Theoretical Studies on Helical Polymer Modeling and Folding

Zur Erlangung des akademischen Grades einer
DOKTORIN DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

von der KIT-Fakultät für Chemie und Biowissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte
DISSERTATION
von

M.Sc. Montserrat Peñaloza-Amion
aus
Santiago, Chile

"What we observe is not nature itself,
but nature exposed to our method of questioning."

W. Heisenberg

"It is all a great mystery... Look up at the sky and
you'll see how everything changes."

Antoine de Saint-Exupéry

Acknowledgments

Chile is an *Aymara* word that means *where the land ends*. From this corner of the world, in South America I came to Germany to reach my dream of doing the PhD. This journey was not easy and it made me realise that leaving everything behind and pursuing my personal dreams is a brave adventure where I had to face fears, insecurities, solitude and failure, a lot of failure but at the end it reaped wonderful rewards and friends.

I am immensely grateful for this journey and the people I met. They participated and supported me during this 5 years in the completion of my PhD studies, contributed to my scientific knowledge and promoted my personal growth.

I would like to specially thank to Prof. Dr. Wolfgang Wenzel for accepting me in his research group and for his supervison and scientific support during my research stay at the Institute of Nanotechnology (INT) at KIT. To Prof. Dr. Marcus Elstner for making my PhD possible and for his scientific insights regarding the induction of induced helical polymers.

Many thanks to the Deutscher Akademischer Austauschdienst (DAAD) for their financial support under the scholarship "Research Grants - Doctoral Programmes in Germany, 2018/19" (91683960) and the project SFB 1176 "Molecular Structuring of Soft Matter" for their financial support and for giving me the opportunity to collaborate with excellent polymer scientists.

Thank you to Dr. Celso R. C. Rêgo, for his mentorship, team work and amazing ability to encourage me to not give up when the results were unexpected. To Clemens Possel and Joshua Futterer. They collaborated with me in the investigation of the folding of oligomers and they showed me a new world of possibilities teaching me about coding and Python. I enjoyed our incredible time working together solving the mysteries of folding and SIMONA.

I would like to thank my family. They have been the main influence of who I am today. To my mother, Paola Amion for her exceptional love, life coaching and for her intrinsic curiosity making her a natural scientist. To my grandparents Maria Angelica Ruiz and Enrique Amion, which supported and loved me. They taught me discipline, commitment and the most important love and perseverance, because: "where there is love the things are made in a perfect way".

To Ashley Smith, my partner, also my lockdown colleague. You were there when the world looked like it was collapsing and supported me in my darkest moments. Your cheerful attitude, your cornucopia of scientific skills and the dynamicity of your kindness gave me the most fun and love I ever had. Your are one of a kind, don't ever change what is good in you.

Thank you to all the members and old members of AG-Wenzel for their support during these years. In particular to Heiken Fliegl, Mariana Kozlowska, Anna Mauri, Modan Liu, Saientan Bag, Yohannes Pramudya, Ka Chun Chan, Srdan Pusara, Deniz Özdemir, Niklas Kappel, David Elsing, Manuel Rommel, Manuel Konrad, Simon Kaiser, Mersad

Abstract

The discovery of the DNA double-helix by Watson and Crick in 1953 was one of the most significant milestones in researching macromolecular structures. This discovery leads to the pursuit of many research lines and applications involving the folding, synthesis, and structural determination of helical structures. Today, it is widely known that helices play an important role in biomolecules, and there is much interest in using this motif in a broader range of applications, such as in bio-medicine and soft materials development. This thesis is based on the latter field and focuses on the design of helical polymers from a theoretical perspective.

Molecular simulation techniques such as Molecular Dynamics and Monte Carlo have already been extensively applied to the research of polymer structures. Frequently, these approaches are applied to helical structures with high inversion barriers or require experimental structural input to reproduce the measurements. In this thesis, we use purely theoretical approaches to understand the driving forces on the induction of dynamic helical polymers and model polymer helical conformations.

First, insights regarding the helical induction of poly 4-carboxyphenyl acetylene by chiral amines were investigated using DFT optimizations, DFT dihedral scans, and Molecular Dynamics. The DFT results show that the geometries of the local interaction of the chiral amines and the polymer carboxy groups play a major role in the induction process. Bader charge analysis reveals different charge transfers upon comparing R and S conformations interacting with clockwise and counterclockwise screw-sense polymers.

In the next piece of work, systematic protocols were created to model helical structures of post-modified poly phenylacetylenes by chiral amines for both para and meta-substituted side chains. Blind and local helical searches based on Monte Carlo simulations and DFT dihedral scans were utilized. The key stabilizing interactions were found to be H-bonds, $\pi - \pi$ stacking interactions, and $\pi - \pi$ T-shaped interactions.

In the last chapter, a *de novo* design of helical structures was carried out using sequences-controlled oligomers. Annealing Monte Carlo simulations were performed, and the helical design was guided by Hydrogen Bond cluster analysis. A variety of classes of sequences were considered with different lengths of backbones and side chain modifications. The best helical sequences were simulated with Molecular Dynamics with an explicit solvent to test their stability.

Kurzfassung

Die Entdeckung der DNA-Doppelhelix bei Watson und Crick im Jahr 1953 war einer der wichtigsten Meilensteine in der Erforschung makromolekularer Strukturen. Diese Entdeckung führte zur Verfolgung zahlreicher Forschungslinien und Anwendungen, die sich mit der Faltung, Synthese und Strukturbestimmung von Helixstrukturen befassen. Heute ist weithin bekannt, dass Helices eine wichtige Rolle in Biomolekülen spielen, und es besteht ein großes Interesse an der Nutzung dieses Motivs in einem breiteren Spektrum von Anwendungen, z. B. in der Biomedizin und der Entwicklung von Soft Materials. Die vorliegende Arbeit basiert auf dem letztgenannten Bereich und konzentriert sich auf das Design helikaler Polymere aus theoretischer Sicht.

Molekulare Simulationstechniken wie Molekulardynamik und Monte Carlo wurden bereits in großem Umfang zur Erforschung von Polymerstrukturen eingesetzt. Häufig werden diese Ansätze auf helikale Strukturen mit hohen Inversionsbarrieren angewandt oder erfordern experimentellen Struktureinsatz, um die Messungen zu reproduzieren. In dieser Arbeit verwenden wir rein theoretische Ansätze, um die treibenden Kräfte bei der Induktion dynamischer helikaler Polymere zu verstehen und helikale Konformationen von Polymeren zu modellieren.

Zunächst wurden Erkenntnisse über die helikale Induktion von Poly-4-carboxyphenylacetylen durch chirale Amine mithilfe von DFT-Optimierungen, DFT-Dihedral-Scans und Molekulardynamik untersucht. Die DFT-Ergebnisse zeigen, dass die Geometrien der lokalen Wechselwirkung zwischen den chiralen Aminen und den Carboxygruppen des Polymers eine wichtige Rolle beim Induktionsprozess spielen. Die Bader-Ladungsanalyse zeigt unterschiedliche Ladungstransfers beim Vergleich von R- und S-Konformationen, die mit rechts- und linksdrehenden schraubenförmigen Polymeren.

In der nächsten Arbeit wurden systematische Protokolle erstellt, um helikale Strukturen von post-modifizierten Polyphenylacetylenen durch chirale Amine zu modellieren. Dies wurde sowohl für para- als auch für meta-substituierte Seitenketten durchgeführt. Es wurden blinde und lokale Helix-Suchen auf der Grundlage von Monte-Carlo-Simulationen und DFT-Dihedral-Scans verwendet. Als wichtigste stabilisierende Wechselwirkungen erwiesen sich H-Bindungen, $\pi - \pi$ Stapelwechselwirkungen und $\pi - \pi$ T-förmige Wechselwirkungen.

Im letzten Kapitel wurde ein *de novo* Design von helikalen Strukturen unter Verwendung von sequenzgesteuerten Oligomeren durchgeführt. Es wurden Monte-Carlo-Simulationen durchgeführt, und das helikale Design wurde durch eine Wasserstoffbrücken-Clusteranalyse geleitet. Es wurden verschiedene Klassen von Sequenzen mit unterschiedlichen Längen der Rückgrate und Seitenkettenmodifikationen berücksichtigt. Die besten helikalen Sequenzen wurden mit Molekulardynamik mit einem expliziten Lösungsmittel simuliert, um ihre Stabilität zu testen.

Contents

1   Introduction

1.1   Polymers and Oligomers

Around 200 million tons of polymers are produced annually in the world and they are a major building block in the design of materials across a multitude of different industries including bio-medicine, semiconductor production, aerospace, and construction. Polymer structures can be subdivided by repeated units known as **monomers**. The assembly of a small number of monomer units (two to low double digits) is called **sequence** and they are typically called **oligomers**. Polymers are characterized by a high molecular weight and the dominant constituent elements include: Carbon, Hydrogen, Oxygen, Nitrogen, Chloride, Fluoride, Sulfur, Phosphorous, and Silicon [1].

Polymers can be classified into different categories (Figure 1.1) such as origin (natural, artificial and synthetic), applications (large-scale, technical, functional) and structure (linear, two-dimensional and three-dimensional). Additionally, independent of the polymer dimension and topology they can be classified as homopolymers and copolymers according to their molecular structure. They are also commonly classified according their physical properties, where Thermoplastic polymers are known for their low glass transition temperature and possess a semi-crystaline or amorphous structure while Thermosetting polymers are highly resistant to high temperatures and have a rigid structure [2].
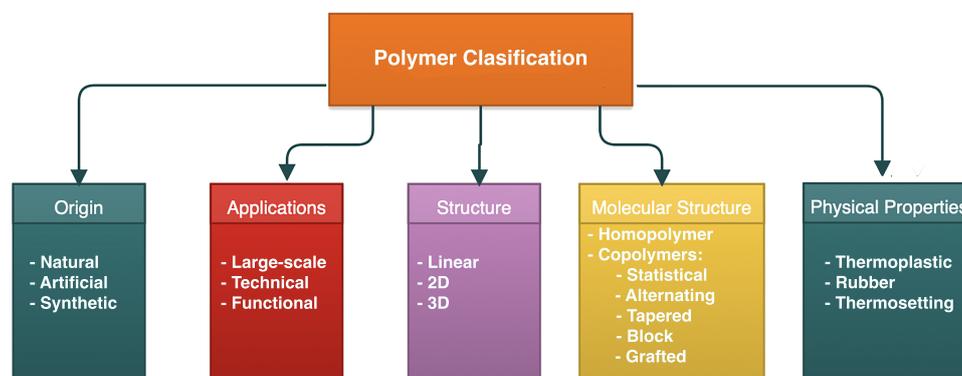
Figure 1.1: Scheme of polymer classification

The size of a polymer can be described by their **degree of polymerization (DP)** which is equal to the number of monomer units of the polymer chain and it is proportional to the molar mass of the polymer. The **molecular weight of a polymer** is the sum of the atomic weights of individual atoms that are part of the polymer. Values of polymer molecular weight are useful to know for experimentalists as they represent the average length of polymer chains in polymer resins. Naturally, not all the polymer chains have the same size and molecular weight there are two important quantities that reflect molecular weight of polymers: number-average molecular weight (M_n)

$$M_n = \frac{\sum_i N_i M_i}{\sum_i N_i} \tag{1.1}$$

and weight average molecular weight(M_w).

$$M_w = \frac{\sum_i N_i M_i^2}{\sum_i N_i} \tag{1.2}$$

In the field of polymer science, there are three main broad areas of study: (i) polymer synthesis, where reactions are studied in order to control the structural selectivity of polymers (regio-, chemo- and stereoselectivity). (ii) Physical properties of polymers such as elasticity and viscosity, and (iii) assembly of polymer chains, which is governed by polymer structure, morphology and self-organization [3, 4]. This thesis is focused on the latter area of study.

1.2   Helical Polymers

A helix is a common structure that can be found in bio-macromolecules such as DNA and proteins. The discovery of α-helix structure [5], DNA double-helix by Watson and Crick [6] have been hallmarks in the elucidation of biomolecular structure, which we know now to be intimately connected to biological function. The fact that helical structures play such an important role in nature suggests that the transfer of this molecular scaffold to synthetic polymers could lead to the discovery of a wealth of new materials.

Some helical polymers have already been synthesized, for instance, helical polypropylene [7] and other helical vinyl polymers [8] were the start of many efforts to study and understand helical conformation in polymers. Helical polymers have high potential for application in materials science in the fields of ferroelectric liquid crystals, nonlinear optical materials, sensing specific molecules, separations of enantiomers and asymmetric catalysis [9]. To date helical polymers such as polyacrylates, polyaldehydes, poly(arylvinyl)s, polyisocyanides, polyisocyanates, polyacetylenes and polysilanes have been synthesized successfully [10].
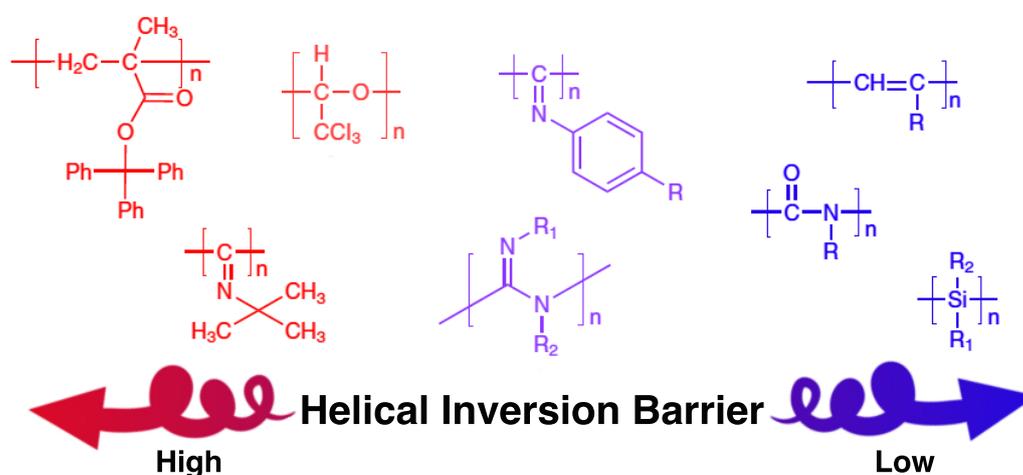
Figure 1.2: Scheme of helical polymers regarding their inversion barrier.

2

The main feature of helical polymers is their intrinsic chirality. Artificial helical polymers synthesis aims to develop chiral functional materials through the control of the screw-sense helicity. Certain helical polymers tend not to maintain a left or right handed conformation but show helix reversals. Polymers like poly(triphenylmethyl metacrylate) and polyisocyanates can be classified according to their high or low helix inversion barrier (Figure 1.2), respectively [11].

1.3   Induced Helical Polymers

Frequently, polymer structural properties can be controlled by polymer synthesis methods and internal structure that can influence the folding of linear polymer chains into helical conformations. Some synthesis strategies are based on chiral monomers in the polymer backbone, while others use chiral side chains [12]. Adding chiral molecules in the solvent, such as chiral amines is a particularly interesting way to promote the formation of a right or left handed helix in an otherwise non-chiral polymer [13–18].

Induced helicity in polymers is a particularly interesting topic where the helicity is transferred via non-covalent interactions of non-optical polymers and chiral compounds. The first polymer showing non-covalent specific chiral acid-base interactions displaying helicity was *cis-transoidal* stereoregular poly(phenyl)acetylene bearing carboxy groups [13]. Later, other poly(phenyl)acetylene derivatives bearing phosphonate [19, 20], amino [14, 16, 21], and borate groups [15, 22] were discovered, which respond to chiral amines, ammoniums, acids, diols and sugars; showing a characteristic induced circular dichroism and specific helicity preferences, respectively. On other hand, it has been observed that helicity can be induced in polyacetylenes [18, 23, 24], polyisocyanates [25–27], poly(phenyl isocyanide) [28] and polyguanidine [29] using chiral acid-base, host-guest, ionic, and hydrogen bonding interactions.
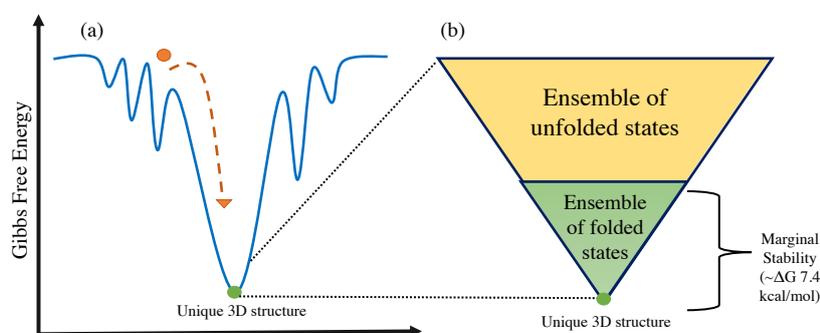
Figure 1.3: (a)Energy landscape of biopolymer folding. (b) Scheme of folded and unfolded states.

1.4   Macromolecular folding and helices

Since the most widely investigated macromolecule system are bio-molecules such as Proteins and DNA due to their early 3D structure determination in the 50's, the research on the folding mechanism appeared as a natural next question in this field. In 1969, Cyrus Levinthal stated that for a polypeptide chain with a large number of degrees of freedom would display an astronomical number of possible conformations [30, 31]. If we take a 100 aminoacid chain with 99 peptide bonds and thus, 198 ϕ and ψ dihedral angles this yields around 10^{300} possible conformations. On the one hand, the large amount of conformations can be correlated with the folding process taking a considerable amount of time on the atomic scale. On the other hand it is well known that folding of proteins takes place on the millisecond and microsecond scale. These two opposite ideas comprise the *Levinthal paradox* [32] and its solution the *folding funnel hypothesis* was proposed [33].

The folding funnel hypothesis states that proteins must have a unique three-dimensional structure where the folding process can be sped up by the formation of local interactions which will lead the protein to a unique low energy state or well-known as *native state*. The graphical representation of funneling is the energy landscape that a extended polypeptide chain must cross to reach the lowest energy conformation state (Figure 1.3 (a)). This principle is the current basis for all the computation approaches on protein structure prediction.

$$\Delta G = \Delta H - T\Delta S \qquad (1.3)$$

Similarly, in 1973 Anfinsen et. al. proposed that the unique three-dimensional structure of proteins was also the lowest free Gibbs energy structure (eq (1.3)). According to Anfinsen, the native state coexists in equilibrium with an ensemble of folded conformations with higher but comparable energy (Figure 1.3 (b)). Studying the folding of polypeptides and comparing them with the folding of ribonuclease they concluded that the native states is marginaly stable and is stabilized by weak interactions such as intra-molecule pairwise interactions and protein-solvent interactions[34]. Those weak but abundant interactions allows the coexistence of highly familiar folded states. For proteins the free energy difference is *ca.* 7.4 kcal/mol.

Figure 1.4: Scheme of classification of helix structures.

The literature on folding of polymers agree that the factors that could influence in the sta- bilization of folded biopolymers are: (i) Hydrogen bonds, (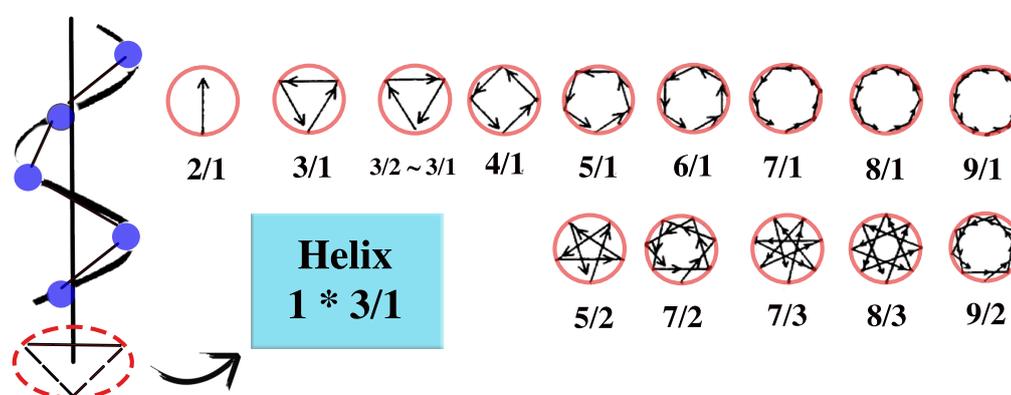ii) van der Waals interactions, (iii) Backbone dihedral angle preferences, (iv) electrostatic interactions, (v) hydropho- bic interactions with the solvent and (vi) chain entropy[35, 36]. When considering the structural hierarchy of proteins secondary structures such as α-helix are one of the most common and stable motifs found in bio-macromolecules. In broad terms, a helix is a result of the synchronization of the ϕ and ψ dihedral angles on proteins and stabilized by Hydrogen Bonds between the amide groups on the protein backbone. A helix structure can be defined as an arrangement of repeated monomer units per turn i.e. $A * \mu/l$; where A is the class according to A chain motifs of the helix with μ motifs per l repeating units (Figure 1.4) [37]. Since proteins and DNA can be thought of as polymers, all the efforts and research on folding of these macromolecules could provide a strong foundation for investigating helical polymers which will be the main focus of this thesis.

1.5   Methods on the structural characterization of polymers

Nowadays, there are many methods to analyze macromolecular structures, such as circu- lar dichroism (CD) spectroscopy, X-ray crystallography, scanning tunneling microscopy (STM), atomic force (AFM) microscopy [38–40] and Dynamic Ligth scattering[41]. These experimental techniques can sometimes struggle to determine 3D structures on some poly- mers, such as low molecular weight polymers or low inversion barrier helical polymers. Molecular modelling methods have been used to contribute to the structural characteriza- tion of helical structures including full-atom and coarse-grained models using conventional molecular dynamics (MD) and Monte Carlo (MC) simulations [42–45]. Christofferson et al. generated models based on reported bond distances, bond angles and internal rotation angles of poly-methylmetacrilate (PMMA) to create different double and triple 10/1 and 9/1 isotactic and syndiotactic helical structures. They performed Molecular Dynamics simulations of several composites of different stereocomplexes of PMMA and tried to re- late their results with X-ray diffraction profiles as blueprints. They confirmed that their all-atom models accurately reproduced their experimental X-ray diffraction profiles with triple helix stereocomplexes with the help of experimental input [46]. However, simulat- ing folding via atomic level simulations alone, without the help of experimental input, is numerically rather infeasible. No simulations of the formation of helices have been re- ported to date for specific polymers. For these reason, accelerated simulations techniques are needed, such as simulated annealing MC which allows to explore the folding reaction coordinate mapping the free energy landscape of the system.

There is currently a lack of theoretical works on the study of the helical induction on non- chiral polymers and the *de novo* design of sequence-controlled oligomers. In this thesis helical polymers are modelled using a variety of simulation methods such as Density Functional Theory (DFT), MD and MC to give insights into the influence of different coupling mechanisms and different schemes for rational design. We believe these efforts will help guide experimentalists in the right direction when synthesizing soft polymer materials.

5

2 Outline and scope

This thesis aims to explore the modeling of induced helical polymers with low inversion barrier by chiral molecules and perform *de novo* helical design of sequence-controlled oligomers.

The structure of this thesis is as follows:

Chapter 4: In this chapter helical structures based on the stereoregular *cis-transoid* polyacetylene with 4-carboxyphenyl bearing groups are modeled. The possible helical conformations this polymer could display are explored via DFT scan calculations and valuable insights regarding the induction mechanism by chiral amines are found using Bader net charge analysis.

Chapter 5: In this chapter the modeling of helical conformations on poly(phenylacetylene)s bearing activated ester moieties was performed. Helical models were investigated using Monte Carlo simulations using blind and local helical searches on polymers with para and meta substitutions and considering R and S chiral components.

Chapter 6: In this chapter *de novo* helical design of sequence-controlled oligomers with backbone and side chain monomer modifications was investigated. Monte Carlo and Molecular Dynamics simulations were employed on different oligomer sequences in order to find structural patterns that could lead to stable helix conformations. Hydrogen bond potential and Hydrogen Bond Cluster analysis were applied and provide valuable insights to propose candidate sequences for further rational polymer synthesis.

3 Methods

This chapter will introduce the main methods and tools applied to the projects within this thesis, such as Density Functional Theory (DFT), Monte Carlo (MC), and Molecular Dynamics (MD). All these methods complement each other when trying to balance the trade-off between accuracy and computational cost when determining the physical and electronic properties of atoms, molecules, and solids.

3.0.1 Schrödinger equation

First lets start with a recap of the time-independent nonrelativistic Schrödinger equation. (eq. (3.1)).

$$\hat{H}\Psi = E\Psi \tag{3.1}$$

$\hat{H}$ is the hamiltonian operator and Ψ is the wave function of the system, which is a set of solutions of the Hamiltonian (eigenstates). Each of the solutions $\Psi(n)$, has an associated eigenvalue $(E(n))$ that satisfies the eigenvalue equation. In a system with multiple electrons (N) interacting with multiple nuclei the problem can quickly become challenging to solve. In its more complete form the Schrödinger equation reads .(3.2) :

$$\left[\frac{h^2}{2m} \sum_{i=1}^{N} \nabla_i^2 + \sum_{i=1}^{N} V(\mathbf{r}_i) + \sum_{i=1}^{N} \sum_{j<i} U(\mathbf{r}_i, \mathbf{r}_j) \right] \Psi = E\Psi \tag{3.2}$$

when m is the electron mass. The terms inside the brackets correspond from left to right to: i) the kinetic energy of the electrons, ii) the interaction energy between the electrons and the nuclei iii) the interaction of each electron with the other surrounding electrons. We can solve this equation at different nuclear coordinates to determine how the energy of the system changes as the atom positions move. The Schrödinger equation only has analytical solutions for a small subset of systems such as the hydrogen atom, in most cases we need to make the well-known Born-Oppenheimer approximation which we discuss in the next section.

3.0.2 Born-Oppenheimer approximation

First lets start with a recap of the idea of a "Potential Energy Surface" (PES). If we have a molecule with M nuclei, their positions will be $\mathbf{R}_1, ..., \mathbf{R}_M$, and the ground-state energy (E) can be expressed as a function of the nuclear positions, which correspond to the "Potential Energy Surface" of the molecule (eq. (3.3)).

$$PES = E(\mathbf{R}_1, ..., \mathbf{R}_M) \tag{3.3}$$

To numerically solve the Schrödinger equation for a given energy state on the PES, we apply the Born-Oppenheimer (BO) approximation. In complete derivation, we find that the first and second-order non-adiabatic coupling terms have the respective form.

$$d_{jj'} = \frac{\left\langle \psi_j \left| \nabla_n \hat{H}_e \right| \psi_{j'} \right\rangle}{(E_j - E_{j'})} \tag{3.4}$$

For more detail explanation of the content discussed in section 3.1 see references [47–50].

and,

$$D_{jj'} = \nabla_n d_{jj'} + \sum_i d_{ji}^* d_{ij'} \qquad (3.5)$$

We can see that whenever E_j is close in value to $E_{j'}$ we can't rely on the BO approximation [51]. These terms are negligible in most cases. However, there are many systems for which these terms can yield a significant and non-negligible contribution, for example, carbon nano-tubes [52]. The result of making the BO approximation can be summarised by the following equations

$$\hat{H}_e \Psi_e(\mathbf{r}; \mathbf{R}) = E_e(\mathbf{R}) \Psi_e(\mathbf{r}; \mathbf{R}) \qquad (3.6)$$

$$[\hat{T}_N + E_e(\mathbf{R})] \Psi_N(\mathbf{R}) = E \Psi_N(\mathbf{R}) \qquad (3.7)$$

the electronic wave function $\Psi_e(\mathbf{r}; \mathbf{R})$ depends explicitly on the electronic coordinates $\mathbf{r}$ but only parametrically on the nuclear coordinates $\mathbf{R}$.

3.1 Density Functional Theory

Density Functional Theory (DFT) is one of the most widely used theories within theoretical chemistry because it enables us to find structures that minimize the energy locally at a fraction of the computational cost of canonical approaches such as Hartree-Fock and coupled cluster theory. Casting our minds back to standard quantum mechanics the wave function is not directly observable, but the probability is. When finding the probability density, we do not care about how the electrons are labeled, just that there is some probability of observing the system there at coordinate $(\mathbf{r}_1, ..., \mathbf{r}_N)$. A closely related quantity is the density of electrons at a particular position in space, $\rho(\mathbf{r})$.

$$\rho(\mathbf{r}) = \sum_i^N |\psi_i(\mathbf{r})|^2 \qquad (3.8)$$

Here, the summation goes over all the individual electron wave functions. The terms inside the summation are the probabilities that an electron in an individual wave function $\Psi(\mathbf{r})$ is located at position $\mathbf{r}$.

3.1.1 Hohenberg-Kohn theorem

The first theorem of Hohenberg and Kohn (H-K) states: *"The ground-state energy from the Schrödinger equation is a unique functional of the electron density"* (eq.(3.9)) [53, 54]. This means that the electron density determines all the ground-state properties of the system.

$$E = E[\rho(\mathbf{r})] \tag{3.9}$$

The second theorem of H-K states: *"The electron density that minimizes the energy of the overall functional is the true electron density corresponding to the full solution of the Schrödinger's equation"*[53, 54].

Conclusion: If an appropriate functional can be found we can solve the Schrödinger equation by finding a function of 3-space variables $\rho(\mathbf{r})$ instead of a the wavefunction which can be complex and multi-variable.

3.1.2 Kohn-Sham equations

Even though, H-K proved that the electron density can be used to solve Schrödinger equation, they didn't give details about identity of the functional. Then Kohn and Sham proposed the Kohn-Sham equations [55] :

$$E[\rho(\mathbf{r})] = E_{known}[\rho(\mathbf{r})] + E_{XC}[\rho(\mathbf{r})] \tag{3.10}$$

$$E_{known}[\psi_i] = \frac{h^2}{m}\sum_i \psi_i^* \nabla^2 \psi_i \mathbf{r} + \frac{e^2}{2}\int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}d^3\mathbf{r}' + \int V(\mathbf{r})\rho(\mathbf{r})d^3\mathbf{r} + E_{ion} \tag{3.11}$$

where $E_{XC}[\rho(\mathbf{r})]$ is a term which we must guess. It is the "Exchange-correlation functional" which can be split into the parts of the exchange and correlation which are only captured partially by $E_{known}[\rho(\mathbf{r})]$. The terms in $E_{known}[\rho(\mathbf{r})]$ are: i) the electron kinetic energies for a fictitious system of non-interacting particles, ii) the Hartree energy which accounts for the Coulomb repulsion between the electrons iii) a term corresponding to the electrons moving around in an effective potential $V(\mathbf{r})$ and iv) the nuclear nuclear term.

When these equations are minimised via a variational procedure [56] this leads to the Kohn-Sham equations

$$\left[\frac{h^2}{2m}\nabla^2 + V(\mathbf{r}) + e^2\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}\quad\mathbf{r}'|}d^3\mathbf{r}' + \frac{\delta E_{\mathrm{XC}}(\mathbf{r})}{\delta n(\mathbf{r})}\right]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}) \tag{3.12}$$

This is a set of N 3-dimensional Schrödinger-like equations which are easily solved whereas before we had a complex N-particle problem which were far harder to solve or perhaps not soluble.

3.1.3   Dispersion corrections

London dispersion forces are long ranged forces which result from small fluctuations in time of the electron density around the atoms. These fluctuations can be thought of as excitations to virtual orbitals. As KS-DFT is a ground-state it does not account dispersion. Simply adding a term E_{disp} accounts for the dispersion

$$E_{\mathrm{DFT}} = E_{\mathrm{KS-DFT}} + E_{\mathrm{disp}} \tag{3.13}$$

The fluctuations in the electronic density around one atom may induce instantaneous dioples and multipoles in nearby atoms. Imagining two molecules in a box at very far away distance they do not interact at all. Move them closer together and they can induce instantaneous multipoles in each-other and hence attract each other. When they are very close the interactions become repulsive. The dispersion forces can be expressed by Grimme's dispersion equation [57] (3.14):

$$E_{\mathrm{disp}} = -\frac{1}{2} \sum_{i=1}^{N_{at}} \sum_{j=1}^{N_{at}} \sum_{\mathbf{L}}{}' \left(f_{\mathrm{damp},6}(r_{ij,L}) \frac{C_{6ij}}{r_{ij,L}^6} + f_{\mathrm{damp},8}(r_{ij,L}) \frac{C_{8ij}}{r_{ij,L}^8} \right) \tag{3.14}$$

3.2 Molecular Dynamics

Molecular Dynamics (MD) is a simulation technique that uses statistical mechanics to estimate equilibrium and dynamics properties of systems. MD simulations are widely used to provide valuable information such as how molecular geometries and energies evolve over time. With sufficiently long simulation times the trajectories can be used to calculate average thermodynamic properties and give insights of how probable it is for the system to access thermally-accessible states. The trajectories are generated applying Newton's equations of motion and Force Fields. MD can also be coupled with experiments to observe the dynamic evolution of systems which can complement and enhance experimental findings.

3.2.1 Integration Schemes

Lets remind ourselves of Newton's equations of motion

$$m_i\mathbf{r}_i = \mathbf{F}_i = -\nabla_{\mathbf{r}_i} E(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad \forall i \in \{1, \dots, N\} \tag{3.15}$$

where N is the number of atoms. One of the most popular ways to solve the Newtonian equations of motion is the Verlet algorithm. To derive the Verlet algorithm it is necessary to start with the Taylor series of the position of a particle in time t and time step Δt (3.16).

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \frac{d\mathbf{r}(t)}{dt} + \frac{d^2\mathbf{r}(t)}{dt^2}\frac{\Delta t^2}{2!} + \frac{d^3\mathbf{r}(t)}{dt^3}\frac{\Delta t^3}{3!} + O(\Delta t^4)) \tag{3.16}$$

The second and third term of the series represent the velocity and the acceleration of the particle, respectively. According to the Newton's second law the second order derivative can be substituted by the force $F(t)$ (3.17).

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m}\Delta t^2 + \frac{d^3\mathbf{r}(t)}{dt^3}\frac{\Delta t^3}{3!} + O(\Delta t^4)) \tag{3.17}$$

And similarly the expression for the backwards step is

$$\mathbf{r}(t - \Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m}\Delta t^2 - \frac{d^3\mathbf{r}(t)}{dt^3}\frac{\Delta t^3}{3!} + O(\Delta t^4)) \tag{3.18}$$

If we sum equations (3.17) and (3.18) and we neglect the higher orders of the Taylor series ($O(\Delta t^4)$) we obtain:

$$\mathbf{r}(t + \Delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{\mathbf{F}(t)}{2m}\Delta t^2 \tag{3.19}$$

where m is the mass and $\mathbf{r}(t)$ is the current position of the particle, $\mathbf{r}(t+\Delta t)$ and $\mathbf{r}(t-\Delta t)$ are the next and previous step, respectively, and $\mathbf{F}(t)$ is the resulting force on the particle. This way the new position of the particle is evaluated in terms of coordinates and forces that are derived from the previous step.

For more detail explanation of the content discussed in section 3.2 see references [36, 58].

An adapted version of the Verlet algorithm is the Leapfrog algorithm. This method uses half-integer time steps on the velocities to calculate the new positions of the particles as is shown in eq. (3.23).

$$\mathbf{v}(t - \frac{\Delta t}{2}) \equiv \frac{\mathbf{r}(t) - \mathbf{r}(t - \Delta t)}{\Delta t} \tag{3.20}$$

$$\mathbf{v}(t + \frac{\Delta t}{2}) \equiv \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t} \tag{3.21}$$

$$\mathbf{v}(t + \frac{\Delta t}{2}) = \mathbf{v}(t - \frac{\Delta t}{2}) + \frac{\mathbf{F}(t)}{m}\Delta t \tag{3.22}$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t + \frac{\Delta t}{2})\Delta t \tag{3.23}$$

3.2.2   Force Field and truncated potentials

MD Force Fields (FF) are used to calculate the interactions between atoms. Generally, a FF is a collection of experimental and/or theoretical parameters which when combined correctly with the right numerical approaches allow us to obtain the potential energy of the system. There are a rich variety of different FFs, some popular names are AMBER, CHARMM or OPLS [59]. They differ in how the interactions are described or the source of their parameters. Commonly, the total potential function is split in bonded and non-bonded interaction parts.

$$E = E_{\text{Bonded}} + E_{\text{Non-bonded}} \tag{3.24}$$

Bonded interactions are described by bond, angle and dihedral potentials and non-bonded potentials include the van der Waals interactions described by the Lennard-Jones function and electrostatic interactions via the Coulomb potential. In this thesis, the General Amber Force Field (GAFF) [60] was employed due to its general applicability to a wide collection of atom types present in organic molecules.

$$E_{\text{GAFF}} = \sum_{\text{bonds}} K_r(\mathbf{r} - \mathbf{r}_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + $$
$$\sum_{dihedrals} \frac{V_n}{2}[1 + cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{\mathbf{R}^{12}} - \frac{B_{ij}}{\mathbf{R}_{ij}^6} + \frac{q_iq_j}{\epsilon\mathbf{R}}\right] \tag{3.25}$$

where $\mathbf{r}_0$ and θ_0 are equilibrium structural parameters and K_r, K_θ, V_n are force constants for bonds, angles and dihedrals respectively. γ and n are dihedral angle phase parameters. A_{ij} and B_{ij} are Lennard-Jones parameters and q_i and q_j are partial atomic charges. ϵ is the dielectric constant and R_{ij} is the distance between atoms i and j.

