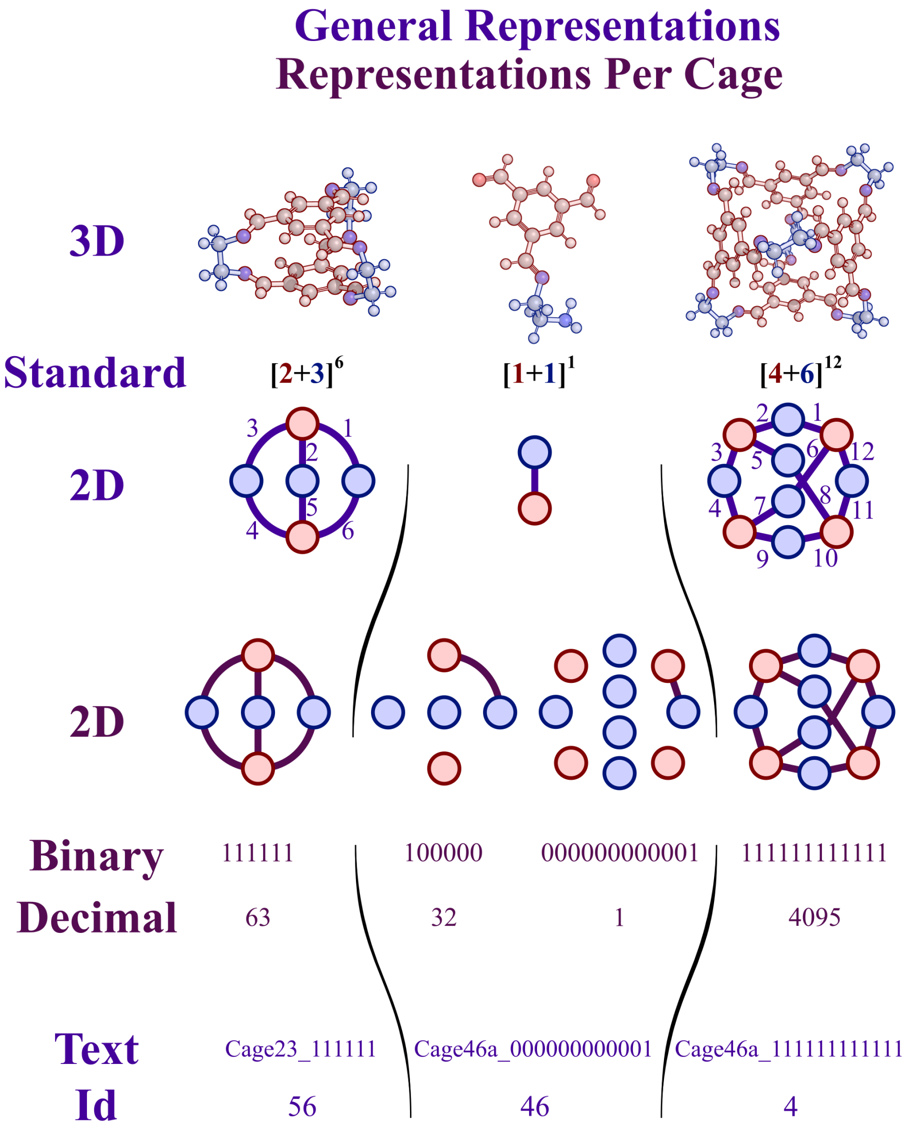

General Representations
Representations Per Cage
3D
Standard
[2+3]6
[1+1]1
[4+6]12
2D
3
1
2
5
4
6
2
1
3
5
6
12
4
7
8
11
9
10
2D
Binary
111111
100000
000000000001
111111111111
Decimal
63
32
1
4095
Text
Cage23_111111
Cage46a_000000000001
Cage46a_111111111111
Id
56
46
4

*Figure 43: Example representations of different molecules*

### 3.2 Compound and Edge Enumeration

In this step we start the Cage breaking process by systematically removing the edges of its 2D representation (equivalent to breaking the imine bonds). This process is illustrated in Figure 44 for a simplified cage with only 4 imine bonds and two monomers of each type.

If we see the binary representation of the cage we can see that generating all the possible intermediates (including adducts) corresponds to generating all the combinations of 4 items with 1s and 0s. The solution to this problem is straightforward when looking at decimal representations. Merely counting from 0 to 31 (in binary) allows us to generate all those combinations.

Now, we are only missing the information related to the reactions. A single edge removal corresponds to a an imine bond hydrolysis. When we remove a single edge, the effect on the binary representation is a switch of the 0 by a 1 in the corresponding position. Following the nomenclature of the previous chapter, the reaction 1111 → 0111 would be equivalent to **[2+2]**$^{4}$ → **[2+2]**$^{3}$.

We use a graph to store these relationships, where each of the nodes is a binary number and a directed edge (arrowed line in the figure) means that the parent node changes a 1 for a 0 in its representation. So, for a binary with two 1s, only nodes with binaries with three 1s can be its parents. We first sort the nodes by the number of 1s and then we proceed to compare them and add the edges to the graph.

The graph that we end up will always start from the binary of the cage (made of only 1s) and will always end with the binary of the free monomers (full of 0s). To differentiate this graph from the molecular representations we will refer to it as the “Cage Tree”.

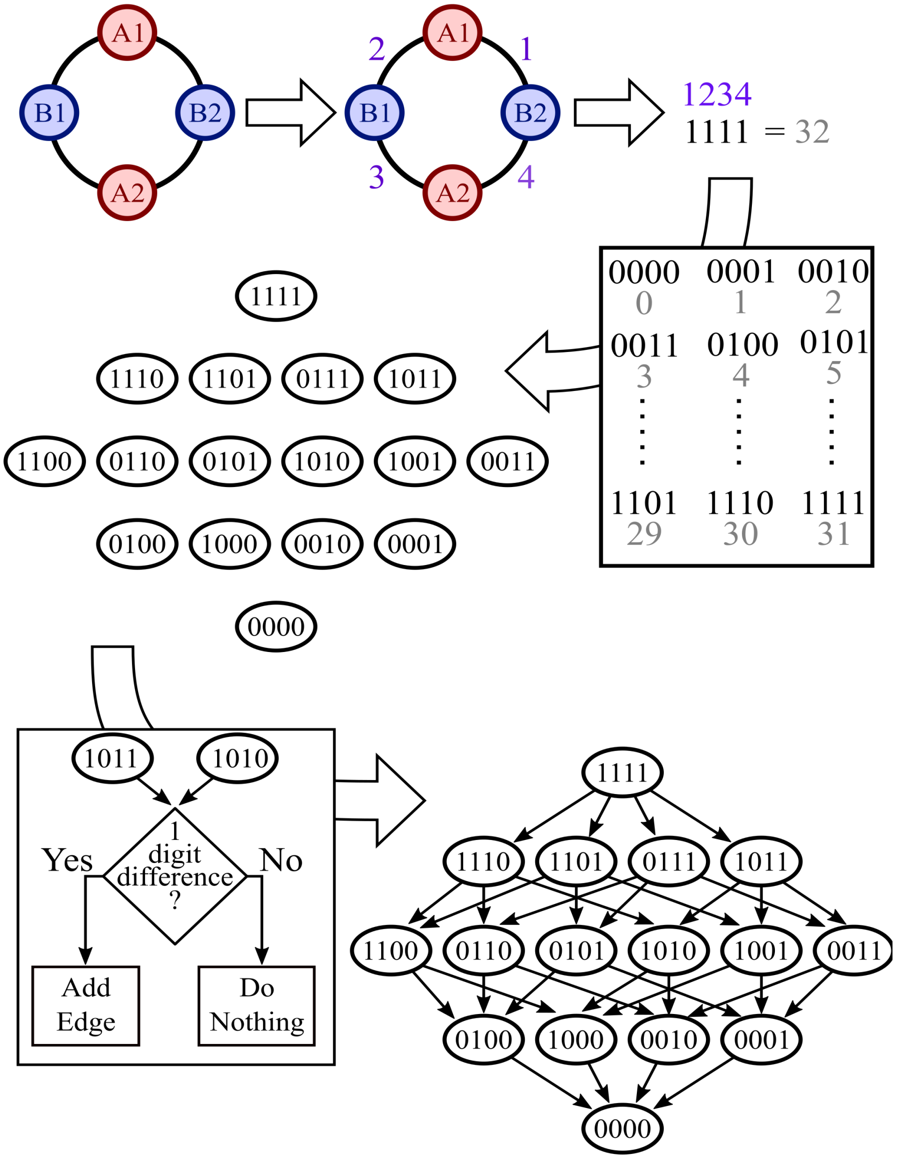

A1
B1
B2
A2
2
1
3
4
1234
1111 = 32
0000 0001 0010
0 1 2
0011 0100 0101
3 4 5
1101 1110 1111
29 30 31
1111
1110 1101 0111 1011
1100 0110 0101 1010 1001 0011
0100 1000 0010 0001
0000
1011 1010
1 digit difference ?
Yes
No
Add Edge
Do Nothing

*Figure 44: Intermediate and reaction enumeration procedure*

### 3.3 Duplicated Compounds Removal

A drawback of the binary representations is that several representations may actually correspond to the same compound (as can be seen in Figure 45). Thus, we need to simplify the Cage Tree by merging the nodes that correspond to the same compound, even in the same cage.

At this point we are forced to do the comparisons at the 2D representation level of the molecules although we can reduce the computational cost by only comparing between the graphs that correspond to binaries with the same number of 1s (also referred as “same generation”). We need to note that from a chemical perspective the two graphs corresponding to 1011 and 0111 in Figure 45 would correspond to the same compound. Intuitively we want to find which graphs we can overlap while matching the colors but not the labels. Such task is equivalent to finding if two labeled graphs are isomorphic in the context of graph theory, which is an already solved problem in that field.

After identifying which intermediates are unique per each generation we use one to represent all of them (we will refer to the process of going from a binary to its unique representative as “standardize a binary”) and merge the nodes of our Cage Tree. Later we will re-use that isomorphism relationship, so we save it. Next, we remove all the edges that are redundant in the Cage Tree.

At the end of the figure we can see the current Cage Tree with the 2D representations of the molecules obtained. In the last three steps we can see the same reaction encoded twice (In binary the sequence 0101 → 0001 → 0000). It is the same reaction because we are deliberately not considering the adducts.

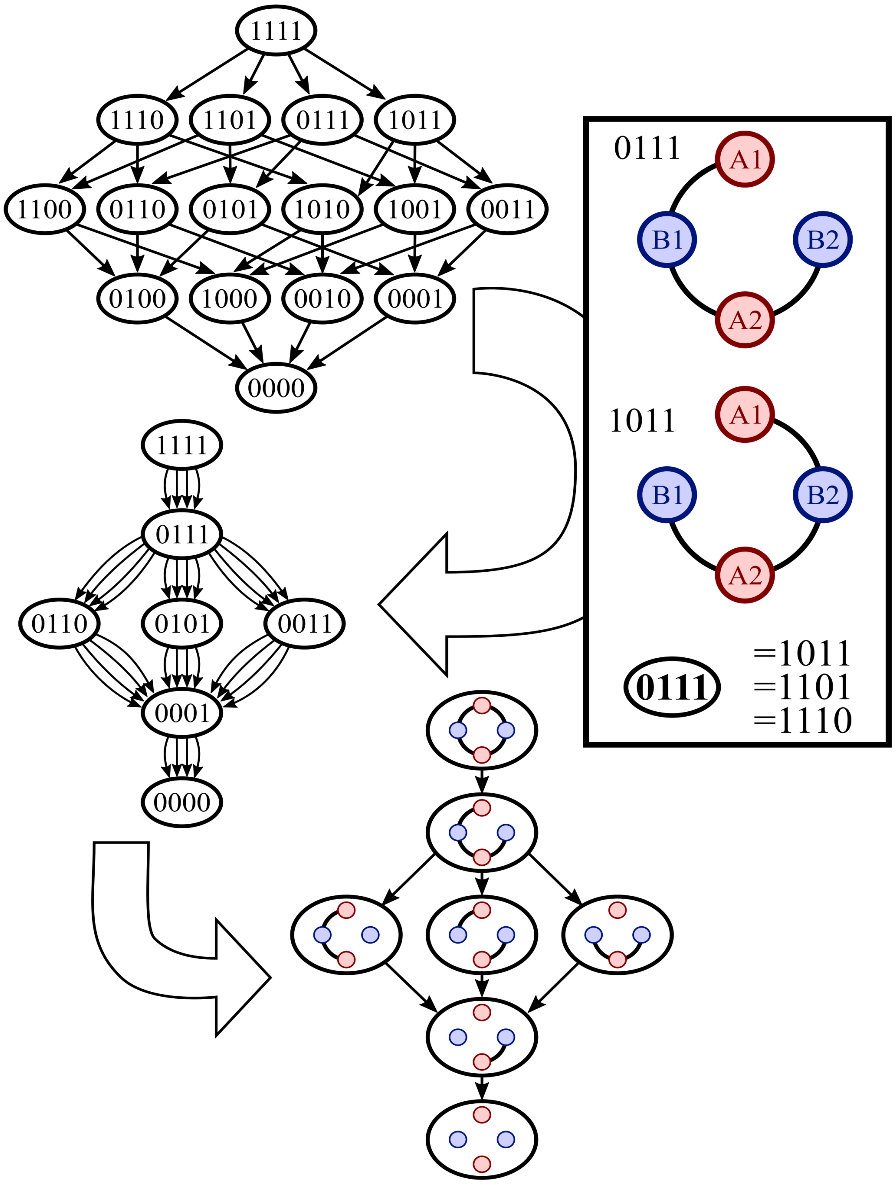

1111
1110
1101
0111
1011
1100
0110
0101
1010
1001
0011
0100
1000
0010
0001
0000
0111
A1
B1
B2
A2
1011
A1
B1
B2
A2
0111
=1011
=1101
=1110
1111
0111
0110
0101
0011
0001
0000

*Figure 45: Example of duplicates removal*

### 3.3.1 Adducts

At this point we will modify the example for the remaining figures to make the following discussion easier to follow. In Figure 46 we can see the same reaction encoded by three different sets of binaries using the binaries corresponding to a **[2+3]**$^{6}$ cage (depicted in the left side for reference).

In the two first cases we can see that we have adducts involved. Thus, we are in need of a way to identify the compounds in the adducts and see if those compounds are repeated at both sides of the reaction.

The solution that we came up with is depicted in Figure 47 which we termed as “binary fragmentation”. First, we identify all the disconnected subgraphs belonging to the graph representation of the molecule for a specific binary.

Next, we ignore any completely disconnected node (as these will not be explicitly detected in the previous step). Afterwards, per each subgraph, we identify which edges belong to it. Knowing those edges we can write the binary representation of each of those subgraphs.

Finally, to store all the compounds that we found while breaking the cage we iterate all the nodes on our Cage Tree. We break their binary representations and standardize the obtained binaries so that we can identify when they are repeated. For each unique binary remaining we create a text representation (Cage origin plus its binary representation within the cage). We discard its unconnected nodes and check if the graph was already stored in our compound database. If it was not stored we assign a new number (decimal representation) to it and store it. If it was already stored we assign the new text representation as an equivalent of the previously stored one.

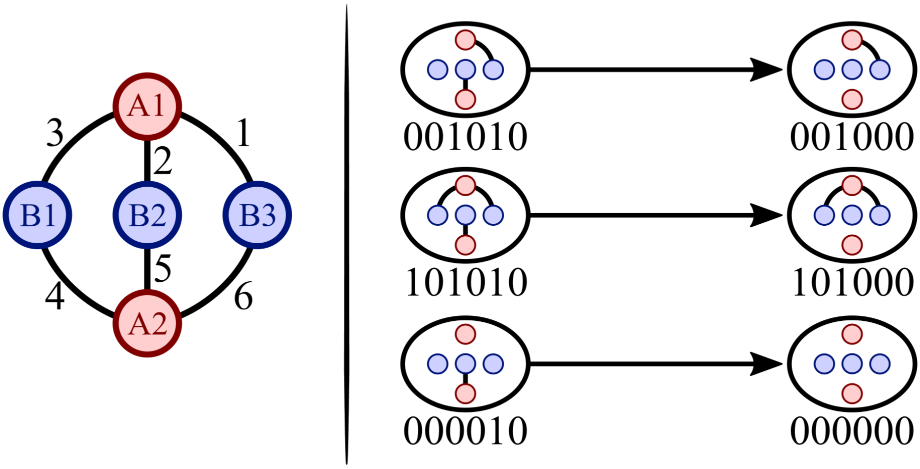

A1
B1
B2
B3
A2
1
2
3
4
5
6
001010
001000
101010
101000
000010
000000

*Figure 46: Examples of differing intermediates encoding the same reaction*

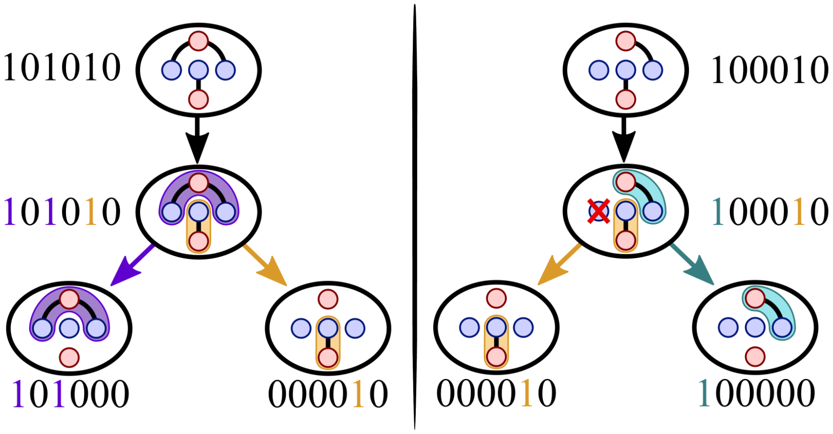

101010
101010
101000
000010
100010
100010
000010
100000

*Figure 47: Binary representation fragment detection*

### 3.4 Edge to Reaction Translation

Finally we need to translate the edges of the Cage Tree to reactions, and represent them in a way that we can easily detect duplicates. This will allow us an easy way of merging the new reactions that we find when breaking different cages. This translation process is illustrated in Figure 48.

First, we start using the previously stored information of isomorphic binaries. We find the two binary representations between the nodes that differ only in one digit (transforming 101010 → 010001 into 101010 → 001010).

Second, we fragment those binaries using the same procedure described before for the adducts. We then standardize each one of the fragments obtained. This allows to easily guess if a compound is repeated at both sides of the edge or not. In this event we proceed to remove it in both sides.

Afterwards we prune the graph that corresponds to each of the remaining binaries. Using the pruned graph representations we look in the compounds database for its decimal representation.

Now we have a simplified reaction in the direction of an imine hydrolysis. We invert the direction of the edge to store the reaction in the imine condensation direction. At this point we check the results to ensure that we only have one product.

Finally, if we have a differing number of nodes on both sides we correct it by adding the appropriate node (we have previously reserved the decimal representations 0 and 1 for each of the monomers) to the reactants. Before storing the reaction we sort the decimal representations of the reactants and the products (separately) which allows easy identification of repeated reactions.

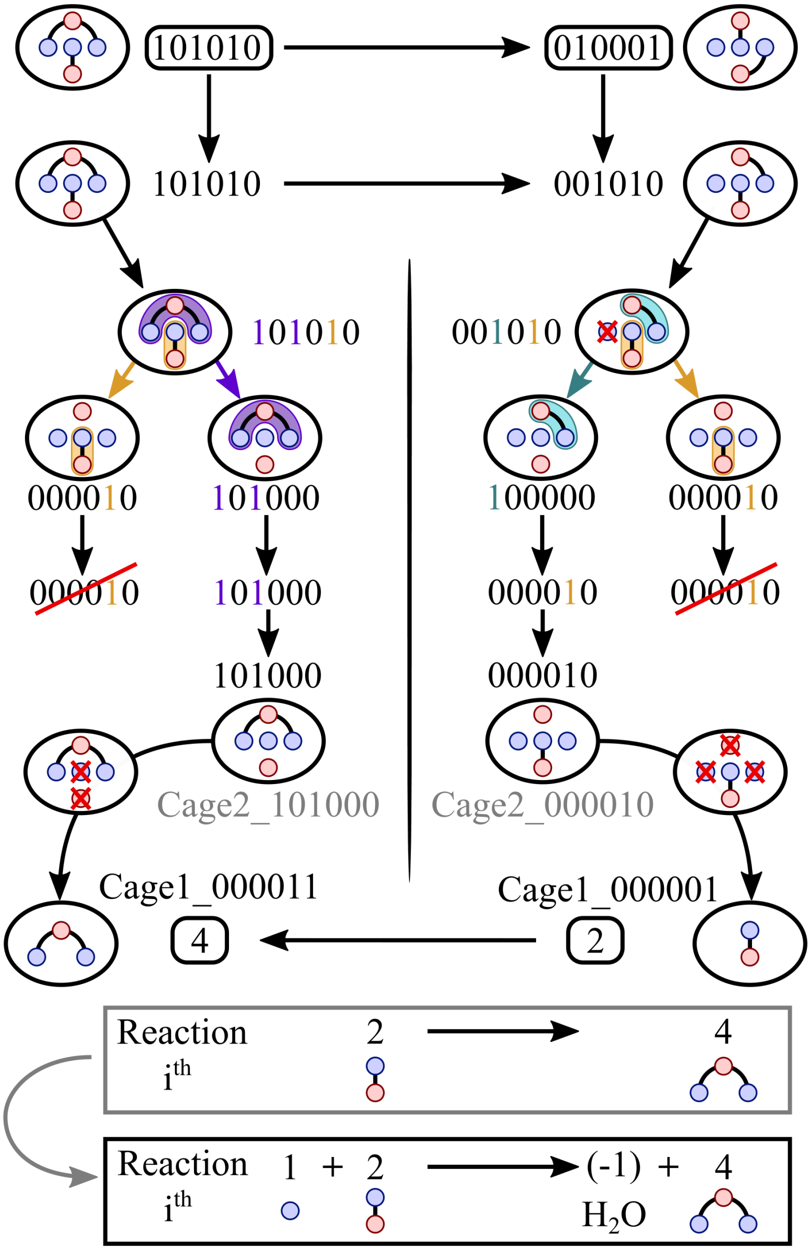

101010
010001
101010
001010
101010
001010
000010
101000
100000
000010
~~000010~~
101000
000010
~~000010~~
101000
000010
Cage2_101000
Cage2_000010
Cage1_000011
Cage1_000001
4
2
Reaction
ith
2
4
Reaction
ith
1 + 2
(-1) + 4
H2O

*Figure 48: Translation from edge break to reaction.*

## 4. Geometry Generation

The next issue to address is the generation of reasonable geometries for the intermediates. We decided that a suitable approach would be to first generate some guess geometries and then proceed to relax those geometries using molecular mechanics calculations.

### 4.1 Generation

We start by acquiring the geometrical information from the cages that will be broken. This process is illustrated in Figure 49. As we know which atoms come from each different monomer, we start by labeling them.

Then, we translate the coordinates of the cage so that its centroid is at the origin. This step simplifies the information and the next step: calculating a local axis per each fragment/monomer. We use the vector that goes from the centroid of the cage (the origin now) to the centroid of the fragment as the direction of the local y axis (green) of the fragment. As we know which atoms are involved in the connections between monomers we use the first two (arbitrarily chosen) as the direction of the z axis (red). Using the cross product of these two vectors we can obtain a third direction perpendicular to both that will serve as the local x axis (blue). After normalizing the vectors we end up with a local axis per each fragment.

Finally we use the vector between the centroids as the translation of the monomer and we calculate the rotation matrix of the local axes with respect to the global axis. In this way, we are representing the position and orientation of each monomer through a translation and rotation. We then store that information attached to each corresponding node.