The calculation of the non-bonded interactions is the most time consuming part of the MD simulation. The number of the intramolecular interactions is proportional to the number of the particles of the system. On the other hand, the non-bonded interactions increase with an order of N^2 on every pair of particles in the system. For this reason, it is sensible to add a cutoff distance, for example, commonly the cutoff applied on the

Lennard-Jones interaction is 2.5 R_{ij}. However, the cutoff strategy provokes a discontinuity in the potential energy, thus a shift function is implemented in order to overcome the discontinuity problem and the total energy is conserved.

3.2.3 Ensembles and temperature-pressure control

Lets move on to describing how the system is coupled to the environment as a whole. Taking a system of N particles in equilibrium with its surrounding environment we should require that the observable thermodynamic quantities $O(t)$ remain as constant as possible over large time frames.

$$O = O(\mathbf{q}^N(t), \mathbf{p}^N(t)) \tag{3.26}$$

The value of observables may fluctuate over short time periods but the over time the system should oscillate around an equilibrium value e.g. a constant temperature. The values of the observable are closely related to the coordinates (q) and the momenta (p) of the particles and the change of the observable in time. We can choose different thermodynamic "ensembles" such as the Microcanonical, Canonical and Isothermal-Isobaric ensembles to study chemistry under different conditions. We assume that the system is erogodic which means that it is able to sample all the microstates if we follow it for long enough time periods.
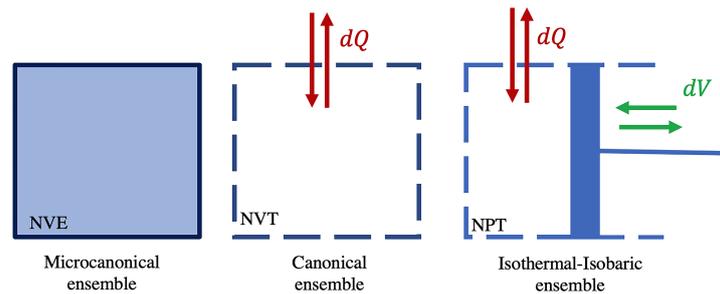
Figure 3.1: Scheme representation of the Microcanonical, Canonical and Isothermal-Isobaric ensambles.

As a general rule of thumb, at starting an MD simulation is recommended to choose the most suitable ensemble for the problem to solve. Each ensemble can be described as follows:

a) **Microcanonical ensemble (NVE):** it considers a system that it is isolated from the surroundings with constant total energy. The micro states generated are degenerate and the resulting thermodynamic potential is equal to the internal energy.

b) **Canonical ensemble (NVT):** the Canonical ensemble allows to the energy fluctuate but keeps the system in equilibrium by controlling the temperature with a fictitious thermal bath. The temperature is controlled by modifying the kinetic energy (K) of the particles (eq. (3.27)). The potential energy obtained from this ensemble is equivalent the Helmholtz free energy.

$$K = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} = \frac{3}{2} N k_b T \tag{3.27}$$

To fulfill the conditions of the Canonical ensemble it is imperative to have a method to maintain constant temperature. Initial velocities are chosen randomly but we must maintain the ideal temperature of the simulation (T_0). To keep the temperature constant during the simulation one of the most used methods is the Berendsen loose coupling technique [61] . The system is coupled to an external heat bath with constant temperature. This change can not be done instantaneously and it must be applied in a proportional scale from v to λv:

$$\lambda = \sqrt{1 + \lambda_c \left(\frac{T_0}{T} - 1 \right)} \tag{3.28}$$

with proportionality constant λ_c a temperature decay deviation in a time constant τ:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{3.29}$$

In this thesis, the MD simulations were using the Velocity-rescaling temperature coupling thermostat [62] implemented in GROMACS [63], that it is in principle a Berendsen thermostat including an additional correction to the kinetic energy (K) distribution using:

$$dK = (K_0 - K)\frac{dt}{\tau} + 2\sqrt{\frac{K K_0}{N_f}}\frac{dW}{\sqrt{\tau}} \tag{3.30}$$

where N_f is the degrees of freedom and dW a Wiener noise. This thermostat method is able to reproduce a canonical ensemble with the advantages over the Berendsen thermostat such as the first order decay of the temperature deviations.

c) Isothermal-Isobaric ensemble (NPT): In this ensemble, temperature and pressure are constant. The volume changes according to a pressure P applied by an external pressure bath re-scaling the position of the particles ($\mathbf{r}_{ij}$) rather than their velocities. We can find the Gibbs free energy using this ensemble. To maintain the conditions required by the NPT ensemble we use temperature coupling and pressure coupling. The instantaneous pressure may be calculated using the virial theorem [64]. In this case via

$$P = \frac{1}{V}\left[k_{\mathrm{B}}T + \frac{1}{3}\sum_{i=1}^{N} \mathbf{r}_i \cdot \mathbf{F}_i \right]. \tag{3.31}$$

The principles we used for Berendsen thermostat can be applied analogously to simulate the coupling to a constant pressure bath via the Berendsen barostat [61]. This results in a pressure change

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} \tag{3.32}$$

Consequently, the atom coordinates and the dimensions of the simulation box have to be scaled by the factor in each simulation step by

$$\mu = \left[1 - \frac{\Delta t}{\tau_P}\left(P_0 - P \right) \right]^{1/3} \tag{3.33}$$

14

Even though the Berendsen barostat generates correct thermodynamic averages, it doesn't provide a true NPT ensemble. The Parinello-Rahman (PR) barostat is a better alternative [65][66]. This method performs a scaling of the simulation box using a matrix transformation. To derive the PR barostat, the atom positions are expressed in terms of to the vectors $\mathbf{u}_i$ that span the simulation box:

$$\mathbf{r}_i = \mathbf{u}_1 \mathbf{s}_{i1} + \mathbf{u}_2 \mathbf{s}_{i2} + \mathbf{u}_3 \mathbf{s}_{i3} = \mathbf{h} \cdot \mathbf{s}_i \tag{3.34}$$

where $0 \leq \mathbf{s}_{ij} \leq 1$ are the positions in the box-system and $\mathbf{h}$ is the transformation matrix between the two systems. The Lagrangian applied on the system has the form

$$\mathcal{L} = T - U = E_{\text{kin}} + \frac{1}{2}W\sum_{ij}\dot{h}_{ij}^2 - E_{\text{internal}} - P_0 V. \tag{3.35}$$

where the kinetic energy associated with the box deformation is incorporated in the second term using a coupling factor W. From the Euler-Lagrange equation follows a differential equation for $\mathbf{h}$ and modifications to the Newtonian equations of motion:

$$W\ddot{\mathbf{h}} = \mathbf{h}^{-1}V(P - P_0) \tag{3.36}$$

$$m_i\ddot{\mathbf{s}}_i = \mathbf{h}^{-1}\mathbf{F}_i - m_i\mathbf{G}^{-1}\dot{\mathbf{G}}\dot{\mathbf{S}}_i \quad \text{with} \quad \mathbf{G} = \mathbf{h}^T\mathbf{h} \tag{3.37}$$

3.2.4   Periodic Boundary Conditions

Boundary conditions are used in MD simulations to conserve the macroscopic behaviour of the system. Including them is of major importance when calculating thermodyanic properties which depend upon the particle number. All the systems simulated in this thesis were performed in a bulk of solvent or solvation box including periodic boundary conditions (PBC). Figure 3.2 illustrates how PBCs are applied in atomic simulations. The central simulation box is replicated in all directions creating an infinite lattice. When particles move in the central box, all the particles in the other boxes move in the same direction. If a particle leaves the box, it is restored in the opposite direction of the box it leaves. This way the number of particles is conserved.
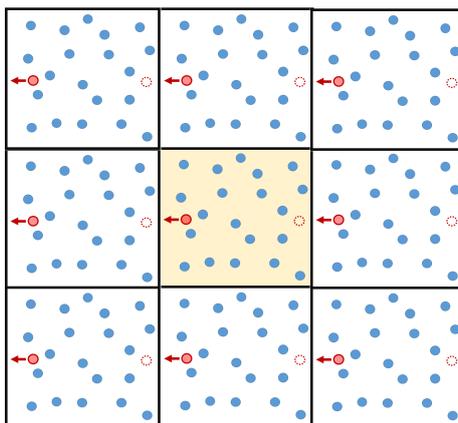
Figure 3.2: Representation of how particles moves in a simulation with periodic boundary conditions. The central simulation box is represented in yellow and the particle crossing the box with red.

3.2.5   Simulation conditions

The simulations were perfomed using the GROMACS [67, 68] package. The MD simulations performed within chapters 4 and 6 were carried out using the Isothermal-Isobaric ensembles with periodic boundary conditions. The specific simulation settings are explained in their corresponding method sections.

3.3 Monte Carlo

Monte Carlo (MC) methods are widely used in physics and chemistry in order to move the system between the different possible micro-states. This is done in a stochastic manner with the structures at each step being the result of the generation of random structural transformations. These transformations are often refereed to as *moves* which can generated via a large variety of different protocols. One of the most widely MC methods used is the Markov chain Monte Carlo (MCMC).

3.3.1 MC detailed balance and move generation

Lets consider the total rate of states Γ for moving the system from from point q to q'. The principle of detailed balance states the forward rate should be equal to the backward probability.

$$\Gamma(q \to q') = \Gamma(q' \to q) \tag{3.38}$$

The transition probability W is equal to the product of the probability a move is constructed π and the the probability the movement is accepted ρ

$$W = \pi(q \to q')\rho(q \to q') \tag{3.39}$$

with $\beta = \frac{1}{k_\mathrm{b}T}$, the transition state from $q \to q'$ can be expressed as

$$\Gamma\left((q) \to (q')\right) = W\left((q) \to (q')\right) \times \frac{1}{\mathcal{Z}}\exp(-\beta E(q)) \tag{3.40}$$

And the principle of detailed balance can be expressed as:

$$\frac{\pi(q \to q')\rho(q \to q')}{\pi(q' \to q)\rho(q' \to q)} = \exp\left(-\beta \Delta E)\right) \tag{3.41}$$

if the moves are generated in a random fashion the approximation that $\pi(q \to q') = \pi(q' \to q)$ is acceptable and then moves can be evaluated using the Metropolis acceptance criteria [69, 70]

$$\rho\left(q \to q'\right) = \begin{cases} \exp\left(-\beta \Delta E\right), & \Delta E > 0 \\ 1, & \Delta E \leq 0 \end{cases} \tag{3.42}$$

3.3.2 The potential terms

The calculation of the potential energy of the conformations generated with MC algorithm were calculated using the non-bonded terms from GAFF [60] explained in the MD simulation section in eq. (3.25) (Lennard-Jones and Coulomb) and the dihedral potential as can be seen in eq.(3.43).

$$E_{\mathrm{GAFF\text{-}MC}} = \sum_{dihedrals} \frac{V_n}{2}[1 + cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{\mathbf{R}^{12}} - \frac{B_{ij}}{\mathbf{R}_{ij}^6} + \frac{q_i q_j}{\epsilon \mathbf{R}}\right] \tag{3.43}$$

Additionally, for polymers which have hydrogen bond (H-Bond) donors and acceptors an extra energy potential term was included to enhance the conformation search for candidate conformations that display stabilizing hydrogen bonds. The explicit hydrogen bond

potential implemented in SIMONA has the form

$$E_{\text{H-bonds}} = \sum_{i \in \text{Acc}} \sum_{j \in \text{Don}} \left[\left(\frac{C_{ij}}{r_{ij}} \right)^{12} - \left(\frac{D_{ij}}{r_{ij}} \right)^{10} - E_0 \right] \cos^2(\theta) \cos^4(\phi) \qquad (3.44)$$

In the above equation, θ, ϕ and r_{ij} are defined as shown in figure 3.3 and only acceptor-donor pairs with a distance below 3.3 Å are considered. The additional offset E_0 is introduced to avoid a discontinuity due to this cutoff. The parameters of the explicit hydrogen bond potential are based on the work of McGuire et al.[71].
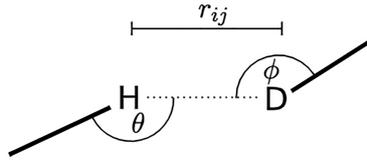

Figure 3.3: Hydrogen bond representation used for HB potential term.

3.3.3 Implicit solvation

The folding of macromolecules can be affected by the surrounding interactions with the solvent. To estimate the solvation free energy of a molecule (ΔG_{solv}) it can be split in three different energy contributions [72]

$$\Delta G_{\text{sol}} = \Delta G_{\text{cav}} + \Delta G_{VdW} + \Delta G_{\text{pol}} \qquad (3.45)$$

ΔG_{cav} represent the work to form the solvent cavity, ΔG_{vdW} represents the long range solute-solvent interactions due to dispersion and ΔG_{pol} accounts for the solute-solvent electrostatic interactions. Since explicit solvation in MC simulation is highly challenging, an implicit solvation method was considered across all the MC simulations performed in this thesis. With this aim the Born Generalized (GB) method [72] was considered to calculate the polar contribution of the free energy of solvation. In general words, this method approximate the solvent to a dielectric environment. The GB model calculates the polar solvation term to evaluate the electrostatic solvent effects via eqs. (3.46)

$$\Delta G_{\text{pol}} = -\frac{\alpha}{2} \left(\frac{1}{\varepsilon_{\text{in}}} - \frac{1}{\varepsilon_{\text{out}}} \right) \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j \exp\left(-r_{ij}^2/(4R_i R_j)\right)}} \qquad (3.46)$$

Where $\alpha = 331.84$ kcal/mol and is a constant, ϵ_{in} and ϵ_{out} are the dielectric constant of the molecule and the solvent respectively and R_i is the effective Born radius that can be obtained from the Coulomb-field approximation

$$\frac{1}{R_i} = \frac{1}{4\pi} \int_{solvent} \frac{dV}{\|r_i - r\|^4} \qquad (3.47)$$

Solving the integral in eq. (3.47) is not easy since the integration of this surface depends of the molecule conformation, SIMONA has implemented the PowerBorn protocol [73] that applies an octree representation to divide up the 3D space. An octree is a tree structure with the nodes being cubic cells with each tree node having eight branches.
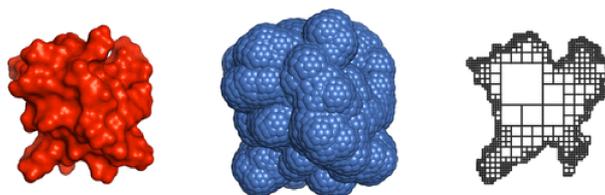
Figure 3.4: Octree representation of PowerBorn protocol. (Figure reproduced from publication of Brieg et al. work [73])

3.3.4   MC sampling and Annealing Simulation

Reliable statistical averages can be obtained when MC simulations can be concentrated and biased towards the regions of most importance. This concept is known as the *importance of sampling*. In this thesis different strategies were applied to guide the MC simulation towards helical conformations using dihedral scans or applying additional potentials to give preference to stabilizing interactions.

One of the most popular strategies is Annealing Simulation (AS). Annealing Simulation MC can be used to find the global minimum of energy states on the PES (fig 3.5). This technique is frequenctly use to refine NMR or crystallograpy models. In contrast to the conventional Metropolis algorithm, in AS the temperature decreases as a function of time, mimicking a cooling process. At high temperatures, the system is free to explore states or conformations that are not easily accessible using techniques such as MD (fig. 3.5). Given a sufficient number of simulation steps and/or enough parallel simulations, the algorithm has a chance of finding a reasonably stable local minimum. Overall, MC simulations are capable of exploring a larger portion of the potential energy surface while MD simulations tend to remain around local minima.
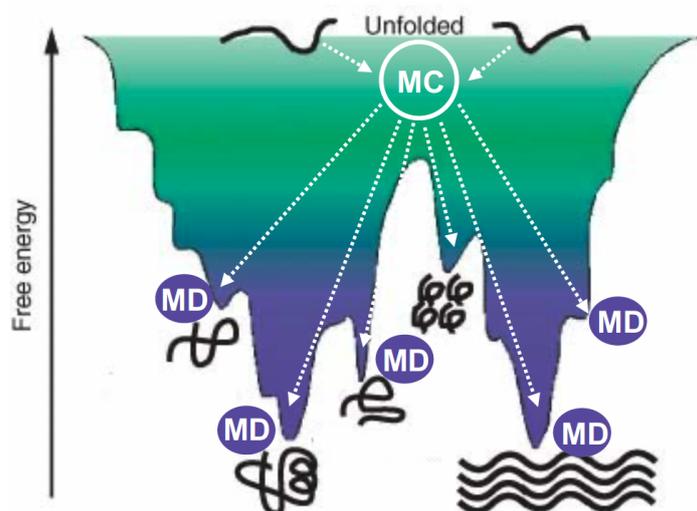

Figure 3.5: Differences between the MC and MD approaches.

3.4   SimStack: Intuitive Workflows and WaNos

SimStack [74] is a tool developed in cooperation of Prof. Wenzel's group (`https://simstack.readthedocs.io/en/latest/`) and Nanomatch GmbH (`https://www.nanomatch.com`) based on a lightweight client-server program. It provides a graphical user interface to construct simulation workflows from individual program or protocol units called Workflow Active Nodes (WaNos) using High Power Computer (HPC) resources. The main idea behind SimStack is to enable to users to build complex simulations protocols and automate its execution. The main advantage of workflow frameworks designed for SimStack is to provide protocols that can be adaptable, reproducible, expandable, transferable and efficient.
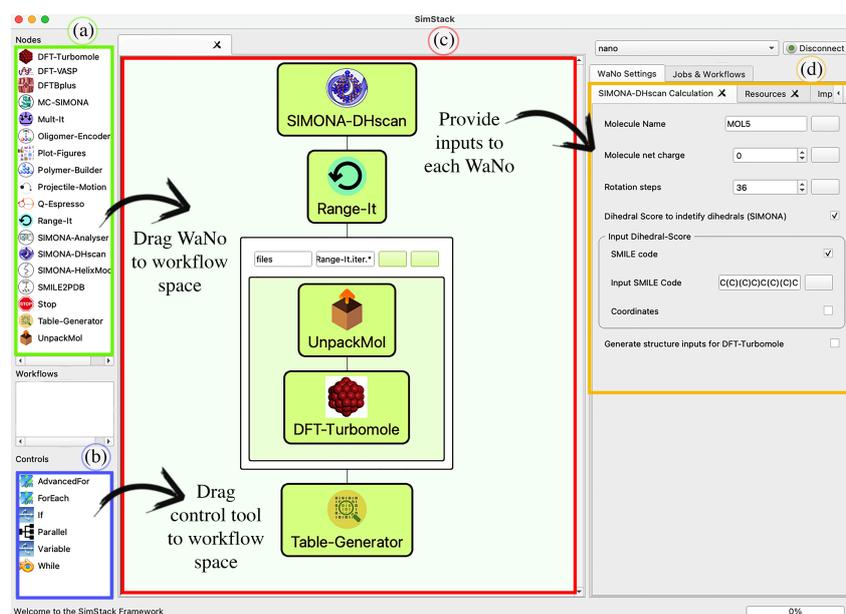

Figure 3.6: Capture of the SimStack workflow server with a dihedral scan workflow as example using SIMONA [75] and TURBOMOLE [76]. (a)list of collection of WaNos available to use, (b) control tools to support workflows with iterations, conditionals, and others flow statements, (c) workflow space were WaNos and control tools are dragged to build the final workflow and (d) user input tabs generated anytime that the user selects a WaNo.

Easy-to-use simulation programs are not always a reality in the field of theoretical chemistry and Soft Matter physics. SimStack uses Extensible Markup Language (XML) format to generate the graphical interface for the settings of each WaNo, which is human readable and user friendly language, even for non-developer users. This approach is able to make closer the relationship between experiment-theory and vice versa, since SimStack can be easily used by non-computer scientists. Many of the protocols generated during the research of this thesis were finally implemented in SimStack. WaNos and Workflows (group of WaNos ordered in a sequential fashion) were created in order to automatize polymer simulations. Further SimStack WaNo implementations will be described in the following chapters.

3.4.1 Highlights on multiscale modelling

Chemistry can be explored at many different atomic scales with a cornucopia of different approaches to solving scientific questions (Figure 3.7). Polymers are macromolecules that can be studied from the electronic scale to coarse-grained models. Here our focus lies both at the electronic and atomistic scales. Some phenomena are better explained by the electronic structure, for example, the charge transfer effects that could play an essential role in polymer folding. However, we can broadly study the inter-molecular interactions with methods such as MD and MC to sample different states on the PES at a lower cost. The following chapters discuss the pros and limitations of various approaches, arguing that combining different approaches within the multi-scale paradigm is suitable for studying helical polymers.
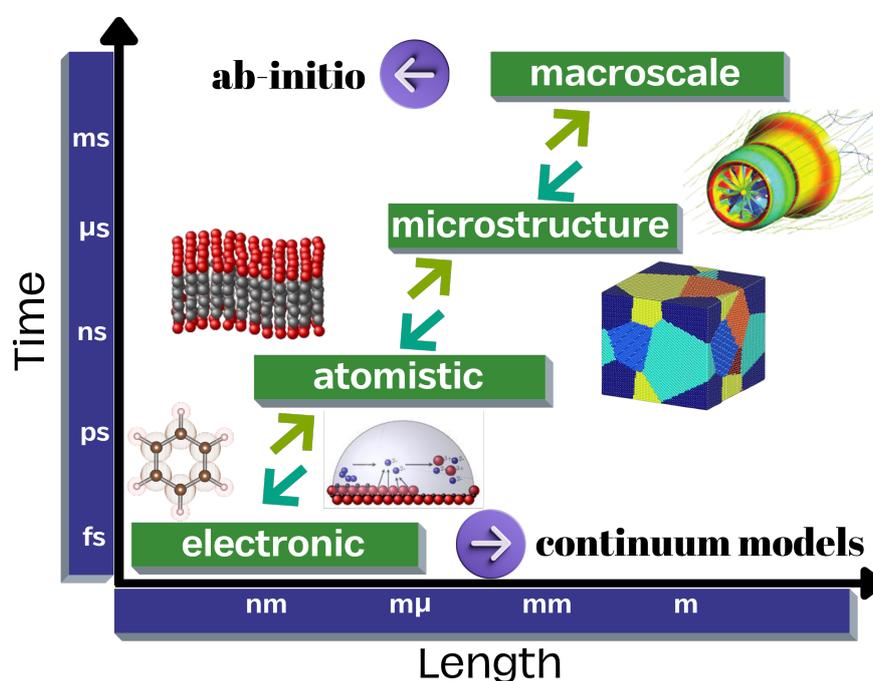

Figure 3.7: Different approaches according to the scale of the systems studied.

4 Theoretical studies on Helical Induction of Cis-Transoid Poly (4-carboxyphenyl) acetylene by Chiral Amines

4.1 Introduction

Helical structures are present throughout nature with famous examples being both DNA and proteins. Taking inspiration from these biomolecules, polymer scientists are highly interested to find synthesis routes and mechanisms to control helix formation in polymers. Synthetic helical polymers can potentially be used in a wide range of potential applications such as chiral recognition, asymmetric catalysis, chiro-optical switches [77] and are garnering increasing interest within nanostructure material design[78–81].

Experimental techniques such as UV, NMR, Circular Dichroism (CD), Atomic Force Microscopy (AFM), and X-ray diffraction (XRD) are regularly used to demonstrate the existence of helical polymers [11]. Depending on the inversion barrier of the helical polymer, these experimental techniques may be able to capture different structural insights. Helical polymers with bulky pendant groups, such as poly(methacrylate)s [82] or poly(isocyanides)s, [83] are detected, for example, using AFM due to their stable helical structures. On the other hand, highly dynamic helical polymers such as Poly(phenylacetylene)s (PPAs) have lower inversion barriers which are only detectable by measuring their optical activity using CD. Therefore, investigating the spatial orientation of dynamic helical polymers via theoretical calculations can be significantly useful tool for understanding existing polymers and designing new materials.

Induced helical polymers (IHPs) exhibit three interesting phenomena: 1) the sergeants and soldiers principle [84] where a small number of chiral molecules could induce a helical conformation in some dynamic helical polymers, 2) the majority rule [85] where the screw-sense of the helical polymer is induced in the presence of a mixture of two enantiomers, and 3) the memory effect [86] where in IHP the helicity can be conserved after removing the chiral compounds and replacing them with nonchiral hosts. The first IHP family discovered was PPAs which are optically inactive and helices are induced by noncovalent interacting ligands was reported by Yashima *et al.* [87, 88]. They found that chiral amines and amino alcohols can induce a prevailing one-handedness in nonoptically active stereoregular *cis-transoid* polyacetylene with 4-carboxyphenyl bearing groups (poly-1) as depicted in Figure 4.1. The chiral effects of the host molecules and chromophore groups are most often measured via induced CD (ICD). The IHP shows mirror CD images for each enantiomer and very characteristic CD signals for the backbone typically in the absorption region between 300 and 500 nm [11, 88].

While there is a considerable number of experimental publications about dynamic helical polymers [11, 77, 89] and works related to the induction mechanism driven by different

---

This chapter is based on the following publication: Penaloza-Amion, M., C. Rego, C. R., & Wenzel, W. (2022). Local Electronic Charge Transfer in the Helical Induction of Cis-Transoid Poly (4-carboxyphenyl) acetylene by Chiral Amines. Journal of Chemical Information and Modeling.
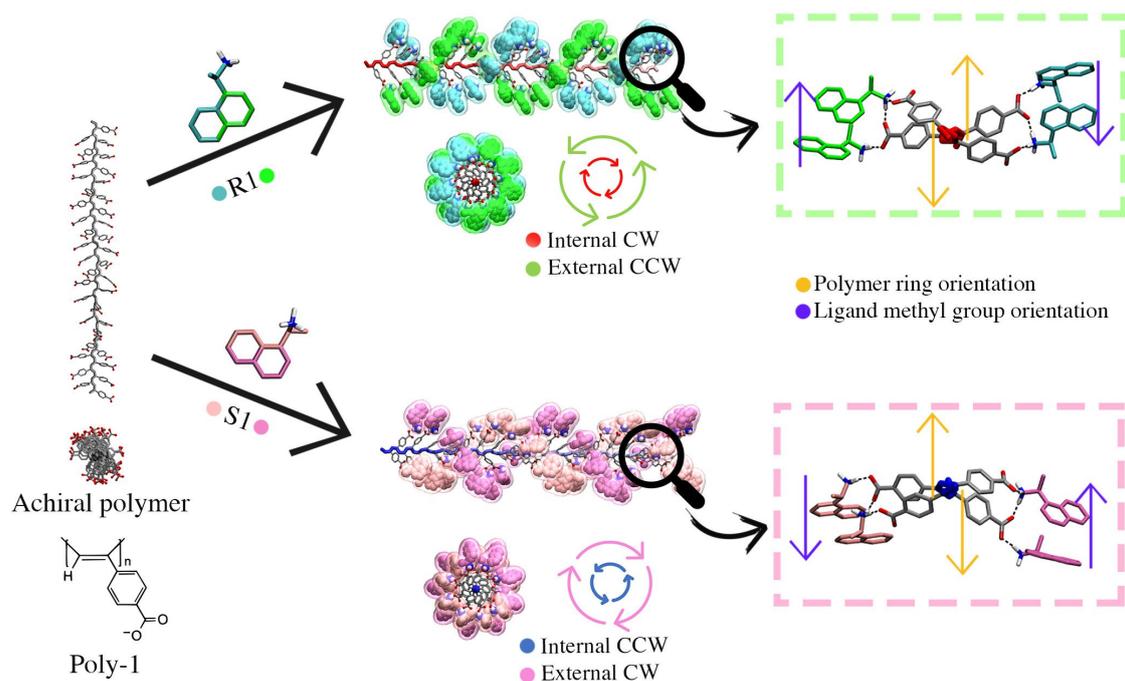
Figure 4.1: Scheme of helical induction of achiral poly-1 interacting with R1 and S1 chiral amines. 40 monomers helical polymer-amine complexes with clockwise (CW) and counterclockwise (CCW) screwsense. They were obtained by studying the interactions with 4 monomer complexes models and DFT calculations (pink and green dash boxes) with implicit solvation. R1 ligands are represented with green and cian while S1 ligands are represented with pink and magenta.

side chain modifications on PPAs [90–95], solvent influence [91, 96], temperature [97], and pH [98], there is still a need for a deeper understanding of the induction process. Currently, there is almost no theoretical work on induced helical polymers or on the induction mechanism. In 1997, Yashima et al. [88]attempted to evaluate the stability of poly-1 using molecular mechanics and molecular dynamics calculations with a model of 20 monomers. They created a left and right-handed optimized model. They reported the existence of a helical conformation that lost its helicity after a few ps of simulation and thus concluded that the reported helical conformation of poly-1 may not be stable in solution.

The focus of this work is the dynamic helix formation in poly-1. We aim shed light on the nature of the induction mechanism in the presence of chiral compounds. We studied the interactions of (R)- and (S)-1-(1naphthyl)ethylamine (R1 and S1) investigated by Yashima et al. [86, 88] with poly-1 clockwise (CW) and counterclockwise (CCW) conformations. We use a multi-scale consisting of Molecular Dynamics (MD) simulations and electronic structure calculations. MD is regularly used to study dynamical processes on helical polymers [46] but relies on a number of approximations for the force field such as the choice of dihedral potential [99]. On the other hand, electronic structure calculations provide useful microscopic information such as the charge transfer in simplified systems [100–102] and provide valuable information regarding the binding affinity [103]. We examine in detail polymer-amine interactions and find affinity trends using Density Functional Theory (DFT) combined with a posteriori energy-dispersion correction. We obtained dissociation energies of polymer-amine complexes and performed Bader charge
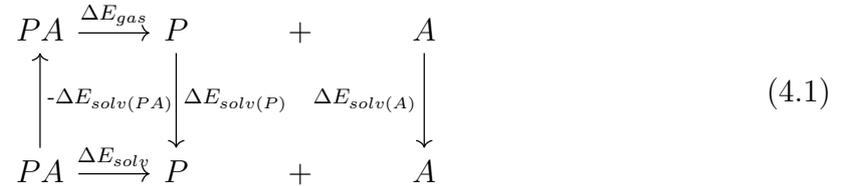
analysis that revealed the charge transfer between the carboxylic and amine group from polymer and chiral amines.

4.2 Computational Details

4.2.1 Density Functional Theory Calculations

To start off we performed electronic structure calculations to determine favorable helix configurations. The formation of the helix is primarily controlled by the values of the backbone dihedral angles. We searched for the angles by performing scan calculations rotating them in the range $0° - 360°$ with steps of $10°$ degrees using TURBOMOLE 7.3 [76, 104, 105] with the hybrid B3LYP functional [106, 107] and def2-SV(P) basis set [108]. To compensate for the poor description of van der Waals (vdW) interactions by local and semilocal XC functionals [109, 110], we employed a posteriori energy-dispersion correction D3-BJ [111, 112]. Minimal polymer-amine complexes consisting of 4 monomers were constructed using the dihedral information obtained from the scan calculations to create CW and CCW conformations for poly-1, and amines were placed interacting with each carboxylic acid. Different orientations of the methyl and naphthalene groups were created to determine the lowest energy complexes. All the complexes were optimized using the same DFT setup and parameters used for the dihedral scan simulation.

The dissociation energies for R1 and S1 chiral compounds around the polymer are computed via DFT (electronic energy) using 4 monomer polymer-amine complexes according to the cycle in (4.1) where PA, P, and A are the polymer-amine complex, polymer, and amine, respectively.

$$
\begin{array}{ccccc}
PA & \xrightarrow{\Delta E_{gas}} & P & + & A \\[2pt]
\big\uparrow{\scriptstyle -\Delta E_{solv(PA)}} & \big\downarrow{\scriptstyle \Delta E_{solv(P)}} & & {\scriptstyle \Delta E_{solv(A)}}\big\downarrow & \\[2pt]
PA & \xrightarrow{\Delta E_{solv}} & P & + & A
\end{array}
\tag{4.1}
$$

The calculation of the dissociation energy of each amine in Dimethyl Sulfoxide (DMSO) (ΔE_{diss}) is equal to the energy variation ΔE_{solv} eq (4.2) as defined in eq (4.1),

$$\Delta E_{diss} = \Delta E_{solv}. \tag{4.2}$$

According to the paths in eq (4.1), ΔE_{solv} is calculated by accounting for the contributions ΔE_{gas} and $\Delta\Delta E_{solv}$ as given in eq (4.3).

$$\Delta E_{solv} = \Delta E_{gas}(PA) + \Delta\Delta E_{solv}(PA) \tag{4.3}$$

Then the energy contributions are obtained as represented in eqs (4.4) and (4.5) for $\Delta E_{gas}(PA)$ and $\Delta\Delta E_{solv}(PA)$, respectively.

$$\Delta E_{gas}(PA) = E_{gas}(P) + E_{gas}(A) - E_{gas}(PA) \tag{4.4}$$

$$\Delta\Delta E_{solv}(PA) = \Delta E_{solv}(P) + \Delta E_{solv}(A) - \Delta E_{solv}(PA) \tag{4.5}$$

We computed all the states for the polymer-amine complex, polymer, and amine represented in eq (4.1) using the same DFT configuration as in our previous studies. Solvated states in DMSO (Dielectric Constant ($\epsilon_{\text{solvent}}$) = 47) were obtained using the COSMO solvation method [113]. The dissociation, ΔE_{diss}, of the amines was calculated one by one per monomer position with CW and CCW screw-senses. To evaluate the influence of the chiral center, in each complex, the chiral methyl group covalently bound to each chiral center was exchanged with the hydrogen atom to convert each monomer from R to S and S to R, respectively.

4.2.2   Bader Net Charge Analysis

We used the Bader analysis to estimate the partial charge for the system under study to observe the amine's influence over the polymer's charge distribution. To compute the Bader charge Q^{B} for every atom in the molecule, we employed VASP [114–117] high-density grids [118]. Thus, we calculated the effective charge by $\Delta Q = Z_{\text{val}} - Q^{\text{B}}$, where Z_{val} and Q_{B} are the valence electron number, extracted from the VASP POTCAR file, and Bader charge on each atom. As in this part of the protocol, we would have to submit many independent tasks for different scenarios in the CW or CCW screw-sense, and to speed up the process, we managed all Bader analysis simulations by using the Workflow Active Nodes DFT-VASP developed within the SimStack workflow framework [119, 120]. We analyzed the orientation of the amines in extended 40 monomer complexes to obtain the initial coordinates for MD simulations. The lowest energy complexes obtained for the 4 monomer polymers with similar energy compared to CW and CCW were used to extract the relative orientation of the amines and transfer these coordinates to CW and CCW 40 monomer helical models.

4.2.3   Molecular Dynamics Simulations

MD simulations of 20 nanoseconds on CW and CCW helical conformations and polymer-amine complexes were carried out using GROMACS 2019 [67, 68]. The polymer and amines were parametrized using AmberTools19 [121] applying the General Amber Force Field (GAFF) [122] and the AM1-BCC method for RESP charges [123]. The 40 monomer helical models were solvated using a pre-equilibrated cubic box of DMSO molecules provided by Caleman et al. and their solvent database [124], and sodium counterions were added to the pure polymer simulations. The systems were minimized and then equilibrated in two steps using the NVT and NPT ensembles considering position restraint using LINCS [125] to constrain all bonds of the polymer. During the NVT step, with 500 ps, a 2 fs time step using the V-scale method [126] was used to control the temperature. The NPT equilibration step was performed under 500 ps with 2 fs time step and the Parrinello– Rahman pressure coupling method [65]. Finally, a 20 nanoseconds MD production with periodic boundary conditions was performed using the V-scale method [126] to keep the temperature at 300 K and the Parrinello–Rahman pressure coupling method [65] to keep the pressure at 1 atm. Electrostatic and van der Waals shortrange interactions were calculated with the Verlet method [127] with a cutoff of 1.4 nm. The long-range electrostatic interaction was taken under the Particle Mesh Ewald (PME) method [128]. MD simulations for 40 monomer polymer-amine complexes were carried out with the same conditions and settings.

4.3   Results and Discussion

4.3.1   Helical Model and Polymer-Amine Complex Stability

Dynamic helical polymers are challenging to study because their helical configurations are difficult to distinguish experimentally. A helix structure can be constructed by arranging repeated monomer units per turn. The notation of A * u/l is frequently used, where A is the class according to A chain motifs of the helix with u motifs per l repeating units [37]. Inspired by the the experimental results obtained by Maeda et al. [96] carboxylate groups ($-COO^-$) in every side chain were considered. Maeda et al. observed intensity changes in IR signals from –COOH to $-COO^-$ groups when the helix is induced suggesting that the helical conformation on poly-1 is not possible without the presence of the negative carboxy charged groups.

DFT scan dihedral calculations on a dimer of a polyacetylene backbone with different degrees of side chain substitution were performed to observe energy differences in the twist of the dihedral formed by the single bond adjacent to two double bonds as displayed in panel (**a**) of Figure 4.2. In panel (**a**) we observe that when we add different 4-carboxyphenyl (4-CaPhe) groups to the polyacetylene backbone, the energy profile shows different local minima resulting in different possible poly-1 conformations. Both *Cis-transoid* and *cis-cisoid* conformations were found. Assuming that the perfect *cis-transoid* confirmation is obtained when the dihedral angles of the backbone are 180° and −180°, this local minimum moves to the values of 145° and −145° when attaching 4-CaPhe side chains to the backbone, resulting in a 2/1 helix (Figure 4.2(a)). On the other hand, *cis-cisoid* conformations for negatively charged poly-1 were found at 54° and −54° resulting in a 3/1 helix shown in panels (**b**) and (**c**) from Figure 4.2. The difference in the backbone dihedral angle between the local minimum of the pure polyacetylene backbone and the dimer with 4-CaPhe substitutions may be due to the steric repulsion of the phenyl rings and the charge repulsion of the negative carboxylate groups. We found a low energy barrier of ca. 0.1 eV which makes inversions in the screw-sense of the polymer from 145° to −145° facile.
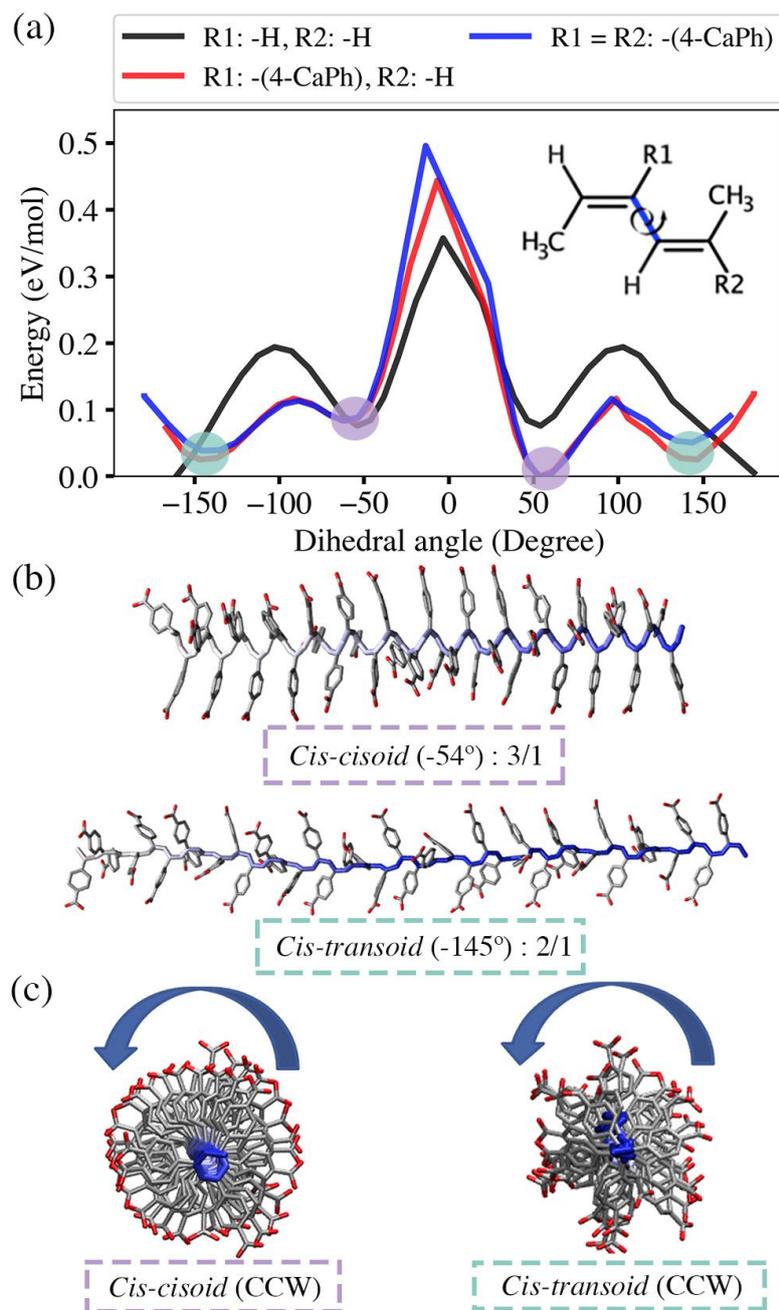
Figure 4.2: (a) Dihedral energy potential obtained from DFT scan calculations with different degrees of substitution of dimer of poly-acetylene. (b) Helical models created from the DFT scan information obtained for CCW screw-sense, *cis-cisoid* and *cis-transoid* conformations. (c) Upper view of the helical models for CCW screw-sense.

In order to test the stability of these conformations, polymer models of *cis-transoid* of 40 monomers for CW and CCW were created using the results from the dihedral scans. Since the transition from *cis-transoid* to *cis-cisoid* is not entropically favorable [97, 129] and the transition was not observed during the helical induction experiments [88], further calculations were limited to the *cis-transoid* conformation. From dihedral scans, we obtained both CW and CCW models. The convergence during the MD simulations of each helical polymer simulation was based on Root Mean Square deviation (RMSD) taking as a reference the initial configuration of each simulation (Figure 4.3).
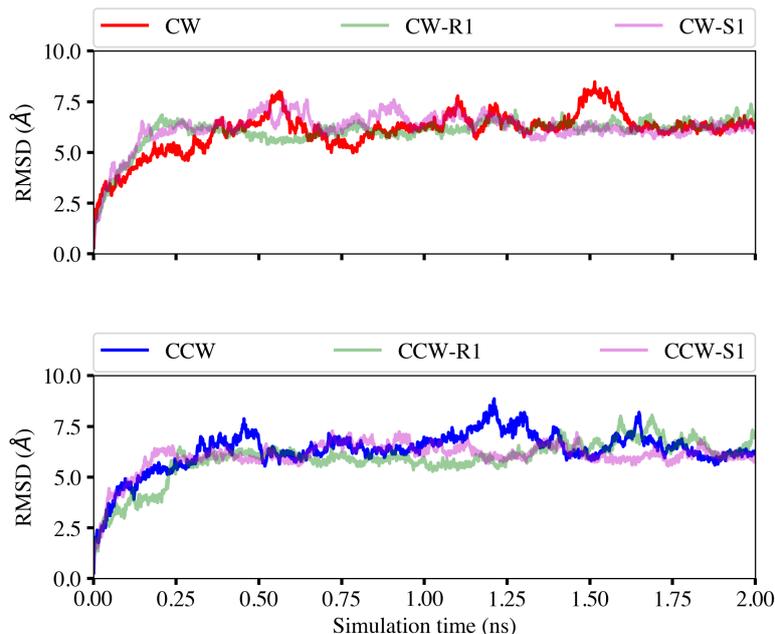

Figure 4.3: RMSD values calculated from the MD trajectories for the CW (red), CCW (blue), CW-R1, CCW-S1, CW-S1 and CCW-S1 systems.

On the other hand, the stability of polymers were tested using Helical Index Error (HIE) (eq (4.6)). This equation considers the deviation of dihedral angles during the simulation relative to an ideal model (ϕ_{model} - $\phi_{simulation}$). The range of angles in the local minimum obtained from the DFT scan studies where the helices start to lose their helicity is $\Delta \phi_{helix}$.

$$\text{HIE} = 1 - \tanh\left(\frac{\phi_{model} - \phi_{simulation}}{\Delta\phi_{helix}}\right)^2 \qquad (4.6)$$

4.3   Results and Discussion
4   Theoretical studies on Helical Induction of
Cis-Transoid Poly (4-carboxyphenyl) acetylene by Chiral Amines

(**a**) in Figure 4.4 shows the HIE values for helical models in the poly-1 CW and CCW screw-senses. Its oscillatory behavior reveals the flexible nature of poly-1, presenting reversals along the polymer backbone, which is in agreement with the results of the polymer simulations by Yashima in 1997 [88] when simulating 20 monomer poly-1 but with neutral carboxylic groups. Additionally, dihedral angles per monomer of the backbone polymer were obtained by taking the average of dihedral values (dihedral values are available on Appendix 7.1.1). The helical sense per dihedral angle is computed by sorting the dihedral angles as CW for positive and CCW for negative. While angles of $145°$ and $-145°$ were obtained from the dihedral scan calculations for the best helical model, during the simulations, most of the dihedral average values moved toward *ca.* $160°$ and $-160°$. The results obtained for the dihedral angle occurrence clearly show the observed inversion of screw-sense during the simulations without showing any preference for the CW or CCW sense.
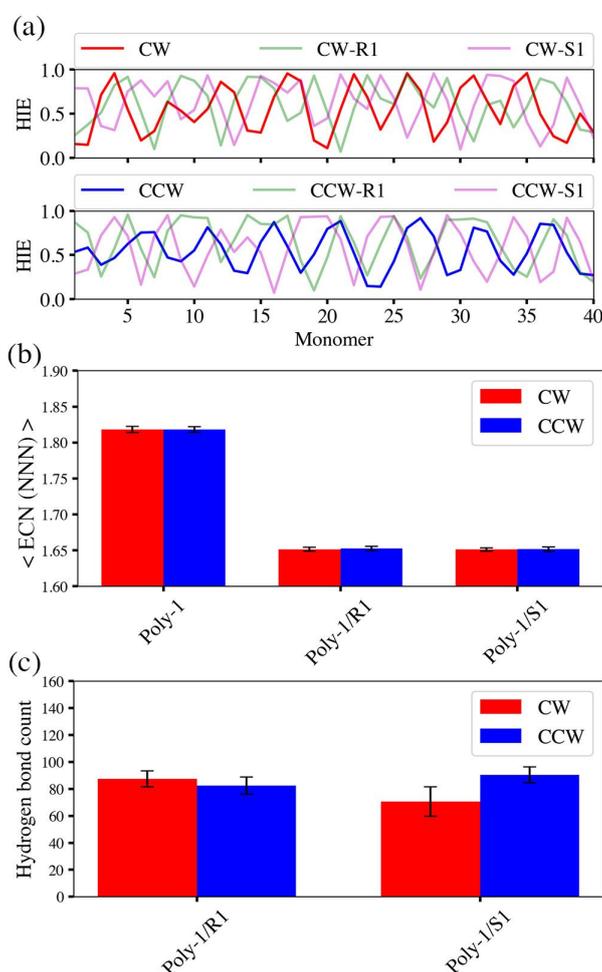

Figure 4.4: (a) Helical Index Error (HIE) obtained for CW (red) and CCW (blue) screw-sense of poly-1. (b) The ECN (NNN) was calculated for each frame in a trajectory of 200 frames from our MD calculations, then the global ⟨**ECN**⟩ was obtained. (c) Hydrogen Bond count performed by amines and polymer during the MD simulations.

To understand the role of chiral amines interacting with poly-1 in the helical stabilization in these chiral compounds, MD simulations of CW and CCW helical models interacting with R1 and S1 chiral compounds were carried out. R1 and S1 amines as depicted in Figure 4.1 were chosen for our simulations since they are chiral compounds with high-intensity ICD bands [96] when interacting with poly-1. They have bulky naphthalene rings as aromatic groups, a feature notified as an essential factor for helical induction [88]. Experimental results by Maeda et al. [96] on helical induction of poly-1 using R1 and S1 revealed that the ion pair formation between the polymer side chains and amines is essential for the helix induction and to obtain one-handedness excess. Additionally, it has been suggested that chiral amines might generate cooperative hydrogen bonds (H-Bonds) or bidentatetype ion-pair interactions after the basic amine group removes the proton from the carboxylic groups of each monomer [88]. In order to incorporate this information and determine the starting conformations of the polymer-amine complexes, several systems of 4 monomer poly-1 complexes with R1 and S1 interacting with poly-1 by ionic H-Bond interaction were optimized using the same DFT setup to determine the lowest energy complexes to use as a reference. Different orientations of the groups attached to the chiral carbon (naphthalene and methyl groups) were considered in the screening. The lowest energy complexes found are CW-R1 and CCW-S1 complexes with a ΔE = 0.012 eV between them. The optimized complex structures obtained showed the bidentate-type ion-pair interaction as described in Figure 4.1, green and pink boxes show CW-R1 and CCW-S1 complexes respectively. These two compounds exhibit opposite orientations of naphthalene rings and methyl groups. The ligands' relative position optimized via DFT was used to create 40 monomer polymer-amine complexes to feed the MD simulations.

The adopted starting configurations for R1 and S1 with 40 mononomer polymer complexes are displayed in Figure 4.1. The simulation using 40 monomer poly-1 amine complexes presents the helical sense performed by the backbone, as well as the ligands around the helical sense going in the opposite direction to the backbone sense (Figure 4.1). As before we performed MD simulations of 20 nanoseconds and the trajectories were analyzed by calculating the HIE in panels **a**) and **b**) of Figure 4.4. Average torsion angles and their frequency of occurrence are shown in Appendix 7.1.1. For both screw senses CW and CCW the HIE values do not show any significant trends where comparing the cases without amines and with R1 and S1 amines respectively. We can conclude from this that for all cases the polymer is highly flexible. These results are very similar to our first MD simulations in the absence of chiral amines attached but distinct from the minimum energy dihedral values obtained from the DFT scan. After our MD simulations, the average dihedral values with their occurrence per screw-sense (Appendix 7.1.1) reveal no difference between polymer-R1 and polymer-S1 complexes indicating a lack of a driving force that pushes the system to one helical sense. The average of dihedral values that fluctuate around 160° and −160° for CW and CCW, respectively (Appendix 7.1.1).

To evaluate the binding of the amines of the polymer, the average of the Effective Coordination Number (ECN) [130, 131], relative to the number of nearest neighbors (NNNs), was calculated for every polymer-amine complex in the MD simulations. As shown in panel **b**) of Figure 4.4, we did not find any significant differences for the ECN values calculated for poly-1 interacting with R1 and S1, showing that their absorption is similar in all simulated complexes. As displayed in panel **c**) Figure 4.4, the count of hydrogen bonds (H-Bonds) reveals that almost as many H-Bonds are present per monomer, showing the high probability of a cooperative H-Bond between the chiral compounds and the polymer. We therefore conclude that the model employed in the MD simulation cannot distinguish between chiral compounds interacting with poly-1. Viscometric and theoretical studies have shown that poly-1 right- and left-handed helical conformations could be rapidly interconvertible, and it has been suggested that even though the chiral amines could perform predominant ion-pair interactions, they could be changing the population of the configuration states of the polymer [96, 132]. It is challenging to model the complete induction process with MD due to the inherent time scale problem. Other limitations, such as the classical force fields by construction, do not explicitly account for the intermolecular interactions that include polarization and charge transfer effects, which are well described for several systems in DFT using the appropriate van der Waals correction [109] or that could be overcome by advanced MD techniques such as Car–Parrinello MD [133, 134] or using non-conventional force fields such as multipolar force fields [135, 136].

4.3.2   Polymer-Amine Dissociation Calculations and Charge Analysis

To overcome the limitations explained in the previous section, we resorted to the electronic structure calculations to evaluate the dissociation energy of every ligand interacting with every monomer using our 4 monomer poly-1 amine complexes obtained before. The influence of the chiral center was accounted for by manually modifying the swap of the methyl groups connected to the chiral carbon by the hydrogen. Using this strategy, we preserve the coordinates of the complex but with the opposite chiral amine identity. We computed the dissociation energy values of all complexes by analyzing the dissociation pathway using a thermodynamic cycle as given in eq (4.1). As can be seen in panels **c** and **d** of Figure 4.5, we denote each ligand as *X-M*, where *X* is R1 or S1 and *M* is the monomer position where the ligand interacts with the 4 monomer polymer.
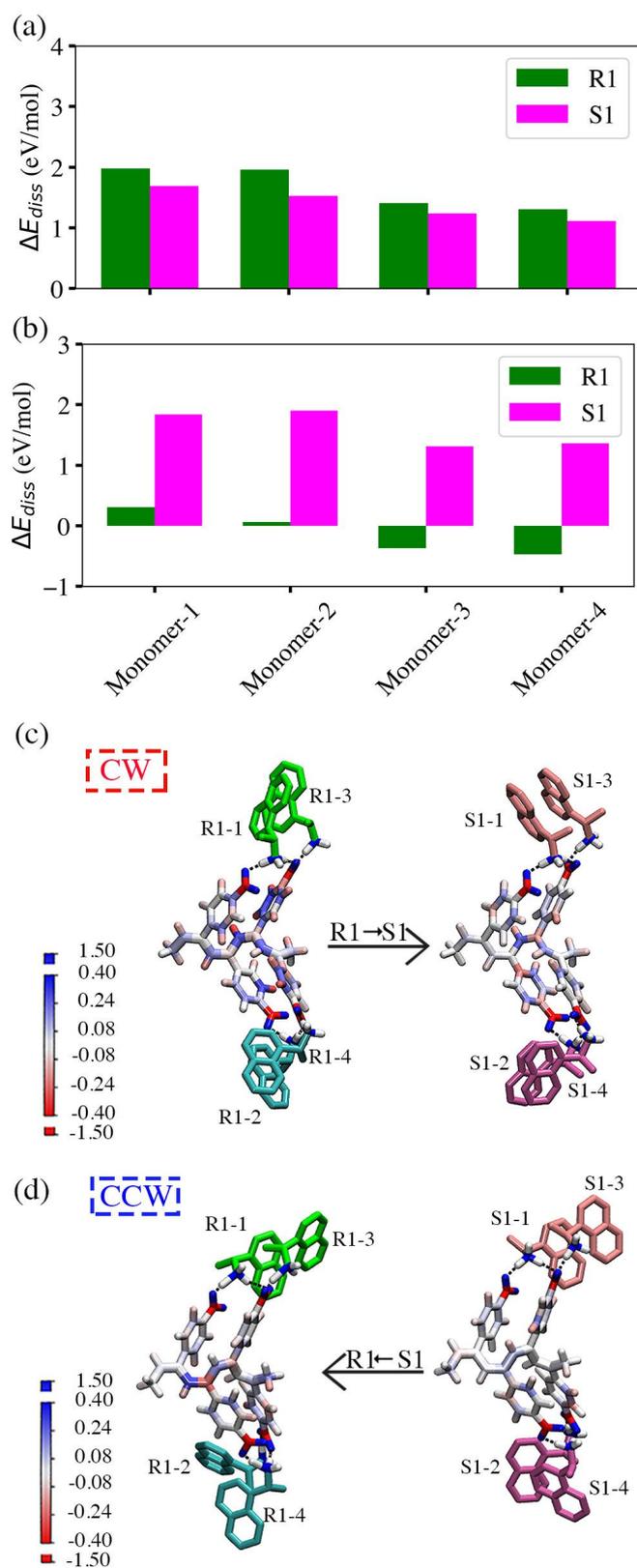
Figure 4.5: Values of ΔE_{diss} calculated with DFT for CW (a) and CCW (b) helix sense. Distribution of Bader net atomic charges on poly-1 CW sense (c), CCW sense (d) at interacting with R1 and S1 chiral amines. R1 ligands are represented with green and cian while S1 ligands are represented with pink and magenta.

The data indicates that the optimized poly-1-amine complexes obtained (Figure 4.5) have the ion-pair interaction between amines and polymer as the main stabilizing interaction, which is in agreement with the high number of H-Bond obtained in our previous MD simulations. The values of the dissociation energy of each complex exhibit an affinity trend of R1 and S1 for CW and CCW helical senses, respectively (panels **a** and **b** of Figure 4.5) with values available in appendix 7.1.2. Such results are intuitive since the dissociation energy values obtained are directly related to the exchange of the methyl groups in the chiral center, meaning that the bidentate-type interaction becomes geometrically less favorable. This agrees with the distance values of hydrogen–oxygen displayed in Table 4.1, where the second H-Bond performed by ligands S1-1 and S1-2 in the CW helical sense and particularly with R1-1 and R1-2 in the CCW helical sense increases compared with the lowest energy complexes found.

Table 4.1: Values of hydrogen bonds between chiral amines and 4 monomer poly-1. (a) Main H-Bond with the monomer. (b) Second H-Bond with the next neighbour monomer.

Amine-monomer	Helix screw-sense	Type of bond distance	First H-Bond[a] ($\mathring{A}$)	Second H-Bond[b] ($\mathring{A}$)
R1-1	CW	N $\cdots$ H	1.11	1.06
		C-O– $\cdots$ H	1.49	1.68
R1-2	CW	N $\cdots$ H	1.11	1.06
		C-O– $\cdots$ H	1.48	1.69
R1-3	CW	N $\cdots$ H	1.12	-
		C-O– $\cdots$ H	1.48	-
R1-4	CW	N $\cdots$ H	1.12	-
		C-O– $\cdots$ H	1.48	-
S1-1	CW	N $\cdots$ H	1.15	1.05
		C-O– $\cdots$ H	1.40	1.77
S1-2	CW	N $\cdots$ H	1.14	1.05
		C-O– $\cdots$ H	1.40	1.77
S1-3	CW	N $\cdots$ H	1.14	-
		C-O– $\cdots$ H	1.44	-
S1-4	CW	N $\cdots$ H	1.14	-
		C-O– $\cdots$ H	1.40	-
R1-1	CCW	N $\cdots$ H	1.56	1.02
		C-O– $\cdots$ H	1.05	2.21
R1-2	CCW	N $\cdots$ H	1.58	1.04
		C-O– $\cdots$ H	1.05	1.93
R1-3	CCW	N $\cdots$ H	1.64	-
		C-O– $\cdots$ H	1.04	-
R1-4	CCW	N $\cdots$ H	1.58	-
		C-O– $\cdots$ H	1.05	-
S1-1	CCW	N $\cdots$ H	1.13	1.05
		C-O– $\cdots$ H	1.43	1.72
S1-2	CCW	N $\cdots$ H	1.10	1.06
		C-O– $\cdots$ H	1.52	1.68
S1-3	CCW	N $\cdots$ H	1.11	-
		C-O– $\cdots$ H	1.54	-
S1-4	CCW	N $\cdots$ H	1.11	-
		C-O– $\cdots$ H	1.50	-

Experimental results from Maeda et al. [96] using titration and IR spectra analysis showed ion-pair interactions play a major role in the helical induction of poly-1. Their conclusions are in good agreement with our DFT results, in particular when analyzing H-Bond distances of the CCW-R1 and CCW-S1 systems. The H-Bond distance values obtained for acceptor-proton (C–O⋯ H) for CCW-R1 are lower than for CCW-S1, hence the nature of the amines and carboxylic groups is closer to neutral species for CCW-R1 than ionic like in the case for CCW-S1.

On the one hand, the differences in the absolute values of ΔE_{diss} obtained are explained by the differences in the H-Bond interactions. In the case of the CCW-R1 complex the differences also result from the aromatic interaction of R1-2 and R1-4 ligands being lost, see panel **d** of Figure 4.5. The dissociation energy shows a clear trend in the affinity of R1 and S1 for the clockwise and counterclockwise screw-senses of poly-1, respectively. However, it does not point out where this interaction happens in the 4 monomer polymer-amine complexes. To clarify this point, we performed a Bader charge analysis to determine the nature of the interaction in polymer-amine in the complexes (Bader charges are available on Appendix 7.1.3). The values obtained for the 4 monomer poly-1 for all the complexes are illustrated in Figure 4.5 with a color scale. Figure 4.5 shows that the Bader charge distribution along with the complexes that R1 and S1 influence different effects over the backbone. We observe that R1 promotes a more significant charge in the backbone carbons of poly-1 than S1. Since this effect is observed in both helix senses, it could contribute to the stabilization of the complex or be part of the helical induction.

On the other hand, the values obtained for the chiral carbon, nitrogen, and oxygen involved in the ion-pair interaction between the polymer and the amines showed significant differences, confirming that the local effects play a prominent role in the stabilization. The values summarized in Table 4.2 show opposite trends when we compare amines in the same helical sense. One notes that nitrogen has a significant charge in CW-R1 and CCW-S1 and the opposite trend in the case of oxygen. These phenomena could be possible to the differences in the cooperative H-Bond performed by the amines explained before. It is well-known that the strength of acids and bases could be affected by aprotic polar solvents like DMSO [137]. However, the basicity of the amines could also be affected by the interaction displayed with different screw-senses of the polymer reflected in the values of the Bader charges, as shown in Table 4.2.

Table 4.2: Bader charge values for atoms involved in the ion pair(chiral carbon, Nitrogen and Oxygen).

Atoms	CW-R1	CW-S1	CCW-R1	CCW-S1
Chiral Carbon	-0.83	-1.05	-1.28	-0.87
Nitrogen	7.13	6.98	6.40	7.00
Oxygen	9.06	9.25	10.70	9.12

4.4 Conclusion

The helical induction on dynamic helical polymers such as the PPA polymer family has been extensively reported in the literature but not deeply understood. MD simulations cannot describe the helical induction nature due to its limitations, as we previously discussed. Despite our MD trajectories not being able to provide useful information regarding the helical induction, the data obtained could provide useful information for future investigations in the field of high dynamic helical polymers. On the other hand, our DFT calculation results show that charge, chiral nature, and spatial orientation are crucial in helical induction because they enhance or diminish the bidentate-type interactions. This agrees with the experimental results discussed by Yashima et al. [88] where they tried different chiral center positions away from the amino group yielding unclear ICD signals. Structural analysis of the contact interactions between the amines and polymer and Bader charge analysis gave essential insights to understand the nature of these complexes' screw sense, affinity, and stabilization. The findings in this work will be of interest to understand other dynamic helical polymers that experimental techniques struggle to probe.

5   Helix Modeling On Poly(phenylacetylene)s Bearing Activated Ester Moieties

5.1   Introduction

In many cases, different stereoisomers of the same compound can have very different physico-chemical and biochemical properties, and it can be challenging to separate different racemic mixtures as the molecules have the same chemical weight and functional groups. Having a method to recognize the chirality of a compound could have critical applications in the discovery of pharmaceutical drugs or the detection of agrochemical substances [138, 139]. Poly phenylacetylenes (PPAs) are a family of highly dynamic helical polymers, which under specific conditions, such as in the presence of chiral molecules, can transition between different helical conformations via induction.
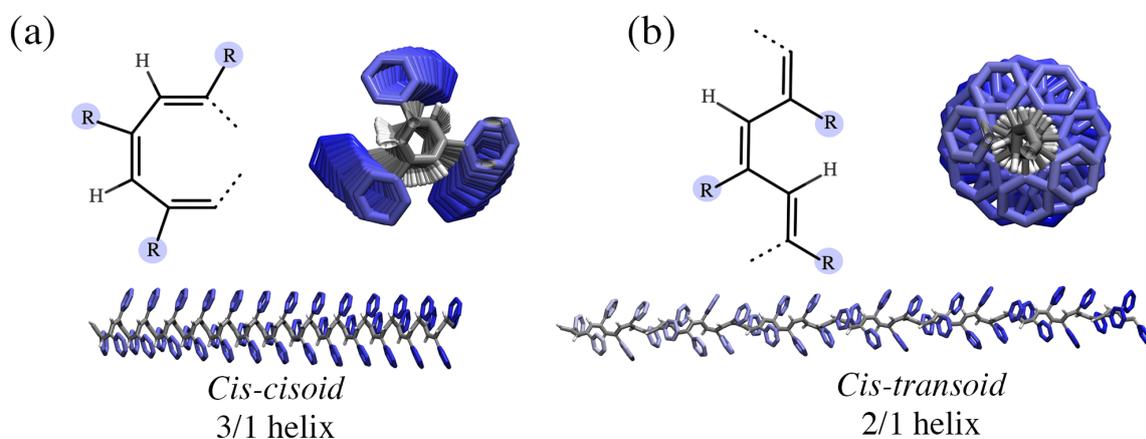
Figure 5.1: (a) Cis-cisoid 2D and 3D representation (b) Cis-transoid 2D and 3D representation.

PPAs are typically π-conjugated systems which tend to adopt helical *cis-transoid* or *cis-cisoid* conformations (see Figure 5.1). An experimental study by Rodriguez et al. revealed that para-, meta- and ortho- substituted PPAs displayed remarkable helical stability differences [140]. They concluded that PPAs exhibit different degrees of dynamicity and helical inversion barriers, which depend on the position of the substitution on the phenyl ring. Para-substituted PPAs showed the highest dynamic behavior where the most persistent helical structures were *cis-transoid* conformations. Meta-substituted PPAs were able to adopt *cis-cisoid* and *cis-transoid* conformations which coexisted in equilibrium. Lastly, ortho-substituted PPAs were difficult to obtain, with the likely explanation being steric hindrance at the ortho-position. Their configurations were found to be nearly planar, and they were one of the most stretched helices reported.

There are a few different strategies to induce helical conformations in PPAs. One of them is post-polymerization modification [141]. This approach involves replacing the leaving groups in the PPA polymer with functional groups such as chiral amines that could induce a helical conformation [142]. Prof. Theato's group at KIT has been researching post-modification strategies using PPA polymers with pentafluorophenyl groups (PFP) attached [143–145].
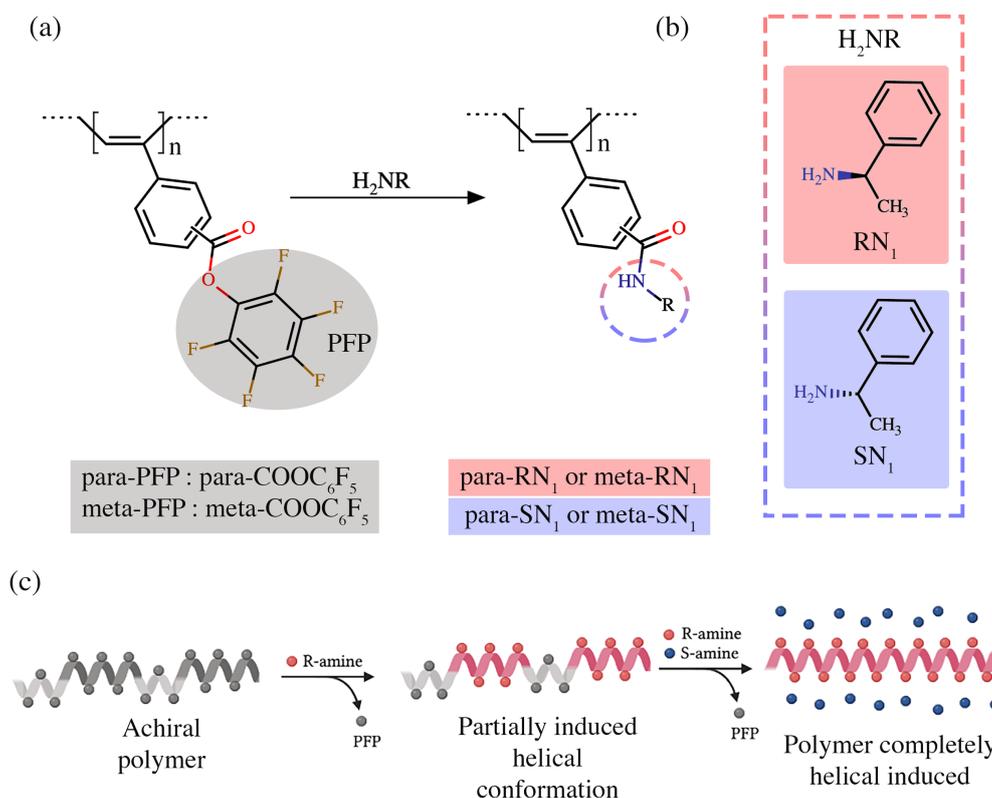

Figure 5.2: (a) Reaction scheme, (b) chiral amines studied and (c) polymer-amine induction scheme.

Theato's group partially post-modified the structures of para-PFP and meta-PFP, and then they mixed the partially induced polymers with chiral amine mixtures of R- and S- 1-phenylethan-1-amine (RN_1 and SN_1, respectively) as is illustrated in Figure 5.2 (c). Surprisingly, the characteristic Induced Circular Dichroism (ICD) signal was observed in the first induction step. Then, after the reaction with the racemic mixture of amines, the ICD signal intensified, leading us to conclude that the polymer prefers to bind the chiral amine that was used in the previous step. The sergeants and soldier's effect can explain this apparent selectivity observed in other helical inductions on PPAs [84].

Due to the dynamic nature of this polymer family, it is imperative to obtain theoretical helical models to gain more robust insights into the helical induction mechanism. In this work, we present an autonomous model PPAs polymers with para- and meta- substitutions inspired by the polymers from our experimental partners from Prof.Theato's group.

5.2   Computational Details

Helical models based on the structure of PPAs structures with both *cis-cisoid* and *cis-transoid* were obtained using a combination of dihedral scan calculations using DFT and Monte Carlo simulations. We created two workflows in SimStack to obtain helical candidates. The first is shown in Figure 5.4 (a).

5.2.1   Backbone Dihedral Scan Studies

Here we wish to explore the possible dihedral values on the backbone of the polymer using SIMONA [75], TURBOMOLE [76], and SIMSTACK [74]. The WaNo SIMONA-DHscan was built to generate molecular structures using SMILES code [146] as input. Then a specific SIMONA [75, 147] scan protocol was individually applied to all possible torsion angles identified. The scan protocol uses the metropolis MC algorithm to perform arbitrary rotations of a selected torsion angle and apply relaxations on adjacent torsion angles. The total energy of each configuration is calculated using the Coulomb, and Lenard-Jones terms from the General Amber forcefield (GAFF) [122].
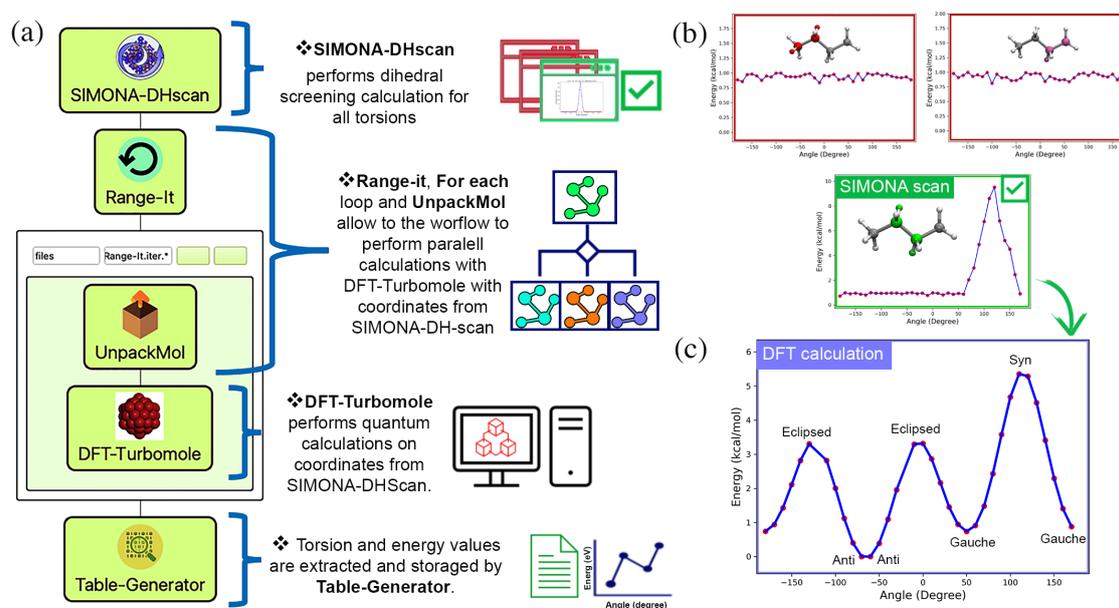

Figure 5.3: Dihedral scan workflow implemented in SimStack [74](caption from publication)

Each scan calculated a score based on the difference between its minimum and maximum energy values. The sorted values are according to their score and dihedral. The largest energy difference is selected as the main dihedral that influences the conformations of the molecule. The coordinates generated from SIMONA-DHscan were used to feed multiple parallel DFT-TURBOMOLE calculations by using the DFT-Turbomole WaNo. The DFT-calculations were performed considering the hybrid B3LYP functional [106, 107] and def2-SV(P) basis set [108] and the D3-BJ dispersion correction [111, 112]. Using *Range-It* and *UnPack* WaNos it is possible to perform parallel DFT calculations. Finally, the data was collected using a Table-Generator. The final dihedral profile obtained reveals local minima and maxima, as can be seen in Figure 5.3 (c).