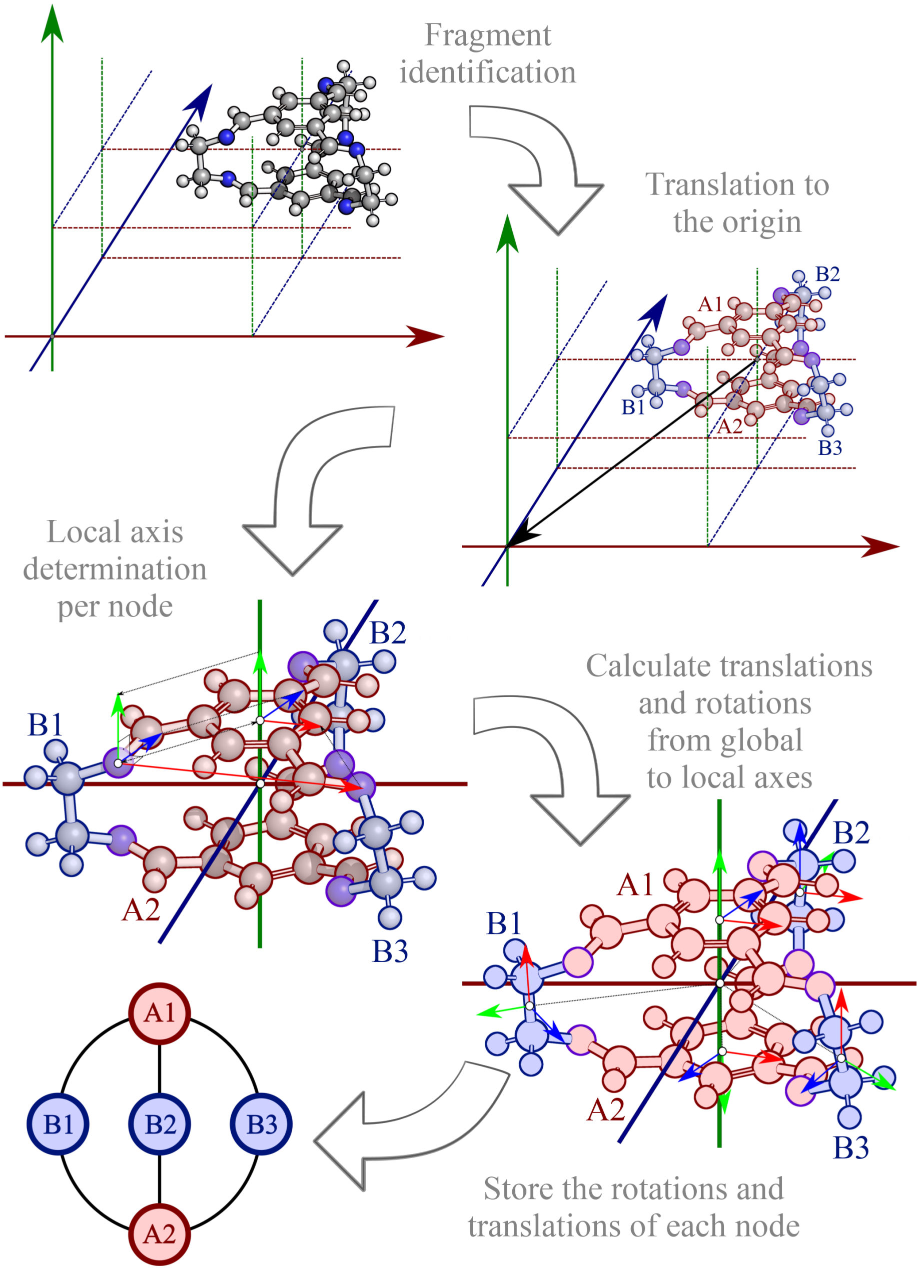

Fragment identification
Translation to the origin
A1
A2
B1
B2
B3
Local axis determination per node
B1
B2
A2
B3
Calculate translations and rotations from global to local axes
A1
B1
B2
A2
B3
A1
B1
B2
B3
A2
Store the rotations and translations of each node

*Figure 49: Scheme of the extraction of geometrical information from the cage*

The backwards process is depicted in Figure 50. For a given intermediate we start from its graph representation, that implicitly includes the cage from where it was found. We start by retrieving the translations and rotations of each one of the nodes. Then, to avoid atom clashes we consider 2 times the translation. After we know the rotations and translations, we use templated geometries for the different states of the fragments which are previously centered around the origin and then rotated to match the local axis criteria described in the previous page.

We refer by state as the different number of connections between fragments. For example, a diamine monomer by itself has 0 connections. After the first imine bond formation, it has one connection to an aldehyde monomer. Finally if a second imine is formed, it would have a second connection. Using this templating approach we can bypass any problem regarding atom removal as well as any issue arising from differences in how the fragment bonds within itself. As commented in Section 4, we remove unconnected nodes. In the event that we are interested into generating the possible adducts we could easily switch the approach by not removing those unconnected nodes and including an appropriate template.

Now, sequentially per each node, we proceed as follows. First, we start by guessing the type of template necessary (encoded in the type of node and the number of edges it has). Next, we apply the rotation to the atoms of the corresponding template. We then translate it to its new coordinates[7] and add the bonds of the template to the molecule that is being generated. Finally we proceed to add the bonds connecting the fragments and write the file with the generated geometry.

---

[7] The actual implementation of the code is adapted to merely substitute the N of the imines by oxygens and thus re-uses the geometries of the aldehyde fragments at the cage.

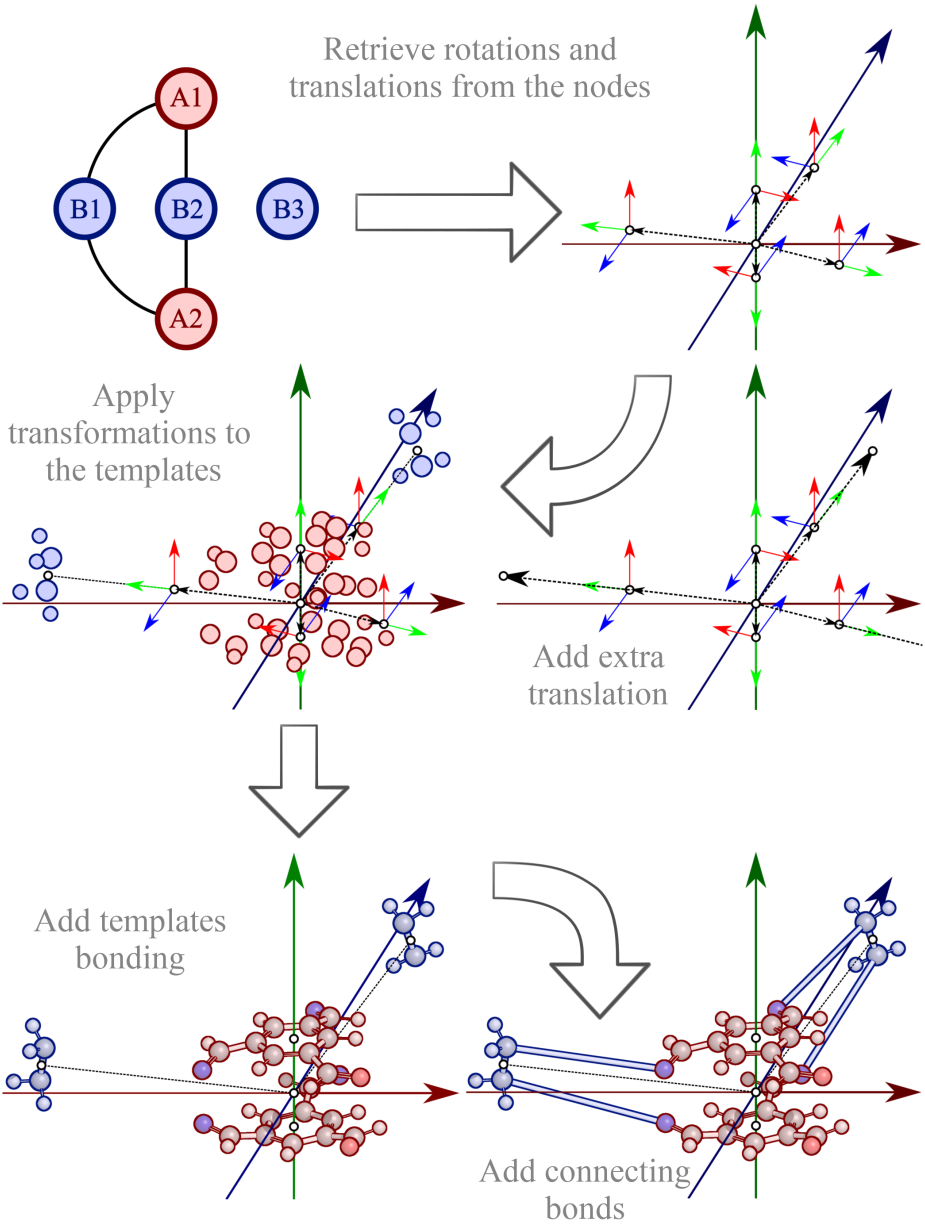

A1
B1
B2
B3
A2
Retrieve rotations and translations from the nodes
Apply transformations to the templates
Add extra translation
Add templates bonding
Add connecting bonds

*Figure 50: Scheme of the geometry generation of an intermediate*

### 4.2 MM Refinement

At this point we have some guess geometries for each intermediate but those geometries have been intentionally distorted. This distortion was enforced taking into account that we were going to follow it with a molecular mechanics relaxation. Such an exaggerated distortion in bond lengths will lead the optimization algorithm into a very specific direction since the bonds are considered explicitly in molecular mechanics calculations. We can see in Figure 51 that direction highlighted by arrows in purple.

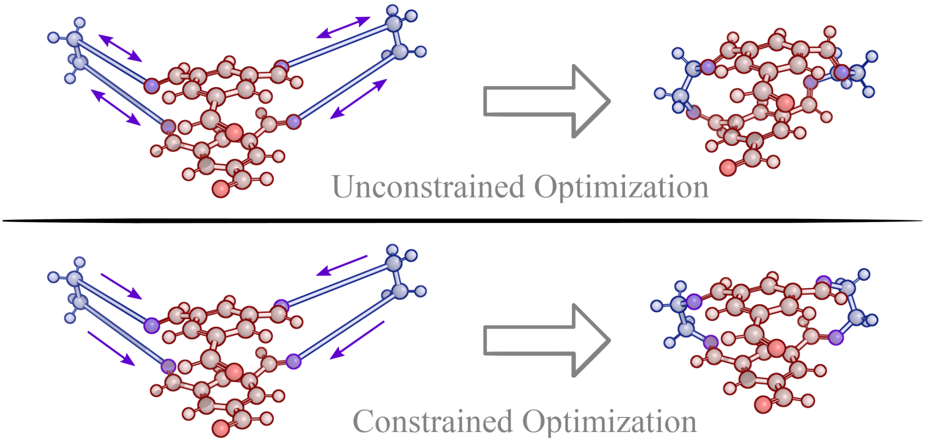


*Figure 51: Results of an unconstrained and a constrained optimization for the same input geometry. The cartesian coordinates of the atoms in purple are frozen for the constrained optimization.*

For the example highlighted in Figure 51 the same number of steps using the same force field have been carried out. What we can see is that a careful selection of the constraints either speeds up the convergence and/or biases the direction of the optimization. By freezing in place the atoms at one side of the distorted bonds we achieve control over the direction in which the molecule optimizes. Thus, we included the generation of these constraints of the final molecule in our procedure. To add such constraints through a simple yet easy to extend manner we decided to add an extra file to each monomer template. As the bonds of the template (Figure 50) are being added, the constraints are simultaneously written to a file with the appropriate indices of the atoms. Therefore we end with a file containing the geometry of each intermediate and a file with the constraints to refine it.

To actually run the optimization we used the Open Babel software[8] which includes a lightweight implementation of force field optimizations. Regarding the force field, we believe that UFF[9] could be acceptable if the proper constraints are set up, opening the door in the future to using cages with metals. However, we decided to use the MMFF94[10] due to the organic nature of the compounds in this work. For this specific project we only froze cartesian coordinates of the nitrogen atoms of the imines (which correspond to the connecting atoms of the aldehyde monomer templates).[11]

---

[8] (a) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminform* **2011**, 3, 33. (b) Open Babel, version 3.1.0, compiled from source code obtained at https://github.com/openbabel/openbabel.

[9] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A. III; Skiff, W. M.; *J. Am. Chem. Soc.* **1992**, 114, 10024-10039.

[10] (a) Halgren, T. A. *J. Comput. Chem.* **1996**, 17, 490-519 (b) Halgren, T. A. *J. Comput. Chem.* **1996**, 17 520-552. (c) Halgren, T. A. *J. Comput. Chem.* **1996**, 17, 553-586. (d) Halgren, T. A.; Nachbar, R. B. *J. Comput. Chem.* **1996**, 17, 587-615 (e) Halgren, T. A. *J. Comput. Chem.* **1996**, 17, 616-641.

[11] The same setup was used for the geometries depicted in Figure 51.

## 5. Energy Calculations

Even small sized molecules have pools of local minima corresponding to different conformations. If the pool is small enough, most optimization algorithms and its human users can efficiently navigate such PES, allowing to either arrive to the global minima or to a local minima not far from it.[12] When the molecules are larger their pool increases, and thus the probability of the optimization algorithm to stop at a local minima far from the global minima. When we have rings in those molecules some local minima become steeper. And although they get less steep as the size of the ring grows larger, the starting geometry becomes important. This holds even more true as we increase the number of cycles in the structure going from a macrocycle to a cage molecule.

To avoid the pitfalls associated to conformational complexity, we have divided the minima optimization in two steps. First, a Conformational Sampling step where we generate a large amount of different starting geometries. This will help us to reduce the influence of the initial geometry. This step was run with GFN2-xTB[13] using the xTB software.[14] Second, we proceed to the Optimization and Frequency calculations, also at the semi-empirical level with GFN2-xTB.

After we have our pool of reasonable global minima candidates, or at least near it, we would ideally refine all the energies with DFT single point calculations. However, the computational cost of DFT calculations is significantly higher than the semi-empirical calculations so we need to reduce the pool of candidates at this point.

---

[12] Jelfs, K. E.; Cooper, A. I. *Curr. Opin. Solid State Mater. Sci.* **2013**, 17 (1), 19 -30

[13] Bannwarth, C.; Ehlert, S.; Grimme, S. *J. Chem. Theory Comput.***2019**, 15, 3, 1652-1671.

[14] Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. *Comput. Mol. Sci.* **2020**, 11, e01493

Therefore, we proceed to a selection process, using a clustering technique to allow us to ensure some geometrical diversity among the selected conformations without direct human intervention. This was carried out using the k-means algorithm[15] as implemented in the python scikit-learn library.[16]

Finally we proceed with the DFT SP calculations to refine the energies and compile and sort all the obtained results. We used Gaussian 09[17] for the DFT calculations.

Although more accurate details of each of the steps will be provided in the following chapter a broad description is now presented along a scheme of the overall procedure (Figure 52).

---

[15] Lloyd, S. *IEEE Trans. Inf. Theory* **1982**, 28, 129-137.

[16] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. *J. Mach. Learn. Res.* **2011**, 12, 2825-2830.

[17] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, revision D.01; Gaussian, Inc.: Wallingford, CT, **2013**.

### 5.1 Conformational Sampling

The conformational sampling algorithm chosen was based on the algorithm used by CREST.[18] We followed their scheme for running metadynamics in xtb using the same parameters than the original publication (see Table 16 at Chapter IV). Thus, we run 14 metadynamic trajectories per each intermediate were we sampled 3000 geometries at intervals of 10ps at a 500K temperature.

### 5.2 Optimizations and Frequency Calculations

After visualizing the trajectories we decided that selecting one after 25 steps per metadynamic plus the last geometry (121 per metadynamic) was representative enough and reduced significantly the number of conformers to calculate (from 42 000 to 1694 per compound). With this pool of conformers we proceeded to minimize each geometry and calculate its frequencies using xtb 6.2.3. We used GFN2-xTB parametrization [19] for the semi-empirical tight binding with implicit solvation GBSA for dichloromethane for optimization (with "tight" convergence thresholds) and frequencies.

### 5.3 Clustering

Next we proceeded to select 10 conformers out of the 1694 conformers optimized. We employed a k-means algorithm,[15] a type of unsupervised machine learning algorithm used for clustering purposes, to enforce a partitioning of the conformational space into 10 groups. We used as input the centered Cartesian coordinates and the potential energies (as potential energies are less sensitive to small changes in geometry). As the aim of the clustering algorithm was to enforce a partition of the conformational space we chose per each cluster the conformer lowest in potential energy instead of the closest to the center of the cluster.

---

[18] Pracht, P.; Bohle, F.; Grimme, S. *Phys. Chem. Chem. Phys.* **2020**, 22, 7169-7192.

[19] Bannwarth, C.; Ehlert, S.; Grimme, S. *J. Chem. Theory Comput.* **2019**, 15, 3, 1652–1671.

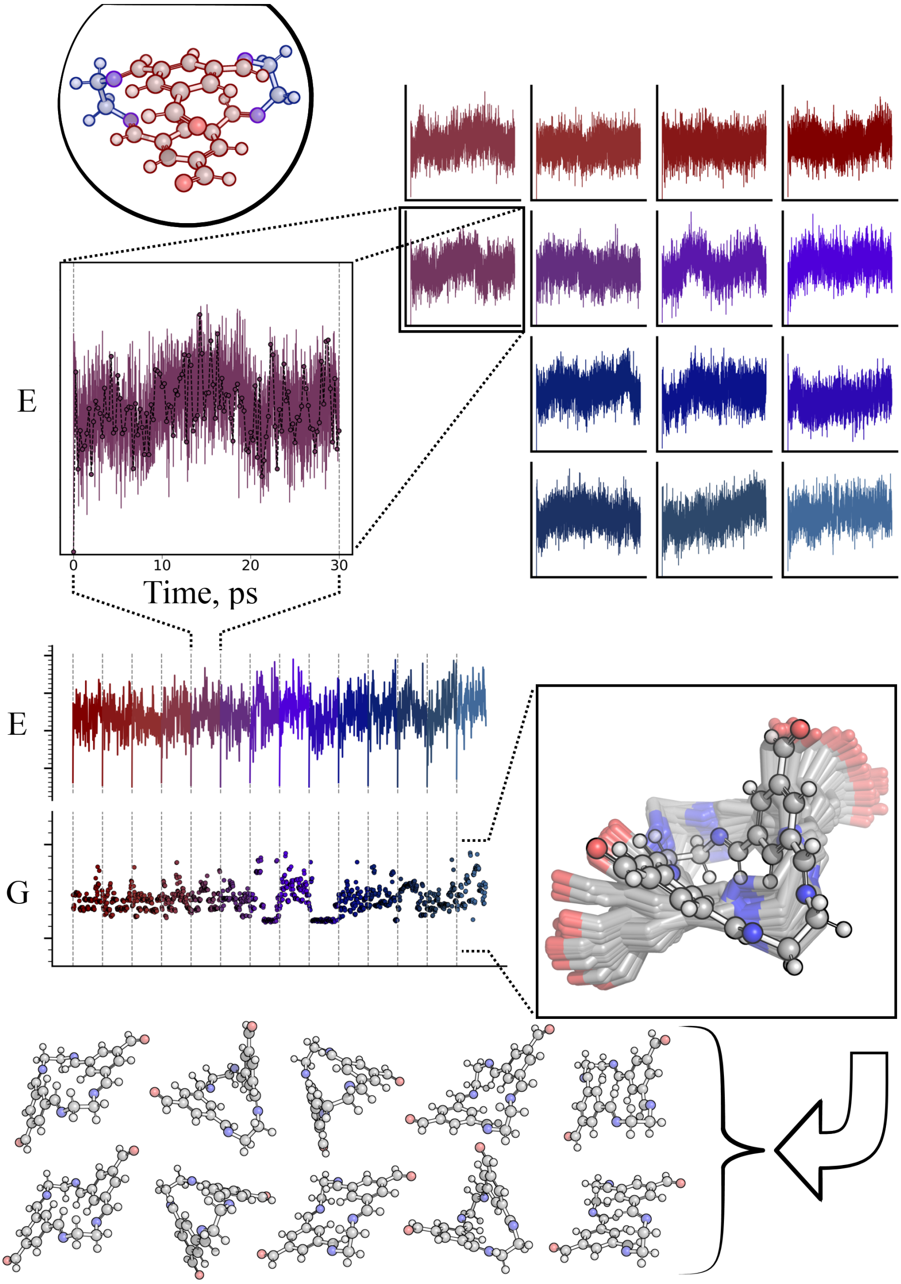

E
0
10
20
30
Time, ps
E
G

*Figure 52: Overview of the steps with semi-empirical calculations.*

### 5.4 DFT Single Point Calculations

Finally, for each of the selected 10 conformers per intermediate we proceeded to run single point calculations at the DFT level. These calculations were set up and their information was collected using the pyssian software developed during the thesis.[20]

We used B3LYP-D3[21] with the 6-311+g(d,p) basis set [22] for all atoms and with SMD[23] implicit solvation for dichloromethane.

## 6. Summary

We have been able to design a procedure for the labeling and selection of possible cages participating in the self-assembly of the different large cages. The procedure relies in three main steps, namely: Cage breaking (Enumeration), Geometry Generation and Energy calculation.

Thanks to the lower dimension representations and their interplay, we have been able to make the procedure as general as possible. This would facilitate adopting a similar strategy to solve other types of chemical problems.

Finally, we have been able to show how this procedure could be tweaked to perform better for our specific problem. Examples of this are the selection of the atoms that belong to each moiety of the cage, the selection of the constraints for the geometry refinement and the distortion induced in the initial geometry generated.

---

[20] Pérez-Soto, R.; Besora, M.; Maseras, F.; **2021**. maserasgroup-repo/pyssian: 1.0.2 (v1.0.2). Zenodo. https://doi.org/10.5281/zenodo.5055860

[21] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, 98, 11623−11627. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, 37, 785−789. (c) Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648−5652. (d) Grimme, S. *J. Comput. Chem.* **2006**, 27, 1787−1799.

[22] Krishnan, R.; Binkley, J.S.; Seeger, R.; Pople, J.A. *J. Chem. Phys.* **1980**, 72, 650-654.

[23] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, 113, 6378-6396.

# Chapter IV. Automated Study of the CC1 Cage Synthesis

## 1. Motivation

In this chapter, we aim to computationally model the self-assembly of the CC1 imine cage in order to understand this process.

For such task, we use the algorithm detailed in Chapter III to obtain a thermodynamic picture and a set of reactions that we then transform into kinetic models. Through the use of these models we will see the effects of the reactor setup (Models 1 and 2) on the concentration profiles and we will try to extract some mechanistic information from them.

Next, we will make our models more realistic by including the experimental temperatures (Model 3) and we will be able to observe how the change in temperature affects the different reactor setups.

In order to improve the model and to understand how the magnitude of the relative energy per imine bond affects the mechanisms we will set up Model 4. Next, we will proceed with Model 5 in an attempt to understand the effect of changing the relative energy differences between the computed species. To set up the model we will use the semi-empirical energies obtained during the application of the algorithm presented in Chapter III.

Finally, as we will aim to obtain an improved qualitative model (Model 6) based on DFT energies with the knowledge gained of the previous models.

## 2. Computational Methods

### 2.1 Thermodynamics

*MM Constrained optimization*

| | |
|---|---|
| **Software** | Open Babel 3.1.0[1] |
| **Calculation type** | Constrained Optimization |
| **Force Field** | MMFF94[2] |
| **Optimization algorithm** | Conjugated Gradients |
| **Optimization goal** | Number of steps. 10 000 steps. |
| **Constraints** | All bonds except the N-C bond connecting the NCCN moieties and the cartesian positions of the N atoms belonging to imine functional groups. |

---

[1] (a) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminform* **2011**, 3, 33. (b) Open Babel, version 3.1.0, compiled from source code obtained at https://github.com/openbabel/openbabel.