5.2.2   Helical Modeling Workflow

A second workflow was designed to build polymer structures and perform helical modeling protocols on polymer chains. The WaNos Polymer-Builder, SIMONA-HelixModeling, and SIMONA-Analyser are used in the order illustrated in Figure 5.4 (a). Polymer-Builder is a protocol that performs two tasks. First, it creates monomer libraries from a PDB file of a minimal subset of the polymer (3-4 monomers) and lists the atom descriptors, such as indexes and names. The monomer libraries are constructed using Ambertools [121] to calculate charges with the AM1-BCC method [123]. Now that the monomer libraries have been obtained, in the second step, Polymer-builder creates polymer chains. Different options are available, such as homopolymers, block polymers, or random copolymers. Then, inputs for Amber, GROMACS, and SIMONA are generated, which can be used to perform any further simulation steps, for example, MD or MC simulations. SIMONA input files are taken from Polymer-Builder by SIMONA-HelixModeling WaNo, where the user can select different options for the *torsion moves* to be performed.
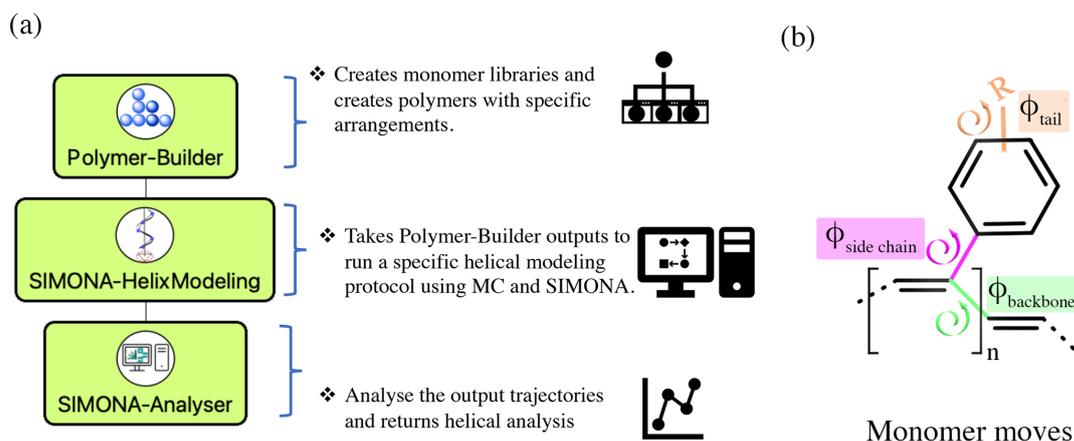

Figure 5.4: (a) Helical Modeling workflow implemented in SimStack [74]. (b) MC moves applied on each polymer monomer, backbone moves (green), side chain moves (magenta), side chain tail moves (orange).

The SIMONA *torsion moves* allowed are arbitrary, relative, and random moves. Arbitrary moves are structural transformations that can be applied on dihedral angles to set a specific torsion value. Relative moves are sequential transformations iteratively setting a specific torsion value, for example, a dihedral scan. Relative random moves are Monte Carlo random moves that are accepted or rejected under the Metropolis energy criteria [69]. In this work, the dihedral torsion in the polymer structure are identified by SIMONA (assigning them an Id) and then sorted by backbone torsions ($\phi_{backbone}$), side chain torsions ($\phi_{sidechain}$), and tail torsions (ϕ_{tail}) as are shown in Figure 5.4 (b).

Two main protocols/screening steps were applied. In the first protocol, the dihedral angles are moved in the following sequence:

1) A Dihedral scan of 36 steps every 10 degrees are performed only on $\phi_{backbone}$ (Figure 5.4 (b) green bond).

2) For each backbone dihedral move, a scan of 36 steps of 10 degrees is applied to $\phi_{sidechain}$ (Figure 5.4 (b) magenta bond).

3) For each side chain dihedral scan step, a relaxation on all the torsion of ϕ_{tail} are performed using 10^5 steps of relative random MC moves at 300K (Figure 5.4 (c) orange bond).

Different protocols can be combined to explore the rich space of different possible helical polymers. Using the above approach moves on $\phi_{backbone}$, $\phi_{sidechain}$, and relaxations of ϕ_{tail} are done together in the same loop rather than sequentially, leading to a much more reasonable shape than if these three parameters were changed independently. The second protocol is derived from the first one, with the main difference being that the backbone dihedral scan is not performed, and a specific dihedral value is set to perform a local helical search. Finally, the trajectories obtained are analyzed by the SIMONA-Analyser WaNo to extract the minimum energy conformation and $\phi_{backbone}$ and $\phi_{sidechain}$ dependence plot (similar to Ramachandran plot on proteins [148, 149]).

5.3 Results and Discussion

5.3.1 DFT Dihedral Scan

To evaluate the torsion values on the backbone of the PPAs studied (para-PFP, para-RN$_1$, para-SN$_1$, meta-PFP, meta-RN$_1$ and meta-SN$_1$) that could lead to the *cis-cisoid* and *cis-transoid*, DFT scan calculations on dimers of all the polymers were performed. Using the Dihedral-Scan workflow on SimStack, the dihedral profiles were obtained as are showed in Figure 5.5.
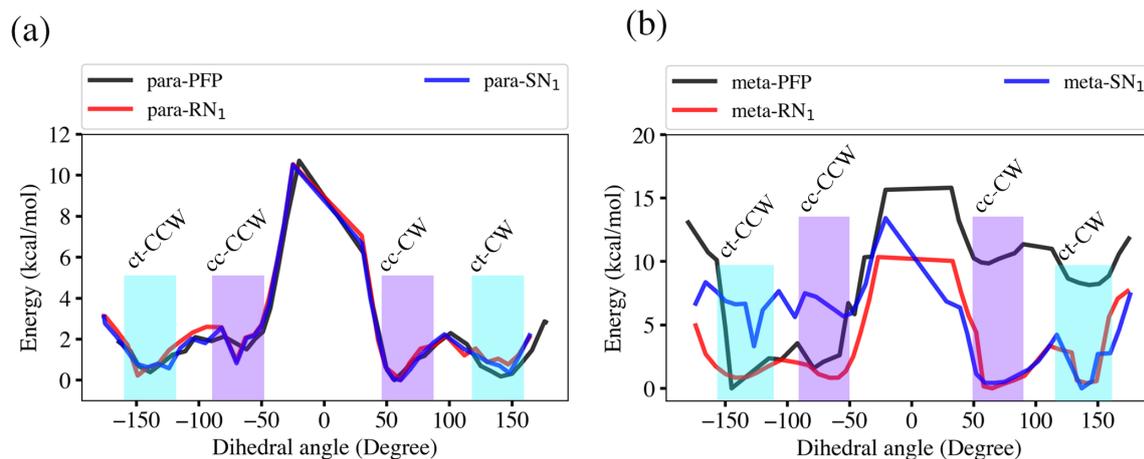

Figure 5.5: (a) para-PFP, para-RN$_1$, para-SN$_1$ dimer dihedral profiles. (b) meta-PFP, meta-RN$_1$ and meta-SN$_1$ dihedral profiles. In both profiles purple and cyan zones are *cis-cisoid* and *cis-transoid*, respectively.

Similar to our previous work on the induction of poly(4-carboxyphenyl)acetylene by chiral amines [150], the para-substituted dihedral profiles displayed in Figure 5.5 (a) formed helices with both clockwise(CW) and counterclockwise(CCW) screw-senses. The torsion values for *cis-cisoid* conformations are ca. 61 ° (cc-CW) and -61° (cc-CCW), while the values obtained for *cis-transoid* are ca. 147 ° (ct-CW) and -147° (ct-CCW).The dihedral profiles obtained for meta-substituted dimers are shown in Figure 5.5 (b). Again we can see similar zones where *cis-cisoid* and *cis-transoid* can be identified (cyan and purple boxes). With the meta position overall, there tended to be less symmetry in the side chains which can be seen by observing the larger local variations in the dihedral profiles. Despite the energetic differences between para-substituted and meta-substituted profiles, the torsion values obtained for both substitution types are still very close to one another (all dihedral values can be found in appendix 7.2.1).

5.3.2   Blind Helical Search

The helical modeling workflow implemented in SimStack was designed to perform "blind" helical searches as a first screening protocol. By blind, we mean we do not have to provide any preliminary information, for example, specific torsion angles to search. This searching method applied in 40 monomer models of para-PFP, para-RN$_1$, para-SN$_1$, meta-PFP, meta-RN$_1$, and meta-SN$_1$ were generated using Polymer-Builder WaNo, and the helix modeling protocol. First, we perform a sequential torsion scan on backbone dihedrals ($\phi_{backbone}$ moves) while also performing a scan on the adjacent side chain dihedral ($\phi_{sidechain}$ moves) and relaxing the rest of the side chain dihedrals with metropolis MC (ϕ_{tail} moves). The dependence of $\phi_{backbone}$ and $\phi_{sidechain}$ was studied for all the trajectories obtained. The helical structures were sorted according to an energy cut-off of 500 kcal/mol. This cut-off may seem high at first glance when thinking about the cut-offs one might use for an individual reaction, but here we are considering 40 monomers. In this scenario, the cut-off energy per monomer is more reasonable. Applying this criterion to 11 million potential conformations were reduced to around 10^4 conformations. The helical conformations obtained exhibited torsion values in the range of -150 to -50 (results are available in appendix 7.2.2). Surprisingly, the lowest energy helical conformations obtained all exhibit CCW conformations. This could be related to the parametrization of the polymer monomers, which were built from CCW tetramer structures.
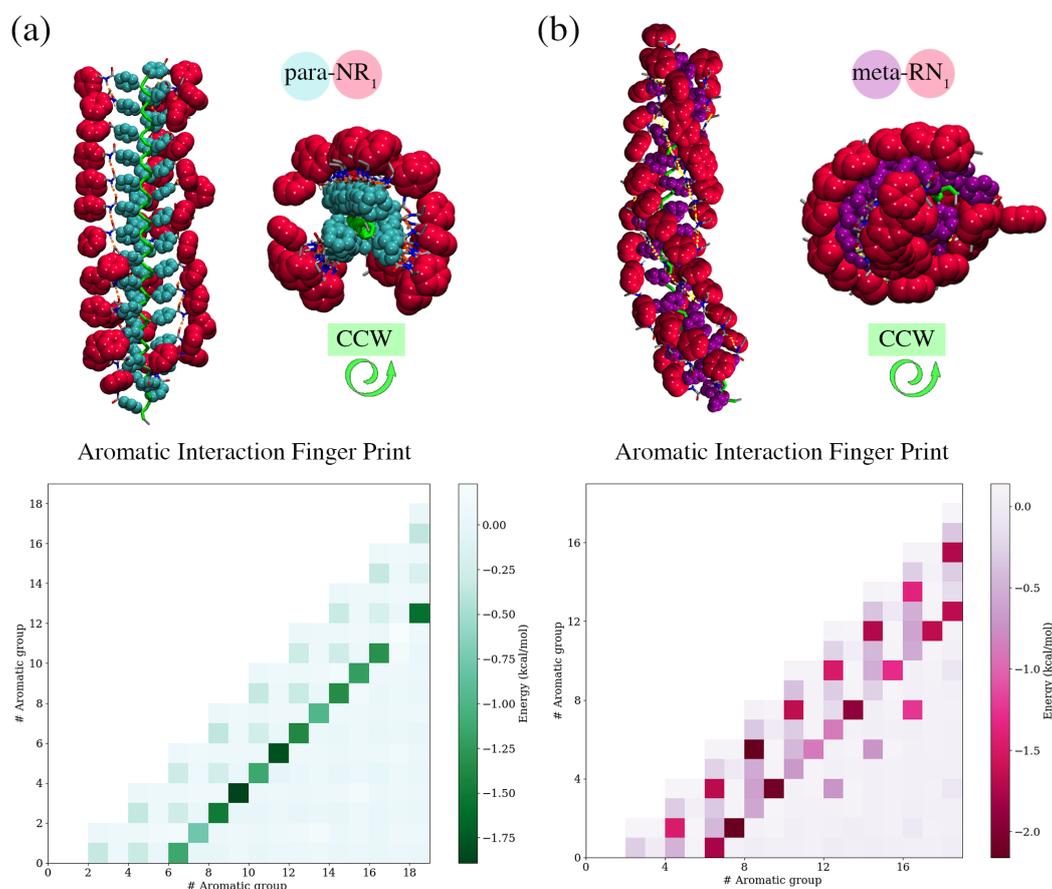

Figure 5.6: Lowest energy conformations found using blind helical search for (a) ccw para-RN$_1$ and (b) ccw meta-RN$_1$ with their corresponding aromatic interaction fingerprints.

The top part of Figure 5.6 (a) and (b) show the lowest energy conformations obtained for para and meta substitutions after a slight relaxation. Both helical conformations correspond to CCW screw-sense with RN$_1$ attached (only the phenyl ring of the amine is shown, and it is colored red). The remaining part of the amine can form stabilizing H-bonds interactions with a large variety of different orientations. In the bottom part of the Figure, it is possible to observe their corresponding aromatic interaction fingerprint profile. Para-RN$_1$ shows a clear pattern of aromatic interactions, with the most substantial interactions due to the phenyl rings attached to the backbone (colored dark green). These rings face one another in each turn of the helix and stabilize each other via $\pi - \pi$ stacking interactions. On the other hand, the phenyl groups belonging to the amine (colored in red) exhibit much weaker interactions, as can be seen from weaker coloring in the fingerprint profile. On the other hand, helical meta-RN$_1$ exhibits similar aromatic interactions for the inner phenyl rings, with the same strong diagonal line in the fingerprint pattern being visible as the Para-RN$_1$ system. The meta-RN$_1$ system also exhibits aromatic interactions between the outer phenyl rings with the next neighbor's inner phenyl ring in a T-shape type interaction.

5.3.3   Local *Cis-transoid* Helical Search

Casting our eye back to Figure 5.3, the most stable structure in the scan was *cis-cisoid* and the blind search tends to result in structures that lie within this zone rather than in the *cis-cisoid* zone. The conformations obtained in the range of the *cis-transoid* tended to be intermediate structures with weaker H-bond interactions than expected. Furthermore, several meta-stable states during the blind sampling increased the challenge of finding *cis-transoid* conformations. For these reasons, we modified our original helical modeling protocol so that moves on $\phi_{backbone}$ were no longer considered and set the torsion backbone values to the ones obtained for *cis-transoid* from the dimer dihedral scans. This way, the *cis-transoid* conformation search is biased specifically to this zone of the dihedral profile. All the possible combinations of screw-sense and para and meta substitutions were sampled. The lowest energy configurations obtained that display the expected H-bonds interactions are displayed in Figure 5.7. The only structures that showed helical conformations with the stabilizing H-bonds were CW para-SN$_1$ and CW meta-RN$_1$ (Figure 5.7 (a) and (b), respectively). Similarly to the previous section, the aromatic interaction profile was obtained and showed similar trends to the *cis-cisoid* helices. For the para-RN$_1$ structure, we see weaker aromatic interaction strength. Surprisingly, this search mainly yielded CW conformations, whereas the blind helical search gave mostly CCW conformations (7.2.2). Our initial expectations were to find both helical screw senses. In the case of the para substitution, the most stable polymer was the para-SN$_1$ conformer. These results suggest that the induction could lead to different screw-senses depending on *cis-cisoid* and *cis-transoid* conformations, and this is intimately influenced not only by the chiral amines but also by the para and meta substitution.
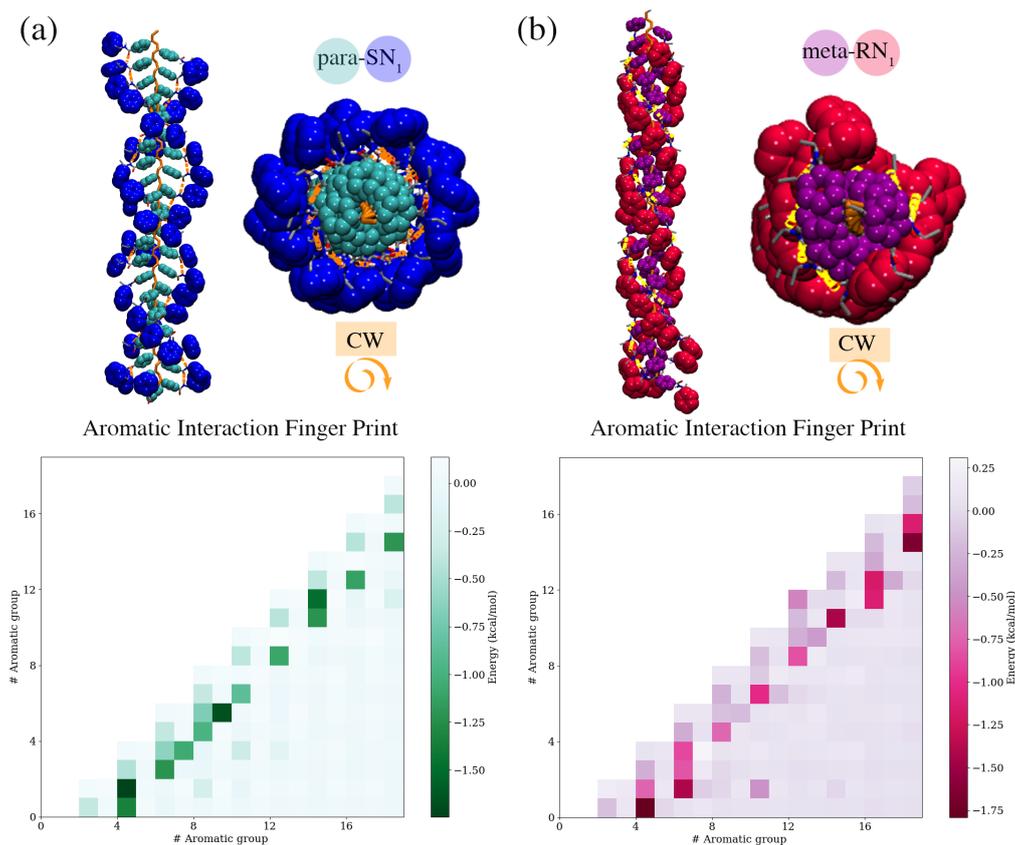
Figure 5.7: Lowest energy conformations found using biased helical search for (a) cw para-SN_1 and (b) cw meta-RN_1 with their corresponding aromatic interaction fingerprints.

5.4   Conclusion

Using combined MC and DFT dihedral scans *cis-cisoid* and *cis-transoid* backbone torsion values were obtained for CW and CCW conformations for a polyphenylacetylene dimers. Next, blind helical searching was done using MC helical modeling protocol leading to *cis-cisoid* conformations. The most stable conformation found was meta-RN_1, stabilized by three types of interactions: H-bonds, face-to-face, and T-shape aromatic interactions. On the other hand, the *cis-cisoid* para-substituted polymer is stabilized by H-bonds between amide groups and face-to-face aromatic interactions between repeated monomers in the helical loop. Overall the spread in energies between the most stable structures found for all classes is relatively small, which provides a possible explanation for why Rodriguez et al. [140] observed *cis-cisoid* and *cis-transoid* in equilibrium in the solution state in their experimental work. Lastly, biased helical modeling combined with MC simulations on side chains was undertaken to try and yield more *cis-transoid* helices. Both screw senses were obtained in this case, with their aromatic interaction fingerprint profiles exhibiting similar but weaker patterns. Obtaining stable helices for highly dynamic helical polymers is challenging, with a large conformational space to search. Many different states with small differences in energy can be found, with the search being further obscured by the presence of meta-stable structures. This study aimed to provide a protocol to search for stable helical polymer conformations which can straightforwardly be applied to a much larger family of helical systems.

6   Sequence-controlled Oligomers: Designing Helical Oligomer Structures

6.1   Introduction

Biological macromolecules such as proteins and DNA are well-known as sequence-defined biopolymers, which have various functions. They are essential for structural support, recognizing compounds, catalysis, and information storage [36]. Inspired by nature, sequence-defined oligomers (SDOs) are molecules with a specific monomer order. They are a new class of polymers that combine the classic definition of a polymer (many monomer repeating units) with features such as monodispersity, chirality, and conformational topology [151, 152]. There has been increasing interest and efforts to synthesize SDOs, with many studies having been published [153–158].

In recent years, the concept of digital polymers has been proposed. Inspired to some degree by biomolecules such as DNA, digital polymers are sequence-controlled polymers that can store binary information. They can potentially have high storage capacity and could conceivably have potential large-scale data storage applications. A notable example is a recent work of Dahlhauser et al., who recently published their efforts to synthesize oligomers with isotope labeling [159]. Within a letter to colleagues, they hid a 256-bit encryption key to unlock a text file of the book "The wonderful wizard of Oz".
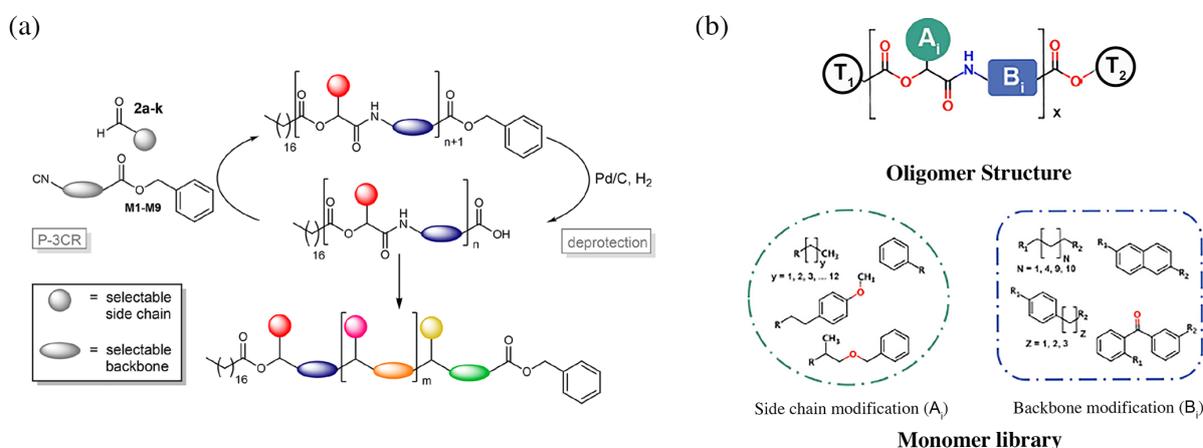

Figure 6.1: (a) Synthesis strategy for dual sequence-defined oligomers. The two-step iterative cycle allows for independent variation of side chain and backbone. Figure reproduced from publication of Wetzel et al. work [160] (b) Scheme of simplified oligomer model and possible library of monomer modifications.

While many experimental polymer scientists aim to obtain high yields during sequence-controlled synthesis and to find efficient decoding protocols, there is a lack of information regarding the 3D arrangement of the oligomer synthesized. Traditionally, techniques such as NMR, atomic force spectroscopy (AFM), and dynamic light scattering are applied to determine structural information about polymers [41]. These techniques can only capture specific insights into a much broader picture. In particular, they struggle to observe or distinguish between the many different polymer conformations. On the other hand, computational approaches such as Molecular Dynamics (MD) have been used to explore the dynamic nature of sequence-controlled oligomers and the aggregation of multiple oligomer chains [157]. A synthesis strategy using the Passerini reaction by Wetzel et al. (figure 6.1 (a)) led to sequence-controlled oligomers inspired by the genetic code and a protocol to decode their sequences using Mass Spectroscopy (MS) [160]. A dual monomer scheme considering side chains and backbone modifications in order to create binary codes was defined (figure 6.1 (b)). Later, using halogen-based isotope tags, Frölich et al. were able to decode 64.5 bits of information decoding hexamers using MS/MS [161]. As many combinations of side-chain and backbone modifications are possible, synthesizing individual sequences for a given backbone and set of side chains is not feasible as there are nearly 10^{22} candidates.
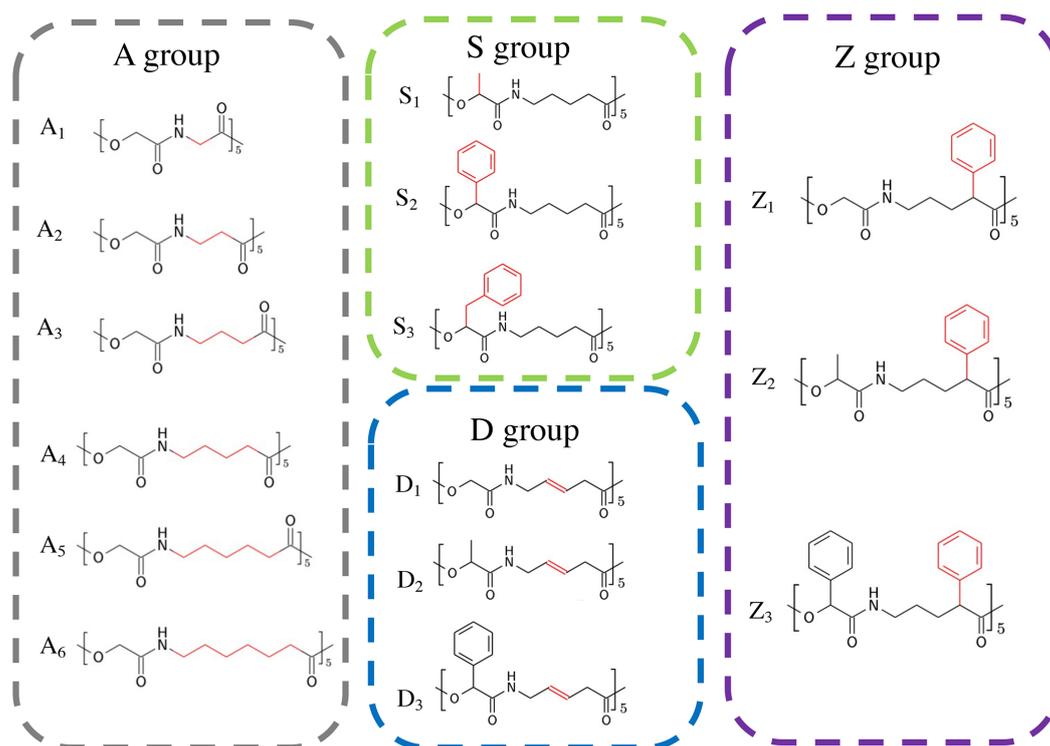

Figure 6.2: Structures of all sequences studied. Group A: Aliphatic group, Group S: side chain modification group, Group D: Double bond backbone modification and Group Z: Phenyl group on backbone modification.

In this work, we simulated oligomer structures based on the work from Wetzel et al. to systematically study different possible conformations on different oligomer sequences. Inspired by proteins, we aimed to design helices since it is a common motif for biopolymers [36]. Our goal is to design stable helical oligomer structures controlling the monomer position and influence in the oligomer chain organization. We used the base structure depicted in (figure 6.1 (b)). To find the optimal length to promote H-Bonds that could stabilize helices, first, we tried different spacings of carbon units on the backbone monomer modification (figure 6.2, A group) carrying out hundreds of Monte Carlo simulated annealing simulations to explore the Potential Energy Surface (PES) and obtained folded conformations. Then different modifications were applied to evaluate the stabilization effects of different groups in specific positions on the oligomer (figure 6.2 showing S, D, and Z sequence groups). Finally, the stability of the best oligomer helical conformations for each sequence was evaluated using Molecular Dynamics (MD) using explicit solvation. We believe that the extensible protocol explained here could help polymer scientist to establish structure-property relationships through rational oligomer design, reducing the prohibitively large search space to feasible sequence candidates for polymer synthesis.

6.2   Computational Details

6.2.1   Oligomer Encoder

To study the possible conformations of oligomer sequences, we created a protocol to systematically create oligomer atomic models from *Oligomer Encoder* (OE) sequences [147]. OE is a tool containing a monomer library of SMILES codes linked to a three-letter code similar to a 3-letter proteins amino acid system (figure 6.3). The OE code system is made in order to create sequences of oligomer chains $\{X_i\}$ (i = 1, . . . , N), where the monomer type is $X_i = A_i B_i$. The head and the tail are Terminal 1 (T_1) and Terminal 2 (T_2), respectively. A_i and B_i are defined as side chains and backbone modifications (figure 6.1 (b)).
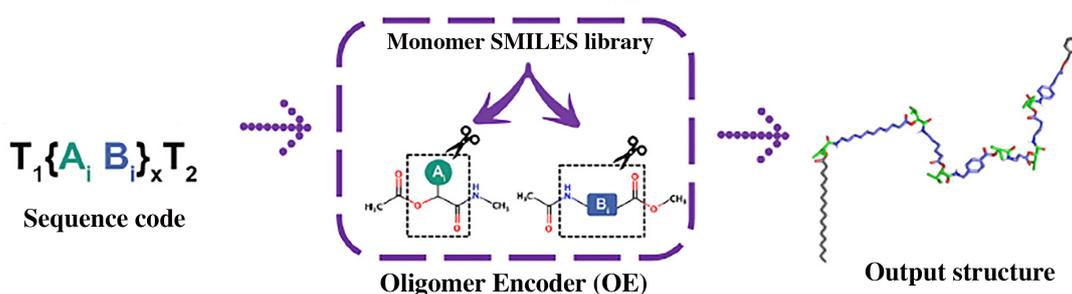

Figure 6.3: OE is able to take a single line monomer sequence and create a 3D sequence atom model.

Internally, OE works in 3 main steps. 1) OE creates monomers atomic models from sequence monomer code input using SMILES code, including small head and tail terminals [146]. These monomers are then optimized at the DFT level using TURBOMOLE [76]. 2) The optimized monomer structures are used to calculate charges using the AM1-BCC method [123] via Ambertools [121]. Parameters from the General Amber FF [122] are extracted with Acpype [162] to generate GROMACS [63] input files. 3) Once GROMACS input files are created, Monte Carlo simulation inputs are generated using the Python

protocol provided by SIMONA [75]. This strategy allows OE to be a highly scalable molecule builder.

6.2.2   Simulation Workflow

After creating all atom-oligomer sequence models using OE and obtaining their parameters, OE outputs are used as inputs for the simulation workflow illustrated in figure 6.4.
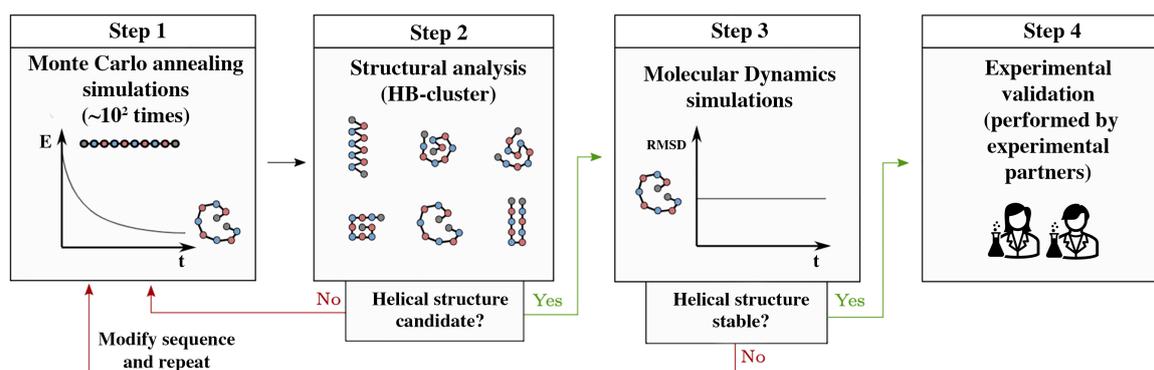

Figure 6.4: Scheme of the simulation Workflow applied. First, 100 MC simulations are performed for a single sequence. Secondly, MC simulations are analysis using the H-bond cluster analysis. The best helical conformations are simulated in the Step3 using MD simulations. If MD simulations reveal stable configurations, the sequences are proposed to our experimental partners.

First, one hundred simulation copies of MC annealing simulation in a temperature range between 900 K to 100 K with one million MC steps are performed (Step 1 in figure 6.4) including implicit solvation using the Generalize Born Method (dielectric constant $\epsilon = 4$). Secondly, the lowest energy conformation of each single MC simulation is analyzed by applying the H-Bond clustering method. Since the entire backbone structure of oligomers studied have amide and ester groups that could display hydrogen bonds (H-Bonds), the H-Bond cluster fingerprint can represent the possible H-bond contacts displayed by the lowest energy conformation in Figure 6.5. This fingerprint is unique for specific conformation, allowing us to identify helical structures quickly.
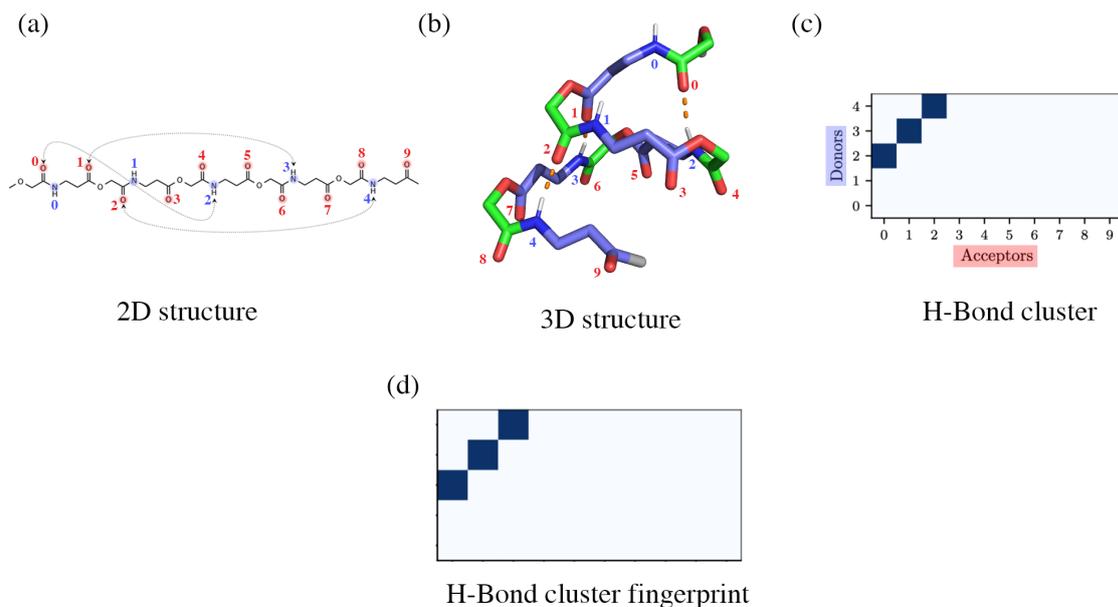
(a)

(b)

(c)

2D structure

3D structure

H-Bond cluster

(d)

H-Bond cluster fingerprint

Figure 6.5: All representations of a sequence structure (a) 2D, (b) 3D , (c) H-Bond cluster and (d) H-Bond fingerprint.

In the third step, the best helical conformations found are simulated using MD using GROMACS 2019 [67, 68]. The systems are solvated with explicit chloroform solvation where the parameters were taken from the database provided by the *virtualchemistry.org* solvent database [124]. The simulations were minimized by using the steepest descent method. Then the systems were equilibrated in two steps, first with a NVT step of 500 ps with 2 fs time step using the V-scale method [126] to keep the temperature constant. Secondly by a NPT equilibration step of 500 ps with 2 fs time step using the Parrinello-Rahman pressure coupling method [65]. The MD simulations were carried out first with 10 ns of production time, including periodic boundary conditions and constant temperature at 300 K under the Isothermal-Isobaric ensemble. Then the MD productions were extended to 100 ns if a stable conformation was obtained via calculating their RMSD, taking the helical confirmation from the first step as a reference. During the simulations 2 fs time step was considered. Electrostatic and van der Waals short-range interactions were calculated with Verlet method [127] with a cut-off of 1.4 nm. The long-range electrostatic interaction was taken care of using the Particle Mesh Ewald method (PME) [128]. Finally, if the sequence candidates present good results for the H-Bond cluster analysis and exhibit stable behavior in the MD simulations, we suggest them as candidates to our experimental partners.