[2] (a) Halgren, T. A. *J. Comput. Chem.* **1996**, 17, 490-519 (b) Halgren, T. A. *J. Comput. Chem.* **1996**, 17 520-552. (c) Halgren, T. A. *J. Comput. Chem.* **1996**, 17, 553-586. (d) Halgren, T. A.; Nachbar, R. B. *J. Comput. Chem.* **1996**, 17, 587-615 (e) Halgren, T. A. *J. Comput. Chem.* **1996**, 17, 616-641.

*XTB Metadynamics*

| | |
|---|---|
| **Software** | XTB 6.2.3[3] |
| **Calculation type** | metadynamic simulations |
| **Method** | GFN2-xTB[4] |
| **Simulation time** | 30 ps |
| **Simulation step** | 1 fs |
| **Dump/report time** | 10 fs |
| **Ensemble** | NVT *nvt=true* |
| **Mass of the hydrogens** | 4 *hmass=4* |
| **xTB accuracy** | 2.0 *sccacc=2.0* |
| **Simulation temperature** | 500 K |
| **Implicit Solvation** | No |
| **alpha and kpush values** | See Table 16. |
| **Others** | velocities not reported. *velo=false*<br>Shake algorithm enabled. *shake=2*<br>maximal rmsd structures, 500. *save=500* |

---

[3] Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S.; *Comput. Mol. Sci.* **2020**, 11, e01493

[4] Bannwarth, C.; Ehlert, S.; Grimme, S. *J. Chem. Theory Comput.***2019**, 15, 3, 1652-1671.

| Metadynamic run id | kpush | alpha |
|---|---|---|
| 1 | 0.00300 | 1.300 |
| 2 | 0.00150 | 1.300 |
| 3 | 0.00075 | 1.300 |
| 4 | 0.00300 | 0.780 |
| 5 | 0.00150 | 0.780 |
| 6 | 0.00075 | 0.780 |
| 7 | 0.00300 | 0.468 |
| 8 | 0.00150 | 0.468 |
| 9 | 0.00075 | 0.468 |
| 10 | 0.00300 | 0.281 |
| 11 | 0.00150 | 0.281 |
| 12 | 0.00075 | 0.281 |
| 13 | 0.00100 | 0.100 |
| 14 | 0.00500 | 0.800 |

*Table 16: Parameters for the GFN2-xTB metadynamic calculations*

The values of the table were extracted from the original CREST algorithm.[5]

---

[5] Pracht, P.; Bohle, F.; Grimme, S. *Phys. Chem. Chem. Phys.***2020**, 22, 7169-7192.

*XTB Optimization and Frequency*

| | |
|---|---|
| **Software** | XTB 6.2.3[3] |
| **Calculation type** | Optimization and Frequencies |
| **Method** | GFN2-xTB[4] |
| **Criteria** | Tight |
| **Reference state** | Default. (25 ºC, 1 bar) |
| **Symmetry** | Disabled. (Maximum number of atoms for symmetry detection set to 0) |
| **Implicit Solvation** | Yes. GBSA(dichloromethane) |
| **Free energy corrections** | RRHO rotor threshold 50 $cm^{-1}$<br>Imaginary frequency threshold 20i $cm^{-1}$<br>Reference state correction applied for the kinetic models. |

*DFT Single Point Calculations*

| | |
|---|---|
| **Software** | Gaussian 09 revision D.01[6] |
| **Calculation type** | Single Point Calculations |
| **Functional** | B3LYP[7] |
| **Basis Set** | 6-311+g(d,p) for all atoms.[8] |
| **Empirical Dispersion** | GD3.[9] *empiricaldispersion=gd3* |
| **Symmetry** | disabled. *nosymm* |
| **Solvation** | Implicit. SMD for dichloromethane[10].<br>*scrf(smd,dichloromethane)* |
| **Integration grid** | 99 radial shells and 590 angular points per shell.[11] *int=(Grid=ultrafine)* |

---

[6] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, revision D.01*; Gaussian, Inc. Wallingford, CT, **2013**.

[7] a) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, 98, 11623−11627. b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, 37, 785−789. c) Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648−5652.

[8] Dunning, T. H. *J. Chem. Phys.* **1989**, 90, 1007.

[9] Grimme, S. *J. Comput. Chem.* **2006**, 27, 1787−1799.

[10] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, 113, 6378−6396.

[11] Lebedev, V. I. *Zh. Vychisl. Mat. Mat. Fiz.* **1976**, 16, 293-306.

## 2.2 Kinetic models

*Initial regime of the batch synthesis.*

| | |
|---|---|
| **Boundaries** | Isothermal. (Constant temperature)<br>Semi-continuous ideal mixture reactor model. with constant reactant inflow. |
| **Temperature** | 298.15K for Model 1 and 273.15K for the other models. |
| **Kinetic constants** | Calculated through Eyring Equation.[12] |
| **Tunneling effect** | Not considered. |
| **Solver type** | Numerical |
| **Solver** | LSODA[13] (from ODEPACK, as used by scipy[14]) |
| **Tolerances** | Relative tolerance (rtol): $10^{-6}$<br>Absolute tolerance (atol): $10^{-12}$ |
| **Other parameters** | scipy defaults. |
| **Report time step** | 1min (largest step for solution 1s) |
| **Simulation time** | 48 h |
| **Initial volume** | 0.85 L |
| **Inflow** | 0.3 mL/min<br>TFB: 0.0201M<br>water: 0.0100M<br>other species: 0.0000M |
| **Initial concentrations** | EtDA: 0.0408M<br>water: 0.0100M<br>other species: 0.0000M |

---

[12] a) Eyring, H. *J. Chem. Phys.* **1935**, 3, 107. b) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, 31, 875-894

[13] Hindmarsh, A. *IMACS Transactions on Scientific Computation* **1983**, 1, 55-64.

[14] a) Oliphant, T. E. *Comp. Sci. Eng.* **2007**, 9, 10-20. b) Millman, K. J.; Aivazis, M. *Comp. Sci. Eng.* **2011**, 13, 9-12.

*Final regime of the batch synthesis.*

| | |
|---|---|
| **Boundaries** | Isothermal. (Constant temperature)<br>Batch reactor model (Closed system)<br>Liquid Phase. (Constant volume) |
| **Temperature** | 298.15K |
| **Kinetic constants** | Calculated through Eyring Equation.[12] |
| **Tunneling effect** | Not considered. |
| **Solver type** | Numerical |
| **Solver** | LSODA[13] (from ODEPACK, as used by scipy[14]) |
| **Tolerances** | Relative tolerance (rtol): $10^{-6}$<br>Absolute tolerance (atol): $10^{-12}$ |
| **Other parameters** | scipy defaults. |
| **Report time step** | 1min (largest step for solution 1s) |
| **Simulation time** | 24 h |
| **Initial concentrations** | Output concentrations of the Initial regime. |

*Flow synthesis*

| | |
|---|---|
| **Boundaries** | Isothermal. (Constant temperature)<br>Plug-flow reactor. (Open system)<br>Stationary state per diferential volume assumed. |
| **Temperature** | 298.15K for Model 2 and 373.15K for the other models. |
| **Kinetic constants** | Calculated through Eyring Equation.[12] |
| **Tunneling effect** | Not considered. |
| **Solver type** | Numerical |
| **Solver** | LSODA[13] (from ODEPACK, as used by scipy[14]) |
| **Tolerances** | Relative tolerance (rtol): $10^{-6}$<br>Absolute tolerance (atol): $10^{-12}$ |
| **Other parameters** | scipy defaults. |
| **Report volume step** | 0.01 mL (largest step for solution 0.01mL) |
| **Total volume** | 10 ml |
| **Inflow** | 1 mL/min<br>TFB: 0.03154 M<br>EtDA: 0.05146 M<br>water: 0.0100M<br>other species: 0.0000M |

*Reactions*

The energy of the transition states for each reaction was modeled by adding a value X (specified in the text for each model) to the free energy of the reactants (if the reaction is exergonic in the direct reaction) or the products (if the reaction is exergonic in the reverse direction). All reactions were modeled as two bimolecular reactions of order 1 on each reactant (one reaction for the direct direction and another one for the reverse direction).

In the following pages a list of all the reactions can be found. A stands for 1,3,5 triformylbenzene (TFB), B for ethylene-1,2-diamine (EtDA), W for water and the numbers to the global IDs of the compounds. Representations of the different compounds together with global IDs and Text representations can be found in Annex I.

$42 + 56 \rightleftarrows 81 + W$
$15 \rightleftarrows 81 + W$
$82 \rightleftarrows 81 + W$
$82 \rightleftarrows 75 + W$
$B + 43 \rightleftarrows 75 + W$
$32 \rightleftarrows 75 + W$
$B + 55 \rightleftarrows 70 + W$
$42 \rightleftarrows 70 + W$
$74 \rightleftarrows 72 + W$
$63 \rightleftarrows 72 + W$
$55 + 56 \rightleftarrows 72 + W$
$35 \rightleftarrows 6 + W$
$B + 11 \rightleftarrows 6 + W$
$44 \rightleftarrows 6 + W$
$52 \rightleftarrows 6 + W$
$3 \rightleftarrows 6 + W$
$71 \rightleftarrows 60 + W$
$84 \rightleftarrows 60 + W$
$B + 62 \rightleftarrows 60 + W$
$79 \rightleftarrows 60 + W$
$67 \rightleftarrows 60 + W$
$83 \rightleftarrows 60 + W$
$3 \rightleftarrows 60 + W$
$81 \rightleftarrows 60 + W$
$75 \rightleftarrows 60 + W$
$71 \rightleftarrows 69 + W$
$67 \rightleftarrows 69 + W$
$56 + 70 \rightleftarrows 69 + W$
$B + 72 \rightleftarrows 69 + W$
$83 \rightleftarrows 69 + W$
$55 + 55 \rightleftarrows 69 + W$
$81 \rightleftarrows 69 + W$
$6 \rightleftarrows 4 + W$
$60 \rightleftarrows 59 + W$
$69 \rightleftarrows 59 + W$

A + 26 ⇄ 29 + W
9 + 46 ⇄ 29 + W
8 + 33 ⇄ 29 + W
A + 50 ⇄ 37 + W
9 + 46 ⇄ 37 + W
55 ⇄ 57 + W
A + 19 ⇄ 51 + W
A + 45 ⇄ 51 + W
46 + 50 ⇄ 51 + W
8 + 31 ⇄ 51 + W
26 + 46 ⇄ 51 + W
B + 29 ⇄ 51 + W
33 + 33 ⇄ 51 + W
A + 45 ⇄ 36 + W
B + 37 ⇄ 36 + W
46 + 50 ⇄ 36 + W
26 + 46 ⇄ 36 + W
9 + 13 ⇄ 36 + W
A + 33 ⇄ 9 + W
8 + 46 ⇄ 9 + W
A + 28 ⇄ 22 + W
B + 29 ⇄ 22 + W
26 + 46 ⇄ 22 + W
9 + 13 ⇄ 22 + W
33 + 33 ⇄ 22 + W
8 + 54 ⇄ 22 + W
B + 18 ⇄ 5 + W
34 + 46 ⇄ 5 + W
13 + 42 ⇄ 5 + W
B + 47 ⇄ 5 + W
41 + 54 ⇄ 5 + W

B + 39 ⇄ 5 + W
A + 18 ⇄ 10 + W
45 + 46 ⇄ 10 + W
B + 51 ⇄ 10 + W
31 + 33 ⇄ 10 + W
A + 18 ⇄ 53 + W
B + 36 ⇄ 53 + W
13 + 50 ⇄ 53 + W
19 + 46 ⇄ 53 + W
9 + 41 ⇄ 53 + W
A + 39 ⇄ 16 + W
45 + 46 ⇄ 16 + W
B + 36 ⇄ 16 + W
28 + 46 ⇄ 16 + W
13 + 26 ⇄ 16 + W
A + 24 ⇄ 40 + W
51 ⇄ 40 + W
36 ⇄ 40 + W
2 + 46 ⇄ 40 + W
22 ⇄ 40 + W
A + 47 ⇄ 25 + W
31 + 33 ⇄ 25 + W
13 + 50 ⇄ 25 + W
45 + 46 ⇄ 25 + W
B + 51 ⇄ 25 + W
28 + 46 ⇄ 25 + W
33 + 54 ⇄ 25 + W
B + 22 ⇄ 25 + W
A + 39 ⇄ 17 + W
B + 51 ⇄ 17 + W
19 + 46 ⇄ 17 + W

13 + 26 ⇄ 17 + W
B + 22 ⇄ 17 + W
33 + 54 ⇄ 17 + W
45 + 46 ⇄ 17 + W
8 + 42 ⇄ 17 + W
A + 47 ⇄ 38 + W
31 + 33 ⇄ 38 + W
9 + 41 ⇄ 38 + W
13 + 26 ⇄ 38 + W
B + 22 ⇄ 38 + W
19 + 46 ⇄ 38 + W
8 + 42 ⇄ 38 + W
A + 34 ⇄ 18 + W
B + 45 ⇄ 18 + W
13 + 31 ⇄ 18 + W
42 + 46 ⇄ 18 + W
B + 19 ⇄ 18 + W
33 + 41 ⇄ 18 + W
A + 42 ⇄ 45 + W
B + 50 ⇄ 45 + W
31 + 46 ⇄ 45 + W
46 + 54 ⇄ 45 + W
B + 26 ⇄ 45 + W
13 + 33 ⇄ 45 + W
B + 45 ⇄ 39 + W
42 + 46 ⇄ 39 + W
13 + 54 ⇄ 39 + W
B + 28 ⇄ 39 + W
B + 19 ⇄ 47 + W
42 + 46 ⇄ 47 + W
B + 28 ⇄ 47 + W

13 + 54 ⇄ 47 + W
13 + 31 ⇄ 47 + W
33 + 41 ⇄ 47 + W
A + 42 ⇄ 19 + W
B + 26 ⇄ 19 + W
13 + 33 ⇄ 19 + W
31 + 46 ⇄ 19 + W
8 + 41 ⇄ 19 + W
A + 54 ⇄ 26 + W
B + 9 ⇄ 26 + W
33 + 46 ⇄ 26 + W
8 + 13 ⇄ 26 + W
A + 46 ⇄ 8 + W
B + 42 ⇄ 34 + W
13 + 41 ⇄ 34 + W
28 + 46 ⇄ 20 + W
B + 22 ⇄ 20 + W
13 + 26 ⇄ 20 + W
33 + 54 ⇄ 20 + W
22 ⇄ 21 + W
46 + 54 ⇄ 28 + W
B + 26 ⇄ 28 + W
13 + 33 ⇄ 28 + W
A + 31 ⇄ 50 + W
33 + 46 ⇄ 50 + W
B + 9 ⇄ 50 + W
B + 47 ⇄ 78 + W
13 + 42 ⇄ 78 + W
31 + 41 ⇄ 78 + W
A + 73 ⇄ 65 + W
46 + 77 ⇄ 65 + W

8 + 55 ⇄ 65 + W
51 ⇄ 65 + W
33 + 56 ⇄ 65 + W
22 ⇄ 65 + W
A + 66 ⇄ 61 + W
46 + 77 ⇄ 61 + W
51 ⇄ 61 + W
66 ⇄ 86 + W
73 ⇄ 86 + W
24 ⇄ 86 + W
B + 14 ⇄ 23 + W
5 ⇄ 23 + W
A + 14 ⇄ 48 + W
10 ⇄ 48 + W
53 ⇄ 48 + W
16 ⇄ 48 + W
B + 40 ⇄ 48 + W
25 ⇄ 48 + W
24 + 46 ⇄ 48 + W
17 ⇄ 48 + W
38 ⇄ 48 + W
A + 5 ⇄ 30 + W
B + 10 ⇄ 30 + W
18 + 46 ⇄ 30 + W
13 + 45 ⇄ 30 + W
B + 25 ⇄ 30 + W
31 + 54 ⇄ 30 + W
39 + 46 ⇄ 30 + W
B + 17 ⇄ 30 + W
33 + 42 ⇄ 30 + W
A + 5 ⇄ 27 + W

B + 53 ⇄ 27 + W
18 + 46 ⇄ 27 + W
B + 16 ⇄ 27 + W
13 + 45 ⇄ 27 + W
46 + 47 ⇄ 27 + W
13 + 19 ⇄ 27 + W
26 + 41 ⇄ 27 + W
A + B ⇄ 46 + W
B + 24 ⇄ 14 + W
39 ⇄ 14 + W
47 ⇄ 14 + W
18 ⇄ 14 + W
A + 5 ⇄ 49 + W
B + 17 ⇄ 49 + W
13 + 19 ⇄ 49 + W
26 + 41 ⇄ 49 + W
B + 38 ⇄ 49 + W
33 + 42 ⇄ 49 + W
18 + 46 ⇄ 49 + W
8 + 34 ⇄ 49 + W
16 ⇄ 12 + W
17 ⇄ 12 + W
24 + 46 ⇄ 12 + W
B + 40 ⇄ 12 + W
20 ⇄ 12 + W
2 + 13 ⇄ 12 + W
25 ⇄ 43 + W
38 ⇄ 43 + W
17 ⇄ 43 + W
20 ⇄ 43 + W
B + 21 ⇄ 43 + W

45 ⇄ 24 + W
19 ⇄ 24 + W
B + 46 ⇄ 13 + W
28 ⇄ 24 + W
B + 2 ⇄ 24 + W
B + 25 ⇄ 32 + W
46 + 47 ⇄ 32 + W
B + 17 ⇄ 32 + W
39 + 46 ⇄ 32 + W
13 + 45 ⇄ 32 + W
33 + 42 ⇄ 32 + W
13 + 28 ⇄ 32 + W
B + 20 ⇄ 32 + W
54 + 54 ⇄ 32 + W
31 + 54 ⇄ 15 + W
26 + 41 ⇄ 15 + W
B + 38 ⇄ 15 + W
46 + 47 ⇄ 15 + W
13 + 19 ⇄ 15 + W
33 + 42 ⇄ 15 + W
13 + 28 ⇄ 15 + W
B + 20 ⇄ 15 + W
39 + 46 ⇄ 7 + W
B + 16 ⇄ 7 + W
13 + 28 ⇄ 7 + W
26 ⇄ 2 + W
B + 13 ⇄ 41 + W
A + 41 ⇄ 31 + W
B + 33 ⇄ 54 + W
13 + 46 ⇄ 54 + W
13 + 46 ⇄ 31 + W

B + 33 ⇄ 31 + W
A + 13 ⇄ 33 + W
B + 31 ⇄ 42 + W
41 + 46 ⇄ 42 + W
13 + 13 ⇄ 42 + W
B + 54 ⇄ 42 + W
46 + 46 ⇄ 33 + W
B + 8 ⇄ 33 + W
B + 64 ⇄ 68 + W
13 + 70 ⇄ 68 + W
B + 58 ⇄ 68 + W
5 ⇄ 68 + W
41 + 55 ⇄ 68 + W
78 ⇄ 68 + W
A + 64 ⇄ 80 + W
46 + 66 ⇄ 80 + W
8 + 70 ⇄ 80 + W
B + 65 ⇄ 80 + W
17 ⇄ 80 + W
33 + 55 ⇄ 80 + W
38 ⇄ 80 + W
B + 66 ⇄ 64 + W
46 + 70 ⇄ 64 + W
B + 73 ⇄ 64 + W
39 ⇄ 64 + W
13 + 55 ⇄ 64 + W
47 ⇄ 64 + W
A + 58 ⇄ 74 + W
B + 65 ⇄ 74 + W
46 + 73 ⇄ 74 + W
33 + 55 ⇄ 74 + W

10 ⇄ 74 + W
31 + 56 ⇄ 74 + W
25 ⇄ 74 + W
A + 78 ⇄ 82 + W
B + 38 ⇄ 82 + W
46 + 47 ⇄ 82 + W
33 + 42 ⇄ 82 + W
B + 25 ⇄ 82 + W
13 + 45 ⇄ 82 + W
31 + 31 ⇄ 82 + W
41 + 50 ⇄ 82 + W
A + 64 ⇄ 76 + W
B + 61 ⇄ 76 + W
46 + 66 ⇄ 76 + W
46 + 73 ⇄ 76 + W
17 ⇄ 76 + W
13 + 77 ⇄ 76 + W
25 ⇄ 76 + W
A + 70 ⇄ 66 + W
46 + 55 ⇄ 66 + W
19 ⇄ 66 + W
B + 77 ⇄ 66 + W
45 ⇄ 66 + W
B + 73 ⇄ 58 + W
13 + 55 ⇄ 58 + W
18 ⇄ 58 + W
41 + 56 ⇄ 58 + W
47 ⇄ 58 + W
A + 55 ⇄ 77 + W
46 + 56 ⇄ 77 + W
26 ⇄ 77 + W

50 ⇄ 77 + W

B + 65 ⇄ 63 + W

46 + 73 ⇄ 63 + W

13 + 77 ⇄ 63 + W

33 + 55 ⇄ 63 + W

25 ⇄ 63 + W

54 + 56 ⇄ 63 + W

20 ⇄ 63 + W

B + 77 ⇄ 73 + W

46 + 55 ⇄ 73 + W

45 ⇄ 73 + W

13 + 56 ⇄ 73 + W

28 ⇄ 73 + W

64 ⇄ 85 + W

B + 86 ⇄ 85 + W

58 ⇄ 85 + W

14 ⇄ 85 + W

A + 23 ⇄ 35 + W

B + 48 ⇄ 35 + W

30 ⇄ 35 + W

27 ⇄ 35 + W

14 + 46 ⇄ 35 + W

49 ⇄ 35 + W

48 ⇄ 11 + W

12 ⇄ 11 + W

43 ⇄ 11 + W

30 ⇄ 44 + W

27 ⇄ 44 + W

B + 48 ⇄ 44 + W

14 + 46 ⇄ 44 + W

13 + 24 ⇄ 44 + W

32 ⇄ 44 + W

15 ⇄ 44 + W

7 ⇄ 44 + W

B + 12 ⇄ 44 + W

27 ⇄ 52 + W

49 ⇄ 52 + W

B + 12 ⇄ 52 + W

13 + 24 ⇄ 52 + W

15 ⇄ 52 + W

2 + 41 ⇄ 52 + W

30 ⇄ 3 + W

49 ⇄ 3 + W

B + 43 ⇄ 3 + W

15 ⇄ 3 + W

32 ⇄ 3 + W

31 ⇄ 55 + W

B + 56 ⇄ 55 + W

54 ⇄ 55 + W

33 ⇄ 56 + W

A + 68 ⇄ 71 + W

B + 80 ⇄ 71 + W

46 + 64 ⇄ 71 + W

33 + 70 ⇄ 71 + W

B + 74 ⇄ 71 + W

30 ⇄ 71 + W

31 + 55 ⇄ 71 + W

82 ⇄ 71 + W

A + 68 ⇄ 84 + W

B + 76 ⇄ 84 + W

13 + 66 ⇄ 84 + W

46 + 58 ⇄ 84 + W

49 ⇄ 84 + W

41 + 77 ⇄ 84 + W

82 ⇄ 84 + W

80 ⇄ 62 + W

76 ⇄ 62 + W

63 ⇄ 62 + W

43 ⇄ 62 + W

46 + 64 ⇄ 79 + W

B + 76 ⇄ 79 + W

13 + 73 ⇄ 79 + W

32 ⇄ 79 + W

33 + 70 ⇄ 67 + W

13 + 66 ⇄ 67 + W

B + 80 ⇄ 67 + W

46 + 64 ⇄ 67 + W

54 + 55 ⇄ 67 + W

15 ⇄ 67 + W

B + 63 ⇄ 67 + W

32 ⇄ 67 + W

B + 74 ⇄ 83 + W

46 + 58 ⇄ 83 + W

B + 63 ⇄ 83 + W

13 + 73 ⇄ 83 + W

54 + 55 ⇄ 83 + W

30 ⇄ 83 + W

42 + 56 ⇄ 83 + W

32 ⇄ 83 + W

31 + 55 ⇄ 81 + W

41 + 77 ⇄ 81 + W

B + 63 ⇄ 81 + W

13 + 73 ⇄ 81 + W

## 3. Thermodynamic Results

We first start by presenting the thermodynamic results obtained from the application of the algorithm explained in Chapter III. In Figures 53 and 54 we can see the relative free energies per bond per each compound at 1 bar and 298.15 K (default reference state of the XTB program). In Figure 53 we have the species that are unique to Cages **[2+3]**$^6$ (Cage23) **[4+6]b** (Cage46b) and the intermediate **[3+5]**$^9$ (R2R3B) and in Figure 54 we have all the species directly derived from breaking the cage **[4+6]a** (Cage46a).