6.2.3   SimStack Workflow

In order to simplify the process of sequence design and automate the search for possible folded conformations, we built a SimStack workflow creating three WaNos (figure 6.6). First, we created a WaNo version of OE to create oligomer all-atom models, just giving a sequence as an input. Then, we used Range-it to create iterative variables to parallelize the simulation of hundreds of copies for the model sequence. This is possible because the control flow tool "ForEach" can take every iteration and execute WaNos inside the loop. Inside the ForEach loop, we created the MC-SIMONA WaNo to run the MC simulation via the SIMONA code. Finally, we created a WaNo called "SIMONA-Analyser" to take the hundreds of output trajectories and perform analysis, e.g., the H-Bond cluster analysis.
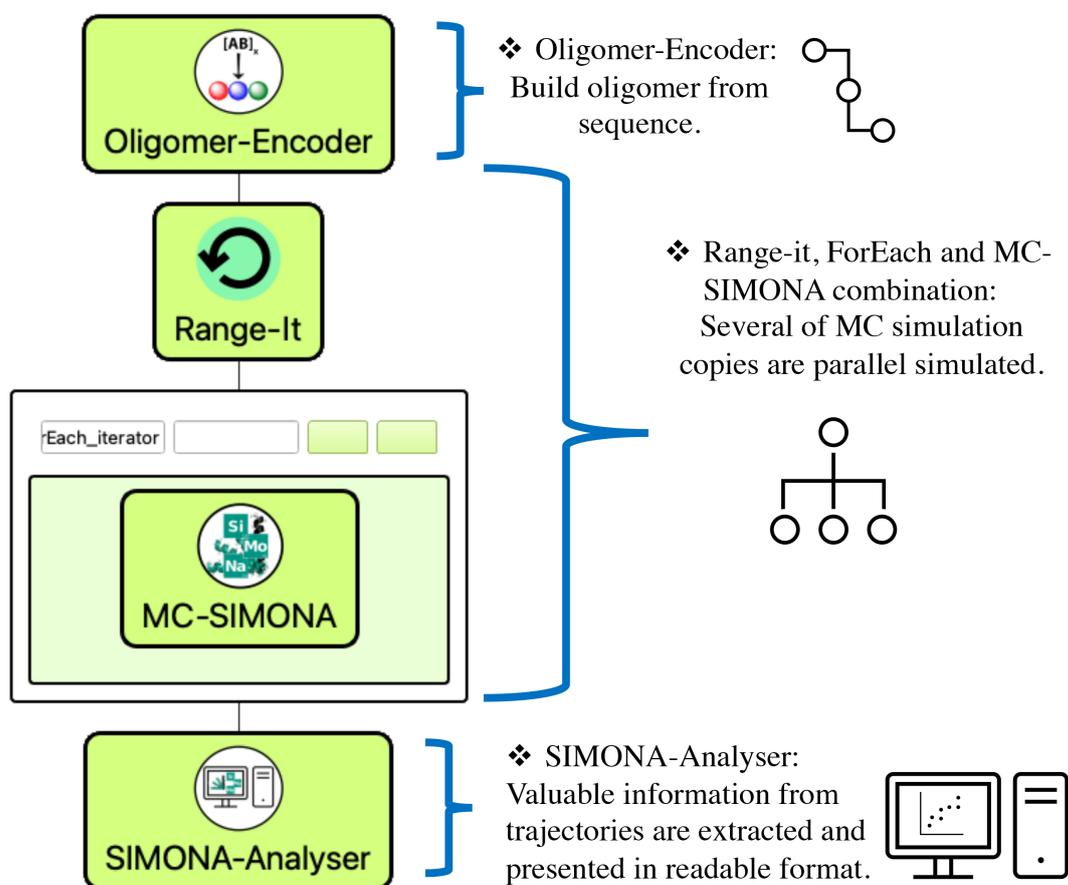
Figure 6.6: SimStack workflow designed to calculate in parallel SIMONA MC simulations fed by OE all-atom sequence models.

6.3   Results and Discussion

6.3.1   Influence of backbone length on helical conformation

The set of all sequences A_n for n = {1,...,6}, S_n for n = {1,2,3}, D_n for n = {1,2,3}, Z_n for n = {1,2,3} considered in this work are shown in Figure 6.2. The unmodified pattern in this kind of oligomer chain is the presence of amide and ester groups due to the Passerini reaction to connect the sidechain and backbone monomers. The resulting hydrogen bonds from these groups are the main stabilizing interaction that could lead to stable helices. Figure 6.7 summarize the results of the MC simulations obtained. Figure 6.7 (a) shows the the sum of all the H-bond fingerprints obtained (isolated H-bond fingerprints are shown in the Appendix 7.3.1). It reveals a clear H-Bonds fingerprint pattern when sequences with an even number of carbons in the backbone monomer modification (A_2, A_4, and A_6) are present. A justification for this is that these sequences have a high H-bond frequency (details of frequency of each cluster in (Appendix 7.3.1) when compared with the uneven sequences (A_1, A_3, A_5).
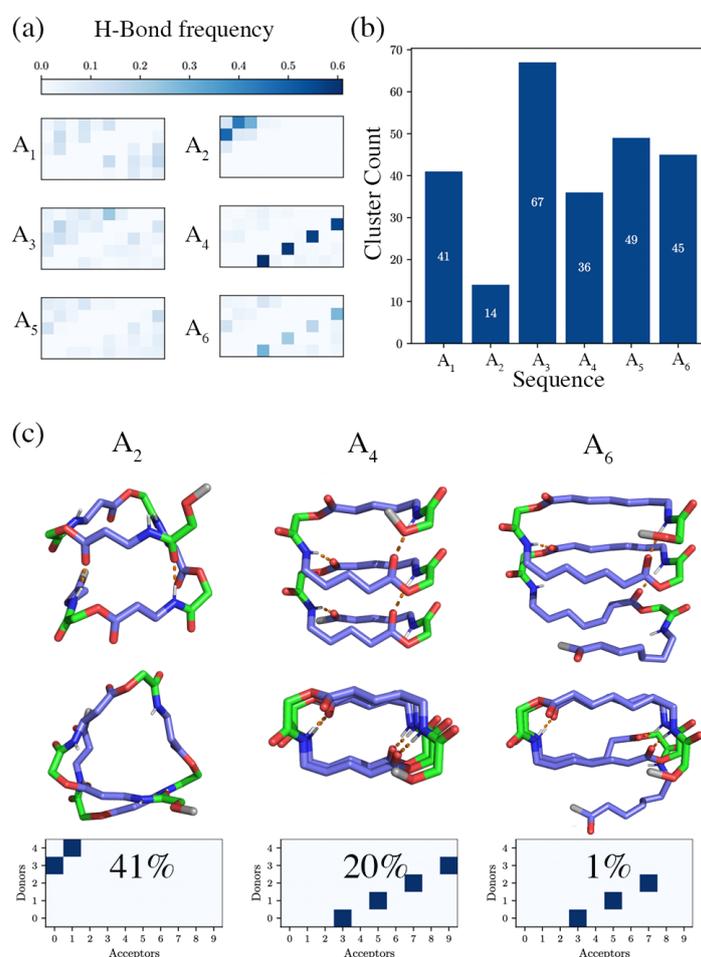

Figure 6.7: (a) H-bond frequencies obtained for helical H-bond fingerprint. (b) Cluster count for all the A group sequences analysed. (c) Structures of the helical conformation found in the A group with their respectively H-bond fingerprint.

The cluster count analysis is shown in Figure 6.7 (b). This shows the number of times a given H-Bond fingerprint is obtained after 100 MC simulations. Sequences A_2, A_4 and A_6 are less prevalent than their uneven counterparts A_1, A_3, A_5 which were not able to lead to a helix conformation. This result is intuitive since oligomers are often observed to behave as random coils rather than a well-defined structure.

The structural conformation for sequences A_2, A_4 and A_6 is shown alongside the corresponding H-Bond fingerprint in Figure 6.7 (c). The conformations exhibit signs of helicity, and the dotted lines indicate when H-Bonds stabilize the structures. Focusing on A_4, it is striking that the formation of H-bonds looks crucial to the formation of the helix. The H-bonds are only performed between amide (H-Bond donor) and ester groups (carbonyl group acceptor), even though amide carbonyl groups are allowed to be H-Bond acceptors. This means that the spacing length and the geometry of the backbone modification play a fundamental role in the helix conformation. Sequence A_2 forms a helix in 41 % the simulations but has a very different conformation to that of A_4 and A_6 which are obtained with frequencies of 20 % and 1 % respectively. Sequences A_4 and A_6 also exhibit a clear diagonal pattern of 4 H-Bonds. Even though A_4 and A_6 share similar H-Bond fingerprints and also a helical conformation (figure 6.7 (c)) A_6 is far less common than A_4.

6.3.2   Structural Modifications On Helical Sequences

The sequence A_4, which presented the best structural features to build a helix, was used as a template to incorporate different modifications to improve the helical conformation's stability. For example, one idea is to put double bonds in the backbone to make it less flexible and reduce how free the oligomer is to access all the conformations. Another idea is to add bulky groups to prefer conformations with a specific steric hindrance. As can be observed in figure 6.2, first we included three side chain monomer substitutions to create sequences S_1 (methyl group), S_2 (phenyl group) and S_3 (Benzyl group). Secondly, we included double bonds modification on the previous S conformation on the backbone modification, namely as D_1, D_2, D_3; finally, we tested adding an extra phenyl group on the backbone modification for sequences Z_1, Z_2, Z_3.

All new sequences were analyzed similarly using the MC protocol and H-Bond clustering. To elucidate the structural properties related to the cluster analysis, we defined five parameters that can be analyzed in Table 6.1. The calculated parameters are the total number of H-bond clusters obtained (N_c), the fraction of conformations that are a helical cluster (H_c), (H_{sub}), and the fraction of conformations that belong to a helical subcluster i.e. their fingerprint exhibits some but not all features its parent cluster, such as the, σ_E, the energy standard deviation of the main helical cluster and the Root Mean Square Deviation ($RMSD$) between the lowest energy conformations.

Table 6.1:   Cluster parameter values obtained from MC simulations.

Sequence	N_c	H_c	H_{sub}	σ_E (kcal/mol)	$RMSD$ (Å)
A_1	41	-	-	-	3.33
A_2	16	0.41	0.48	1.89	2.41
A_3	65	-	-	-	3.75
A_4	37	0.2	0.71	1.74	2.76
A_5	49	-	-	-	4.41
A_6	42	0.01	0.23	-*	3.89
S_1	28	0.54	0.78	2.00	1.89
S_2	29	0.13	0.54	3.34	3.13
S_3	39	0.05	0.58	4.22	3.32
D_1	54	0.14	0.44	1.72	3.06
D_2	39	0.09	0.31	1.40	2.71
D_3	24	0.04	0.36	1.60	3.28
Z_1	29	0.06	0.57	1.31	4.09
Z_2	28	0.43	0.66	2.31	2.02
Z_3	17	0.43	0.82	1.73	1.99

*: A_6 have one main helical cluster of 1%, thus there is no value for σ_E.

Sequences S_1, S_2, and S_3 were studied by performing 100 annealing simulations per sequence, and the structural parameters were obtained. All S sequences could be identified as helical conformations from their respecting H-Bond cluster if their patterns match with the fingerprint pattern obtained for sequence A_4.

A reduction of N_c was observed for S_1 and S_2 while the N_c value of sequence S_3 with the benzyl ring side chain substitution is comparable with sequence A_4. On the other hand, the values of σ_E are reduced for sequences S_1 and S_2 contrary to sequence S_3. These results tell us that with more steric hindrance in this position of the oligomer with the bulky group closer to the backbone, we get an increased probability of obtaining a helix conformation with our MC protocol. The lowest helical energy conformations obtained for sequences S_1, S_2, S_3 are illustrated in figure 6.8. In this figure, we note that the H-Bond pattern is conserved. While sequence $S1$ obtained the lowest amount of clusters, it has the largest value of H_c (more than 50 %). These values suggest that the presence of the methyl increases the likelihood of a helical conformation due to the bulky groups since they prevent obtaining other conformations. In the case of the aromatic side chain modification sequences, it was expected they would perform π stabilizing interactions that can be observed in figure 6.8. However, the effect of the steric hindrance over the oligomer conformation reduces the chance to obtain a helical structure according to the low values of H_c.
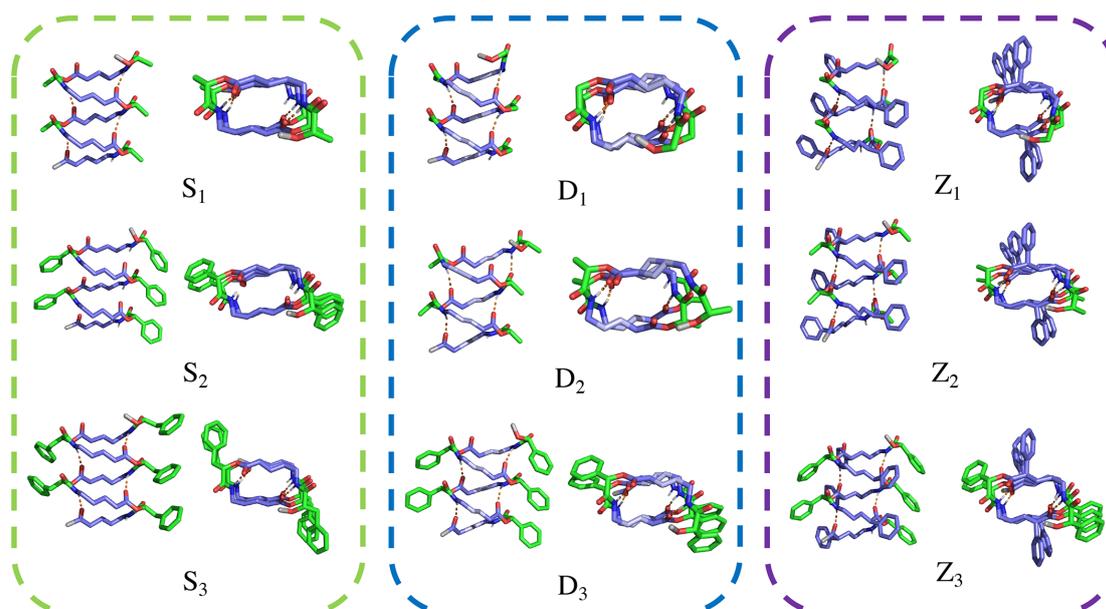
Figure 6.8: Most stable conformations for the S, D and Z groups.

Now lets turn to the D_1, D_2, D_3 groups. Double bond modifications between the second and third carbon of the aliphatic chain of the backbone were considered to evaluate the flexibility of the backbone monomers. This time the presence of the hydrogen, methyl, and phenyl groups in the side chain was considered. We observe a reduced value of N_c when bulky side chain groups are attached. In comparison with the sequence A_4, the results do not improve significantly, and an explanation could be the difference between the possible geometries allowed by C_{sp3} and C_{sp2}. The dihedral angles for the neighboring atoms of nitrogen in the amide groups (C_{sp3}-C_{sp3}-C_{sp3}-N) in sequence A_4 are close to 60°. This is expected since they correspond to one of the local minimum energy values of the C_{sp3} [163], and they promote an optimal orientation for the H-Bond donors in the sequence. On the contrary, the dihedral angles adjacent to the H-Bond donor groups (C_{sp2}-C_{sp2}-C_{sp3}-N) on sequences from the D group are around 16° where nitrogen atom and the third carbon atom are closely clashing leading to a lower strength of the H-Bond interactions (detailed view on figure 6.10). Thus, a lower H-Bond frequency on our simulations (figure 6.9).
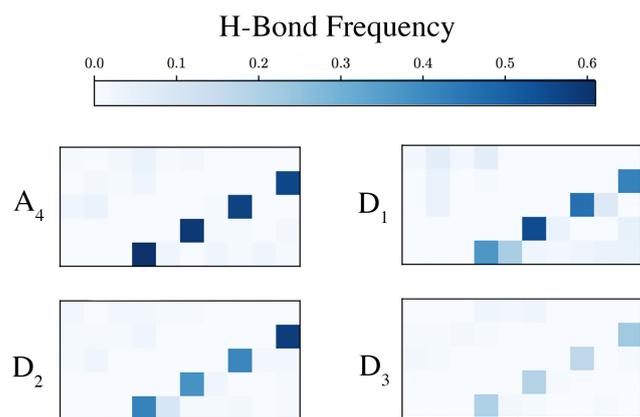
H-Bond Frequency

Figure 6.9: H-Bond frequencies of A_4, D_1, D_2 and D_3 sequences.
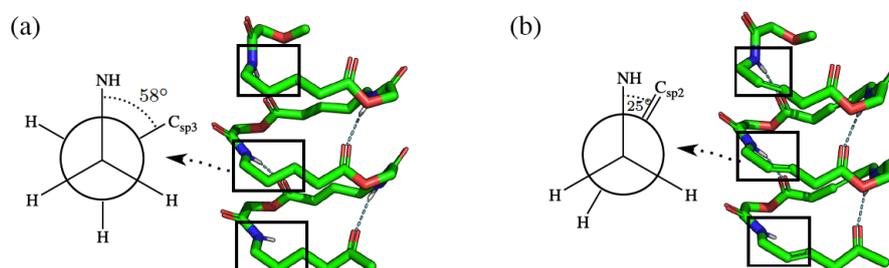
Figure 6.10: Newman projections that illustrate the difference between the adjacent dihedral angle between A_4 (a) and D (b) group sequences.

Finally, the sequences from the Z group incorporate an additional phenyl substitution in the backbone monomer modification. We observe a significant reduction in the values of N_c and H_c compared with the original sequence. The best sequence of this set was Z_3, with the fraction of helical clusters being 43 % and the fraction of helical subclusters being 82 % which exceeds all other sequences simulated. This is a remarkable finding since the phenyl side chain on sequence S_2 decreased both values. This is due to the high number of favorable π-interactions. As is shown in figure 6.8 for sequences of Z group having only side chain or only backbone phenyl groups leads to three off-centered π-interactions each. In particular, combining these features for the sequence Z_3 results in eight additional T-shaped interactions between side chains and backbones. A total of fourteen aromatic interactions can be observed in figure 6.11 where the aromatic rings' interaction energies are displayed and show a specific pattern.
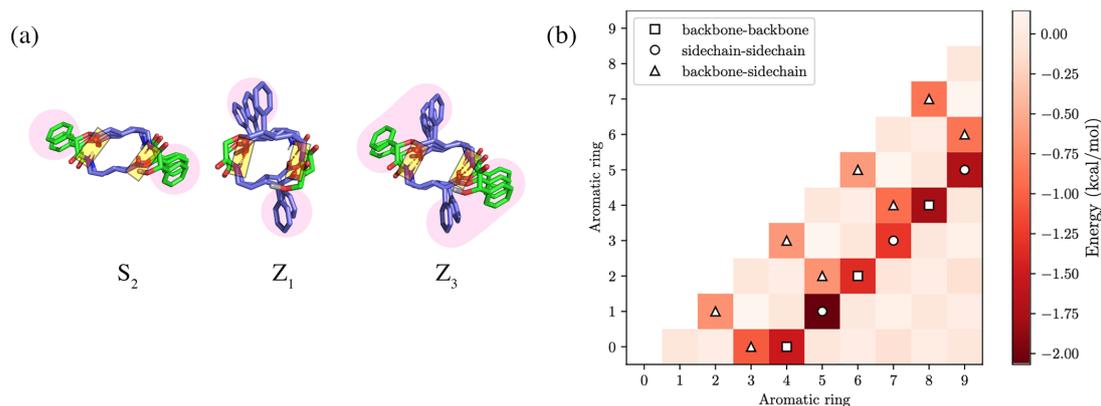
Figure 6.11: (a) Helical conformation obtained for sequences S_2, Z_1 and Z_3 where the main interactions are highlighted. Yellow shading for H-Bond interactions and pink for aromatic interactions. (b) Contact map of aromatic interactions for sequence Z_3. Energies were calculated using the non-bonded energy term from GAFF between aromatic rings. The rings are enumerated in order of occurrence in the sequence. Even numbers correspond to backbone rings, while uneven numbers correspond to side chain rings.

6.3.3   Helical Stability

Since MC simulations are suitable for exploring the PES for probable low energy conformations, MD simulations are widely accepted to evaluate the local stability of the low energy candidates obtained. Helical conformations of sequences A_4, S_1, S_2, S_3, D_1, D_2, D_3, Z_1, Z_2 and Z_3 were solvated in a cubic box of explicit chloroform molecules and simulated with MD simulations at a constant temperature of 300K. Each simulation was repeated 5 times, and the trajectories were followed by calculating the RMSD taking the initial helical conformation as a reference (detail results are in Appendix 7.3.2). To evaluate the stability of the helical conformations, we considered the unfolding time as the last simulation step where the RMSD compared to the initial coordinates is below 2 Å and has not yet exceeded 6 Å. Most of the sequences simulated with MD presented different unfolding time averages 6.12, but the large standard deviation makes it difficult to establish a straight comparison.
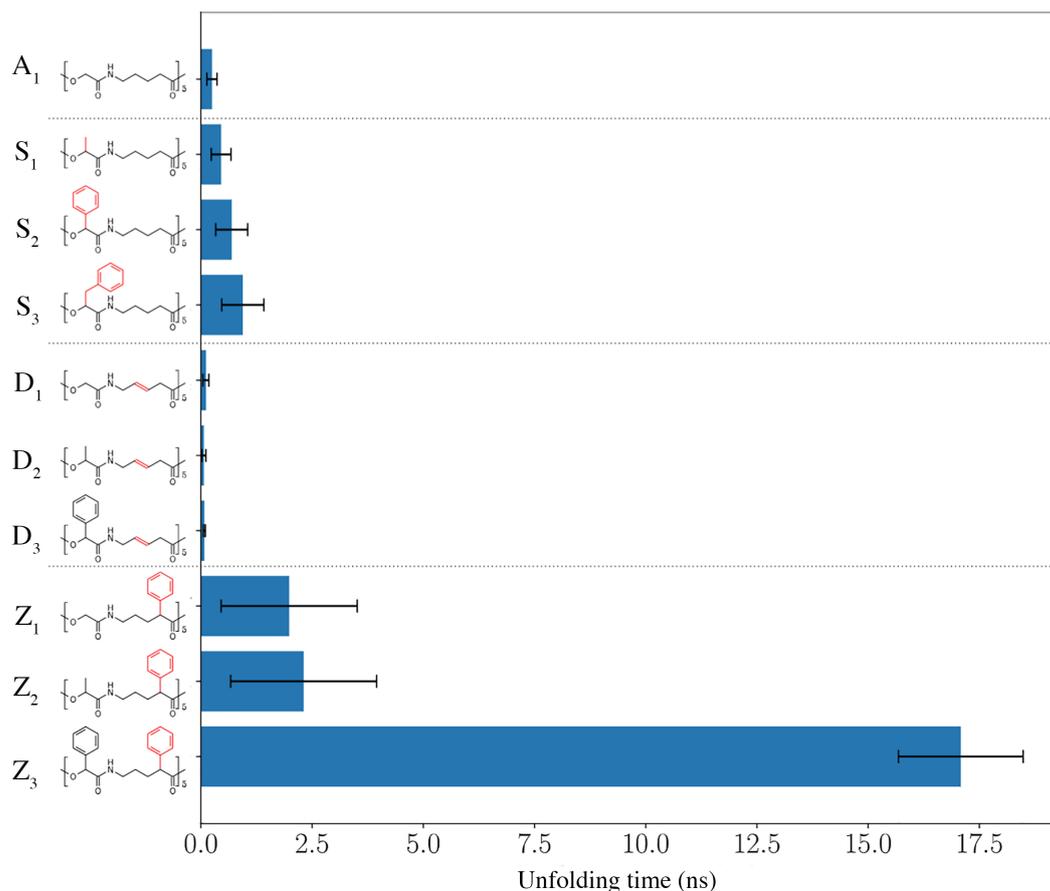
Figure 6.12: Average of unfolding times from MD simulations for all helical sequences.

In general, sequences from the A and D groups unfold much faster than the sequences from the Z group. Surprisingly, in the case of the sequence Z_3, it displays a considerably larger unfolding time (figure 6.12), meaning that the stabilizing interactions are more persistent for this confirmation. The main interactions that characterize Z_3 are the π-π interactions performed by the phenyl ring in the backbone monomers and the phenyl side chain interactions. In Figure 6.13, they are shown as purple and green. Following the trajectory in part (a), it is possible to observe that at 10 ns (figure 6.13) the oligomer is partially unfolded and folded back at 5 ns. Finally, Z_3 is completely unfolded after ca. 20 ns (c) and (d)). This behavior was also observed in the other 4 MD simulation copies. The H-Bonds and the aromatic interactions keep the oligomer partially folded in the helical conformation; when a sufficient number of bonds are broken, the structure unfolds entirely. When the H-Bonds are re-established, the conformation gets packed again.
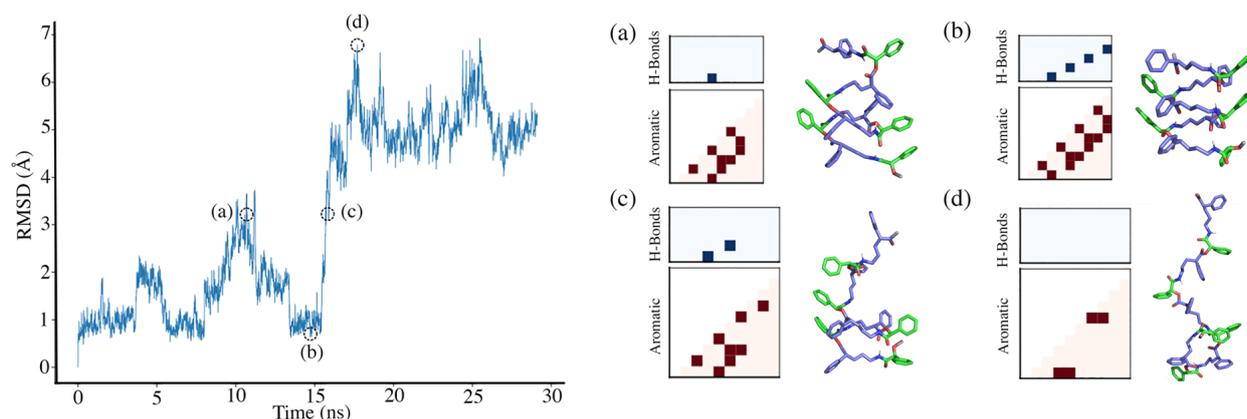
Figure 6.13: MD trajectory for sequence Z_3, points (a), (b), (c) and (d) displays structural views of the unfolding process with their respectively H-bond and aromatic interaction fingerprints.

Li et al. investigated the folding of octamers using long MD simulations, and implicit solvation of different sequence-defined oligomers with amide groups in the backbone and aromatic side chains [157]. Similarly to our results, they conclude that H-Bonds are weakly stabilizing interactions and the aromatic interactions played a significant role in stabilizing their globular conformations. Our results suggest that the conformation stability does not critically depend on individual bonds but is instead the result of many additive interactions. These observations agree with our H-Bond cluster and aromatic contact map fingerprints and the differences in the unfolding times of the sequences studied.

6.4   Conclusion

Using a combined MC-MD approach, possible oligomers which lead to helical conformations have been explored. The results were analyzed by applying the H-Bond cluster analysis and calculating five metrics. The MC annealing simulations sampled six candidate sequences from the group A with different backbone modification lengths. It was found that a backbone spacing of four carbon atoms yields a well-defined hydrogen bonding helical pattern. Then, specific modifications to the backbone and side chain modification monomers were applied to sequence A_4, resulting in helical conformations but with different numbers of H-Bond clusters. From the analysis of the child sequences, it was found that sequence Z_3 that combines aromatic side chains and backbones with an aromatic ring attached to next to ester groups achieves the best clustering metrics.

Molecular Dynamics simulations were applied to evaluate the stability of the oligomer conformations obtained in the MC protocol by measuring their unfolding time. The oligomer sequences examined were highly flexible and required additional stabilizing interactions to remain folded for an extended period, such as the π-interactions between side chains and backbone monomers. The sequence that achieved the best metrics in the MC approach also has the most stable conformation and is, so far, the most promising candidate for a sequence that can fold under experimental conditions.

Our results suggest that H-Bonds and aromatic interactions play the most significant role in stabilizing the helices. The results presented here are the first step in the quest for *de novo* design of helical oligomers for this class of polymers, and the systematic protocol achieved using SimStack workflows and WaNos could speed up the research on oligomer folding. We believe this contribution will interest synthetic polymer scientists and could lead to a new family generation of folding polymers.

7   Final outlook and conclusions

The main goal of this thesis was to find automatic protocols to model and design helical polymers and to understand the driving forces on the atomic scale that could lead to helical conformations. The main tools employed are DFT, Monte Carlo-based protocols, and MD simulations.

Chapter 4 studied the critical interactions involved in the helical induction of poly (4-carboxyphenyl) acetylene by chiral amines. This complex phenomenon has been extensively studied experimentally, but the atomic scale mechanism remains unclear. As the first step of this study, helical models for a 4-monomer chain were created based on DFT dihedral scan calculations with both clockwise and counterclockwise conformations being obtained. Next, the interactions between this 4-monomer chain and the amines were studied by calculating the dissociation energies via DFT. The results revealed a clear trend in the affinities with R and S amine conformations, preferring clockwise and counterclockwise orientations, respectively. Bader charge analysis revealed the presence of local charge transfers effects at the contact point between the amines and the polymer, which agreed well with and further elucidated the affinity results.

In Chapter 5, the poly phenylacetylene polymer family has been studied again in the presence of R and S amine conformations. This study aimed to extend the modeling approaches from Chapter 4 by modeling much larger 40-monomer chains and exploring para and meta-side chain modifications. This polymer family also exhibits macroscopic behaviors such as helical induction, the Sergeant's and soldiers' effect, and the majority effect. These polymers are highly dynamic, and determining their structure remains extremely challenging. The workflow framework SimStack was heavily utilized with blind and local biased helical searches for low-energy conformations. The main simulation methods used were Monte Carlo simulations and DFT dihedral scans. Both *cis-cisoid* and *cis-transoid* conformations were obtained and the para and meta side chain substitutions were found to strongly influence the stability mainly via H-bonds, $\pi - \pi$ stacking interactions and $\pi - \pi$ T-shaped interactions.

Chapter 6 *de novo* helical structure design was explored using sequence controlled-oligomers. Sequence-controlled oligomers are gaining more and more attention due to their potential applications in data storage systems. There are currently many synthesis strategies for encoding and decoding protocols, but a lack of attention has been given to how these polymers fold. In this final piece of work, an automatic protocol to generate sequence-controlled oligomers using text line sequence inputs coupled with annealing Monte Carlo simulations was implemented in SimStack. Different backbone lengths from 1-6 carbons and various side chain modifications were investigated. H-bond fingerprint patterns were found to be a useful tool for the rational design of helical structures. The annealing Monte Carlo simulations plus further Molecular Dynamics simulations demonstrated that the van der Waals interactions are vital in conserving the helical structure at key positions on each helical turn.

Chapter 4 studied the helical induction of poly (4-carboxyphenyl) acetylene for a 4-monomer chain with different R and S amine conformations. In Chapter 5, we studied poly phenylacetylene extending the size of the system to 40-monomer chains and exploring para and meta side chain modifications. Lastly, in Chapter 6, we investigated *de novo* helical structure design of sequence-controlled oligomers varying the backbone length and applying side chain modifications.

Together, these studies illustrate that various methods are needed to paint the fullest possible picture of helical induction mechanisms and polymer folding. On the one hand, due to the scale of the systems studied, accelerated sampling techniques such as Monte Carlo simulations are needed to gain a fuller picture of the entire conformational space, with millions of different structures being searched through in chapters 5 and 6. On the other hand, Chapter 4 clearly showed us that it was essential to go to the DFT level of theory to describe the local charge transfer effects. An intelligent and well-chosen combination of cheaper force field-based methods such as Molecular Dynamics and Monte Carlo Simulations and more expensive electronic structure-based methods such as DFT needs to be employed.

The MC simulations performed here primarily relied upon non-bonded force field terms. Future studies should incorporate additional bonded force field terms. Another refinement could be to couple the MC trajectory generation to DFTb calculations to capture the van der Waals interactions, which were vital in all our polymer systems. Such modifications may enable us to refine our search for low-energy helical conformations while still working at similar scales to the work of this thesis. Our eventual goal is to extend the systematic protocols outlined in chapters 5 and 6 so that the *in Silico* design and exploration of helical polymers can become much more widespread.

References

[1] Y. Gnanou and M. Fontanille, *Organic and physical chemistry of polymers.* John Wiley & Sons, 2008.

[2] C. E. Carraher Jr, *Introduction to polymer chemistry.* CRC press, 2017.

[3] D. M. Teegarden, *Polymer chemistry: introduction to an indispensable science.* NSTA Press, 2004.

[4] M. Kotelyanskii and D. N. Theodorou, *Simulation methods for polymers.* CRC Press, 2004.

[5] L. Pauling, R. B. Corey, and H. R. Branson, "The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain," *Proceedings of the National Academy of Sciences*, vol. 37, no. 4, pp. 205–211, 1951.

[6] J. D. Watson and F. H. Crick, "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[7] G. Natta, P. Pino, P. Corradini, F. Danusso, E. Mantica, G. Mazzanti, and G. Moraglio, "Crystalline high polymers of α-olefins," *Journal of the American Chemical Society*, vol. 77, no. 6, pp. 1708–1710, 1955.

[8] H. Tadokoro, "Structure and properties of crystalline polymers," *Polymer*, vol. 25, no. 2, pp. 147–164, 1984.

[9] E. Yashima, K. Maeda, H. Iida, Y. Furusho, and K. Nagai, "Helical polymers: synthesis, structures, and functions," *Chemical reviews*, vol. 109, no. 11, pp. 6102–6211, 2009.

[10] T. Nakano and Y. Okamoto, "Synthetic helical polymers: conformation and function," *Chemical Reviews*, vol. 101, no. 12, pp. 4013–4038, 2001.

[11] E. Yashima, "Synthesis and structure determination of helical polymers," *Polymer Journal*, vol. 42, pp. 3–16, Jan. 2010.

[12] T. Kajitani, K. Okoshi, S.-i. Sakurai, J. Kumaki, and E. Yashima, "Helix-sense controlled polymerization of a single phenyl isocyanide enantiomer leading to diastereomeric helical polyisocyanides with opposite helix-sense and cholesteric liquid crystals with opposite twist-sense," *Journal of the American Chemical Society*, vol. 128, no. 3, pp. 708–709, 2006.

[13] E. Yashima, T. Matsushima, and Y. Okamoto, "Poly ((4-carboxyphenyl) acetylene) as a probe for chirality assignment of amines by circular dichroism," *Journal of the American Chemical Society*, vol. 117, no. 46, pp. 11596–11597, 1995.

[14] E. Yashima, Y. Maeda, and Y. Okamoto, "A stereoregular polyphenylacetylene derivative bearing an amino group as a probe for chirality assignments of acids by circular dichroism," *Chemistry letters*, vol. 25, no. 11, pp. 955–956, 1996.

[15] E. Yashima, T. Nimura, T. Matsushima, and Y. Okamoto, "Poly ((4-dihydroxyborophenyl) acetylene) as a novel probe for chirality and structural assignments of various kinds of molecules including carbohydrates and steroids by circular dichroism," *Journal of the American Chemical Society*, vol. 118, no. 40, pp. 9800–9801, 1996.

[16] E. Yashima, T. Matsushima, and Y. Okamoto, "Chirality assignment of amines and amino alcohols based on circular dichroism induced by helix formation of a stereoregular poly ((4-carboxyphenyl) acetylene) through acid- base complexation," *Journal of the American Chemical Society*, vol. 119, no. 27, pp. 6345–6359, 1997.

[17] E. Yashima, Y. Maeda, T. Matsushima, and Y. Okamato, "Preparation of polyacetylenes bearing an amino group and their application to chirality assignment of carboxylic acids by circular dichroism," *Chirality*, vol. 9, no. 5-6, pp. 593–600, 1997.

[18] E. Yashima, H. Goto, and Y. Okamoto, "Induced helix of an aliphatic polyacetylene detected by circular dichroism," *Polymer journal*, vol. 30, no. 1, pp. 69–71, 1998.

[19] H. Onouchi, K. Maeda, and E. Yashima, "A helical polyelectrolyte induced by specific interactions with biomolecules in water," *Journal of the American Chemical Society*, vol. 123, no. 30, pp. 7441–7442, 2001.

[20] H. Onouchi, T. Miyagawa, A. Furuko, K. Maeda, and E. Yashima, "Enantioselective esterification of prochiral phosphonate pendants of a polyphenylacetylene assisted by macromolecular helicity: Storage of a dynamic macromolecular helicity memory," *Journal of the American Chemical Society*, vol. 127, no. 9, pp. 2960–2965, 2005.

[21] K. Maeda, S. Okada, E. Yashima, and Y. Okamoto, "Helix formation of poly (phenylacetylene) derivatives bearing amino groups at the meta position induced by optically active carboxylic acids," *Journal of Polymer Science Part A: Polymer Chemistry*, vol. 39, no. 18, pp. 3180–3189, 2001.