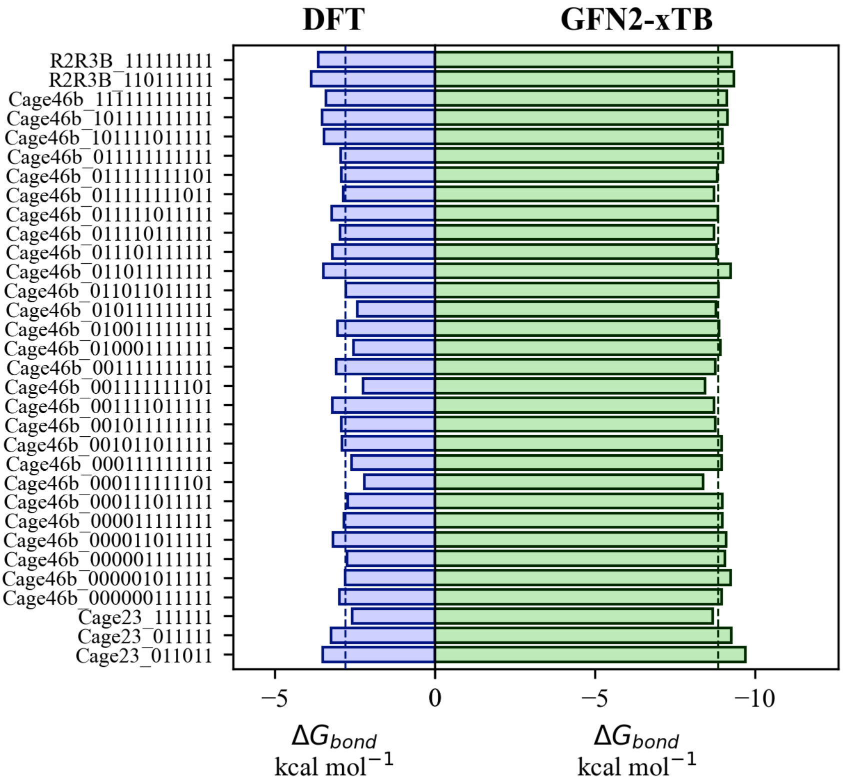


*Figure 53: DFT and GFN2-xTB relative free energies per imine bond for the species derived from opening and fragmentation of intermediate* **[3+5]**$^9$*,* **Cage [2+3]**$^6$ *and* **Cage [4+6]b***. Average values in dashed lines.*

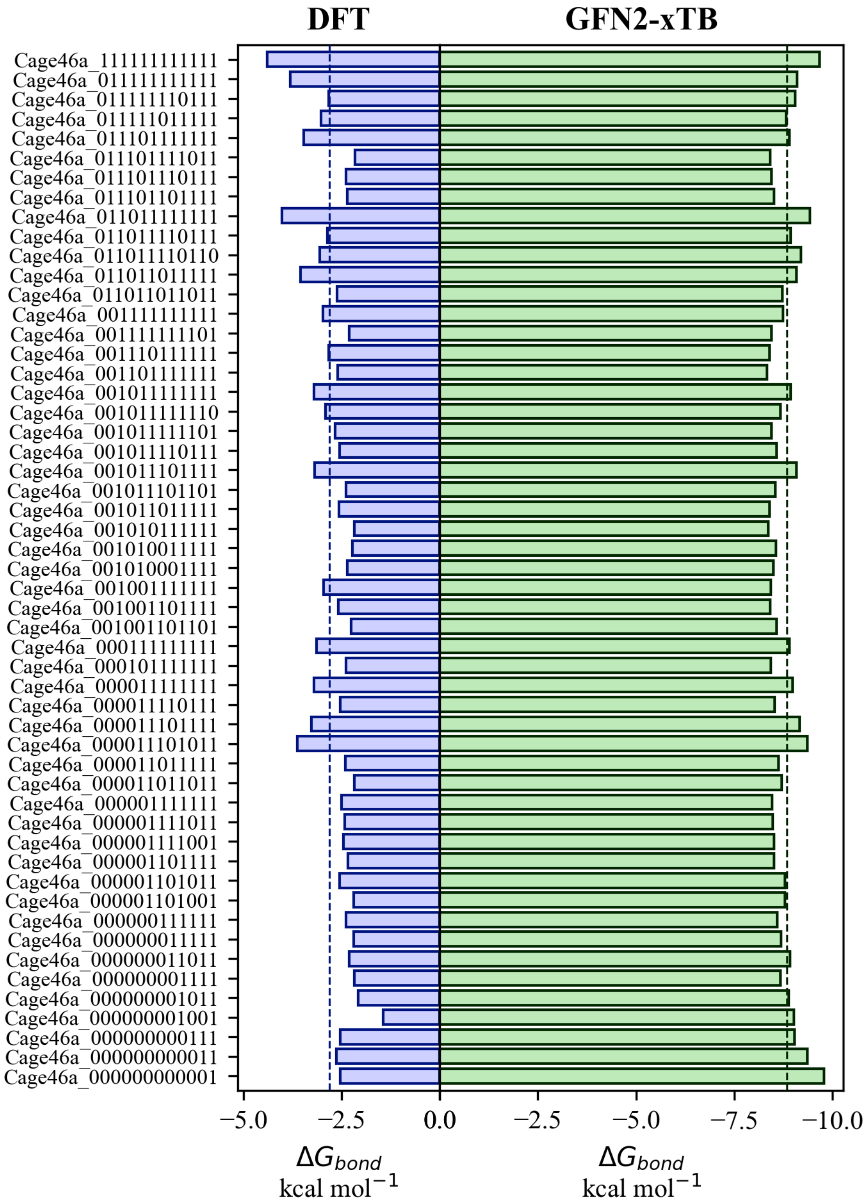


*Figure 54: DFT and GFN2-xTB relative free energies per imine bond for the species derived from* **Cage [4+6]a**. *Average values in dashed lines.*

From these figures we can observe two clear facts. The first one is that the $\Delta G_{bond}$ values obtained with GFN2-xTB, in general, deviate less from the average than the energies obtained from DFT. The second, is that the $\Delta G_{bond}$ obtained from GFN2-xTB are significantly lower (larger negative values) than the $\Delta G_{bond}$ obtained with DFT, meaning that DFT predicts less exergonic reactions than GFN2-xTB. In general, we can see that in both sets of energies, deviation of the individual values from the average is similar in magnitude for the two methods. The most notable exceptions can be seen at the lower end of Figure 54, which corresponds to the initial imine intermediates: **[1+1]$^1$**, **[1+2]$^2$**, **[1+3]$^3$**, **[2+1]$^2$**...

The effect of the difference in magnitude of the $\Delta G_{bond}$ between DFT and GFN2-xTB can be seen in the magnitude of the relative free energies. Figures 55 and 56 show the relative free energies per molecule instead of the above presented $\Delta G_{bond}$. We can see that the free energy surface predicted by GFN2-xTB is much steeper than the one predicted with DFT. We can also see how the deviation from the average energy per bond leads to larger differences in energy in DFT, compared with energies of GFN2-xTB.

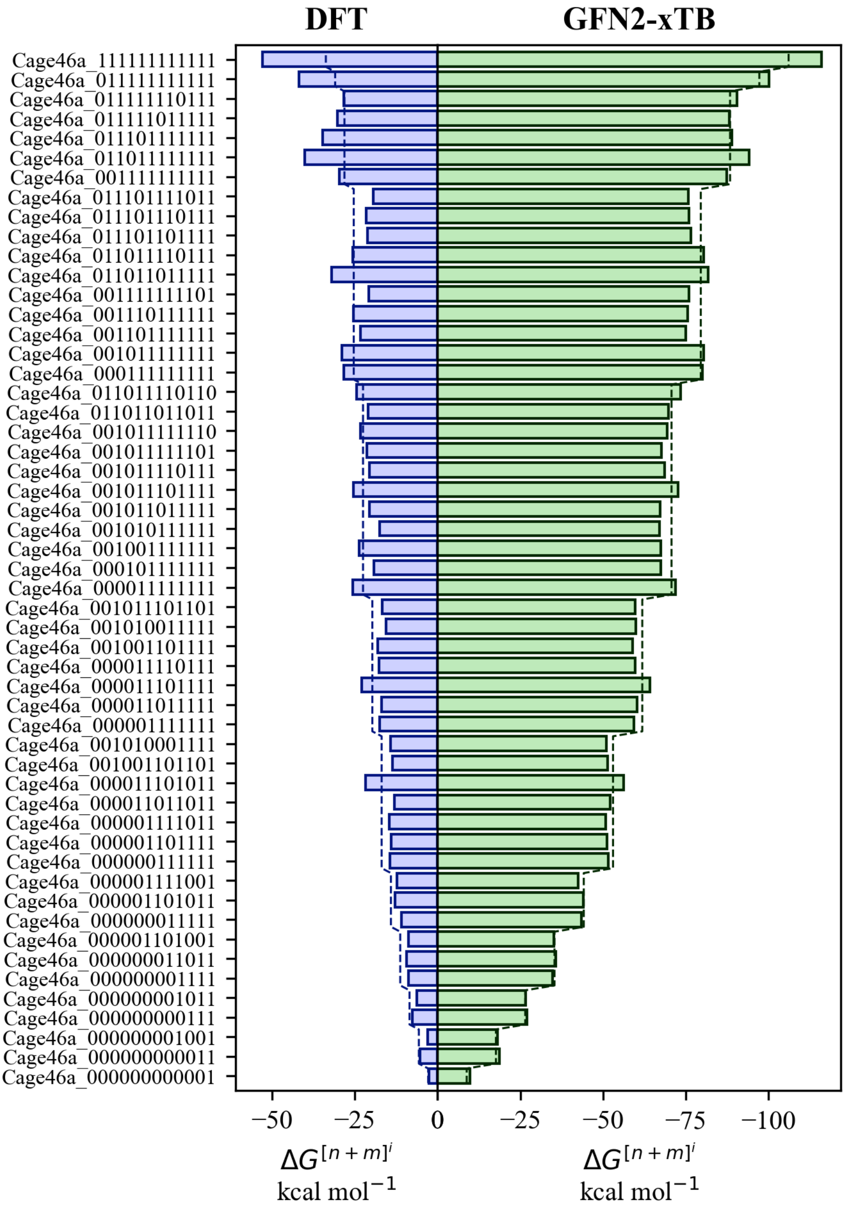


*Figure 55: DFT and GFN2-xTB relative free energies for the species derived from* **Cage [4+6]a**. *The average energy per bond times the number of imines of each species is depicted in dashed lines.*

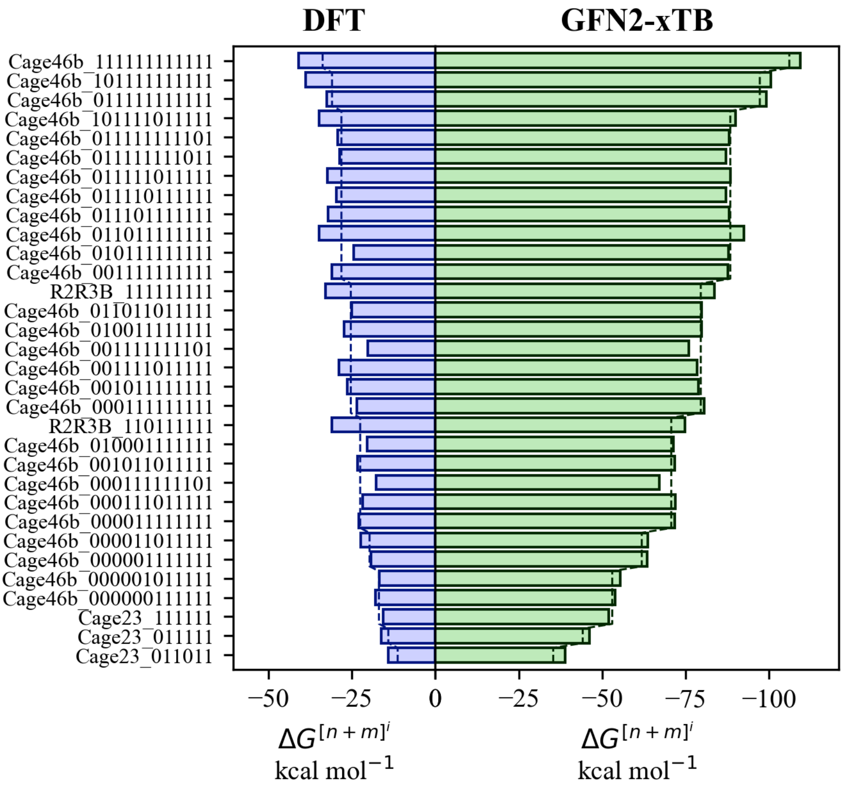


*Figure 56: DFT and GFN2-xTB relative free energies for the species derived from opening and fragmentation of intermediate* **[3+5]**$^{9}$*,* **Cage [2+3]**$^{6}$ *and* **Cage [4+6]b***. The average energy per bond times the number of imines of each species is depicted in dashed lines.*

## 4. Model 1, RT

Our first kinetic model corresponds to the simulation of the batch synthesis under the assumption that the temperature is constant, and equal to 25ºC, during the whole synthetic process (see scheme in Figure 57)

This model, although more simplistic, allows a smoother introduction to the intricacies of the kinetic simulations of the batch synthesis. As the transition states have not been explicitly computed through DFT or semi-empirical means we assigned energies to those hypothetical transition states. The criteria that we followed was to add the same value, X, to all reactions in the exergonic direction and we simulated various models utilizing different values of X. During the chapter we will refer to those models as “model of barrier of X kcal mol$^{-1}$ ” but please note that we do not mean that all barriers were considered to have that value, just the barriers of the exergonic direction of each reaction. This modeling choice allows us to get a measure of how the overall reaction behaves in relation to how dynamic the imine bond formation is.

Moving to the actual results depicted in Figure 57 we have the different concentration profiles of the amine monomer for the different simulated models with barriers ranging from 5 to 25 kcal mol$^{-1}$. The first thing to note is the decrease of the concentration of Ethylene-1,2-diamine (EtDA) even when no reaction is produced. This dilution effect will be present in all the models used to simulate the batch synthesis. The second thing to note is the clear discontinuity at 48h. It is due to the change of regime of the reactor. After two days, the addition of the aldehyde is stopped and as a consequence, the volume stops increasing and the dilution effect disappears.

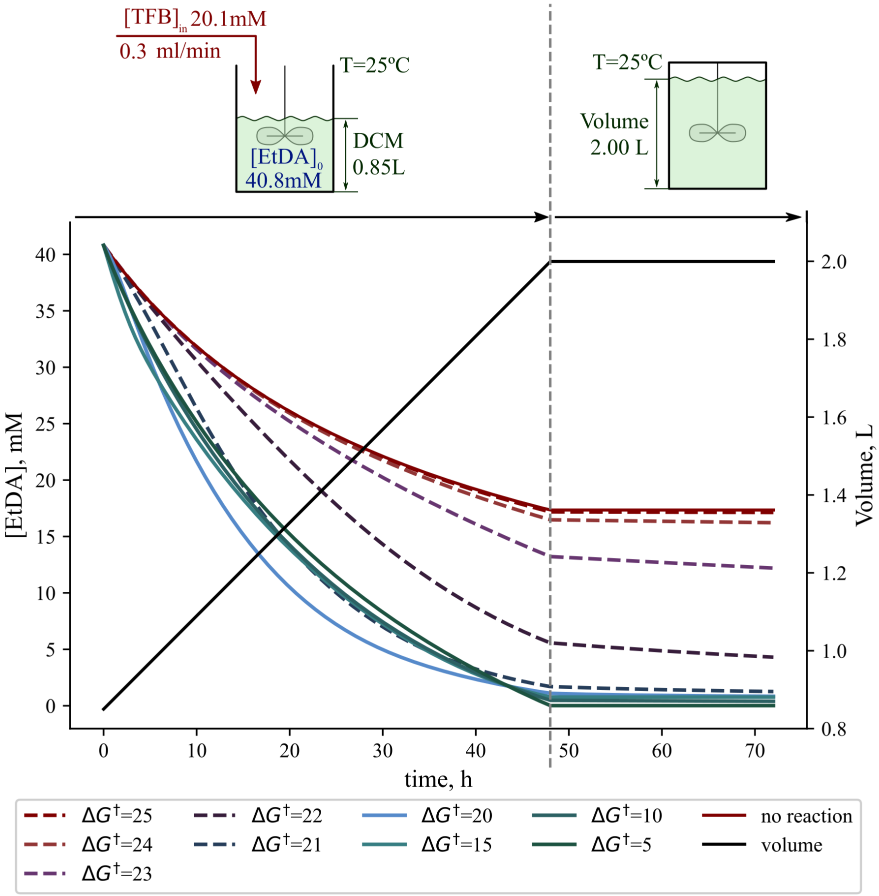


*Figure 57: Kinetic model 1 which simulates the batch synthesis of CC1. Scheme and effect of the barrier over the concentration of the amine monomer. Energies in kcal mol$^{-1}$.*

As we lower the barriers from 25 kcal mol$^{-1}$ to 20 kcal mol$^{-1}$ we see how the amine starts to get consumed during the simulation time. At about 20 kcal mol$^{-1}$ most of the amine is consumed after three days but we can see that as we further decrease the barrier the slope becomes less steep in the first ~7h, compared to the model with barriers of 20 kcal mol$^{-1}$, but ends up with a higher conversion at the end of the reaction.

Two different effects are happening here. First, accounting for the higher conversion, is the acceleration of the reaction due to the decrease of the barrier. As a consequence, we start accessing intermediates with a larger number of imine bonds, which explains why we observe a larger conversion as we decrease the barrier below 20 kcal mol$^{-1}$. Second, related to the increase of the reversibility of the imine bond formation, the system is becoming more dynamic. With a barrier of 20 kcal mol$^{-1}$ we require at least 48h to get most of the amine to react. As we saw in the previous section the formation of all intermediate species is exergonic with respect to the unreacted monomers, which will lead to barriers higher than 20 kcal mol$^{-1}$ in the non-exergonic direction. As we decrease the barriers, these reverse reactions start to become more competitive, and coupled with the effect of the dilution, the speed at which the amine is consumed becomes slower during the first ~ 20h than for barriers a bit higher.

With this brief introduction to the different effects governing the model, we can now move to the outcome of the self-assembly that we are interested in. Looking at Figure 58 we can see, grouped by general categories, the concentration profiles of the different intermediates and cages. In the top row we have the most dynamic model with barriers of 5 kcal mol$^{-1}$. We can see the major species present in the mixture in the axis with linear y scale on the left column and the species with a minor presence in axis of the right column represented with a logarithmic y scale. What we observe here is that we obtain the CC1 imine cage (**[4+6]a**) almost from the beginning, with a delay period induced by the accumulation of the **[4+5]**$^{9}$ intermediate (R2R3B_111111111). This evidences that the CC1 imine cage is the major thermodynamic sink out of the modeled species.

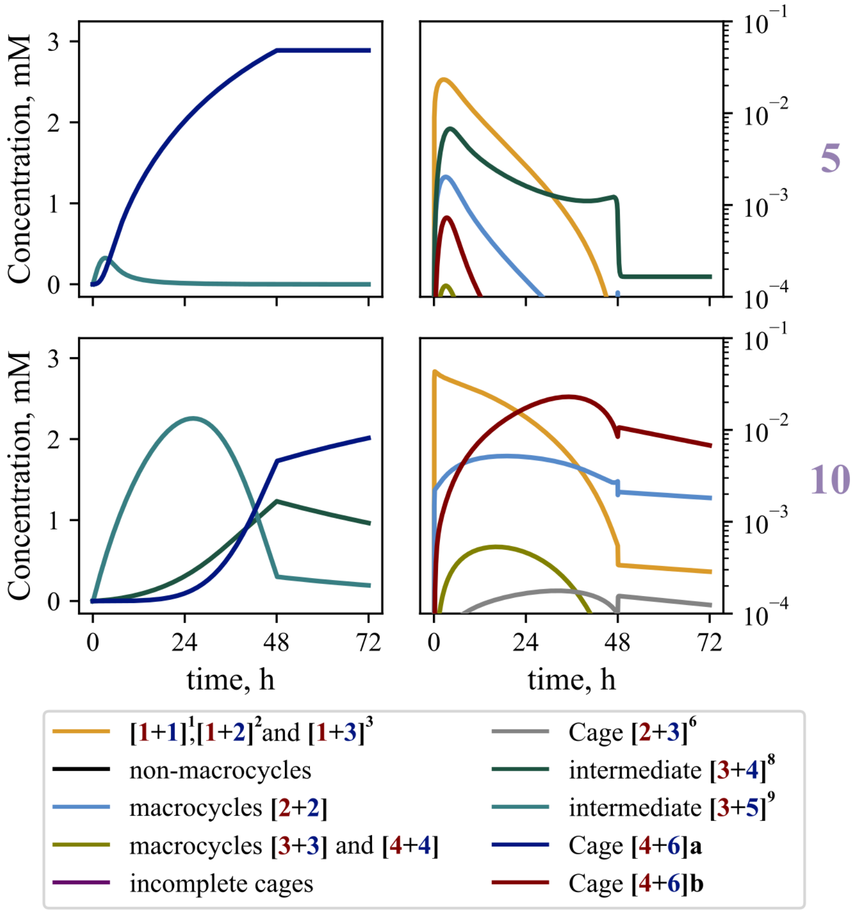


*Figure 58: Simulation of the Kinetic models with barriers of 5 and 10 kcal mol*$^{-1}$ *(top and bottom rows respectively). The groups of species whose concentration did not exceed 0.1 mM are plotted in a logarithmic scale.*

We can see that out of the other intermediates (in logarithmic scale), the early stages of the reaction (intermediates **[1+1]**$^{1}$, **[1+2]**$^{2}$, **[1+3]**$^{3}$) are the major components of the reaction mixture. These are followed by the intermediate **[3+4]**$^{8}$ (R2R3B_110111111) which in the early stages of the reaction, due to the high concentration of the amine monomer, has its equilibrium shifted to form the **[3+5]**$^{9}$. Next, we observe the species containing the smallest imine macrocycles (macrocycles **[2+2]**).

As we shift our attention to the model with barriers of 10 kcal mol$^{-1}$ we observe how the conversion to the **[4+6]a** cage has dropped as the delay period has increased. We now see that the **[3+4]**$^{8}$ intermediate ends up in higher concentration than the **[3+5]**$^{9}$ intermediate. If we go back to Figure 53 (Page 192) we can see how this intermediate has a lower energy per bond compared with its **[3+5]**$^{9}$ counterpart. We start realizing that the equilibrium of **[3+4]**$^{8}$ and **[3+5]**$^{9}$ is going to be governed by concentration effects. Amongst the species in low concentrations present in the mixture (shown in logarithmic scale) we can see that the **[4+6]b** cage becomes slightly more relevant.

Now, we move to the models with barriers of 15 and 20 kcal mol$^{-1}$ depicted in Figure 59. We observe that the CC1 imine cage has disappeared from the left column. We also see how, as they are not so reactive, there is some accumulation of the **[1+1]**$^{1}$, **[1+2]**$^{2}$, **[1+3]**$^{3}$ intermediates. As their maximum concentration raises, these species are more likely to form non-macrocycles. Nonetheless, the **[2+2]** macrocycles seem to play a major role compared with the non-macrocycles or with larger macrocycles. It is actually due to the accumulation of these species that **[4+6]b** becomes more accessible, which coupled to the loss of the dynamic nature of the bond, leads to some accumulation of this compound in the model of the 20 kcal mol$^{-1}$ barriers.