[22] H. Kawamura, K. Maeda, Y. Okamoto, and E. Yashima, "Molecular recognition of nucleosides and nucleotides based on circular dichroism induced by helix formation of poly [(4-dihydroxyborophenyl) acetylene]," *Chemistry Letters*, vol. 30, no. 1, pp. 58–59, 2001.

[23] K. Maeda, H. Goto, and E. Yashima, "Stereospecific polymerization of propiolic acid with rhodium complexes in the presence of bases and helix induction on the polymer in water," *Macromolecules*, vol. 34, no. 5, pp. 1160–1164, 2001.

[24] J. Tabei, R. Nomura, F. Sanda, and T. Masuda, "Induction of one-handed helix sense in achiral poly (n-propargylamides)," *Macromolecules*, vol. 36, no. 23, pp. 8603–8608, 2003.

[25] M. M. Green, C. Khatri, and N. C. Peterson, "A macromolecular conformational change driven by a minute chiral solvation energy," *Journal of the American Chemical Society*, vol. 115, no. 11, pp. 4941–4942, 1993.

[26] K. Maeda, N. Yamamoto, and Y. Okamoto, "Helicity induction of poly (3-carboxyphenyl isocyanate) by chiral acid-base interaction," *Macromolecules*, vol. 31, no. 17, pp. 5924–5926, 1998.

[27] R. Sakai, T. Satoh, R. Kakuchi, H. Kaga, and T. Kakuchi, "Macromolecular helicity induction for novel optically inactive poly (phenyl isocyanate) bearing crown ether based on the host- guest complexation," *Macromolecules*, vol. 36, no. 10, pp. 3709–3713, 2003.

[28] M. Ishikawa, K. Maeda, and E. Yashima, "Macromolecular chirality induction on optically inactive poly (4-carboxyphenyl isocyanide) with chiral amines: a dynamic conformational transition of poly (phenyl isocyanide) derivatives," *Journal of the American Chemical Society*, vol. 124, no. 25, pp. 7448–7458, 2002.

[29] D. S. Schlitzer, "Trapped kinetic states, chiral amplification and molecular chaperoning in synthetic polymers: chiral induction in polyguanidines through ion pair interactions," *J. Am. Chem. Soc.*, vol. 120, pp. 2196–2197, 1998.

[30] C. Levinthal, "How to fold graciously," *Mossbauer spectroscopy in biological systems*, vol. 67, pp. 22–24, 1969.

[31] C. Levinthal, "Are there pathways for protein folding?," *Journal de chimie physique*, vol. 65, pp. 44–45, 1968.

[32] R. Zwanzig, A. Szabo, and B. Bagchi, "Levinthal's paradox.," *Proceedings of the National Academy of Sciences*, vol. 89, no. 1, pp. 20–22, 1992.

[33] K. A. Dill and H. S. Chan, "From levinthal to pathways to funnels," *Nature structural biology*, vol. 4, no. 1, pp. 10–19, 1997.

[34] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 4096, pp. 223–230, 1973.

[35] K. A. Dill and J. L. MacCallum, "The protein-folding problem, 50 years on," *science*, vol. 338, no. 6110, pp. 1042–1046, 2012.

[36] T. Schlick, *Molecular modeling and simulation: an interdisciplinary guide*, vol. 2. Springer, 2010.

[37] B. Wunderlich, *Macromolecular physics. vol. 1: Crystal structure morphology, defects.* Academic Press, 1973.

[38] B. S. Li, K. K. Cheuk, L. Ling, J. Chen, X. Xiao, C. Bai, and B. Z. Tang, "Synthesis and hierarchical structures of amphiphilic polyphenylacetylenes carrying l-valine pendants," *Macromolecules*, vol. 36, no. 1, pp. 77–85, 2003.

[39] S.-i. Sakurai, K. Kuroyanagi, K. Morino, M. Kunitake, and E. Yashima, "Atomic force microscopy study of helical poly (phenylacetylene) s on a mica substrate," *Macromolecules*, vol. 36, no. 25, pp. 9670–9674, 2003.

[40] V. Percec, J. G. Rudick, M. Wagner, M. Obata, C. M. Mitchell, W.-D. Cho, and S. N. Magonov, "Afm visualization of individual and periodic assemblies of a helical dendronized polyphenylacetylene on graphite," *Macromolecules*, vol. 39, no. 21, pp. 7342–7351, 2006.

[41] Y. Liu, Z. Wang, and X. Zhang, "Characterization of supramolecular polymers," *Chemical Society Reviews*, vol. 41, no. 18, pp. 5922–5932, 2012.

[42] B. Adisa and D. A. Bruce, "Molecular dynamics simulations of helix-forming, amine-functionalized m-poly (phenyleneethynylene) s," *The Journal of Physical Chemistry B*, vol. 109, no. 15, pp. 7548–7556, 2005.

[43] M. J. Williams and M. Bachmann, "Stabilization of helical macromolecular phases by confined bending," *Physical review letters*, vol. 115, no. 4, p. 048301, 2015.

[44] M. J. Williams and M. Bachmann, "Significance of bending restraints for the stability of helical polymer conformations," *Physical Review E*, vol. 93, no. 6, p. 062501, 2016.

[45] X. Liu, Q. S. Yang, K. M. Liew, and X. Q. He, "Superstretchability and stability of helical structures of carbon nanotube/polymer composite fibers: coarse-grained molecular dynamics modeling and simulation," *Carbon*, vol. 115, pp. 220–228, 2017.

[46] A. J. Christofferson, G. Yiapanis, J. M. Ren, G. G. Qiao, K. Satoh, M. Kamigaito, and I. Yarovsky, "Molecular mapping of poly(methyl methacrylate) super-helix stereocomplexes," *Chemical Science*, vol. 6, no. 2, pp. 1370–1378, 2015.

[47] C. J. Cramer, *Essentials of computational chemistry: theories and models.* John Wiley & Sons, 2013.

[48] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory.* Mineola: Dover Publications, Inc., first ed., 1996.

[49] F. Jensen, *Introduction to Computational Chemistry.* Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.

[50] W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory.* Weinheim - New York: Wiley - VCH, 2nd edition, 2001.

[51] C. R. C. Rêgo, *Ab-initio studies of adsorbate-surface interactions.* PhD thesis, Instituto de Física de São Carlos, 2017.

[52] A. W. Bushmaker, V. V. Deshpande, S. Hsieh, M. W. Bockrath, and S. B. Cronin, "Direct observation of born- oppenheimer approximation breakdown in carbon nanotubes," *Nano letters*, vol. 9, no. 2, pp. 607–611, 2009.

[53] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev*, vol. 136, pp. 864–871, Nov 1964.

[54] D. S. Sholl and J. A. Steckel, *Density functional theory: a practical introduction.* John Wiley & Sons, 2011.

[55] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical review*, vol. 140, no. 4A, p. A1133, 1965.

[56] K. Capelle, "A bird's-eye view of density-functional theory," *Braz. J. Phys*, vol. 36, pp. 1318 – 1343, 12 2006.

[57] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu," *J. Chem. Phys*, vol. 132, no. 15, p. 154104, 2010.

[58] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Oxford university press, 2017.

[59] J. W. Ponder and D. A. Case, "Force fields for protein simulations," *Advances in protein chemistry*, vol. 66, pp. 27–85, 2003.

[60] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," vol. 25, no. 9, pp. 1157–1174.

[61] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," vol. 81, no. 8, pp. 3684–3690.

[62] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *The Journal of chemical physics*, vol. 126, no. 1, p. 014101, 2007.

[63] H. Berendsen, "vd s., d.; van drunen, r," *GROMACS: A message-passing parallel molecular dynamics implementation. Computer Physics Communications*, vol. 91, no. 1-3, pp. 43–56, 1995.

[64] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Oxford University Press, second edition ed.

[65] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, pp. 7182–7190, Dec. 1981.

[66] S. Nosé and M. Klein, "Constant pressure molecular dynamics for molecular systems," vol. 50, no. 5, pp. 1055–1076.

[67] H. J. Berendsen, D. van der Spoel, and R. van Drunen, "Gromacs: a message-passing parallel molecular dynamics implementation," *Computer physics communications*, vol. 91, no. 1-3, pp. 43–56, 1995.

[68] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1-2, pp. 19–25, Sept. 2015.

[69] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[70] K. Binder, *Monte Carlo and molecular dynamics simulations in polymer science*. Oxford University Press, 1995.

[71] R. F. McGuire, F. A. Momany, and H. A. Scheraga, "Energy parameters in polypeptides. v. empirical hydrogen bond potential function based on molecular orbital calculations," vol. 76, no. 3, pp. 375–393.

[72] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, "Semianalytical treatment of solvation for molecular mechanics and dynamics," vol. 112, no. 16, pp. 6127–6129.

67

[73] M. Brieg and W. Wenzel, "Powerborn: a barnes–hut tree implementation for accurate and efficient born radii computation," *Journal of Chemical Theory and Computation*, vol. 9, no. 3, pp. 1489–1498, 2013.

[74] C. R. C. Rêgo, J. Schaarschmidt, T. Schlöder, M. Penaloza-Amion, S. Bag, T. Neumann, T. Strunk, and W. Wenzel, "Simstack: An intuitive workflow framework," *Frontiers in Materials*, vol. 9, 2022.

[75] T. Strunk, M. Wolf, M. Brieg, K. Klenin, A. Biewer, F. Tristram, M. Ernst, P. Kleine, N. Heilmann, I. Kondov, *et al.*, "Simona 1.0: An efficient and versatile framework for stochastic simulations of molecular and nanoscale systems," *Journal of computational chemistry*, vol. 33, no. 32, pp. 2602–2613, 2012.

[76] "TURBOMOLE V7.3 2018, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.."

[77] E. Yashima, K. Maeda, H. Iida, Y. Furusho, and K. Nagai, "Helical Polymers: Synthesis, Structures, and Functions," *Chemical Reviews*, vol. 109, pp. 6102–6211, Nov. 2009.

[78] B. Jeong and A. Gutowska, "Lessons from nature: stimuli-responsive polymers and their biomedical applications," *Trends in Biotechnology*, vol. 20, pp. 305–311, July 2002.

[79] T. Miyata, T. Uragami, and K. Nakamae, "Biomolecule-sensitive hydrogels," *Advanced Drug Delivery Reviews*, vol. 54, pp. 79–98, Jan. 2002.

[80] A. Kikuchi and T. Okano, "Intelligent thermoresponsive polymeric stationary phases for aqueous chromatography of biological compounds," *Progress in Polymer Science*, vol. 27, pp. 1165–1193, July 2002.

[81] E. Gil and S. Hudson, "Stimuli-reponsive polymers and their bioconjugates," *Progress in Polymer Science*, vol. 29, pp. 1173–1222, Dec. 2004.

[82] Y. Okamoto, K. Suzuki, K. Ohta, K. Hatada, and H. Yuki, "Optically active poly(triphenylmethyl methacrylate) with one-handed helical conformation," *Journal of the American Chemical Society*, vol. 101, pp. 4763–4765, Aug. 1979.

[83] R. J. M. Nolte, A. J. M. Van Beijnen, and W. Drenth, "Chirality in polyisocyanides," *Journal of the American Chemical Society*, vol. 96, pp. 5932–5933, Sept. 1974.

[84] M. M. Green, M. P. Reidy, R. D. Johnson, G. Darling, D. J. O'Leary, and G. Willson, "Macromolecular stereochemistry: the out-of-proportion influence of optically active comonomers on the conformational characteristics of polyisocyanates. The sergeants and soldiers experiment," *Journal of the American Chemical Society*, vol. 111, pp. 6452–6454, Aug. 1989.

[85] M. M. Green, B. A. Garetz, B. Munoz, H. Chang, S. Hoke, and R. G. Cooks, "Majority Rules in the Copolymerization of Mirror Image Isomers," *Journal of the American Chemical Society*, vol. 117, pp. 4181–4182, Apr. 1995.

[86] E. Yashima, K. Maeda, and Y. Okamoto, "Memory of macromolecular helicity assisted by interaction with achiral small molecules," *Nature*, vol. 399, pp. 449–451, June 1999.

[87] E. Yashima, T. Matsushima, and Y. Okamoto, "Poly((4-carboxyphenyl)acetylene) as a Probe for Chirality Assignment of Amines by Circular Dichroism," *Journal of the American Chemical Society*, vol. 117, pp. 11596–11597, Nov. 1995.

[88] E. Yashima, T. Matsushima, and Y. Okamoto, "Chirality Assignment of Amines and Amino Alcohols Based on Circular Dichroism Induced by Helix Formation of a Stereoregular Poly((4-carboxyphenyl)acetylene) through Acid–Base Complexation," *Journal of the American Chemical Society*, vol. 119, pp. 6345–6359, July 1997.

[89] K. Morino, N. Watase, K. Maeda, and E. Yashima, "Chiral Amplification in Macromolecular Helicity Assisted by Noncovalent Interaction with Achiral Amines and Memory of the Helical Chirality," *Chemistry - A European Journal*, vol. 10, pp. 4703–4707, Oct. 2004.

[90] H. Onouchi, D. Kashiwagi, K. Hayashi, K. Maeda, and E. Yashima, "Helicity Induction on Poly(phenylacetylene)s Bearing Phosphonic Acid Pendants with Chiral Amines and Memory of the Macromolecular Helicity Assisted by Interaction with Achiral Amines in Dimethyl Sulfoxide," *Macromolecules*, vol. 37, pp. 5495–5503, July 2004.

[91] H. Onouchi, T. Hasegawa, D. Kashiwagi, H. Ishiguro, K. Maeda, and E. Yashima, "Helicity Induction in Charged Poly(phenylacetylene)s Bearing Various Acidic Functional Groups in Water and Its Mechanism," *Macromolecules*, vol. 38, pp. 8625–8633, Oct. 2005. Publisher: American Chemical Society.

[92] H. Onouchi, T. Miyagawa, A. Furuko, K. Maeda, and E. Yashima, "Enantioselective Esterification of Prochiral Phosphonate Pendants of a Polyphenylacetylene Assisted by Macromolecular Helicity: Storage of a Dynamic Macromolecular Helicity Memory," *Journal of the American Chemical Society*, vol. 127, pp. 2960–2965, Mar. 2005.

[93] K. Nagai, K. Maeda, Y. Takeyama, K. Sakajiri, and E. Yashima, "Helicity Induction and Chiral Amplification in a Poly(phenylacetylene) Bearing N , N - Diisopropylaminomethyl Groups with Chiral Acids in Water," *Macromolecules*, vol. 38, pp. 5444–5451, June 2005.

[94] K. Morino, M. Oobo, and E. Yashima, "Helicity Induction in a Poly(phenylacetylene) Bearing Aza-18-crown-6 Ether Pendants with Optically Active Bis(amino acid)s and Its Chiral Stimuli-Responsive Gelation," *Macromolecules*, vol. 38, pp. 3461–3468, Apr. 2005. Publisher: American Chemical Society.

[95] S. Kobayashi, K. Morino, and E. Yashima, "Macromolecular helicity inversion of an optically active helical poly(phenylacetylene) by chemical modification of the side groups," *Chemical Communications*, pp. 2351–2353, June 2007. Publisher: The Royal Society of Chemistry.

69

[96] K. Maeda, K. Morino, Y. Okamoto, T. Sato, and E. Yashima, "Mechanism of Helix Induction on a Stereoregular Poly((4-carboxyphenyl)acetylene) with Chiral Amines and Memory of the Macromolecular Helicity Assisted by Interaction with Achiral Amines," *Journal of the American Chemical Society*, vol. 126, pp. 4329–4342, Apr. 2004.

[97] V. Percec, J. G. Rudick, M. Peterca, M. Wagner, M. Obata, C. M. Mitchell, W.-D. Cho, V. S. K. Balagurusamy, and P. A. Heiney, "Thermoreversible Cis – Cisoidal to Cis – Transoidal Isomerization of Helical Dendronized Polyphenylacetylenes," *Journal of the American Chemical Society*, vol. 127, pp. 15257–15264, Nov. 2005.

[98] E. Suárez-Picado, E. Quiñoá, R. Riguera, and F. Freire, "Poly(phenylacetylene) Amines: A General Route to Water-Soluble Helical Polyamines," *Chemistry of Materials*, vol. 30, pp. 6908–6914, Oct. 2018.

[99] J. Wildman, P. Repiščák, M. J. Paterson, and I. Galbraith, "General Force-Field Parametrization Scheme for Molecular Dynamics Simulations of Conjugated Materials in Solution," *Journal of Chemical Theory and Computation*, vol. 12, pp. 3813–3824, Aug. 2016.

[100] W. Liu, X. Luo, Y. Bao, Y. P. Liu, G.-H. Ning, I. Abdelwahab, L. Li, C. T. Nai, Z. G. Hu, D. Zhao, B. Liu, S. Y. Quek, and K. P. Loh, "A two-dimensional conjugated aromatic polymer via C–C coupling reaction," *Nature Chemistry*, vol. 9, pp. 563–570, June 2017.

[101] X. Huang, X. Zhang, G.-K. Ren, J. Jiang, Z. Dan, Q. Zhang, X. Zhang, C.-W. Nan, and Y. Shen, "Non-intuitive concomitant enhancement of dielectric permittivity, breakdown strength and energy density in percolative polymer nanocomposites by trace Ag nanodots," *Journal of Materials Chemistry A*, vol. 7, no. 25, pp. 15198–15206, 2019.

[102] T. L. M. Pham, K. T. Phung, T. V. Ho, T. A. Le, and A. T. T. Nguyen, "Density functional theory insights into the bonding of CH3OH and CH3O with Ir(111) surface," *Science and Technology Development Journal*, vol. 24, Feb. 2021.

[103] U. Ryde and P. Söderhjelm, "Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods," *Chemical Reviews*, vol. 116, pp. 5520–5566, May 2016.

[104] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, "Turbomole," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, pp. 91–100, Mar. 2014.

[105] S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, C. van Wüllen, V. K. Voora, F. Weigend, A. Wodyński, and J. M. Yu, "TURBOMOLE: Modular program suite for *ab initio* quantum-chemical and condensed-matter simulations," *The Journal of Chemical Physics*, vol. 152, p. 184107, May 2020.

[106] A. D. Becke, "A new mixing of Hartree–Fock and local density-functional theories," *The Journal of Chemical Physics*, vol. 98, pp. 1372–1377, Jan. 1993.

[107] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *The Journal of Chemical Physics*, vol. 98, pp. 5648–5652, Apr. 1993.

[108] J. Zheng, X. Xu, and D. G. Truhlar, "Minimally augmented Karlsruhe basis sets," *Theoretical Chemistry Accounts*, vol. 128, pp. 295–305, Feb. 2011.

[109] C. R. C. Rêgo, L. N. Oliveira, P. Tereshchuk, and J. L. F. D. Silva, "Comparative study of van der waals corrections to the bulk properties of graphite," *Journal of Physics: Condensed Matter*, vol. 27, p. 415502, sep 2015.

[110] C. R. C. Rêgo, L. N. Oliveira, P. Tereshchuk, and J. L. F. D. Silva, "Corrigendum: Comparative study of van der waals corrections to the bulk properties of graphite (2015j. phys.: Condens. matter27415502)," *Journal of Physics: Condensed Matter*, vol. 28, p. 129501, feb 2016.

[111] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," *The Journal of Chemical Physics*, vol. 132, p. 154104, Apr. 2010.

[112] S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," *Journal of Computational Chemistry*, vol. 32, pp. 1456–1465, May 2011.

[113] A. Schäfer, A. Klamt, D. Sattel, J. C. W. Lohrenz, and F. Eckert, "COSMO Implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems," *Physical Chemistry Chemical Physics*, vol. 2, no. 10, pp. 2187–2193, 2000.

[114] P. E. Blöchl, "Projector augmented-wave method," *Physical Review B*, vol. 50, pp. 17953–17979, dec 1994.

[115] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Physical Review B*, vol. 59, pp. 1758–1775, jan 1999.

[116] G. Kresse and J. Furthmüller, "Efficient iterative schemes forab initiototal-energy calculations using a plane-wave basis set," *Physical Review B*, vol. 54, pp. 11169–11186, oct 1996.

[117] J. Hafner, "Ab-initio simulations of materials using VASP: Density-functional theory and beyond," *Journal of Computational Chemistry*, vol. 29, pp. 2044–2078, oct 2008.

[118] W. Tang, E. Sanville, and G. Henkelman, "A grid-based bader analysis algorithm without lattice bias," *Journal of Physics: Condensed Matter*, vol. 21, p. 084204, jan 2009.

[119] C. R. C. Rêgo, "Kit-workflows/dft-vasp: Dft-vasp, available from `https://github.com/kit-workflows/dft-vasp`," 2021.

[120] J. Schaarschmidt, J. Yuan, T. Strunk, I. Kondov, S. P. Huber, G. Pizzi, L. Kahle, F. T. Bölle, I. E. Castelli, T. Vegge, *et al.*, "Workflow engineering in materials design within the battery 2030+ project," *Advanced Energy Materials*, p. 2102638, 2021.

[121] R. Duke, T. Giese, H. Gohlke, A. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. Lee, and others, "AmberTools 16," *University of California, San Francisco*, 2016.

[122] G. A. Ozpinar, W. Peukert, and T. Clark, "An improved generalized AMBER force field (GAFF) for urea," *J Mol Model*, vol. 16, pp. 1427–40, Sept. 2010.

[123] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. am1-bcc model: I. method," *Journal of computational chemistry*, vol. 21, no. 2, pp. 132–146, 2000.

[124] C. Caleman, P. J. van Maaren, M. Hong, J. S. Hub, L. T. Costa, and D. van der Spoel, "Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant," *Journal of Chemical Theory and Computation*, vol. 8, pp. 61–74, Jan. 2012.

[125] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, "LINCS: a linear constraint solver for molecular simulations," *Journal of computational chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997. Publisher: Wiley Online Library.

[126] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *The Journal of Chemical Physics*, vol. 126, p. 014101, Jan. 2007.

[127] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters," *The Journal of Chemical Physics*, vol. 76, pp. 637–649, Jan. 1982.

[128] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N –log(N) method for Ewald sums in large systems," *The Journal of Chemical Physics*, vol. 98, pp. 10089–10092, June 1993.

[129] S. Wang, X. Feng, Z. Zhao, J. Zhang, and X. Wan, "Reversible *Cis-Cisoid* to *Cis-Transoid* Helical Structure Transition in Poly(3,5-disubstituted phenylacetylene)s," *Macromolecules*, vol. 49, pp. 8407–8417, Nov. 2016.

[130] R. Hoppe, "Effective coordination numbers (ECoN) and mean active fictive ionic radii (MEFIR)," *Zeitschrift für Kristallographie*, vol. 150, pp. 23–52, jan 1979.

[131] J. L. F. D. Silva, "Effective coordination concept applied for phase change (GeTe)m(sb2te3)n compounds," *Journal of Applied Physics*, vol. 109, p. 023502, jan 2011.

[132] Y. Ashida, T. Sato, K. Morino, K. Maeda, Y. Okamoto, and E. Yashima, "Helical structural change in poly ((4-carboxyphenyl) acetylene) by acid- base complexation with an optically active amine," *Macromolecules*, vol. 36, no. 9, pp. 3345–3350, 2003.

[133] J. Hutter, "Car–parrinello molecular dynamics," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 4, pp. 604–612, 2012.

[134] J. Stare, J. Mavri, J. Grdadolnik, J. Zidar, Z. B. Maksic, and R. Vianello, "Hydrogen bond dynamics of histamine monocation in aqueous solution: Car–parrinello molecular dynamics and vibrational spectroscopy study," *The Journal of Physical Chemistry B*, vol. 115, no. 19, pp. 5999–6010, 2011.

[135] D. Koner, S. M. Salehi, P. Mondal, and M. Meuwly, "Non-conventional force fields for applications in spectroscopy and chemical reaction dynamics," *The Journal of Chemical Physics*, vol. 153, no. 1, p. 010901, 2020.

[136] P. Mondal, P.-A. Cazade, A. K. Das, T. Bereau, and M. Meuwly, "Multipolar force fields for amide-i spectroscopy from conformational dynamics of the alanine-trimer," *arXiv preprint arXiv:2106.10142*, 2021.

[137] E. V. Anslyn and D. A. Dougherty, *Modern physical organic chemistry*. University science books, 2006.

[138] S. W. Smith, "Chiral toxicology: it's the same thing... only different," *Toxicological sciences*, vol. 110, no. 1, pp. 4–30, 2009.

[139] E. M. Ulrich, C. N. Morrison, M. R. Goldsmith, and W. T. Foreman, "Chiral pesticides: identification, description, and environmental implications," *Reviews of Environmental Contamination and Toxicology Volume 217*, pp. 1–74, 2012.

[140] R. Rodríguez, E. Quiñoá, R. Riguera, and F. Freire, "Architecture of Chiral Poly(phenylacetylene)s: From Compressed/Highly Dynamic to Stretched/Quasi-Static Helices," *Journal of the American Chemical Society*, vol. 138, pp. 9620–9628, Aug. 2016. Publisher: American Chemical Society.

[141] K. A. Günay, P. Theato, and H.-A. Klok, "History of post-polymerization modification," *Functional Polymers by Post-Polymerization Modification*, pp. 1–44, 2013.

[142] P. Theato and H.-A. Klok, *Functional polymers by post-polymerization modification: concepts, guidelines and applications*. John Wiley & Sons, 2013.

[143] A. C. Pauly and P. Theato, "Control of reactivity of constitutional isomers of pentafluorophenyl ethynylbenzoates for the synthesis of functional poly (phenylacetylenes)," *Polymer Chemistry*, vol. 3, no. 7, pp. 1769–1782, 2012.

[144] A. C. Pauly and P. Theato, "Toward photopatternable thin film optical sensors utilizing reactive polyphenylacetylenes," *Macromolecular rapid communications*, vol. 34, no. 6, pp. 516–521, 2013.

[145] A. C. Pauly, C. D. Varnado Jr, C. W. Bielawski, and P. Theato, "Electrochromic poly (acetylene) s with switchable visible/near-ir absorption characteristics," *Macromolecular rapid communications*, vol. 35, no. 2, pp. 210–213, 2014.

[146] D. Weininger, "Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules," *Journal of chemical information and computer sciences*, vol. 28, no. 1, pp. 31–36, 1988.

[147] M. Penaloza-Amion, E. Sedghamiz, M. Kozlowska, C. Degitz, C. Possel, and W. Wenzel, "Monte-carlo simulations of soft matter using simona: A review of recent applications," *Frontiers in Physics*, vol. 9, p. 635959, 2021.

[148] G. J. Kleywegt and T. A. Jones, "Phi/psi-chology: Ramachandran revisited," *Structure*, vol. 4, no. 12, pp. 1395–1400, 1996.

[149] G. t. Ramachandran and V. Sasisekharan, "Conformation of polypeptides and proteins," *Advances in protein chemistry*, vol. 23, pp. 283–437, 1968.

[150] M. Penaloza-Amion, C. R. C. Rego, and W. Wenzel, "Local electronic charge transfer in the helical induction of cis-transoid poly (4-carboxyphenyl) acetylene by chiral amines," *Journal of Chemical Information and Modeling*, 2022.

[151] N. Badi and J.-F. Lutz, "Sequence control in polymer synthesis," *Chemical Society Reviews*, vol. 38, no. 12, pp. 3383–3390, 2009.

[152] J.-F. Lutz, M. Ouchi, D. R. Liu, and M. Sawamoto, "Sequence-controlled polymers," *Science*, vol. 341, no. 6146, p. 1238149, 2013.

[153] P. Espeel, L. L. Carrette, K. Bury, S. Capenberghs, J. C. Martins, F. E. Du Prez, and A. Madder, "Multifunctionalized sequence-defined oligomers from a single building block," *Angewandte Chemie International Edition*, vol. 52, no. 50, pp. 13261–13264, 2013.

[154] S. C. Solleder, D. Zengel, K. S. Wetzel, and M. A. Meier, "A scalable and high-yield strategy for the synthesis of sequence-defined macromolecules," *Angewandte Chemie International Edition*, vol. 55, no. 3, pp. 1204–1207, 2016.

[155] P. Nanjan, A. Jose, L. Thurakkal, and M. Porel, "Sequence-defined dithiocarbamate oligomers via a scalable, support-free, iterative strategy," *Macromolecules*, vol. 53, no. 24, pp. 11019–11026, 2020.

[156] H. Yu, S. Li, K. E. Schwieter, Y. Liu, B. Sun, J. S. Moore, and C. M. Schroeder, "Charge transport in sequence-defined conjugated oligomers," *Journal of the American Chemical Society*, vol. 142, no. 10, pp. 4852–4861, 2020.

[157] J. Li, M. Leclercq, M. Fossepré, M. Surin, K. Glinel, A. M. Jonas, and A. E. Fernandes, "Discrete multifunctional sequence-defined oligomers with controlled chirality," *Polymer Chemistry*, vol. 11, no. 24, pp. 4040–4046, 2020.

[158] C. Mertens, M. Soete, M. L. Śleczkowski, A. R. Palmans, E. Meijer, N. Badi, and F. E. Du Prez, "Stereocontrolled, multi-functional sequence-defined oligomers through automated synthesis," *Polymer Chemistry*, vol. 11, no. 26, pp. 4271–4280, 2020.

[159] S. D. Dahlhauser, C. D. Wight, S. R. Moor, R. A. Scanga, P. Ngo, J. T. York, M. S. Vera, K. J. Blake, I. M. Riddington, J. F. Reuther, *et al.*, "Molecular encryption and steganography using mixtures of simultaneously sequenced, sequence-defined oligourethanes," *ACS Central Science*, 2022.

[160] K. S. Wetzel, M. Frölich, S. C. Solleder, R. Nickisch, P. Treu, and M. A. Meier, "Dual sequence definition increases the data storage capacity of sequence-defined macromolecules," *Communications Chemistry*, vol. 3, no. 1, pp. 1–10, 2020.

[161] M. Frölich, D. Hofheinz, and M. A. Meier, "Reading mixtures of uniform sequence-defined macromolecules to increase data storage capacity," *Communications Chemistry*, vol. 3, no. 1, pp. 1–10, 2020.

[162] A. W. S. Da Silva and W. F. Vranken, "Acpype-antechamber python parser interface," *BMC research notes*, vol. 5, no. 1, pp. 1–8, 2012.

[163] M. A. Murcko, H. Castejon, and K. B. Wiberg, "Carbon- carbon rotational barriers in butane, 1-butene, and 1, 3-butadiene," *The Journal of Physical Chemistry*, vol. 100, no. 40, pp. 16162–16168, 1996.

7.1 Chapter 4 Additional Information

7.1.1 Polymer structural analysis from MD simulations

To obtain detail information of the behaviour of the dihedral angles along the polymer during the MD simulations we calculated the average values of the backbone dihedral angle per monomer displayed in Figure SI2. Additionally, we calculated the occurrence of the screw-sense (Figure SI3) of the polymer sorting every dihedral value as CW and CCW for positive and negative values, respectively in order to observe if there was a trend or force to stabilize a certain helical conformation.
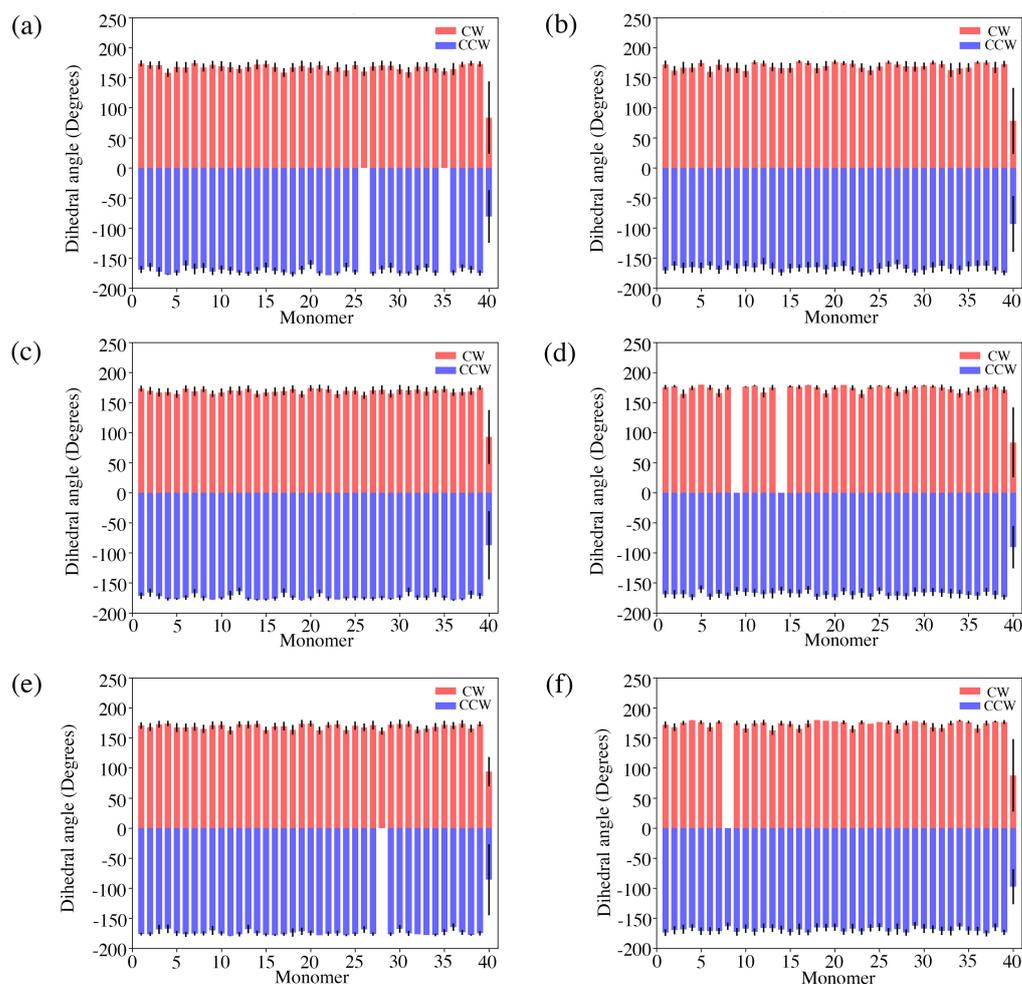
Figure SI1. Average of dihedral angles values per monomer unit for MD simulations of CW (A), CCW (B), CW-R1 (C), CCW-S1(D), CW-S1(E) and CCW-S1(F) systems
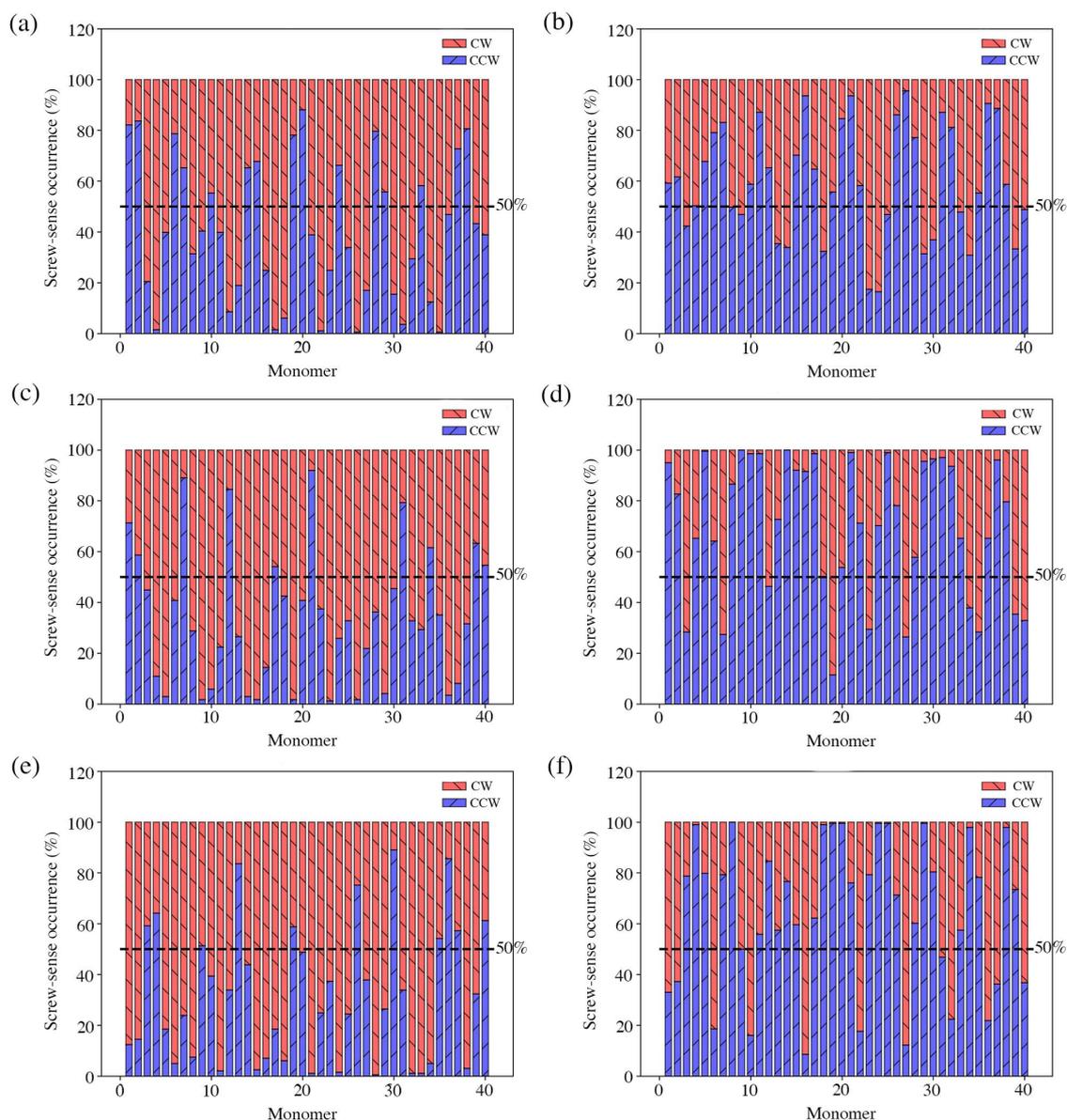
Figure SI2. Occurrence of dihedral angles values per monomer unit sorted by CW or CCW for MD simulations with CW (A), CCW (B), CW-R1 (C), CCW-S1(D), CW-S1(E) and CCW-S1(F) systems. Dash lines represent the 50 percent of occurrence.