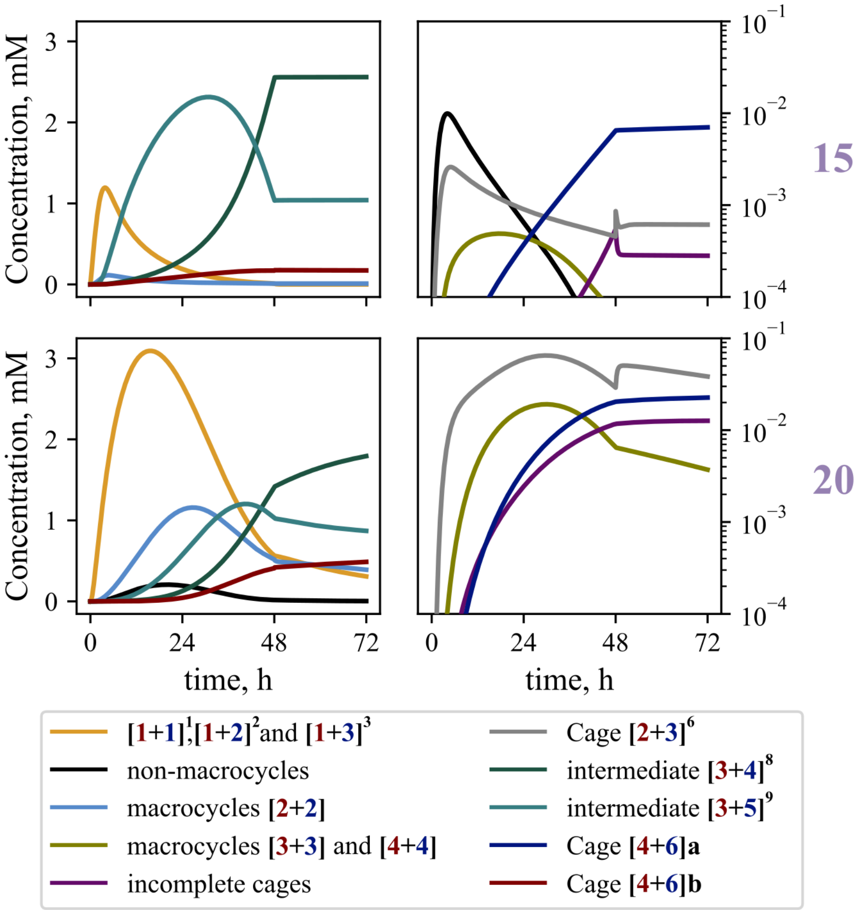


*Figure 59: Simulation of the Kinetic models with barriers of 15 and 20 kcal mol$^{-1}$ (top and bottom rows respectively). The groups of species whose concentration did not exceed 0.1 mM are plotted in a logarithmic scale.*

## 5. Model 2, Reactor Effect

After the initial models that simulate the batch synthesis, we proceed to simulate the flow synthesis. The reaction is the same and the chemical species present in the media are the same, so a complete kinetic model should be compatible with both syntheses.

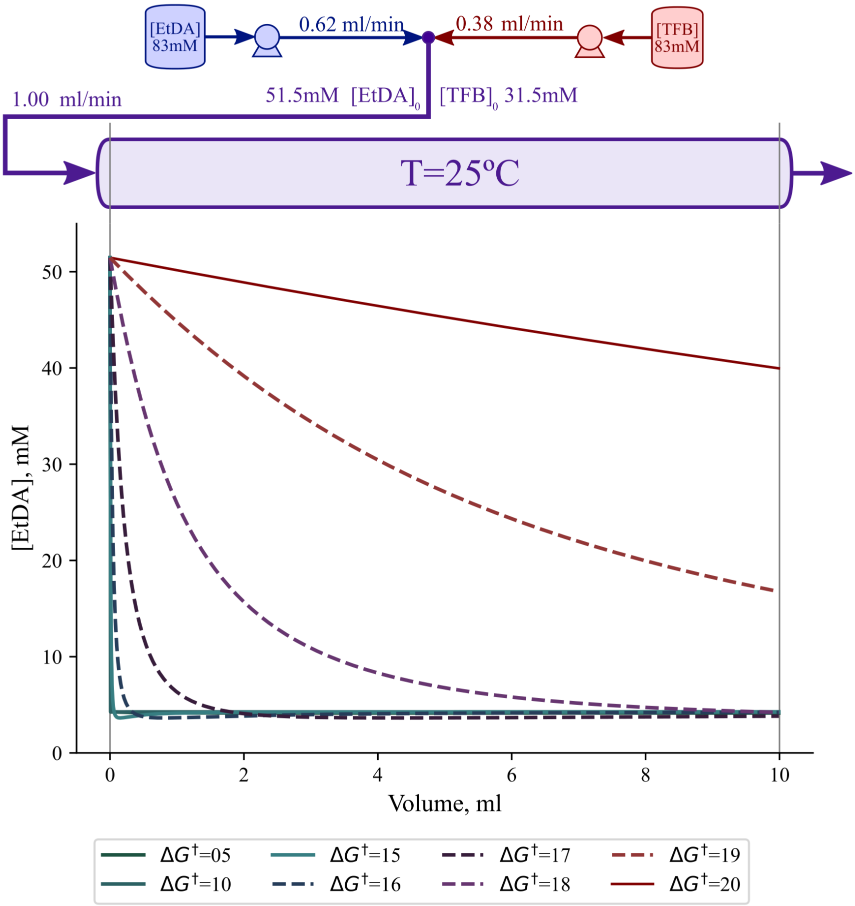


*Figure 60: Kinetic model 2 which simulates the flow synthesis of CC1. Scheme and effect of the barrier over the concentration of the amine monomer. Energies in kcal mol$^{-1}$.*

Although the flow synthesis is carried out at 100ºC, we proceeded to model the flow synthesis at the same temperature (25ºC) to observe how the change of reactor setup affects. The simulation of plug-flow reactors leads to a concentration profile across the reactor instead of the time, nonetheless, the interpretation of the concentration profiles remains similar. In our case, due to the lack of some technical details (length and/or diameter) of the specific reactor used in the synthesis we used the volume as independent variable for the simulation of the reactor.

Compared with Figure 57 we can see a close to full conversion of the EtDA up to barriers of 18 kcal mol$^{-1}$ in Figure 60. The residence time of the EtDA in the flow reactor is 10 min, compared with the 72 h of the previous reactor, however the high concentration of the TFB and the lack of the dilution effect play a major role, allowing large conversions at higher barriers than the batch synthesis.

In Figures 61 and 62 we move to the concentration profiles of the different species present in the mixture. For all models but the one with barriers of 20 kcal mol$^{-1}$ (lower row Figure 62) we observe a really fast reaction upon mixture. In the model with barriers of 5 kcal mol$^{-1}$ the **[3+4]**$^{\mathbf{8}}$ and the **[3+5]**$^{\mathbf{9}}$ are initially formed and are transformed, almost directly, to the cage **[4+6]a**.

In case of the model with barriers of 10 kcal mol$^{-1}$, **[3+4]**$^{\mathbf{8}}$ and **[3+5]**$^{\mathbf{9}}$ are observed as final products, however an initial peak concentration of **[4+6]b** is observed which was not relevant in the equivalent model for the batch synthesis (lower row of Figure 58). Amongst the species with lower concentrations we can see the final cage **[4+6]a**, the **[2+2]** macrocycles and the initial intermediates (**[1+1]**$^{\mathbf{1}}$, **[1+2]**$^{\mathbf{2}}$ and **[1+3]**$^{\mathbf{3}}$).

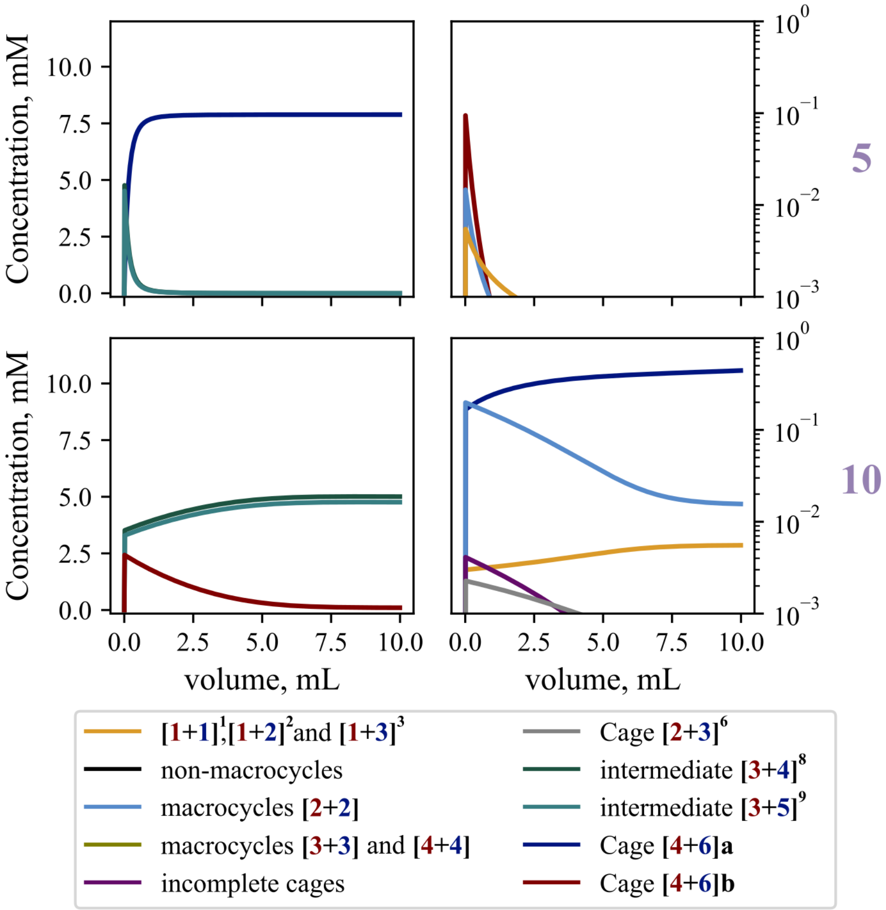


*Figure 61: Simulation of the Kinetic models for the flow synthesis with barriers of 5 and 10 kcal mol$^{-1}$ (top and bottom rows respectively). The groups of species whose concentration did not exceed 1 mM are plotted in a logarithmic scale.*

We can see how these species become major actors in the model with barriers of 15 kcal mol$^{-1}$ although they are immediately consumed. Not only that but we also find the non-macrocycles achieving to concentrations over 1mM. Besides, we observe how the system appears to have reached a stationary state (with respect to the volume) with the **[3+4]**$^{8}$, **[3+5]**$^{9}$ and **[4+6]b** as the major products.

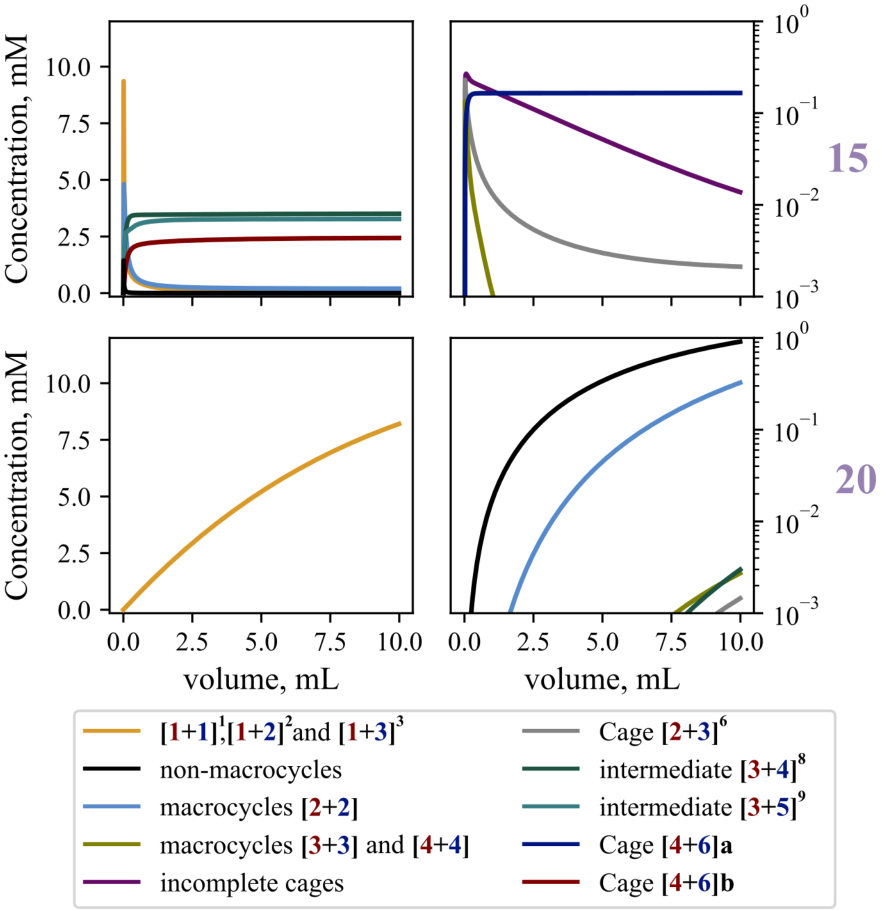


*Figure 62: Simulation of the Kinetic models for the flow synthesis with barriers of 15 and 20 kcal mol*$^{-1}$ *(top and bottom rows respectively). The groups of species whose concentration did not exceed 1 mM are plotted in a logarithmic scale.*

To understand these results it is important to remember that initially a high concentration of aldehyde is available. As a consequence the formation of **[2+m]** species is favored, which in turn leads to the formation of **[2+2]** macrocycles. As the concentration of these macrocycles increases, the likeliness of them dimerizing and forming intermediates that can later close to form the **[4+6]b** also increases. As a consequence, even if the **[4+6]a** cage is more stable, the barrier to overcome the transition from **[4+6]b** to **[4+6]a** requires at least to break 2 imine bonds, making the **[4+6]b** a kinetically trapped product. Because of the dilution effect and the generally low concentration of the **[2+2]** macrocycles this issue was not observed for the batch system.

Finally, in the model with barriers of 20 kcal mol$^{-1}$ we observe that from the EtDA (see Figure 60) that is consumed in the reaction mostly yields the early imine species. From the other species shown we can guess that the early imine species initially go to the non-macrocycles and these convert into the **[2+2]** macrocycles.

## 6. Model 3, Temperature effect

Model 1 (Batch RT) was able to match the fact that the CC1 cage is the major outcome of the reaction but only up to barriers of 10 kcal mol$^{-1}$ but Model 2 (Flow RT) was unable to do so for the same barriers. We were aware that these models were far from reality since we were not taking into account the actual temperatures used in both syntheses.

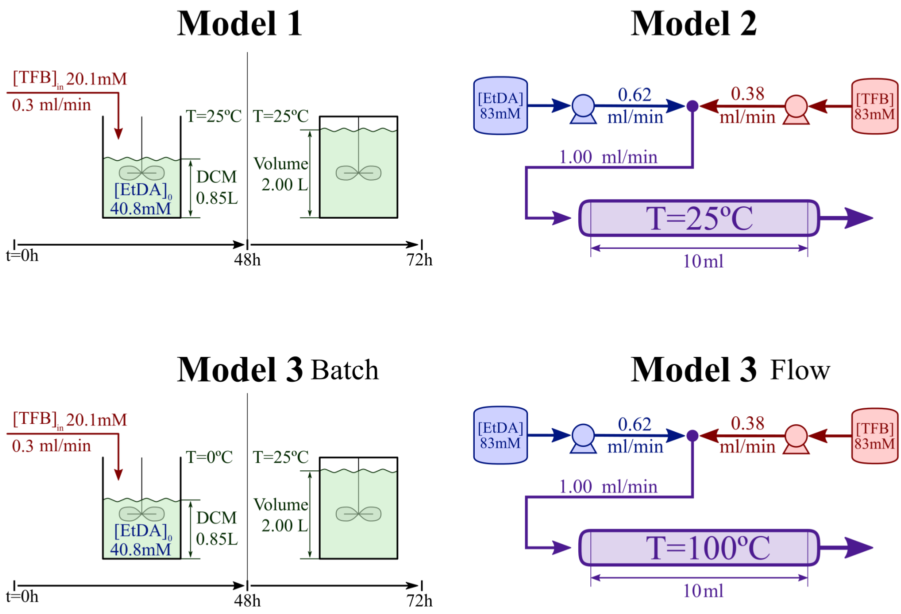

Model 1
[TFB]in 20.1mM
0.3 ml/min
T=25ºC
T=25ºC
[EtDA]0 40.8mM
DCM 0.85L
Volume 2.00 L
t=0h
48h
72h
Model 2
[EtDA] 83mM
0.62 ml/min
0.38 ml/min
[TFB] 83mM
1.00 ml/min
T=25ºC
10ml
Model 3 Batch
[TFB]in 20.1mM
0.3 ml/min
T=0ºC
T=25ºC
[EtDA]0 40.8mM
DCM 0.85L
Volume 2.00 L
t=0h
48h
72h
Model 3 Flow
[EtDA] 83mM
0.62 ml/min
0.38 ml/min
[TFB] 83mM
1.00 ml/min
T=100ºC
10ml

*Figure 63: Reactor setups of models 1, 2 and 3.*

Now we move to an improved version of these models, Model 3 for the batch synthesis (Batch in the following figures) where during the first 48h the temperature is kept constant at 0ºC and 25ºC for the remaining 24h. In this model we are assuming that the change of temperature in the flask and the media is instantaneous.

Although we are aware that it is not realistic we settled with this model due to three major reasons. The first reason is that we lack the data to reproduce the exact temperature profile. The second reason, from a technical perspective, is that we should recalculate the energies, barriers and constants at each step of the simulation which would significantly increase the computational cost of the simulations. The third and final reason is that, as we have already studied Model 1 (fully at 25ºC), the approximation should be acceptable unless we observe drastic changes in Model 3. In such event we would easily identify that the model needs to be improved.

In order to calculate the free energies at 0ºC we considered that the enthalpy and entropy remain constant in this interval. We proceed similarly with the improved version of the Model 3 for the flow synthesis (termed as Flow in the following figures) which assumes a constant temperature of 100ºC across the reactor.

Finally, moving to the results, we decided to look first at the differences in the major self-assembly outcomes. In Figure 64 we can see the concentration profiles for all four kinetic models with barriers of 5, 10, 15 and 20 kcal mol$^{-1}$. For this figure (and similar figures that will appear during the chapter) we decided to group intermediates **[3+4]**$^{8}$ and **[3+5]**$^{9}$ as it gives a better overall picture. We observe a minimum effect of the temperature for the batch system. As we could expect, the reaction slows down in general which can be observed in the model with barriers of 10 kcal mol$^{-1}$ as the yield of the experimentally found cage, **[4+6]a**, is significantly reduced.

On the other hand, we can observe how the increase in temperature favors the kinetics in the flow synthesis, as could be expected. Such acceleration allows the system to form the **[4+6]a** cage at barriers of 10 kcal mol$^{-1}$ for this reactor.

As a consequence of the speed up, the apparent stationary state found for the FlowRT model with barriers of 15 kcal mol$^{-1}$ is now observed in the Flow model with barriers of 20 kcal mol$^{-1}$. However, for the first time in all of the presented models a quantitative accumulation of the **[2+3]**$^{6}$ cage can be observed.

As commented in the discussion of Model 2, we require at least two imine bond breaks to be able to access **[4+6]a** cage from the **[4+6]b** cage. If we focus on the results from the model of Flow with barriers of 15 kcal mol$^{-1}$ we can gain more insight into the mechanism. Here it can be seen, that the cage **[4+6]b** (containing **[2+2]** and **[4+4]** macrocycles) does not go directly to the **[4+6]a** cage (that contains **[3+3]** and **[4+4]** macrocycles) but to the **[3+4]**$^{8}$ and **[3+5]**$^{9}$ intermediates (containing **[2+2]** and **[3+3]** macrocycles) instead. This points towards a favored opening of the **[4+4]** macrocycle (that we also observed in Chapter II).

The different distribution of species between the reactors indicates that concentration effects play a role in the outcome of the reaction. However, the similar distribution across temperatures highlights temperature does not produce a significant change of mechanism. The only exception to this, is the **[2+3]**$^{6}$ cage which seems to act as a reservoir of **[2+2]** macrocycles, indicating that it is also a compound that can act as a thermodynamic sink.

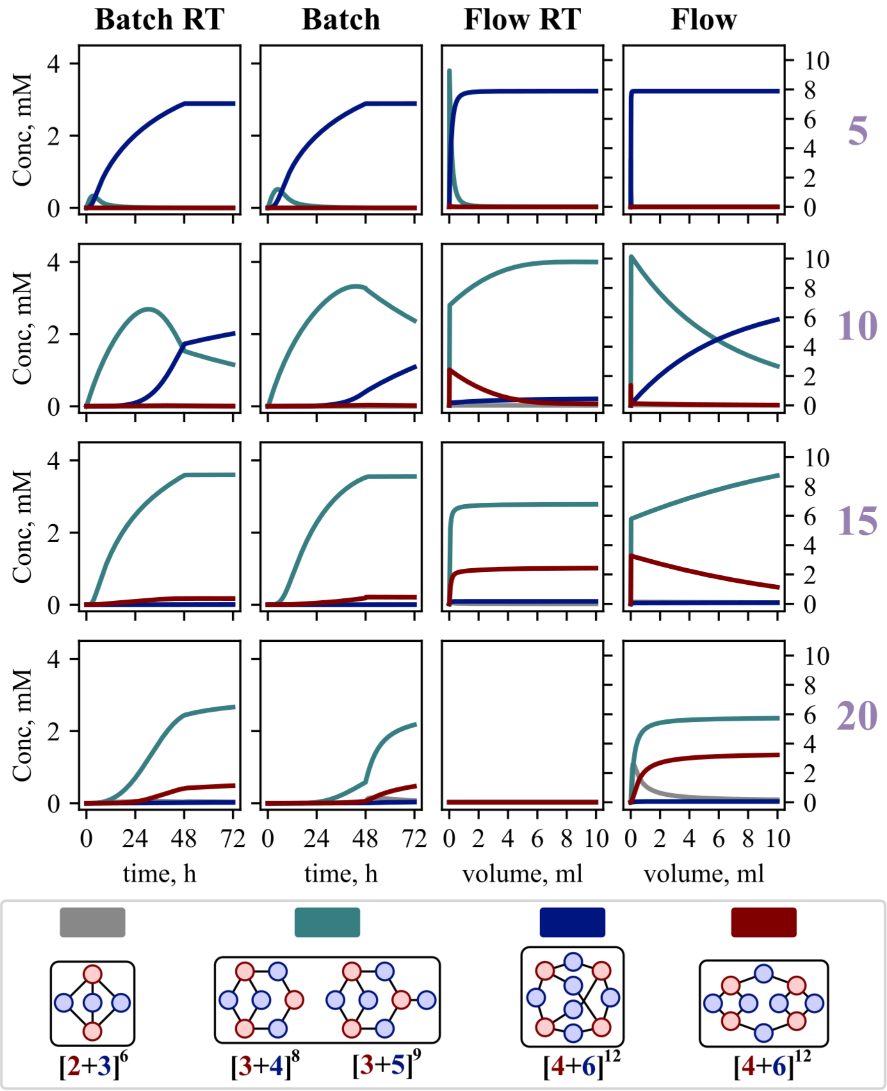


*Figure 64: Effect of the temperature on kinetic models. Numbers in purple are the barrier used for the kinetic models of the same row in kcal mol$^{-1}$.*

## 7. Model 4, Energy per bond

A more precise model is needed as we can see that including more realistic temperatures lead to non-matching models. At room temperature the batch synthesis allowed quantitative amounts of the correct cage with low barriers of 10 kcal mol$^{-1}$ but the flow synthesis could not. At the more realistic temperatures the opposite happens. A possibility is that the DFT functional that we are using underestimates or overestimates the relative energy per imine bond of the species. As we saw in Figures 53 and 54 there is a clear gap between the relative energy per imine bond with respect to GFN2-xTB. Not only that but we also notice something similar in computational results on imine cages where the steepness of the overall free energy profile is highly dependent on the functional or the method used[15] or even the nature of the cage.[16]

We decided to test the effect of scaling the relative energy per bond for all species on the kinetic simulations. We computed 4 new sets of energies by scaling the average relative energy per bond of our set of energies of DFT to the values -2.0, -4.0, -6.0 and -8.8 (this last one matches the average relative energy per bond of GFN2-xTB).