7.1.2 Dissociation of energy values obtained from DFT calculations for 4 monomer polymer–amine complexes

The values obtained for the dissociation of each amine (ΔE_{diss}) for all DFT 4 monomer polymer-amine complexes are presented in Table SI1 with their respective differences between R1 and S1. The affinity values were obtained using the termodynamic cycle of the dissociation of one amine in the polymer-amine complex (**??**) where PA, P, and A are complex polymer-amine, polymer, and amine, respectively.

$$PA \xrightarrow{\Delta E_{gas}} P \quad + \quad A$$

$$\Bigg\uparrow{\text{-}\Delta E_{solv(PA)}} \Bigg\downarrow{\Delta E_{solv(P)}} \quad \Delta E_{solv(A)} \Bigg\downarrow$$

$$PA \xrightarrow{\Delta E_{solv}} P \quad + \quad A \tag{7.1}$$
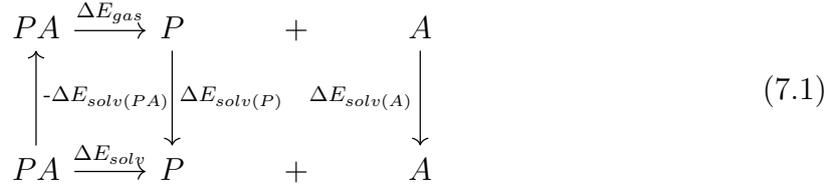
Table SI1. Values of dissociation energies per amines for each polymer-amine complex.

Amine	Helix screw-sense	ΔE_{diss} (eV/mol)	ΔE_{R-S} (eV)
R1-1	CW	1.98	0.29
S1-1	CW	1.69	
R1-2	CW	1.96	0.42
S1-2	CW	1.53	
R1-3	CW	1.41	0.18
S1-3	CW	1.24	
R1-4	CW	1.31	0.20
S1-4	CW	1.11	
R1-1	CCW	0.31	-1.53
S1-1	CCW	1.84	
R1-2	CCW	0.06	-1.84
S1-2	CCW	1.90	
R1-3	CCW	-0.37	-1.68
S1-3	CCW	1.31	
R1-4	CCW	-0.47	-1.83
S1-4	CCW	1.36	

7.1.3 Values of Bader atomic charges for 4 monomer polymer–amine complexes

Polymer and amine structures use in the DFT calculations with their corresponding labels per monomer unit are displayed in Figure SI4. Bader charges calculated for each labeled atom in Figure SI3 are collected in Table SI2 and Table SI3.

Poly-1   Ligand R1 / S1
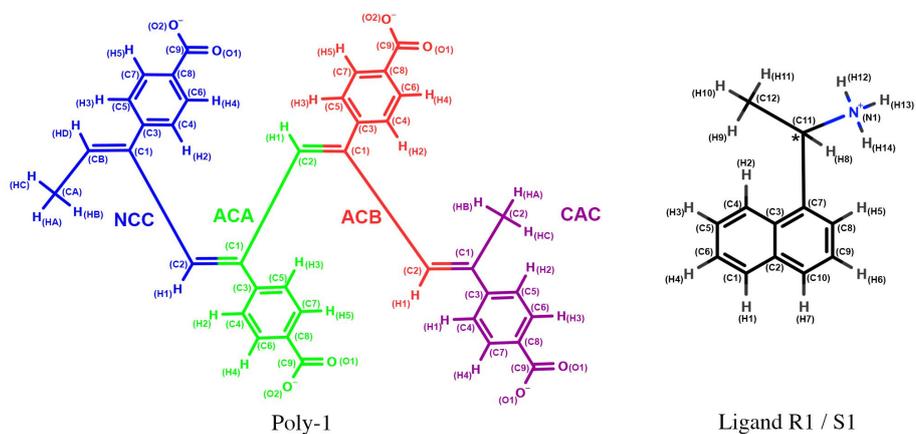
Figure SI3. Scheme of atom labels in each monomer on poly-1 (NCC, ACA, ACB and CAC) and ligand R or S (LIG) mentioned in Table SI2 and Table SI3 for Bader net atomic charges. Carbon with * is the chiral center of R1 and S1.

Table SI2. Bader net atomic charges per monomer (NCC, ACA, ACB and CAC) in 4 monomer complex polymer-amine.

Monomer number	Monomer name	Atom name	Bader net atomic charge			
			CW-R1	CW-S1	CCW-R1	CW-S1
1	NCC	CA	0.108	0.020	0.036	0.122
1	NCC	HA	-0.071	-0.015	-0.077	-0.006
1	NCC	HB	-0.006	-0.020	-0.020	-0.051
1	NCC	HC	-0.060	-0.021	-0.009	-0.066
1	NCC	CB	0.119	0.071	0.063	-0.080
1	NCC	HD	-0.084	-0.059	-0.055	-0.029
1	NCC	C1	-0.003	0.037	0.059	0.036
1	NCC	C3	0.070	-0.029	0.082	-0.028
1	NCC	C4	0.141	0.043	0.126	-0.001
1	NCC	C6	-0.063	0.167	0.060	0.096
1	NCC	C8	0.138	0.016	-0.048	0.098
1	NCC	C7	0.016	0.154	0.101	0.086
1	NCC	C5	-0.015	-0.118	-0.072	0.095
1	NCC	H3	-0.064	0.015	-0.032	-0.089
1	NCC	H5	-0.099	-0.064	-0.130	-0.126
1	NCC	C9	-1.463	-1.556	-1.493	-1.529
1	NCC	O1	1.118	1.153	1.572	1.180
1	NCC	O2	1.161	1.167	1.142	1.154
1	NCC	H4	-0.020	-0.073	-0.094	-0.099
1	NCC	H2	-0.101	-0.043	-0.131	-0.037
1	NCC	C2	0.169	0.051	0.368	0.073
1	NCC	H1	-0.136	-0.074	-0.150	-0.064
2	ACA	C1	-0.061	-0.028	-0.283	0.137
2	ACA	C3	-0.008	-0.195	-0.052	-0.070
2	ACA	C4	0.021	0.191	0.213	0.213
2	ACA	C6	0.221	0.103	-0.064	-0.001
2	ACA	C8	0.077	-0.130	0.025	-0.049
2	ACA	C7	0.130	0.191	0.068	0.121
2	ACA	C5	0.091	0.061	0.140	-0.120
2	ACA	H3	-0.074	-0.003	-0.026	0.043
2	ACA	H5	-0.028	-0.128	-0.079	-0.064
2	ACA	C9	-1.483	-1.266	-1.547	-1.341
2	ACA	O1	1.129	1.137	1.566	1.146
2	ACA	O2	1.062	1.115	1.098	1.072
2	ACA	H4	-0.211	-0.103	-0.156	-0.049
2	ACA	H2	-0.093	-0.127	-0.060	-0.086
2	ACA	C2	0.263	0.259	0.207	0.143
2	ACA	H1	-0.232	-0.158	0.014	-0.120
3	ACB	C1	-0.076	-0.149	-0.140	-0.035
3	ACB	C3	0.160	0.115	0.156	0.016
3	ACB	C4	-0.015	-0.092	0.043	0.020
3	ACB	C6	0.191	0.077	0.024	0.095
3	ACB	C8	-0.137	0.026	-0.037	0.099
3	ACB	C7	0.240	0.096	0.078	-0.082
3	ACB	C5	0.212	0.082	0.061	0.247
3	ACB	H3	-0.236	-0.042	-0.093	-0.200
3	ACB	H5	-0.200	-0.071	-0.096	-0.062
3	ACB	C9	-1.523	-1.539	-1.347	-1.502
3	ACB	O1	1.203	1.172	1.476	1.150
3	ACB	O2	1.173	1.209	1.121	1.198
3	ACB	H4	-0.137	-0.040	-0.118	-0.159
3	ACB	H2	-0.111	-0.171	-0.169	0.002
3	ACB	C2	0.238	0.174	-0.018	-0.053
3	ACB	H1	-0.051	-0.096	-0.073	-0.034
4	CAC	C1	0.052	0.004	0.118	0.062
4	CAC	C2	0.087	0.072	0.228	0.065
4	CAC	HA	0.010	0.034	-0.068	0.006
4	CAC	HB	-0.079	-0.097	-0.195	-0.105
4	CAC	HC	-0.128	-0.090	-0.011	-0.014
4	CAC	C3	-0.105	-0.019	0.086	0.022
4	CAC	C5	0.184	0.170	0.059	-0.057
4	CAC	C6	0.186	0.023	0.176	-0.009
4	CAC	H3	-0.198	-0.025	-0.263	-0.057
4	CAC	H2	-0.143	-0.124	-0.095	0.043
4	CAC	C4	0.108	0.139	-0.162	0.212
4	CAC	H1	-0.114	-0.085	-0.081	-0.153
4	CAC	C7	-0.017	-0.071	0.124	-0.190
4	CAC	H4	-0.046	-0.070	-0.051	-0.031
4	CAC	C8	0.014	0.029	0.130	0.100
4	CAC	C9	-1.380	-1.472	-1.553	-1.321
4	CAC	O1	1.080	1.151	1.497	1.103
4	CAC	O2	1.141	1.140	1.217	1.120

Table SI3. Bader net atomic charges per ligand (LIG) in 4 monomer complex polymer-amine.

Ligand number	Ligand name	Atom name	Bader net atomic charge			
			CW-R1	CW-S1	CCW-R1	CW-S1
5	LIG	C2	-0.001	0.027	-0.107	-0.038
5	LIG	C1	-0.210	0.049	0.015	-0.037
5	LIG	H1	-0.051	-0.122	-0.061	-0.010
5	LIG	C6	0.165	0.043	0.139	-0.061
5	LIG	H4	-0.022	-0.014	-0.140	0.027
5	LIG	C5	0.157	0.161	0.133	0.148
5	LIG	H3	-0.118	-0.109	-0.117	-0.072
5	LIG	C4	0.007	0.030	0.215	0.203
5	LIG	H2	0.037	0.064	-0.172	-0.225
5	LIG	C3	0.045	-0.210	-0.015	0.035
5	LIG	C7	-0.061	0.068	0.140	-0.112
5	LIG	C11	-0.286	-0.177	-0.354	-0.157
5	LIG	N1	2.249	1.667	1.830	1.978
5	LIG	H12	-1.000	-1.000	-1.000	-0.609
5	LIG	H13	-1.000	-0.464	-1.000	-0.721
5	LIG	H14	-0.664	-0.597	-0.557	-1.000
5	LIG	C12	0.131	0.160	0.244	0.099
5	LIG	H9	-0.094	-0.081	-0.123	-0.079
5	LIG	H10	0.020	-0.012	-0.118	-0.079
5	LIG	H11	-0.120	-0.104	-0.057	-0.020
5	LIG	H8	0.010	-0.153	-0.030	-0.124
5	LIG	C8	0.104	0.059	-0.151	0.263
5	LIG	H5	-0.129	-0.035	0.071	-0.128
5	LIG	C9	-0.134	0.221	0.040	0.028
5	LIG	H6	-0.018	-0.192	-0.080	0.010
5	LIG	C10	0.322	0.007	0.161	0.108
5	LIG	H7	-0.129	-0.069	-0.023	-0.149
6	LIG	C2	-0.099	0.077	0.088	0.038
6	LIG	C1	0.023	-0.051	0.029	0.306
6	LIG	H1	-0.115	-0.015	-0.045	-0.223
6	LIG	C6	0.198	0.091	0.125	-0.004
6	LIG	H4	-0.056	-0.064	-0.070	-0.109
6	LIG	C5	0.070	0.089	-0.122	0.122
6	LIG	H3	-0.077	-0.103	-0.006	-0.045
6	LIG	C4	0.225	0.157	0.158	0.033
6	LIG	H2	-0.110	-0.058	-0.049	-0.044
6	LIG	C3	-0.110	0.060	0.051	0.034
6	LIG	C7	0.060	-0.163	-0.182	0.009
6	LIG	C11	-0.293	-0.371	-0.346	-0.287
6	LIG	N1	2.119	1.645	1.343	2.213
6	LIG	H12	-1.000	-1.000	-0.470	-0.563
6	LIG	H13	-0.882	-0.445	-1.000	-1.000
6	LIG	H14	-0.624	-0.583	-0.552	-1.000
6	LIG	C12	0.204	0.094	0.388	0.206
6	LIG	H9	-0.049	0.010	-0.181	-0.091
6	LIG	H10	-0.097	-0.077	-0.085	-0.104
6	LIG	H11	-0.111	-0.012	-0.131	-0.078
6	LIG	H8	-0.052	0.005	-0.030	-0.064
6	LIG	C8	0.215	0.066	0.095	-0.049
6	LIG	H5	-0.162	-0.139	-0.193	0.010
6	LIG	C9	-0.077	0.207	0.265	0.200
6	LIG	H6	-0.062	-0.073	-0.034	-0.131
6	LIG	C10	0.252	-0.079	-0.063	-0.076
6	LIG	H7	-0.174	-0.069	-0.077	-0.081

			Bader net atomic charge			
Ligand number	Ligand name	Atom name	CW-R1	CW-S1	CCW-R1	CW-S1
7	LIG	C2	-0.184	0.044	0.028	-0.073
7	LIG	C1	0.135	0.186	0.053	0.134
7	LIG	H1	-0.104	-0.105	-0.150	-0.093
7	LIG	C6	0.264	-0.025	0.194	0.107
7	LIG	H4	-0.298	-0.044	-0.094	-0.098
7	LIG	C5	0.104	0.045	0.051	0.016
7	LIG	H3	-0.178	-0.125	-0.154	-0.060
7	LIG	C4	0.210	0.097	0.181	0.132
7	LIG	H2	-0.145	-0.087	-0.141	-0.154
7	LIG	C3	0.186	0.020	-0.003	0.038
7	LIG	C7	-0.009	-0.004	-0.006	0.085
7	LIG	C11	-0.014	-0.318	-0.219	-0.351
7	LIG	N1	1.186	2.265	1.242	1.566
7	LIG	H12	-0.519	-1.000	-0.491	-0.609
7	LIG	H13	-0.469	-1.000	-0.450	-0.593
7	LIG	H14	-0.602	-0.626	-1.000	-0.630
7	LIG	C12	0.103	0.231	0.148	0.078
7	LIG	H9	-0.130	-0.142	0.020	-0.133
7	LIG	H10	-0.126	-0.048	-0.061	-0.050
7	LIG	H11	0.013	-0.096	-0.137	0.022
7	LIG	H8	-0.217	-0.025	-0.127	-0.073
7	LIG	C8	0.097	-0.087	0.121	-0.050
7	LIG	H5	-0.210	-0.003	-0.132	-0.014
7	LIG	C9	0.138	0.175	0.066	-0.087
7	LIG	H6	-0.117	-0.153	0.015	-0.092
7	LIG	C10	0.094	0.092	-0.042	0.258
7	LIG	H7	-0.047	-0.078	0.007	-0.087
8	LIG	C1	-0.064	0.087	0.031	0.083
8	LIG	N1	1.569	1.645	1.987	1.243
8	LIG	C2	-0.039	0.046	0.084	0.207
8	LIG	C3	0.088	-0.066	-0.023	-0.201
8	LIG	C4	-0.080	0.107	0.098	0.360
8	LIG	C5	0.089	0.098	0.311	-0.005
8	LIG	C6	0.089	-0.065	-0.062	-0.009
8	LIG	C7	-0.082	-0.036	0.102	-0.027
8	LIG	C8	0.353	0.012	-0.100	0.239
8	LIG	C9	-0.112	0.128	-0.221	-0.010
8	LIG	C10	0.305	0.168	0.426	-0.029
8	LIG	C11	-0.237	-0.188	-0.364	-0.076
8	LIG	C12	0.160	-0.045	0.148	0.054
8	LIG	H1	-0.030	-0.076	-0.049	-0.094
8	LIG	H2	-0.015	-0.037	-0.065	-0.102
8	LIG	H3	-0.150	-0.101	-0.265	-0.153
8	LIG	H4	-0.034	-0.071	-0.053	-0.132
8	LIG	H5	-0.154	0.023	-0.107	-0.143
8	LIG	H6	-0.084	-0.115	-0.038	-0.010
8	LIG	H7	-0.085	-0.148	-0.080	-0.043
8	LIG	H8	-0.103	-0.217	0.027	-0.049
8	LIG	H9	-0.107	0.062	0.017	-0.005
8	LIG	H10	-0.156	0.019	-0.041	-0.106
8	LIG	H11	-0.013	-0.053	-0.174	-0.023
8	LIG	H12	-0.401	-0.393	-0.681	-0.489
8	LIG	H13	-0.536	-0.540	-1.000	-0.600
8	LIG	H14	-1.000	-0.789	-1.000	-0.528

7.2 Chapter 5 Additional Information

7.2.1 Local minimum torsion values obtained for para- and meta- substituted dimers

Table 7.1: Torsion values found at performing DFT dihedral scan for all dimer polymers studied.

Dimer	Screw-sense	Conformation	Dihedral value (Degree)
Para-PFP	cw	ct	141
		cc	56
	ccw	ct	-131
		cc	-62
Para-RN1	cw	ct	147
		cc	61
	ccw	ct	-149
		cc	-70
Para-SN1	cw	ct	140
		cc	61
	ccw	ct	-141
		cc	-70
Meta-PFP	cw	ct	143
		cc	62
	ccw	ct	-145
		cc	-79
Meta-RN1	cw	ct	141
		cc	65
	ccw	ct	-142
		cc	-66
Meta-SN1	cw	ct	137
		cc	66
	ccw	ct	-142
		cc	-49

7.2.2 Backbone and side chain torsion dependence plots from blind helical search

Para-PFP

Para-RN$_1$

Meta-PFP

Para-SN$_1$

Meta-RN$_1$ Meta-SN$_1$
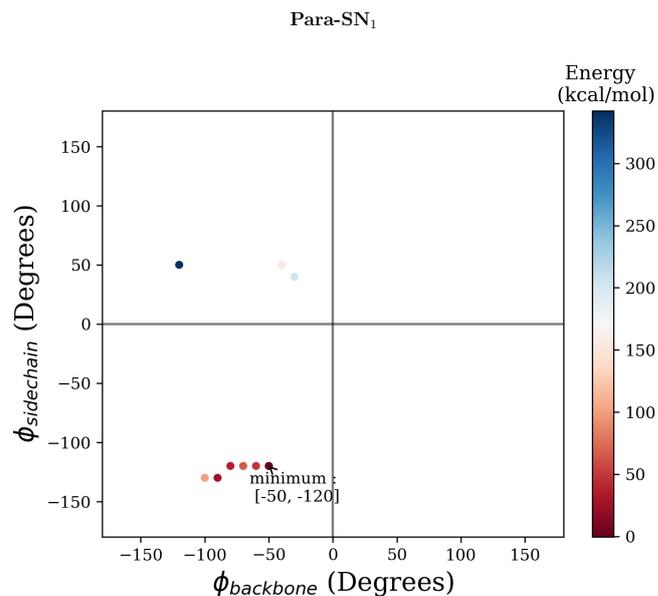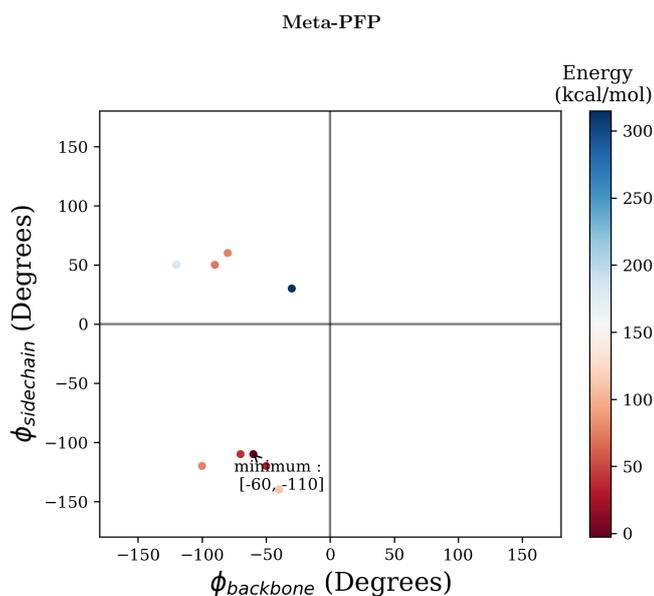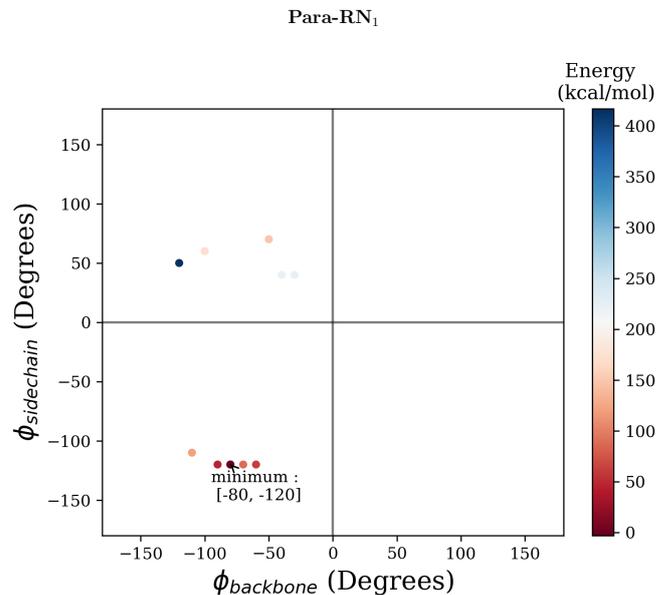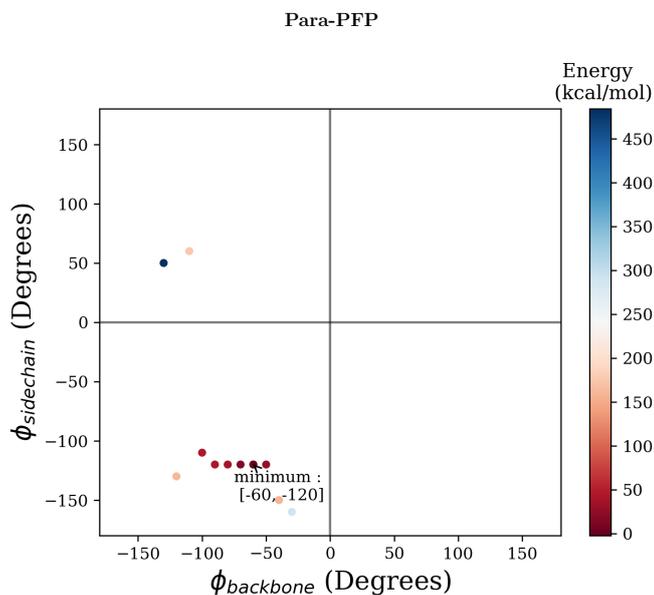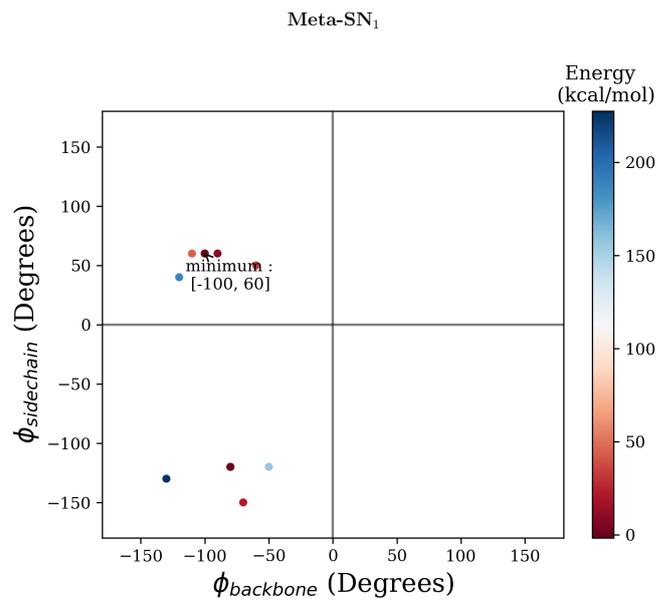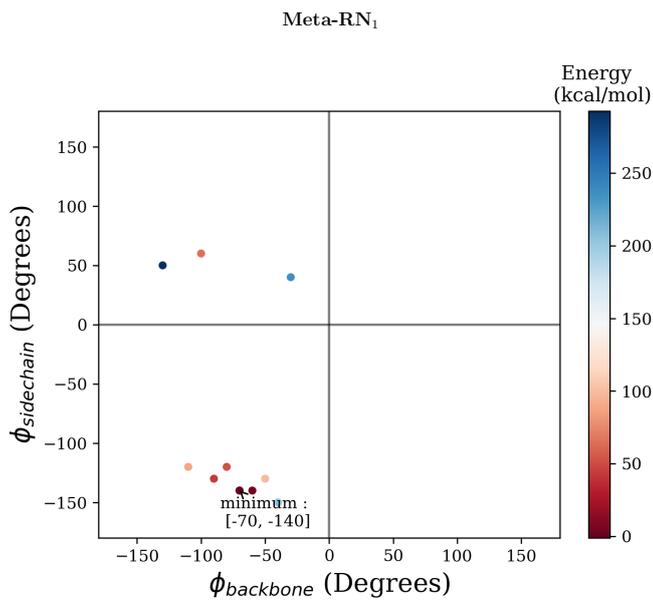
7.2.3   Aromatic fingerprints obtained for lowest helical conformations found

Aromatic Fingerprint *cis-cisoid* para-RN$_1$ Aromatic Fingerprint *cis-cisoid* meta-RN$_1$
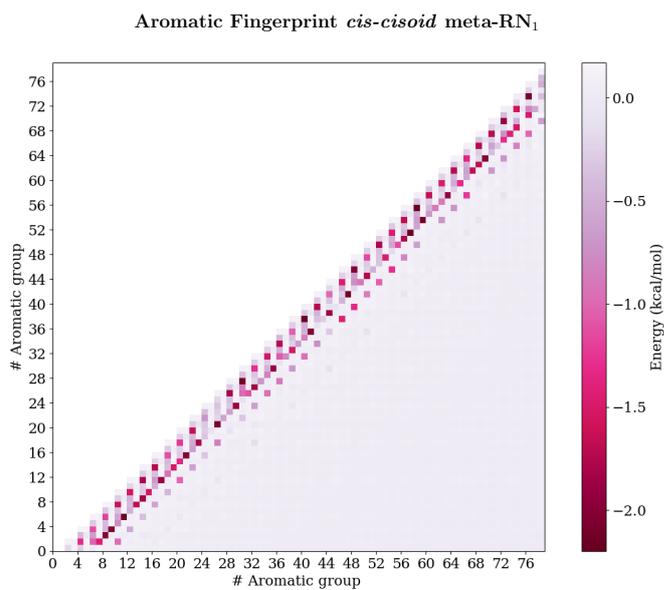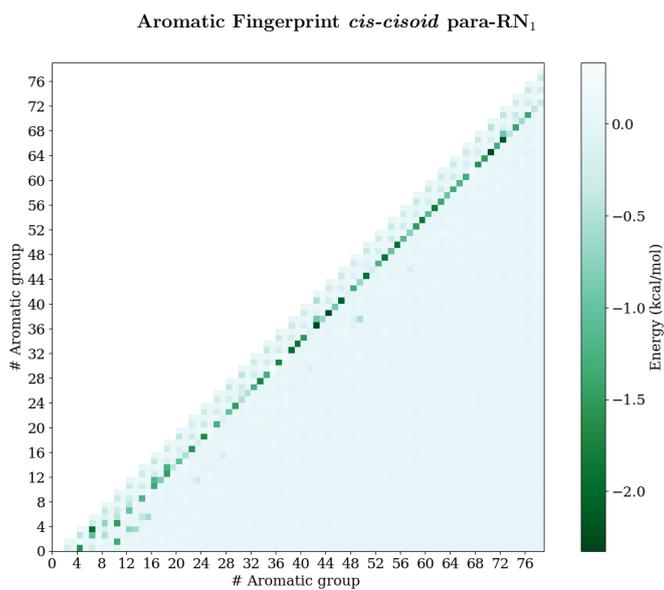
Aromatic Fingerprint *cis-transoid* para-RN$_1$
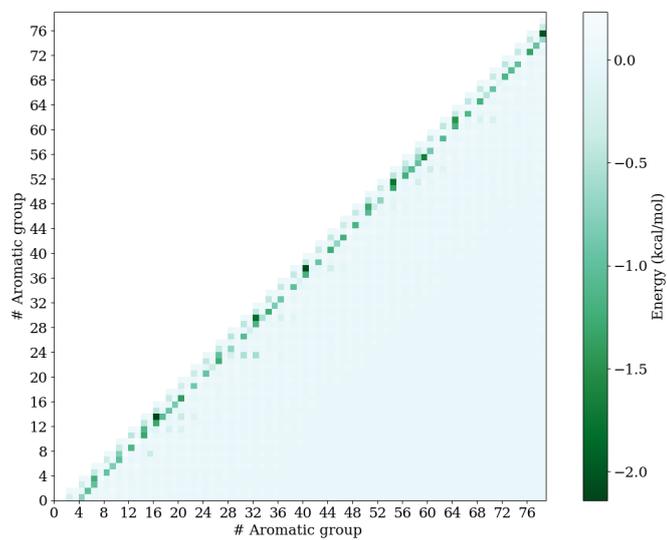
Aromatic Fingerprint *cis-transoid* meta-RN$_1$
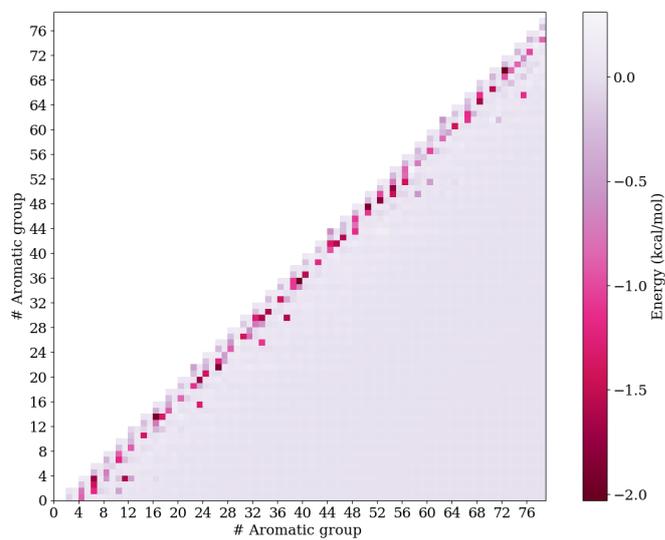
7.3 Chapter 6 Additional Information