The effect of this scaling is illustrated in Figure 65. On the axis at the right hand side of the figure we have the relative energies per bond. On the left hand side of the figure we have the relative free energies of formation of each compound, which allows to see the effect of the scaling simultaneously. In essence, it shows how it affects the steepness of the free energy surface.

---

[15] Zhu, G.; Liu, Y.; Flores, L.; Lee, Z. R.; Jones, C. W.; Dixon, D. A.; Sholl, D. S.; Lively, R. P. *Chem. Mater.* **2018**, 30, 1, 262-272.

[16] Greenaway, R. L.; Santolini, V.; Bennison, M. J.; Alston, B. M.; Pugh, C. J.; Little, M. A.; Eden-Rump, E. G. B.; Clowes, R.; Shakil, A.; Cuthbertson, H. J.; Armstrong, H.; Briggs, M. E.; Jelfs, K. E.; Cooper , A. I. *Nat. Commun.* **2018**, 9, 2849.

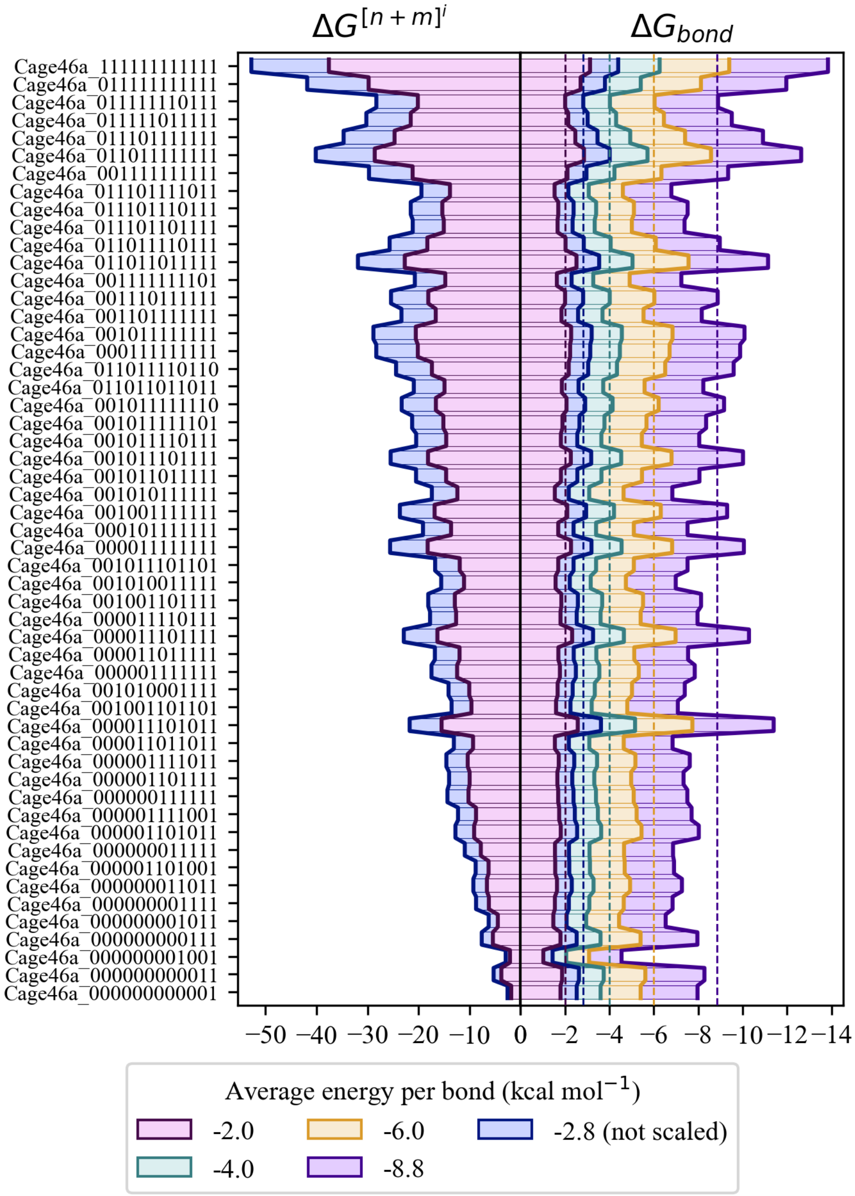


*Figure 65: Effect on the relative energies of the scalings of the energy per bond used in the kinetic models. Subset of energies of species derived from* **Cage [4+6]a**. *Energies in kcal mol$^{-1}$.*

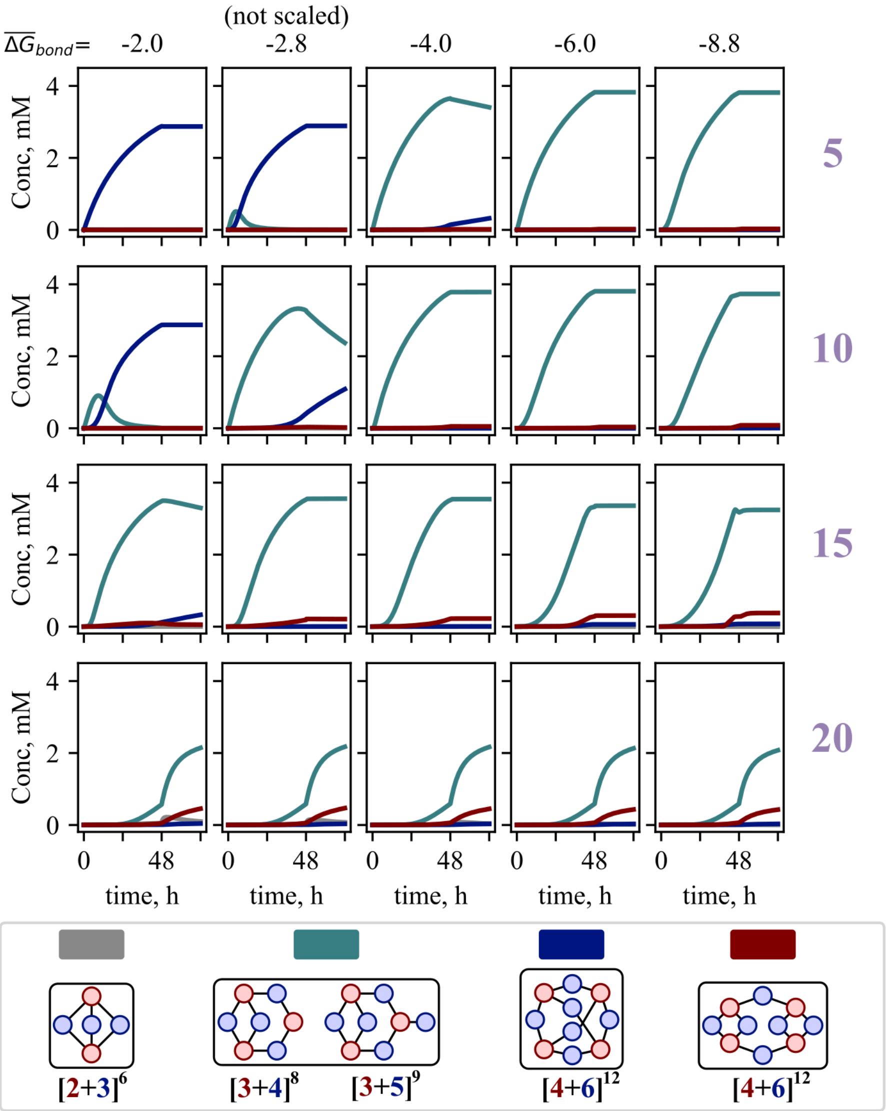


*Figure 66: Batch synthesis. Effect over the kinetic models of the scaling of the energies per bond. Numbers in purple are the barrier used for the kinetic models of the same row in kcal mol$^{-1}$.*

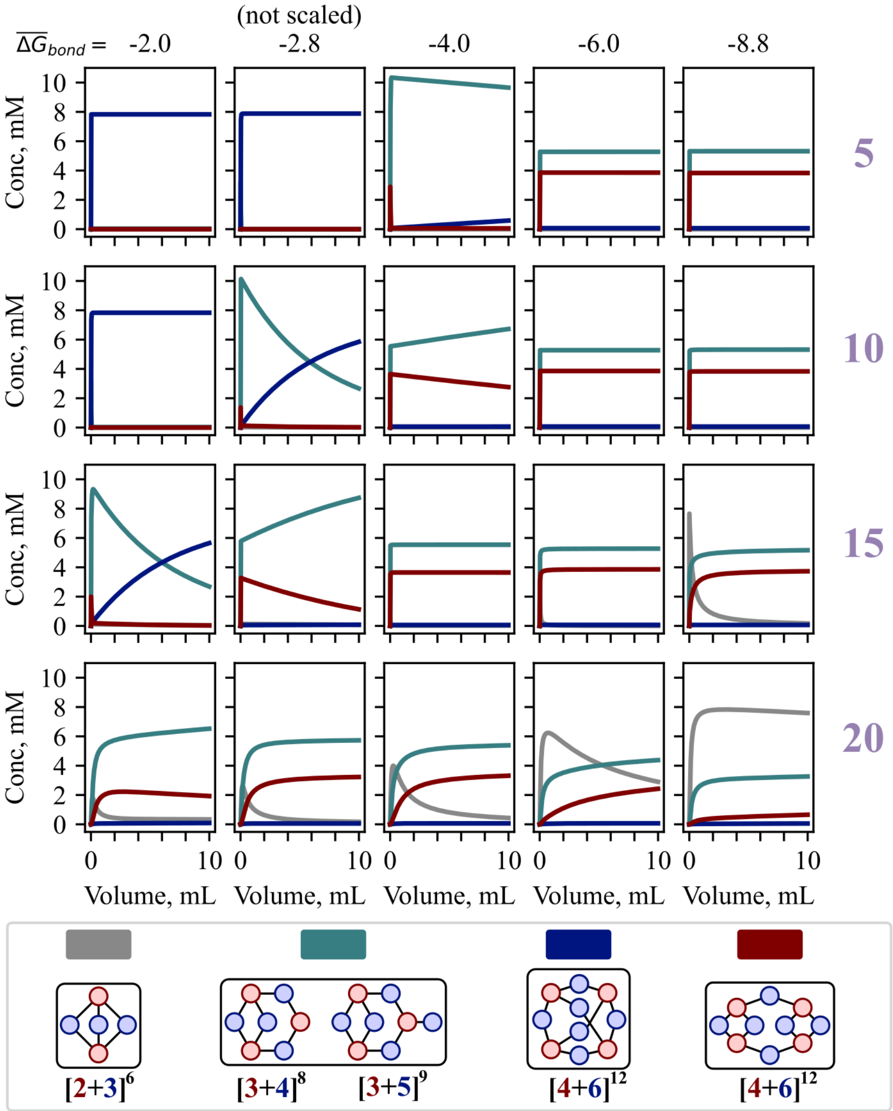


*Figure 67: Flow synthesis. Effect over the kinetic models of the scaling of the energies per bond. Numbers in purple are the barrier used for the kinetic models of the same row in kcal mol$^{-1}$.*

Now we move to the results of these kinetic simulations presented in Figures 66 (for the batch synthesis) and 67 (for the flow synthesis). What we observe is that increasing the steepness of the profile (decreasing the average energy per bond, moving to the right hand side of the figures) leads to models where the CC1 imine cage is not achievable even in the models with barriers of 5 kcal mol$^{-1}$ for both syntheses.

Moreover, we observe the opposite behavior as we reduce the average energy per imine bond (leftmost column in both figures), CC1 becomes a major product in the model with 10 kcal mol$^{-1}$ for the batch synthesis (Figure 66) and even in the model with barriers of 15 kcal mol$^{-1}$ for the flow synthesis (Figure 67). Nonetheless, we are far from yields over 90% yield for the same model, with barriers comparable to the ones obtained in Chapter I (~19 kcal mol$^{-1}$ for the *n*-butylamine and the benzaldehyde[17]) in both syntheses.

Looking at the results from the flow reactors we see how the **[2+3]$^{6}$** cage starts to become more relevant as we decrease the average energy per imine bond (increasing in absolute value), becoming the largest product at barriers of 20 kcal mol$^{-1}$. Interestingly, we can see a small yield of this cage for the batch synthesis only when going in the opposite direction, increasing the average energy per bond to -2.0 kcal mol$^{-1}$ (and barriers of 20 kcal mol$^{-1}$).

---

[17] Value obtained from model calculated with B3LYP in Chapter I with a systematic correction (H2) of 3.2 kcal mol$^{-1}$ as its microkinetic model was the one that fitted the best the experimental data.

## 8. Model 5, Relative Energies GFN2-xTB

During the previous sections, we have seen how the model of the reactor, the temperature and even the magnitude of the free energies affects the kinetics. One other variable to take into account is how sensitive these kinetic models are to the relative energies between the species. In order to account for this effect we decided to build our kinetic models using the free energies obtained with GFN2-xTB. From the results of the previous section, we know that there is a major effect on the scale of the $\Delta G_{bond}$. Therefore, we decided to also scale up the $\Delta G_{bond}$ obtained with GFN2-xTB so that the average relative energy per bond matches the average $\Delta G_{bond}$ of the energies corrected with DFT.

In Figure 68 we can observe the effect of such scaling on the relative energies per imine bond. It becomes even more obvious, now that we have the values obtained with GFN2-xTB at the same scale as the results with DFT side by side, that GFN2-xTB has a very low deviation with respect to the mean value.

Moving to the simulations of the batch synthesis (Figure 69) we can observe several differences with the DFT based simulations. Starting on the unscaled GFN2-xTB models, we observe that as we increase the barrier the **[2+3]**$^6$ cage is more prone to be formed, which was always almost non-existent in the batch synthesis of the previous models (Models 1, 3 and 4). We can easily understand this if we look back at the relative energies in Figure 56 (page 196). With DFT the open cage (**[2+3]**$^5$) is more stable than the closed cage (**[2+3]**$^6$) whilst in GFN2-xTB it is the other way around. Other than that, in the kinetic model we see that the CC1 cage almost disappears, as we previously saw with the scaled DFT energies.

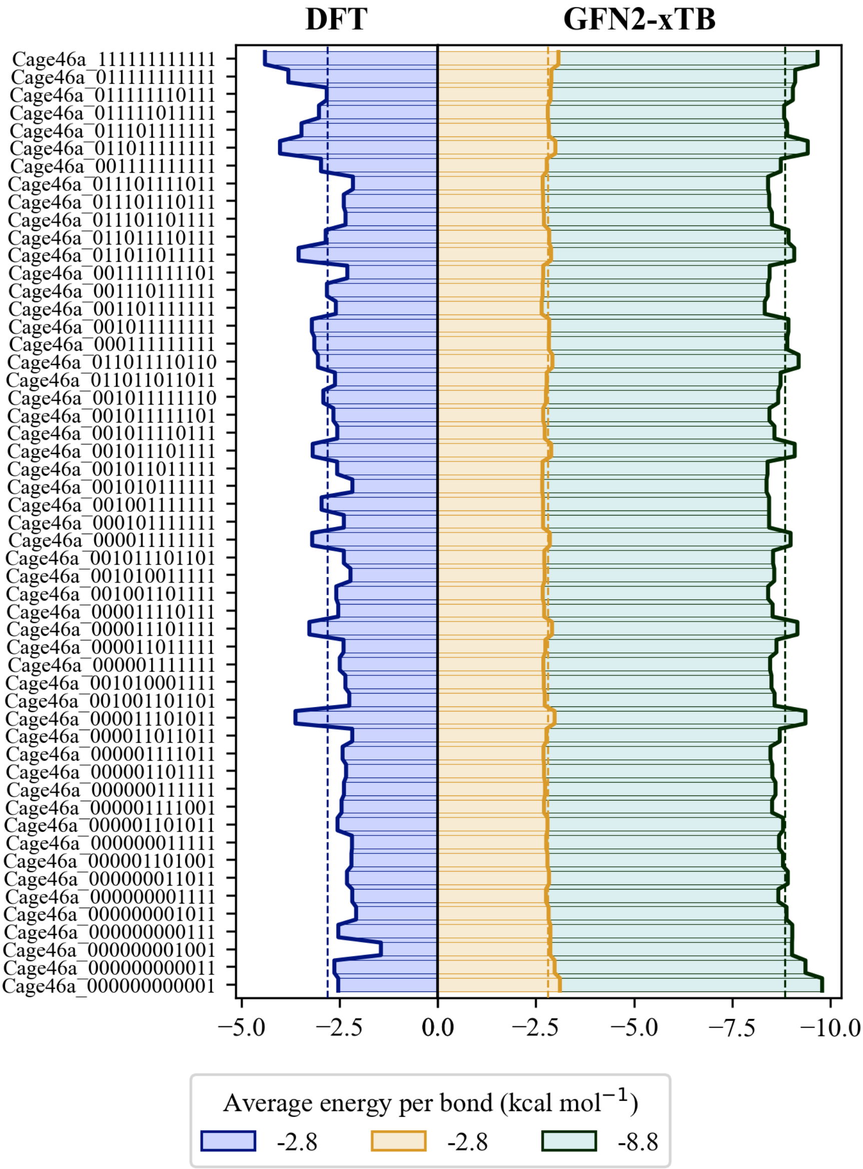


*Figure 68: Subset of raw and scaled relative energies per bond for different intermediates. Energies in kcal mol.*

On the other hand the scaled GFN2-xTB presented above (see Figure 68) shows more promising results (see Figure 69). We observe the CC1 cage as the main product even in the model with barriers of 15 kcal mol$^{-1}$ even if it is at the cost of reducing the final yield compared with the highest yield of the DFT based models.

When we look at the simulations of the flow synthesis (Figure 70) we again observe major changes compared with the results of DFT. As we observed for the batch synthesis the unscaled GFN2-xTB models show almost no yield of the CC1 cage independently of the barriers. The main outcome of these models is that these models highlight the major species that act as thermodynamic sinks (**[2+3]**$^{6}$, **[3+4]**$^{8}$ and **[3+5]**$^{9}$, and **[4+6]b**) and hint at some equilibrium between them.

Such equilibrium can also be appreciated for the scaled GFN2-xTB models of the plug-flow system, with the major difference that these models predict the **[4+6]a** cage as the major product up to barriers of 15 kcal mol$^{-1}$. The models with barriers of 5 and 10 kcal mol$^{-1}$ look almost identical on the batch synthesis (Figure 69) which also happens on the flow synthesis (Figure 70). If we go into further detail to see if there is different speciation between those two models we see that they are still identical. Only the speciation of the mechanistically relevant models is presented in Figures 71 and 72.

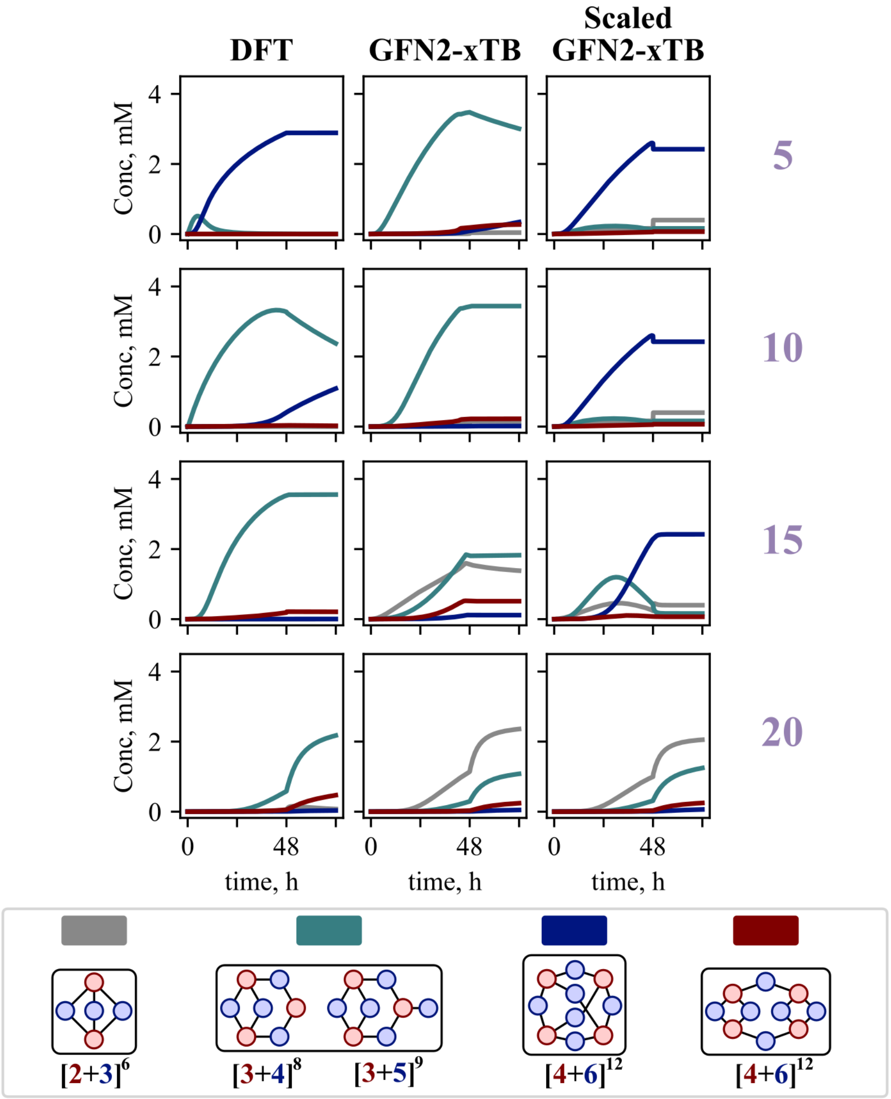


*Figure 69: Kinetic models for the batch synthesis with the energies obtained with GFN2-xTB. Numbers in purple are the barrier used for the kinetic models of the same row in kcal mol$^{-1}$.*

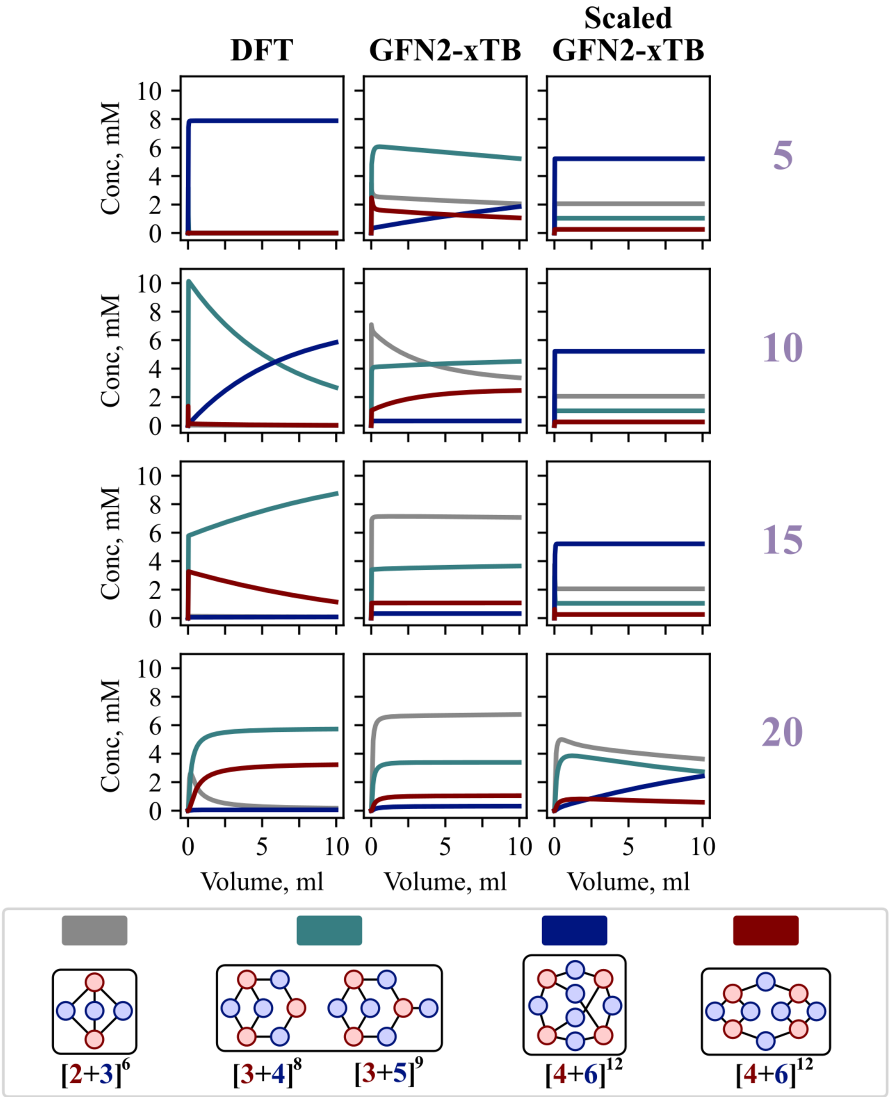


*Figure 70: Kinetic models for the flow synthesis with the energies obtained with XTB. Numbers in purple are the barrier used for the kinetic models of the same row in kcal mol*$^{-1}$.

A curious result that we can find for the batch synthesis is that there are two discontinuities in the slope of the concentration profiles (which is easier to observe on the logarithmic scale). The first discontinuity comes as a consequence of the concentration of the amine monomer being depleted in an almost irreversible manner, before the change of regime occurs, and the second discontinuity is a consequence of such change of regime.

The reason for such fast depletion comes from the stability of the early imine intermediates ( **[1+1]**$^{1}$, **[1+2]**$^{2}$, **[1+3]**$^{3}$ ). Which we can see that corresponds to the largest concentration during the first 24h. If we go back to Figure 54 (page 193) we can see that GFN2-xTB predicts these species with a larger (in absolute value) relative energy per bond than average, contrary to the DFT results. This, combined with the overall ratios of aldehyde and amine, **2:3**, versus the ratios at the **[1+3]**$^{3}$ intermediate, **1:3**, lead to an effectively irreversible consumption of the amine monomer, and that discontinuity appears when the amine monomer is completely depleted.

If we compare these results with previous results for the batch synthesis we can see that the "thermodynamic sink behavior" has shifted from the **[3+4]**$^{8}$ and **[3+5]**$^{9}$ to the early imine intermediates, but both are still observable. Although it is harder to appreciate it in the model with barriers of 10 kcal mol$^{-1}$, we can see it clearly in the model with barriers of 15 kcal mol$^{-1}$ that we have the same mechanistic interpretation as we initially had. After the formation of the first imine intermediates, they grow to **[2+m]**$^{i}$ species and as these form they immediately generate the macrocycles of **[2+2]**. These then grow until a **[3+3]** ring closure is possible forming the **[3+4]**$^{8}$ and **[3+5]**$^{9}$ that will end up arriving to the **[4+6]a** cage.

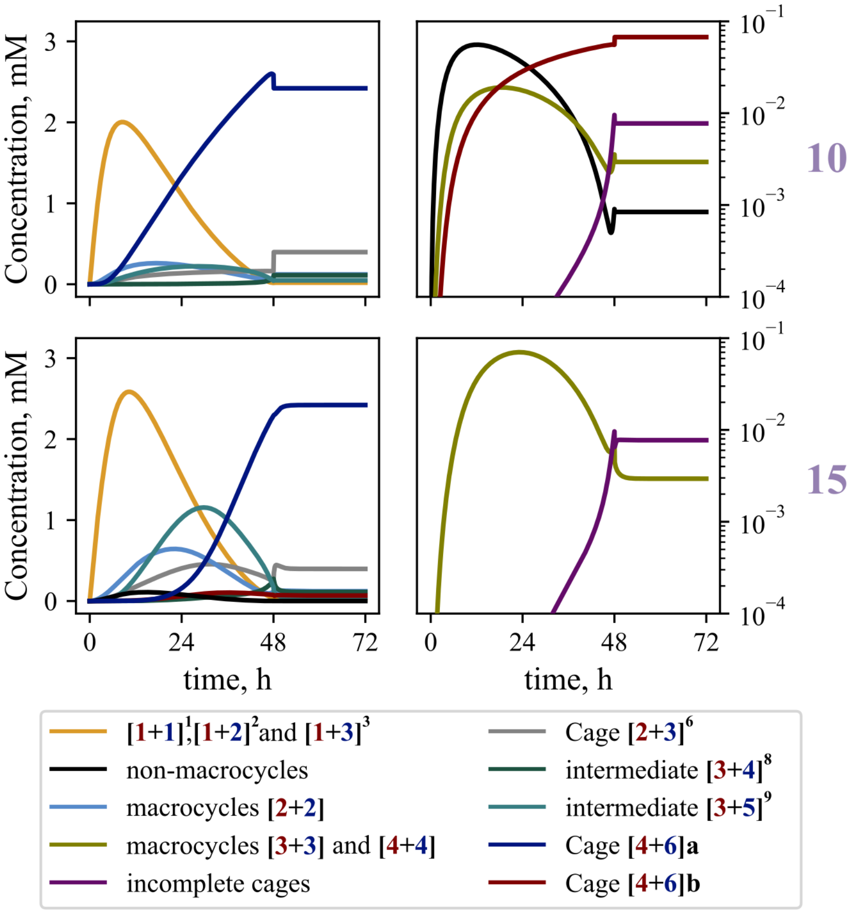


*Figure 71: Detailed results of the kinetic models of the batch synthesis using the scaled GFN2-xTB energies with a barrier of 10 and 15 kcal mol$^{-1}$.*

A similar overall mechanism is observed in the flow synthesis (Figure 72) with the major difference of the accumulation of the **[2+3]**$^{6}$ cage. The reason behind this difference lays on the synthetic procedure. In the batch models, due to the synthetic procedure, we drive the reaction through concentration effects to form the **[1+3]**$^{3}$ intermediate. On the other hand in the flow reactor, due to the stoichiometry and concentrations we let the the **[1+1]**$^{1}$ and **[1+2]**$^{2}$ intermediates accumulate.

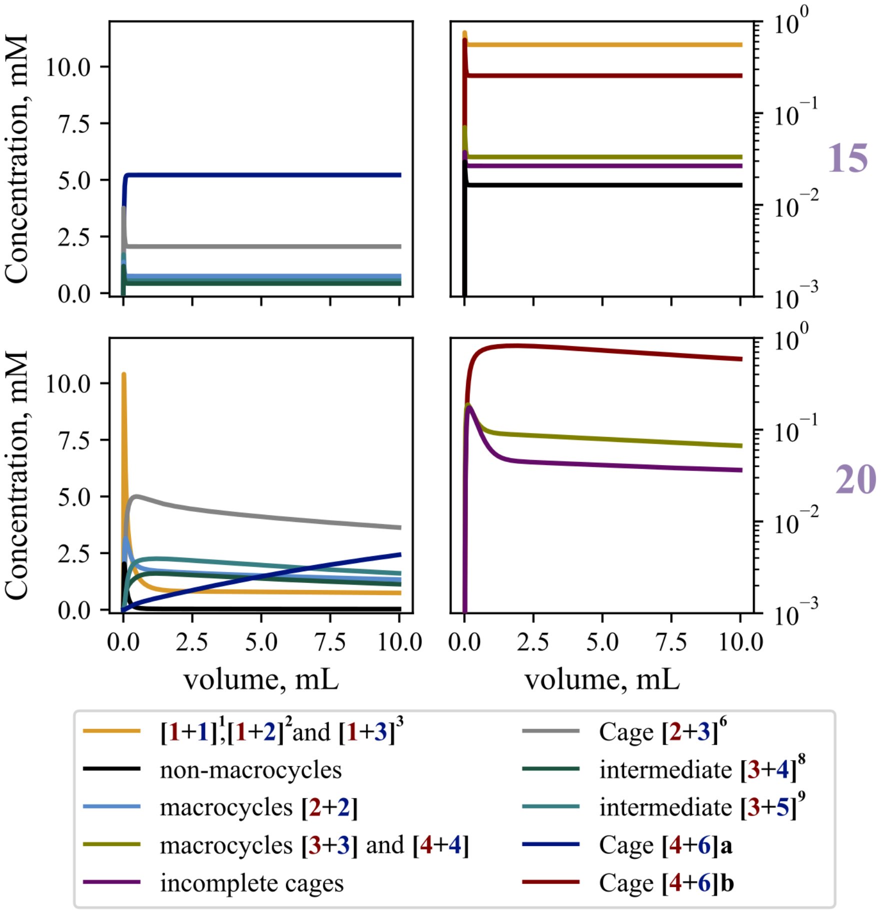


*Figure 72: Detailed results of the kinetic models of the flow synthesis using the scaled GFN2-xTB energies with a barrier of 15 and 20 kcal mol$^{-1}$.*

In the batch models, the most likely compound to react with **[1+3]**$^{3}$ will be the **[1+2]**$^{2}$ intermediate, which then will form a non-macrocyclic intermediate that can undergo a **[2+2]** ring closure. The result of that closure will be the

**[2+4]**$^6$ macrocycle which is unable to directly close to the **[2+3]**$^6$ cage. On the other hand, in the flow reactor, we favor the formation of the **[1+2]**$^2$ intermediate and, due to concentrations, it will be more likely to react with the **[1+1]**$^1$ than with itself or the **[1+3]**$^3$ intermediate. As a consequence it is more likely to reach the **[2+3]**$^5$ macrocycle before arriving to the **[2+4]**$^6$ which shortens the pathway leading to the **[2+3]**$^6$ cage.

## 9. Model 6, Finding the difference between GFN2-xTB and DFT

We could settle with the previous model, which yields the correct cage in both syntheses with an overall mechanistic picture more compatible with the current experimental knowledge on imine chemistry and imine cages, but there are very important questions to answer: Why does the model for the batch synthesis with the scaled GFN2-xTB energies (with the same average relative energy per imine bond) yield the CC1 cage with higher barriers than the respective DFT models? Is it due to the temperature? Was it a coincidence? How sensitive is the scaling? Can we improve the agreement with the experimental yields?

As it is the leitmotif of the current chapter we answered these questions using, yet again, kinetic models. First, we tested the effect of the scaling factor. What happens if instead of enforcing to match the average relative energy per bond we scale the values to match the relative free energy of the first monomer condensation (**[1+1]**$^1$ or Cage46a_000000000001)? What if instead we match the relative energy of the CC1 imine cage? These scalings are illustrated in Figure 73 and the results of the kinetic simulations can be found in Figure 74.

When we scale it to match the imine **[1+1]**$^1$ we observe, in both syntheses, that the overall reaction speeds up, at the cost of sacrificing the final yield of the cage. The batch synthesis models with barriers of 15 kcal mol$^{-1}$ are still far from the reported yield, but we see that for the models with barriers of 10 kcal mol$^{-1}$ adjusted to match the relative energy of the CC1 cage get yields around 80% which are significantly closer to the experimentally reported ones *ca.* 90%.

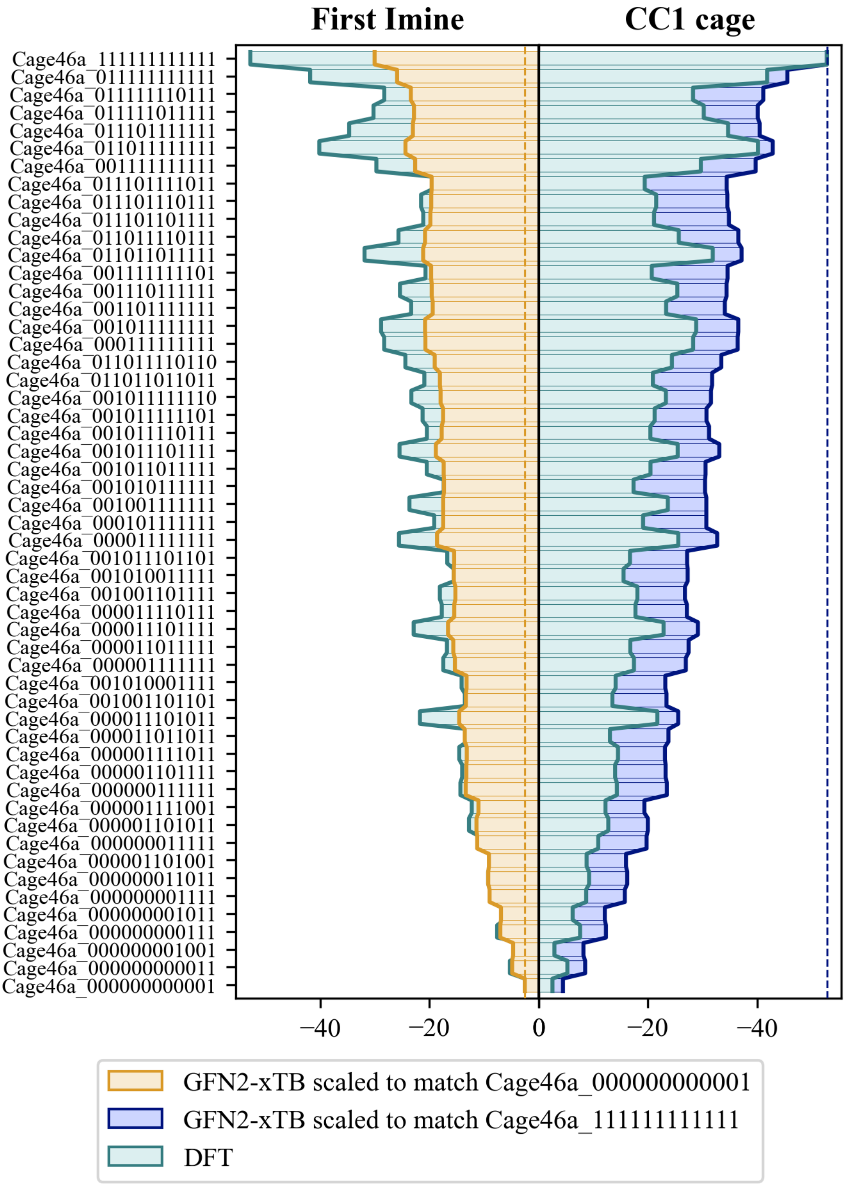


*Figure 73: Scaled subset of relative energies used in the kinetics. Energies in kcal mol. Dashed lines mark the value that was enforced to match with the DFT values.*

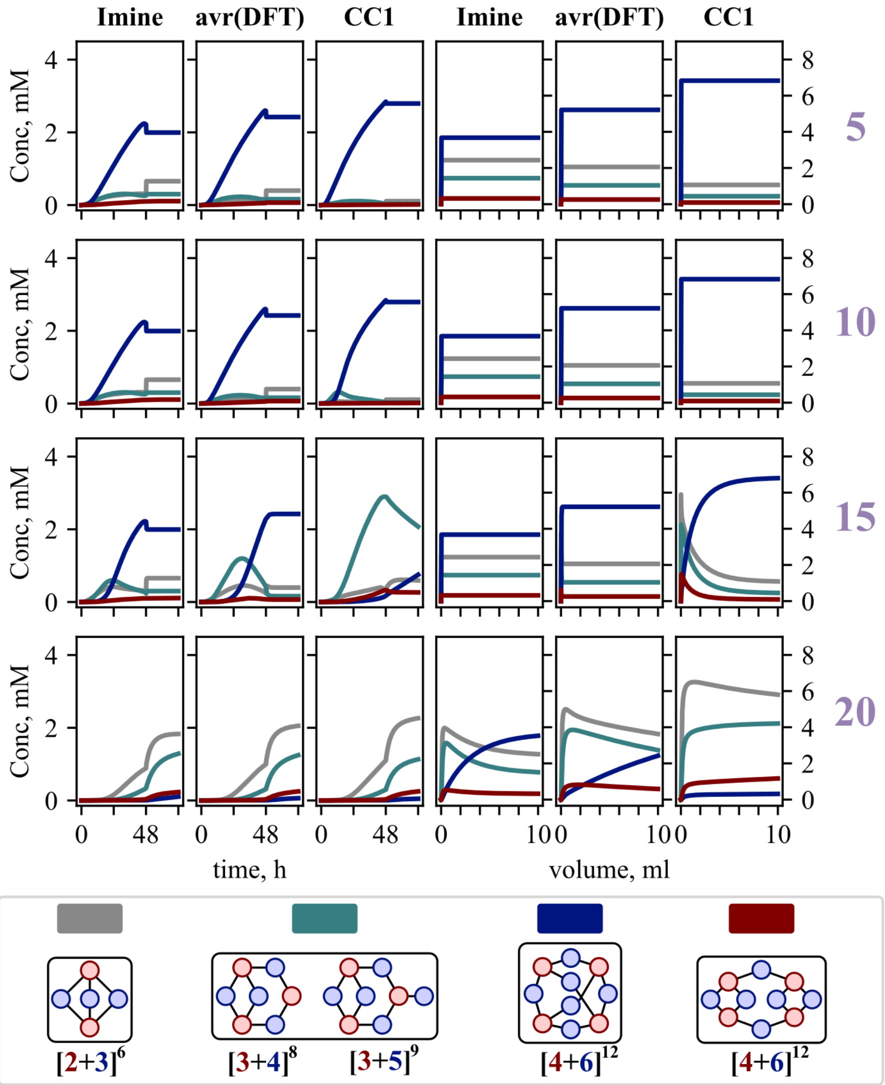


*Figure 74: Kinetic models with different scalings for the XTB energies. Numbers in purple are the barrier used for the kinetic models of the same row in kcal mol$^{-1}$.*

What we can obtain from this is that our computed DFT energy of the CC1 may be reasonable. As a consequence we wonder if it is possible to improve our DFT based models, by correcting the energies of a reduced set of compounds.

To see the differences between both sets of relative energies (DFT *vs* the scaled GFN2-xTB matching the DFT average relative energy per bond) we proceeded to plot them side by side, sorted by the difference between them (see Figure 75). From this image we formulated two hypotheses:

h1) The cause of the differing kinetic behavior is in the species that DFT understabilizes with respect to the scaled GFN2-xTB.

h2) The cause of the differing kinetic behavior is in the species that DFT overstabilizes with respect to the scaled GFN2-xTB.

Then we proceeded to test these hypotheses for models with barriers of 10 and 15 kcal mol$^{-1}$ (see Figure 76). To test the hypotheses we constructed the models using the relative energies of the affected species, for **h2** we de-stabilized the species to match the GFN2-xTB (species in the upper right side of Figure 75) and for **h1** we stabilized the affected species (lower left side of Figure 75).

Looking at the simulations (Figure 76), we can see for **h1** an increase in the yield of both synthesis with barriers of 10 kcal mol$^{-1}$ and a faster formation of the **[3+4]**$^{8}$ and **[3+5]**$^{9}$ intermediates for the model with 15 kcal mol$^{-1}$ barriers without formation of the CC1 cage. On the other side, in the models of **h2** with barriers of 15 kcal mol$^{-1}$ we see how the Cage CC1 becomes the major product for both reactor models.

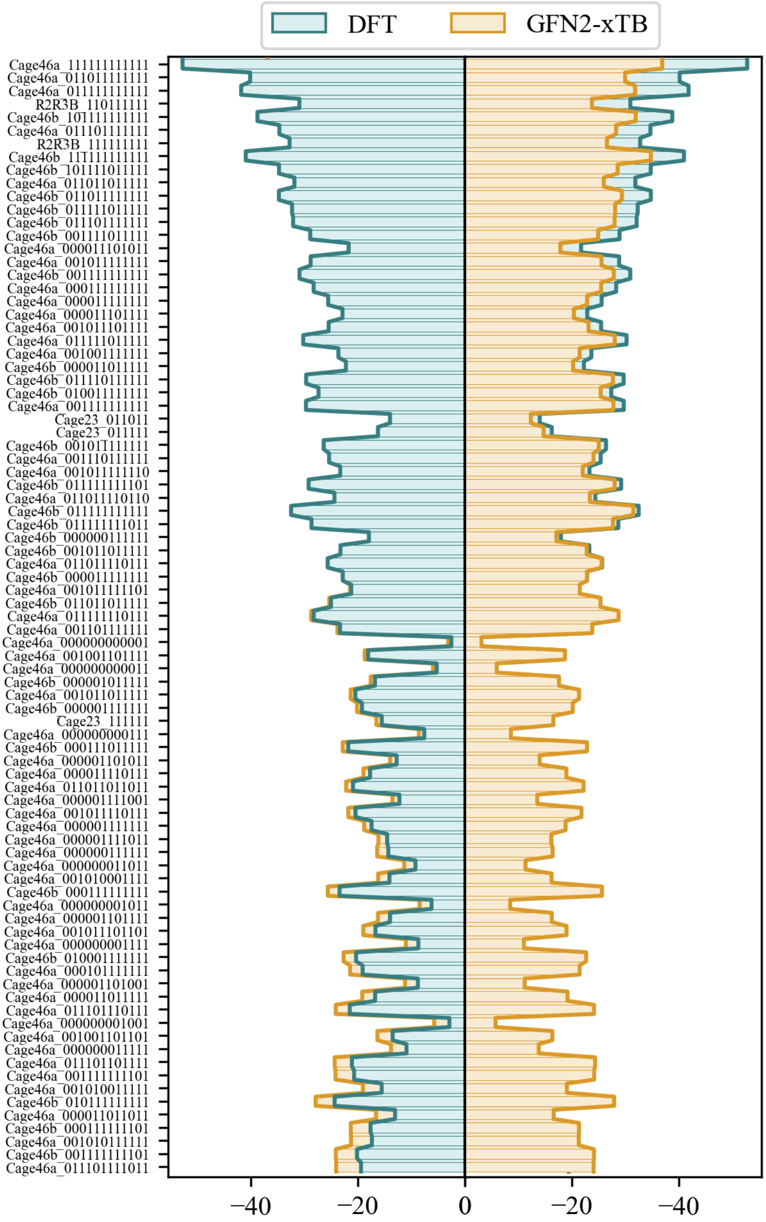


*Figure 75: Relative free energies of all species for DFT and XTB scaled to match the average relative free energy per bond of DFT.*

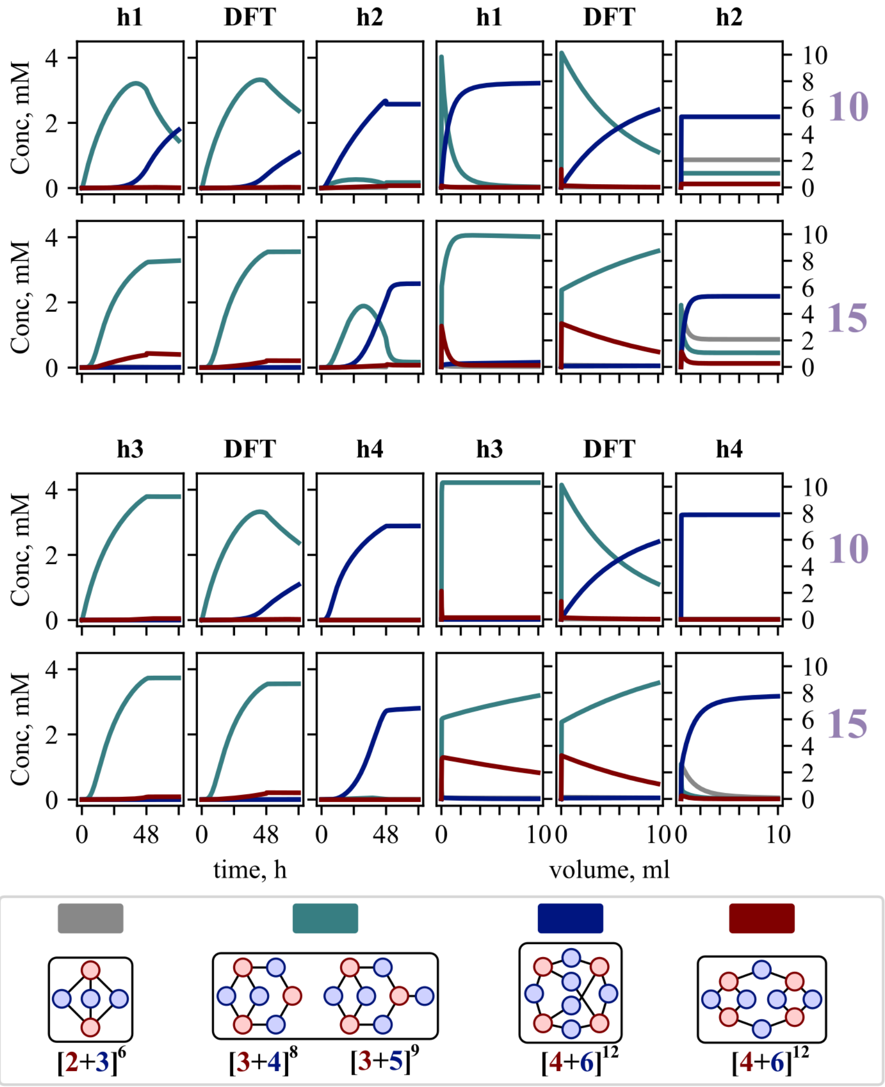


*Figure 76: Effect of the different hypotheses, h1-h4, on the kinetic model based on the energies of DFT.*

With all the knowledge gathered in previous sections and the results from **h1** and **h2** we attempted to further reduce the set of species that might be the cause of this differing behavior. We reduced the set to 5 key species that were amongst the species modified in **h2** that we considered more likely to influence the kinetics:

The **[3+4]**$^{8}$ intermediate (R2R3B_110111111), the **[3+5]**$^{9}$ intermediate (R2R3B_111111111) the **[4+6]b** cage and two of its previous intermediates (Cage46b_101111111111 and Cage46b_101111011111). We then formulated and tested the following two hypotheses.

h3) The cause of the differing kinetic behavior is in the species that DFT overstabilizes that do not belong to this reduced set.

h4) The cause of the differing kinetic behavior is in the species belonging to this reduced set of overstabilized species by DFT.

As we can see in Figure 76, as we destabilize the species of **h3** we end up increasing the gap between them and the reduced set of species. As a consequence. the concentration profiles show very little variation. However, when we destabilize the reduced set of species (**h4**) we observe an increase in yield, similar to the models of **h2** but without sacrificing the overall equilibrium of the CC1 imine cage.

## 10. Summary

In this chapter we have applied the algorithm developed in Chapter III to calculate all relevant species involved in the synthesis of the CC1 imine cage as well as to enumerate all possible reactions connecting those species. With that information we have been able to set up kinetic models to simulate the syntheses of the CC1 cage in two different reactor setups.

We started by analyzing the effect of the reactor setup on the concentration profiles under the same temperature conditions. Afterwards, we made the models more realistic by including the experimental temperatures. We observed very little change between the batch synthesis models as we included the temperature but we saw a more pronounced effect in the flow synthesis models that we were able to rationalize.

Next, we set up kinetic models to observe the effect of the magnitude of the average energy per imine bond which not only may be relevant for the selection of the appropriate computational methodology of future studies but will also allow to rationalize the possible mechanistic differences between cages of different monomers. Following those results we proceeded to test the effect of changing the distribution of the relative energies by setting up models based on the semi-empirical energies (and their scaled versions) instead of DFT energies.

Finally, as the DFT model corrected with some semi-empirical energies was able to qualitatively explain both syntheses whereas the pure DFT based ones were unable to account for the batch synthesis. We dug in search of the their differences and found that by merely correcting the DFT energies of 5 species the model has better agreement with experiments.

In all the models the same mechanistic interpretation (Figure 77) was found and the energy of several species was identified as critical for the reaction outcome.

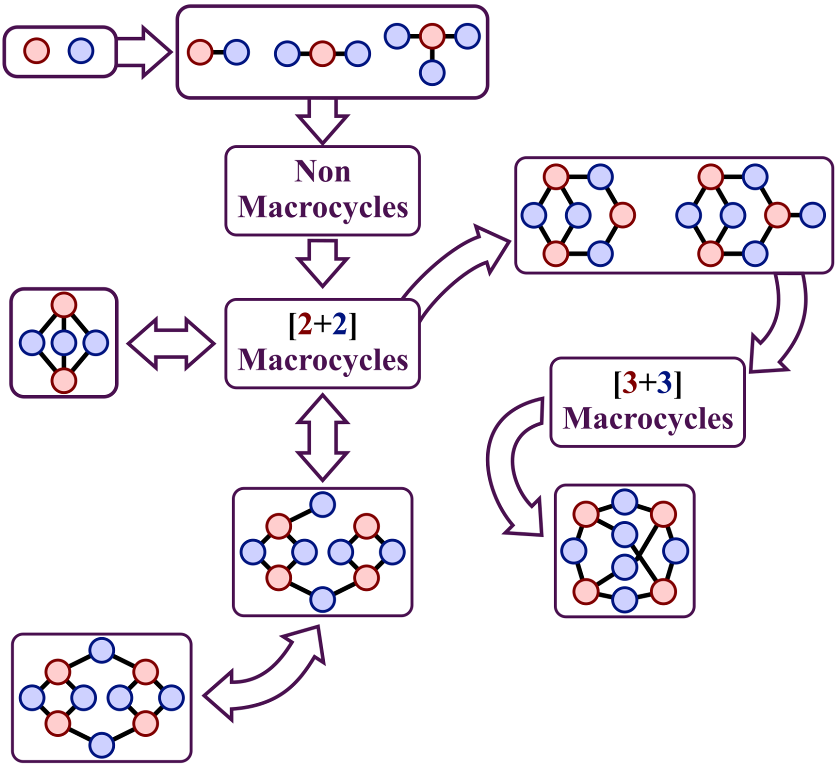


*Figure 77: Scheme summarizing the mechanism derived from the kinetic models.*

In the early stages the two monomers condense together forming three species ( **[1+1]**[1], **[1+2]**[2] and **[1+3]**[3] ) whose ratio depends on the reaction setup and whose relative stability (and consequent accumulation or not during the synthesis) becomes key to access the different main cages. After these species form, some short lived non-macrocyclic species appear, which then undergo a ring closure to form macrocycles with the **[2+2]** motiv. Depending on the relative stability of these compounds as well as the relative stabilities of **[2+3]**$^{6}$ and **[4+6]b** cage these macrocycles will form either of those two cages. Another two key species will be the **[3+4]**$^{8}$ and **[3+5]**$^{9}$ (containing a **[2+2]** and **[3+3]** macrocycle) that seem to be the major gateway for the formation of the **[4+6]a** cage (CC1).

We have also seen that increasing the speed of the imine bond formation reaction (similar to the effect of adding an acid catalyst) may not necessarily lead to the thermodynamic products, as sometimes concentrations coupled with the relative energies of some key compounds may lead to kinetic products instead.

Finally further work is needed in order to identify the minimal set of reactions and species capable of reproducing the results of the different syntheses. Some work will also be necessary to understand how the accumulation of water may affect the barriers of the reaction. And some work comparing simulated *vs* experimental concentration profiles will be necessary to discard or accept an "equal behavior" of all the imine condensation reactions that has been a design hypothesis of the current study.

# General Conclusions

In this thesis we have improved the current understanding of the self-assembly of organic cage molecules through a systematic approach, setting a computational precedent for future mechanistic studies on the self-assembly of these compounds.

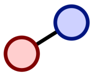
In Chapter I we obtained detailed mechanistic knowledge on the imine condensation reaction that governs the self-assembly of imine cages. We benchmarked computational methods against experimental kinetic results. We found a computational methodology that not only matched qualitatively but also quantitatively the experimental results.

In Chapter II we found no mechanistic change between the model system of the previous chapter and the first monomer condensation reaction of the self-assembly of the CC1 imine cage. Based on that result, we discarded the diaza-Cope rearrangement as a competitive side reaction. We also found that cage closures and ring closures are favored over chain growth reactions. We also found, in agreement with previous observations, that there is a kinetic preference for the condensation of the amine monomer over the aldehyde monomer in chain growth reactions.

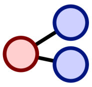
In Chapter III we designed an algorithm of tunable cost/accuracy for the enumeration and generation of the intermediates as well as the enumeration of the reactions. This algorithm was successfully applied in Chapter IV for the extended study of the self-assembly of the CC1 imine cage.

In Chapter IV we elucidated a general pathway for the self-assembly of the CC1 imine cage. We identified which species are key for the prediction of the reaction outcome based on free energies, and described the effects of temperature and reactor setup on the general mechanism. Finally, we obtained a kinetic model based on quantum chemistry calculations that qualitatively matches the current experimental knowledge in two very different reactor setups.

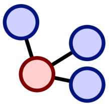
# Annex: List of Intermediates of the Self-Assembly of CC1



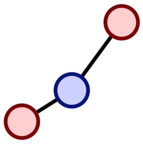
Cage46a_000000000001
[1+1]$^{1}$
46



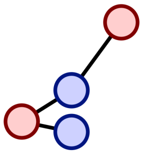
Cage46a_000000000011
[1+2]$^{2}$
13



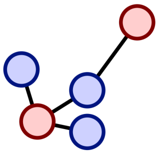
Cage46a_000000000111
[1+3]$^{3}$
41



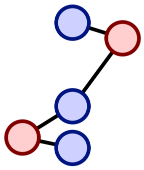
Cage46a_000000001001
[2+1]$^{2}$
8



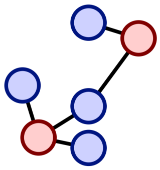
Cage46a_000000001011
[2+2]$^{3}$
33



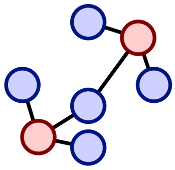
Cage46a_000000001111
[2+3]$^{4}$
31



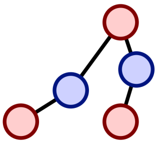
Cage46a_000000011011
[2+3]$^{4}$
54



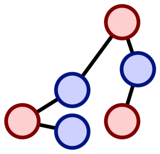
Cage46a_000000011111
[2+4]$^{5}$
42



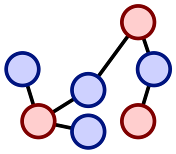
Cage46a_000000111111
[2+5]$^{6}$
34



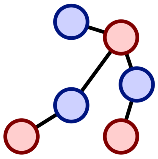
Cage46a_000001101001
[3+2]$^{4}$
9



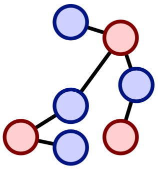
Cage46a_000001101011
[3+3]$^{5}$
26



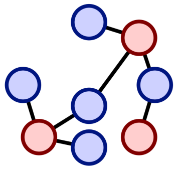
Cage46a_000001101111
[3+4]$^{6}$
19



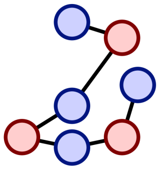
Cage46a_000001111001
[3+3]$^{5}$
50



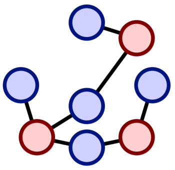
Cage46a_000001111011
[3+4]$^{6}$
45



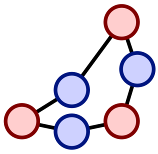
Cage46a_000001111111
[3+5]$^{7}$
18



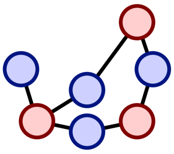
Cage46a_000011011011
[3+4]$^{6}$
28



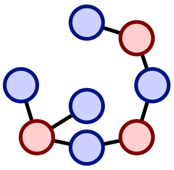
**Cage46a_000011011111**
[3+5]$^{7}$
39



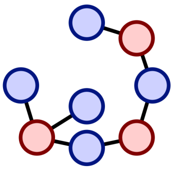
**Cage46a_000011101011**
[3+3]$^{6}$
2



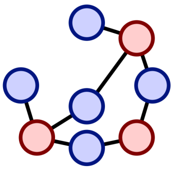
**Cage46a_000011101111**
[3+4]$^{7}$
24



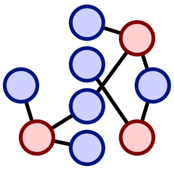
**Cage46a_000011110111**
[3+5]$^{7}$
47



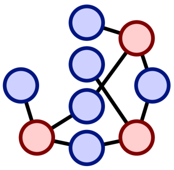
**Cage46a_000011110111**
[3+5]$^{7}$
47



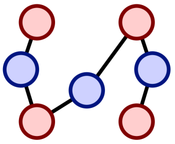
**Cage46a_000011111111**
[3+5]$^{8}$
14



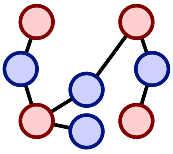
**Cage46a_000101111111**
[3+6]$^{8}$
5



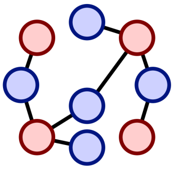
**Cage46a_000111111111**
[3+6]$^{9}$
23



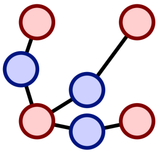
**Cage46a_001001101101**
[4+3]$^{6}$
29



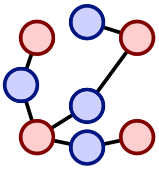
**Cage46a_001001101111**
[4+4]$^{7}$
51



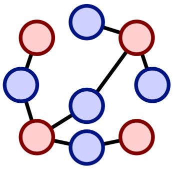
**Cage46a_001001111111**
[4+5]$^{8}$
10



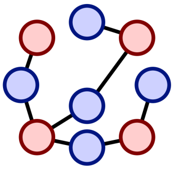
**Cage46a_001010001111**
[4+3]$^{6}$
37



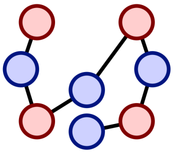
**Cage46a_001010011111**
[4+4]$^{7}$
36



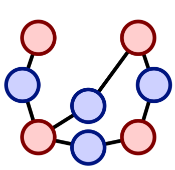
**Cage46a_001010111111**
[4+5]$^{8}$
53



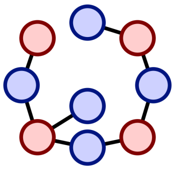
**Cage46a_001011011111**
[4+5]$^{8}$
16



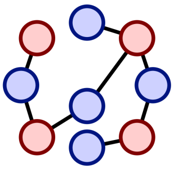
**Cage46a_001011101101**
[4+4]$^{7}$
22



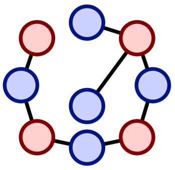
**Cage46a_001011101111**
[4+4]$^{8}$
40



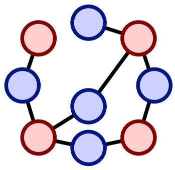
**Cage46a_001011110111**
[4+5]$^{8}$
25



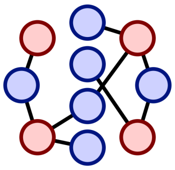
**Cage46a_001011111101**
[4+5]$^{8}$
17



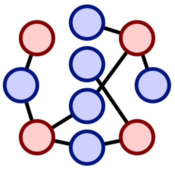
**Cage46a_001011111110**
[4+5]$^{8}$
38



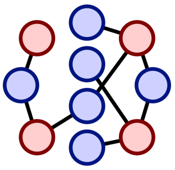
**Cage46a_001011111111**
[4+5]$^{9}$
48



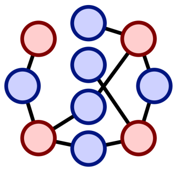
**Cage46a_001101111111**
[4+6]$^{9}$
30



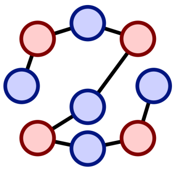
**Cage46a_001110111111**
[4+6]$^{9}$
27



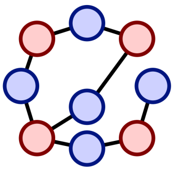
**Cage46a_001111111101**
[4+6]$^{9}$
49



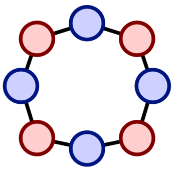
**Cage46a_001111111111**
[4+6]$^{10}$
35



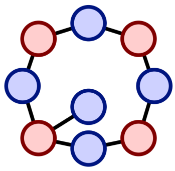
**Cage46a_011011011011**
[4+5]$^{8}$
20



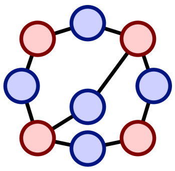
**Cage46a_011011011111**
[4+5]$^{9}$
12



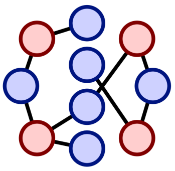
**Cage46a_011011110110**
[4+4]$^{8}$
21



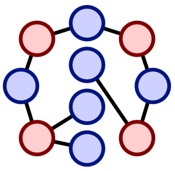
**Cage46a_011011110111**
[4+5]$^{9}$
43



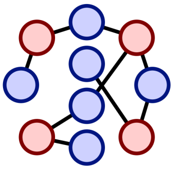
**Cage46a_011011111111**
[4+5]$^{10}$
11



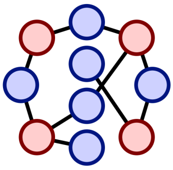
**Cage46a_011101101111**
[4+6]$^{9}$
32



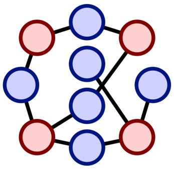
**Cage46a_011101110111**
[4+6]$^{9}$
15



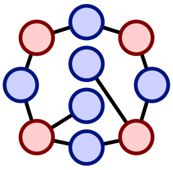
**Cage46a_011101111011**
$[4+6]^{9}$
**7**



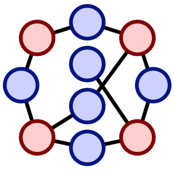
**Cage46a_011101111111**
$[4+6]^{10}$
**44**



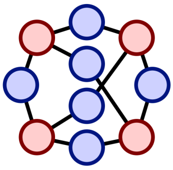
**Cage46a_011111011111**
$[4+6]^{10}$
**52**



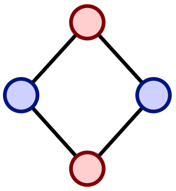
**Cage46a_011111110111**
$[4+6]^{10}$
**3**



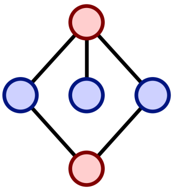
**Cage46a_011111111111**
$[4+6]^{11}$
**6**



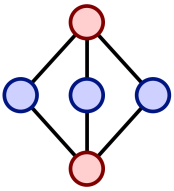
**Cage46a_111111111111**
$[4+6]^{12}$
**4**



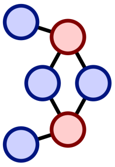
**Cage23_011011**
$[2+2]^{4}$
**56**



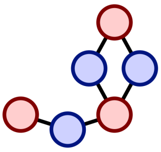
**Cage23_011111**
$[2+3]^{5}$
**55**



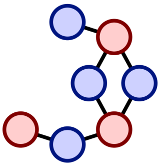
**Cage23_111111**
$[2+3]^{6}$
**57**



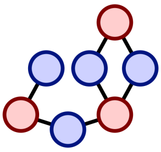
**Cage46b_000000111111**
$[2+4]^{6}$
**70**



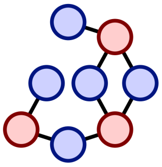
**Cage46b_000001011111**
$[3+3]^{6}$
**77**



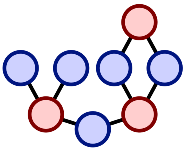
**Cage46b_000001111111**
$[3+4]^{7}$
**66**



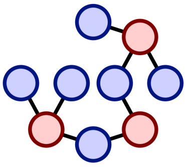
**Cage46b_000011011111**
$[3+4]^{7}$
**73**



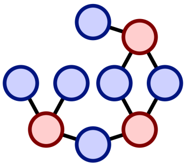
**Cage46b_000011111111**
$[3+5]^{8}$
**64**

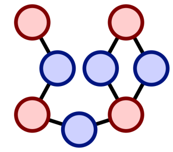


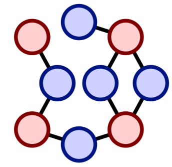
**Cage46b_000111011111**
$[3+5]^{8}$
**58**

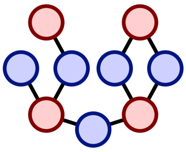


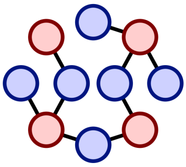
**Cage46b_000111111101**
$[3+6]^{8}$
**78**

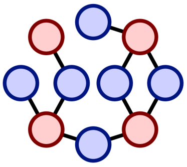

**Cage46b_000111111111**
$[3+6]^{9}$
**68**

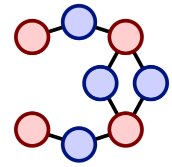

**Cage46b_001011011111**
$[4+4]^{8}$
**65**

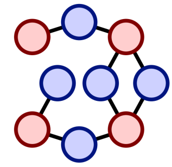

**Cage46b_001011111111**
$[4+5]^{9}$
**80**

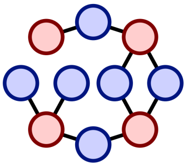

**Cage46b_001111011111**
$[4+5]^{9}$
**74**

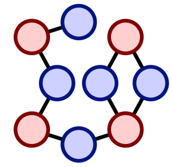

**Cage46b_001111111101**
$[4+6]^{9}$
**82**

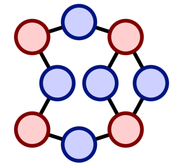

**Cage46b_001111111111**
$[4+6]^{10}$
**71**

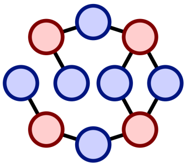

**Cage46b_010001111111**
$[4+4]^{8}$
**61**

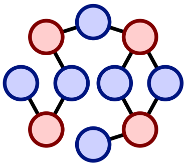

**Cage46b_010011111111**
$[4+5]^{9}$
**76**

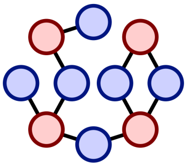

**Cage46b_010111111111**
$[4+6]^{10}$
**84**

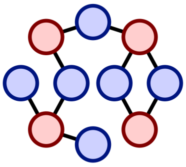

**Cage46b_011011011111**
$[4+5]^{9}$
**63**

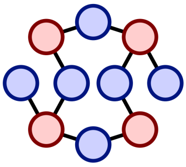

**Cage46b_011011111111**
$[4+5]^{10}$
**62**

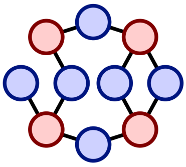

**Cage46b_011101111111**
$[4+6]^{10}$
**79**


**Cage46b_011110111111**
$[4+6]^{10}$
**67**

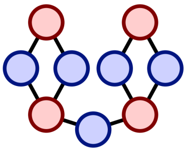

**Cage46b_011111011111**
$[4+6]^{10}$
**83**


**Cage46b_011111111011**
$[4+6]^{10}$
**81**

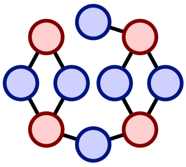

**Cage46b_011111111101**
$[4+6]^{10}$
**75**



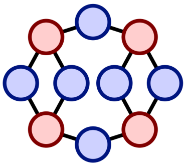
**Cage46b_011111111111**
$[4+6]^{11}$
**60**



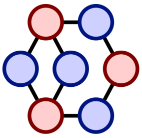
**Cage46b_101111011111**
$[4+5]^{10}$
**72**



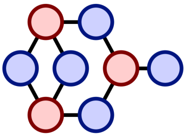
**Cage46b_101111111111**
$[4+6]^{11}$
**69**



**Cage46b_111111111111**
$[4+6]^{12}$
**59**



**R2R3B_110111111**
$[3+4]^{8}$
**86**



**R2R3B_111111111**
$[3+5]^{9}$
**85**