7.3.1 H-Bond Cluster results

Sequence A_1
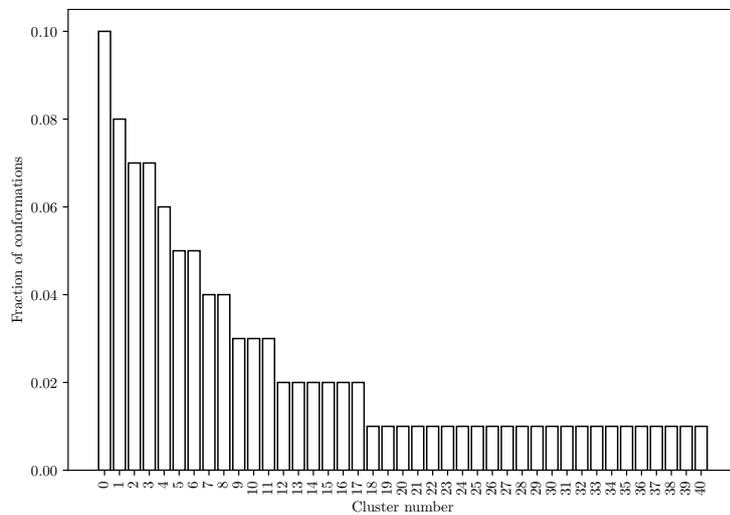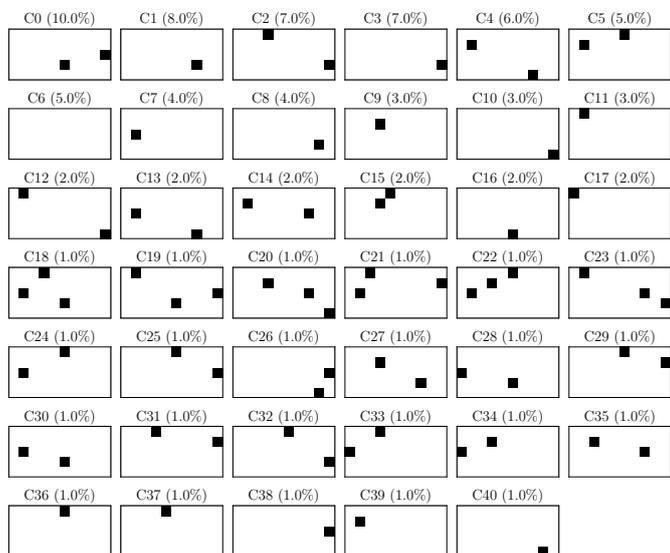
Sequence A_2
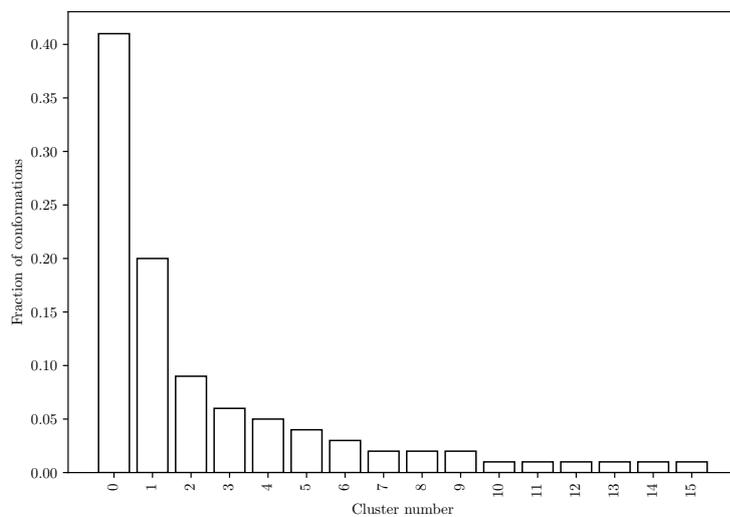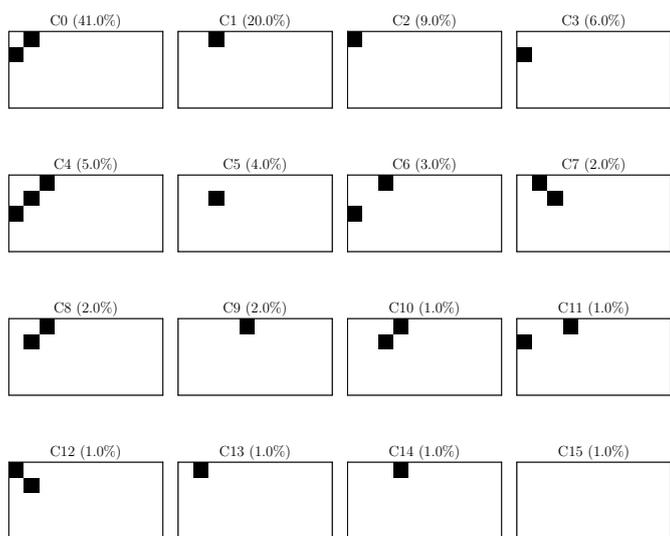
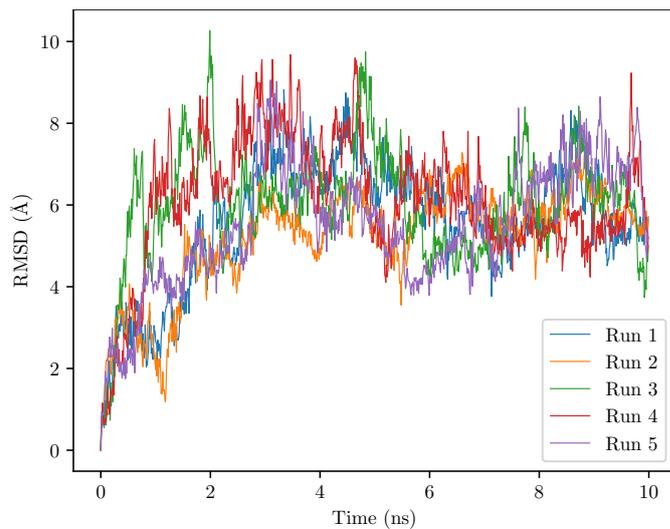## Sequence A_3
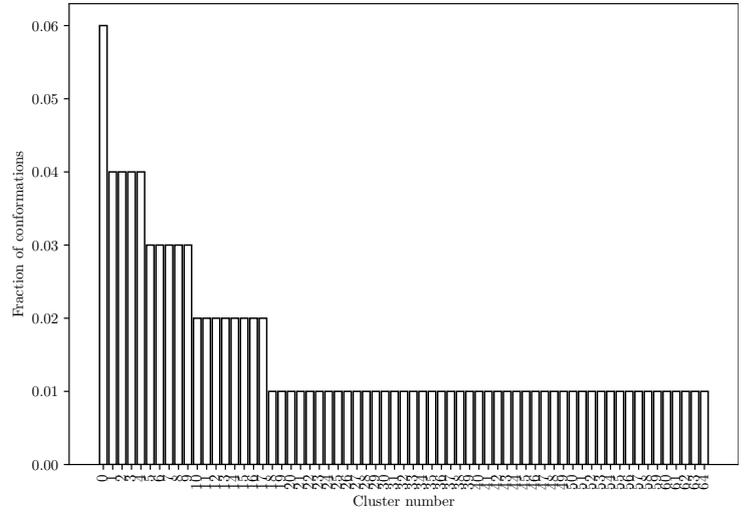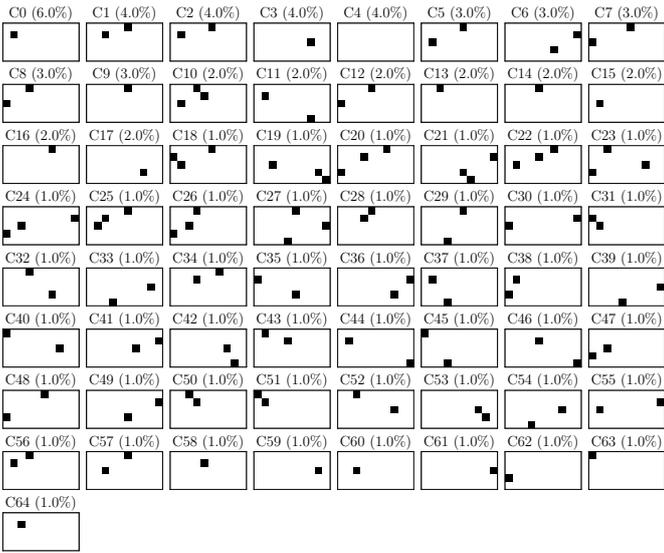
Sequence A_4
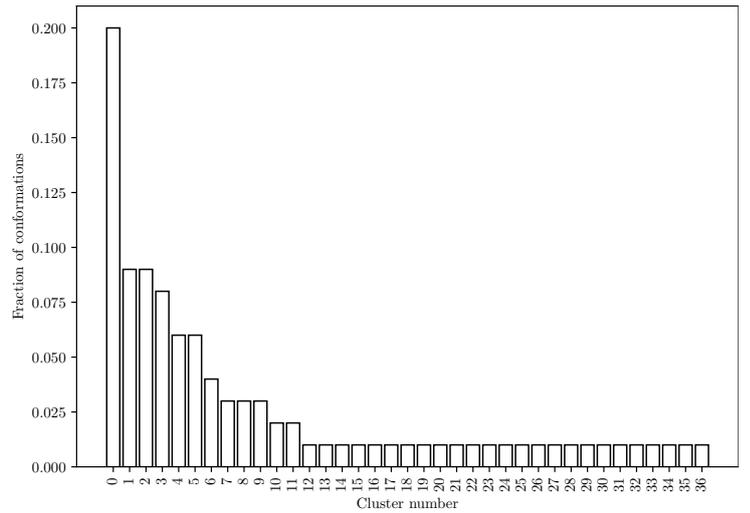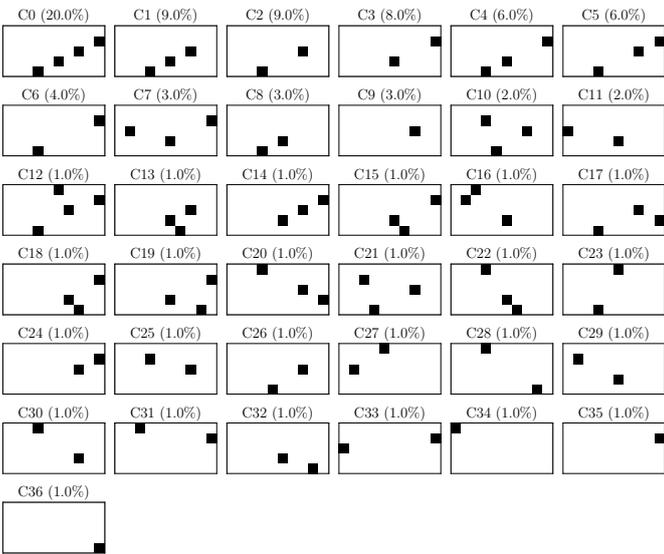
Sequence A_5
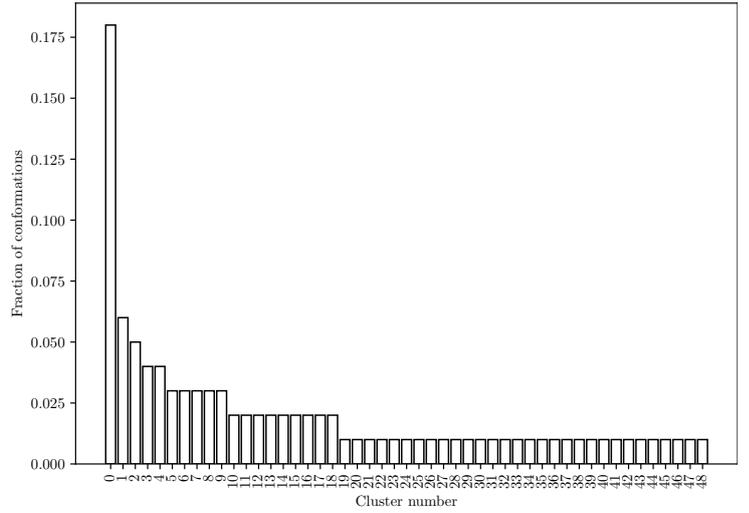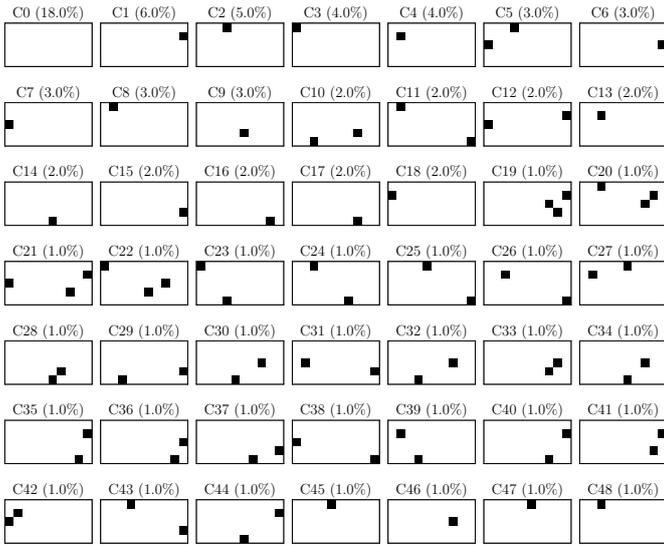
Sequence A_6
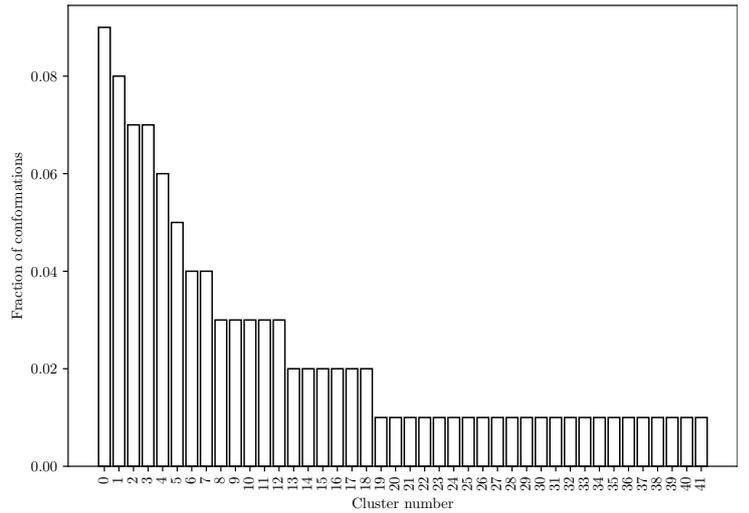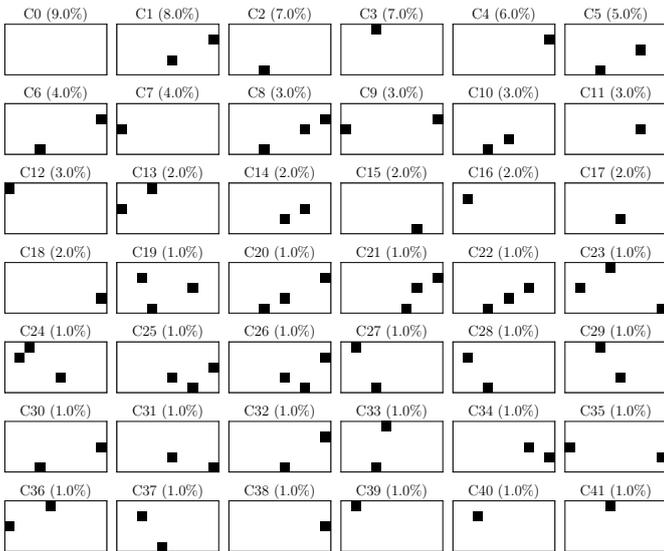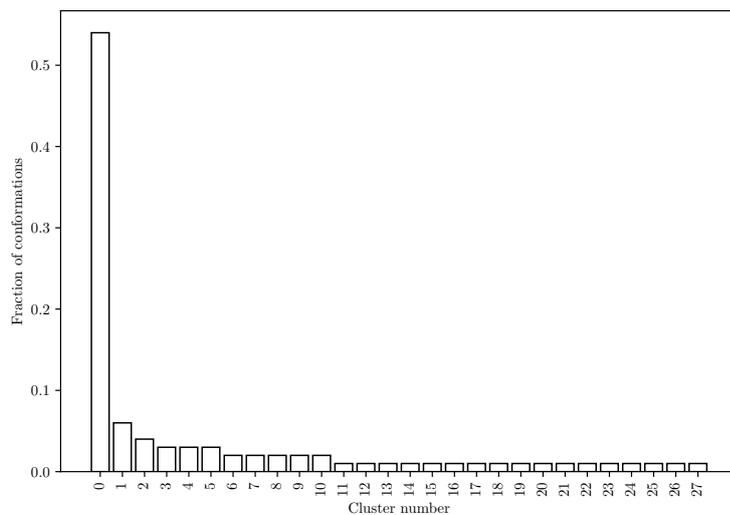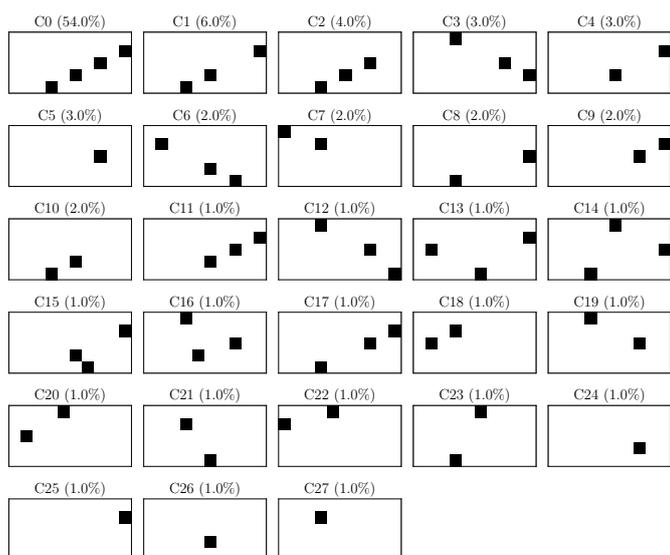
89

Sequence S_1
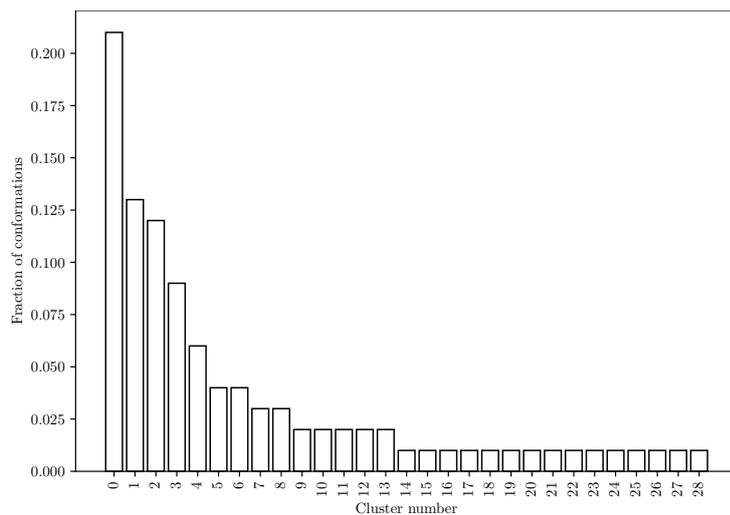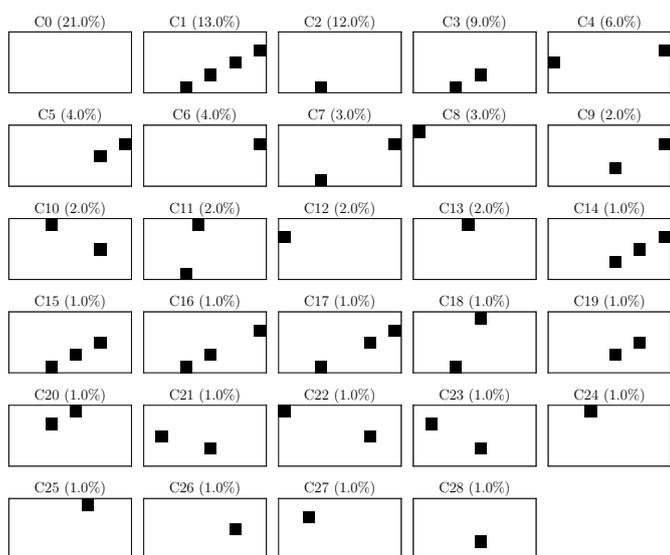
Sequence S_2
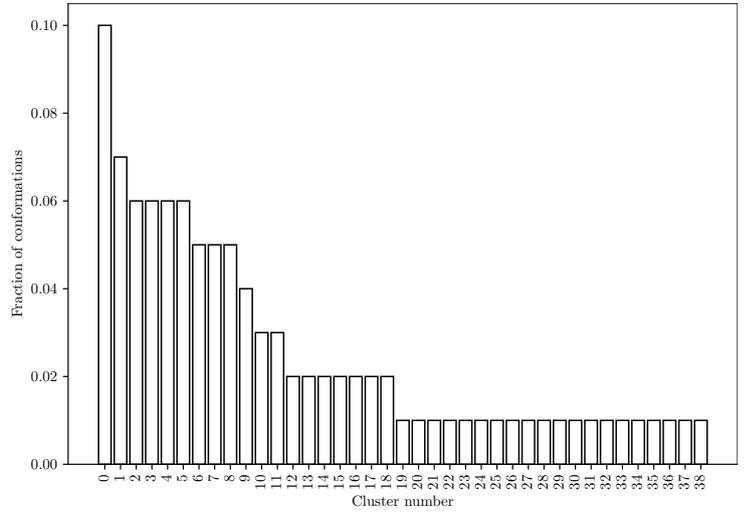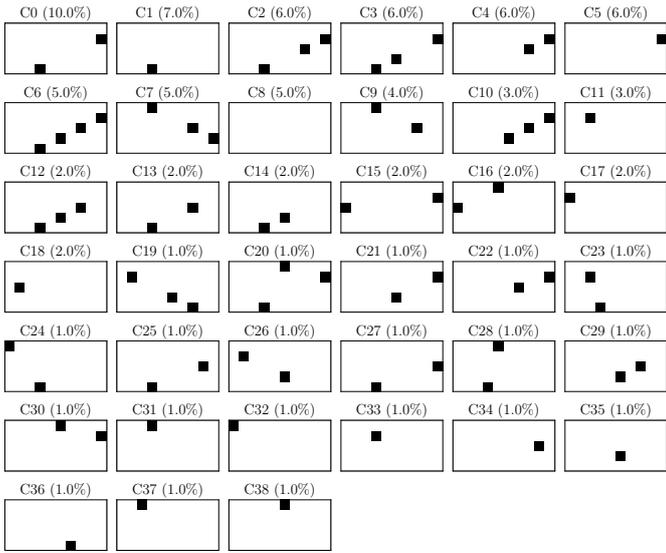
Sequence S_3
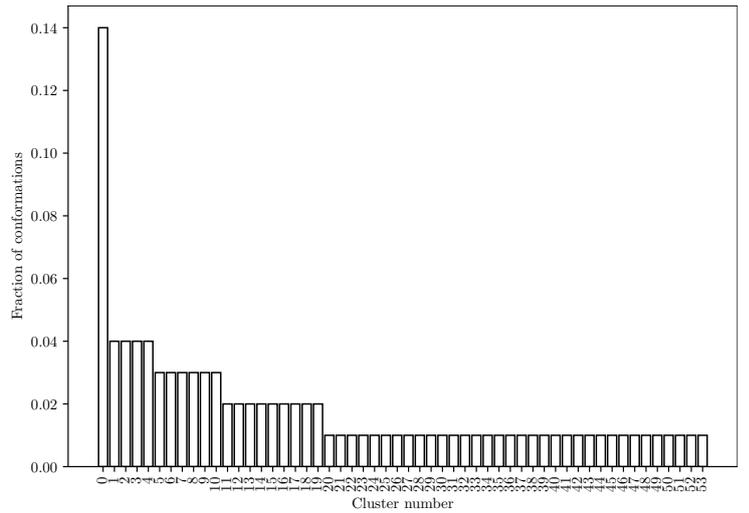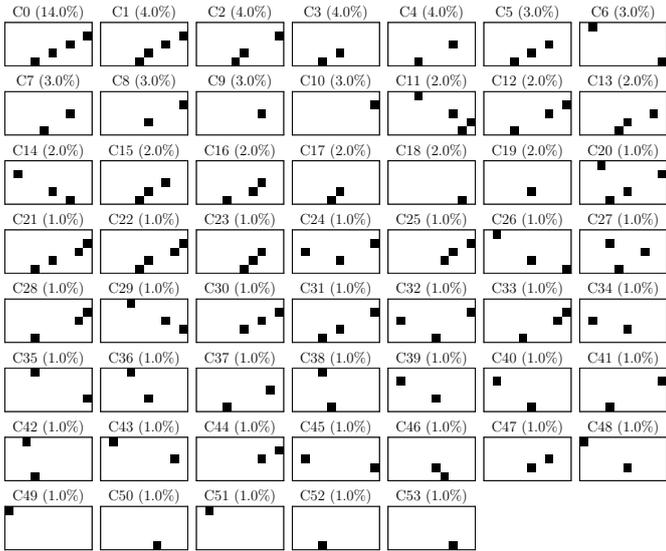
Sequence D_1
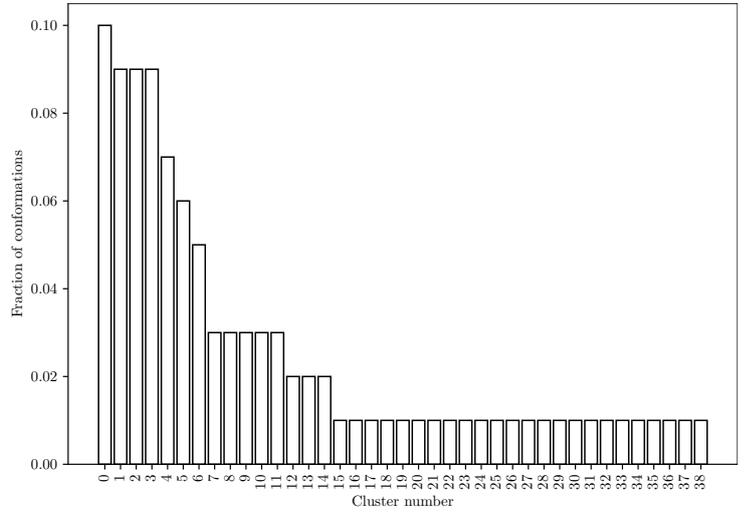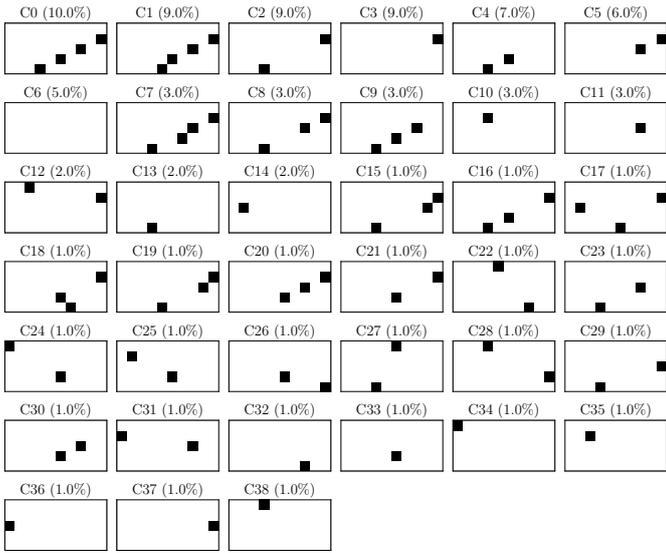
Sequence D_2
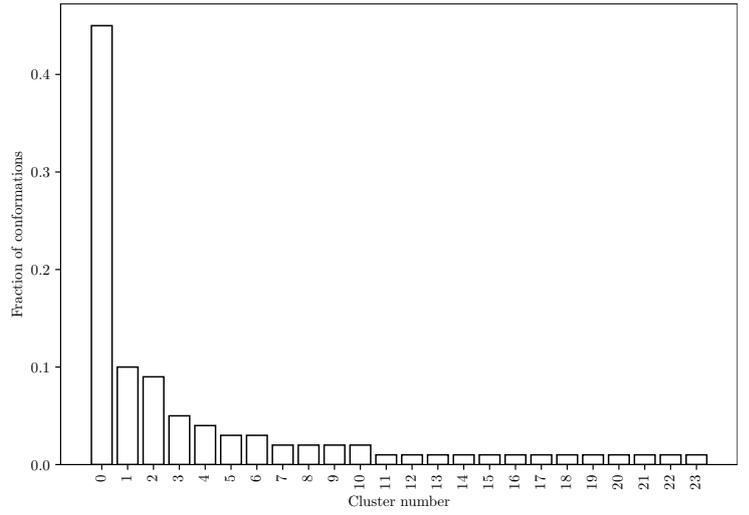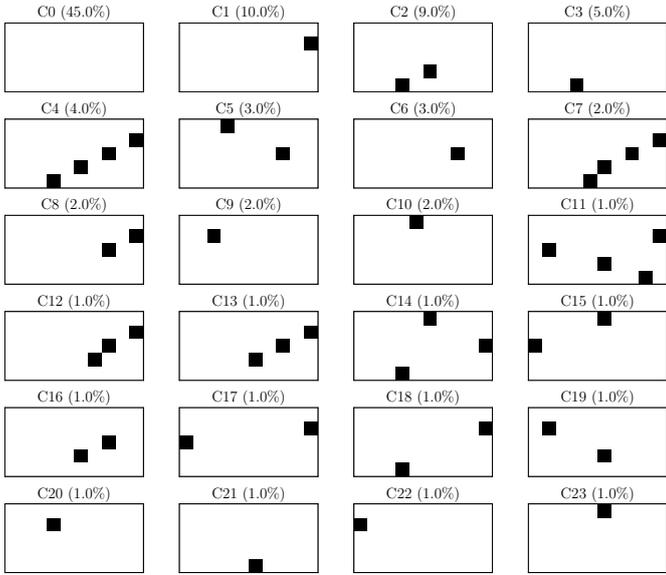
Sequence D_3
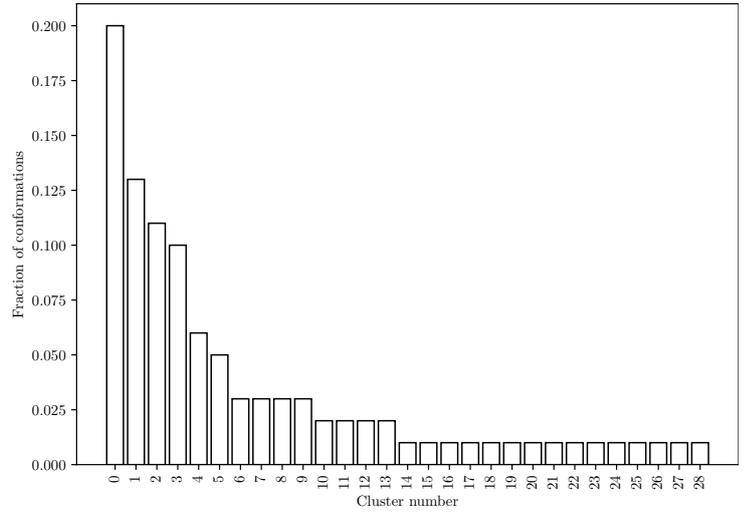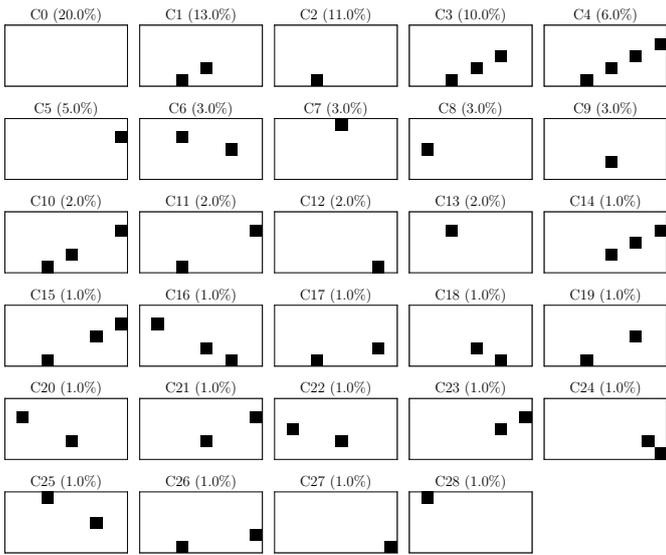
Sequence Z_1
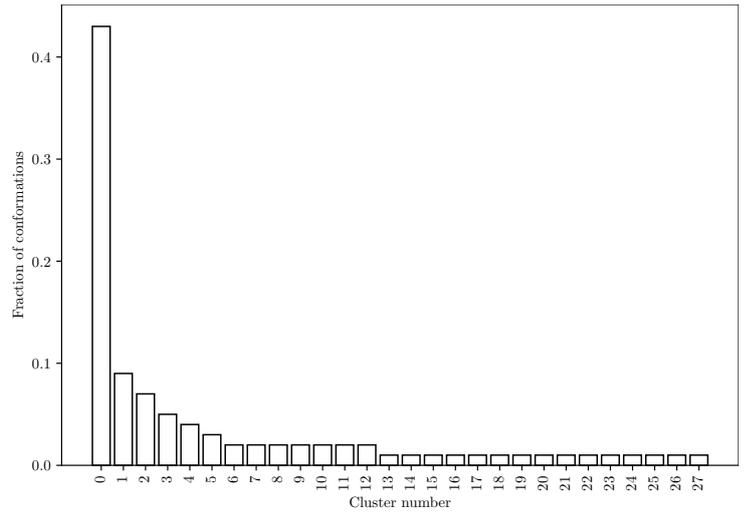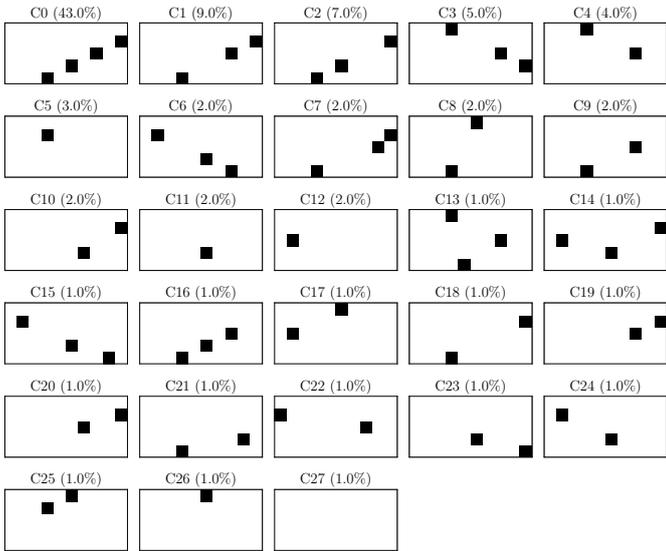
Sequence Z_2
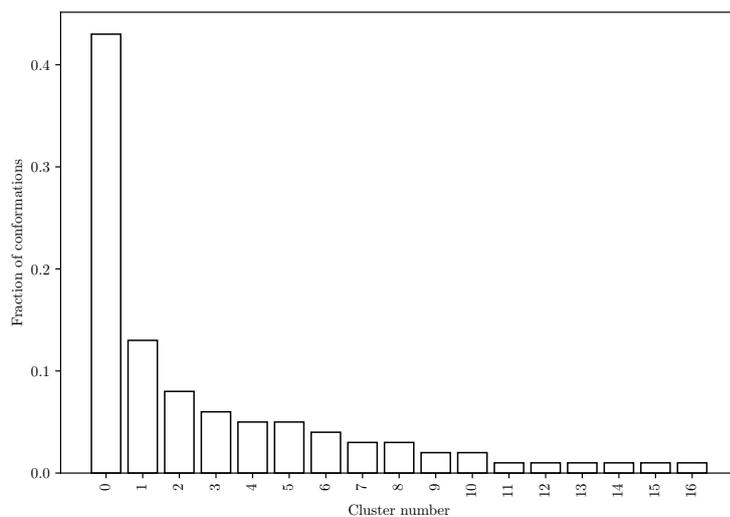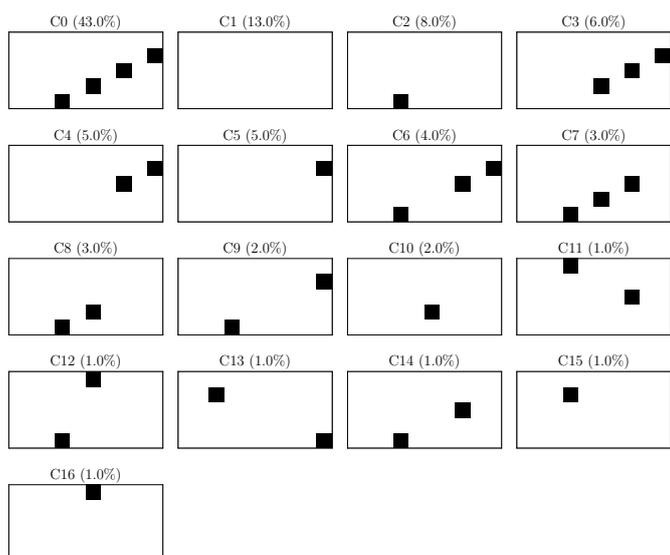
Sequence Z_3

7.3.2 Molecular Dynamics trajectory results
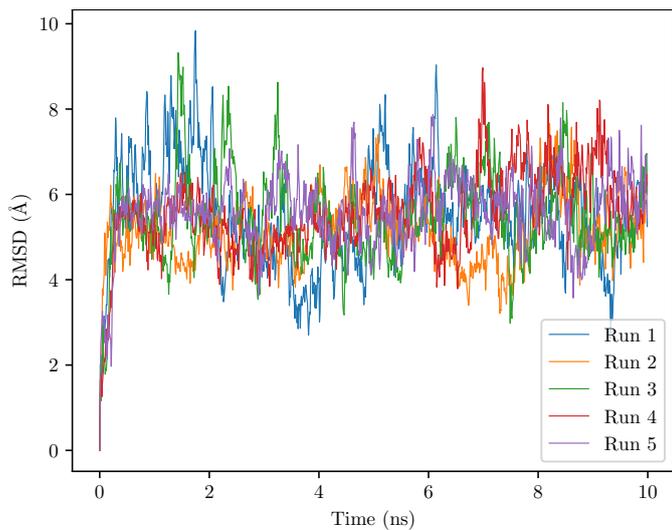
Sequence A_4
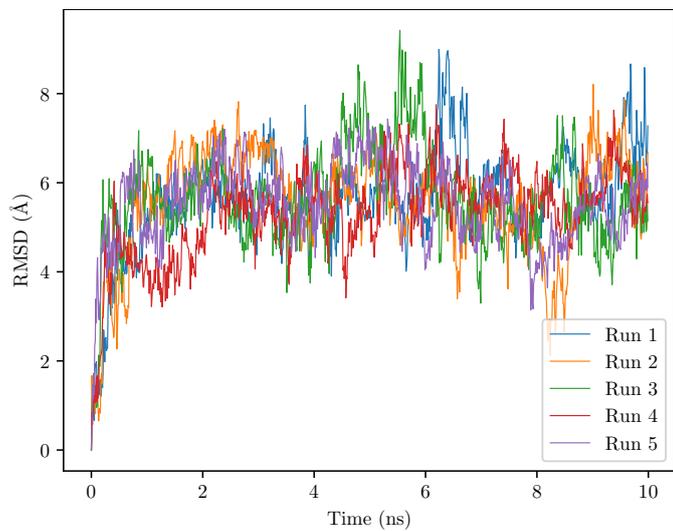
Sequence S_1
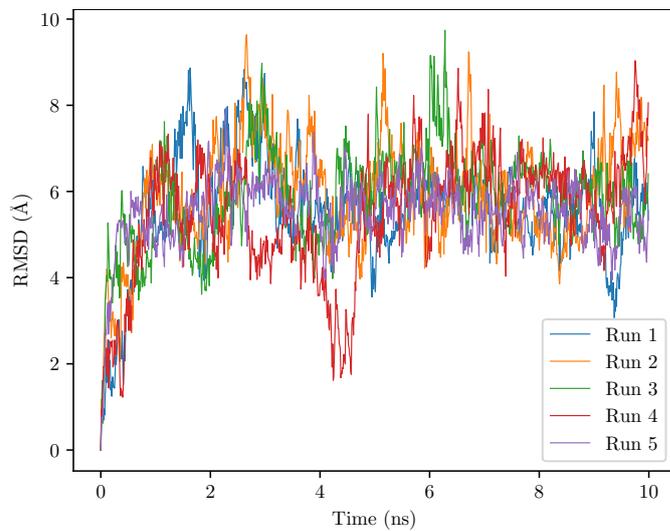
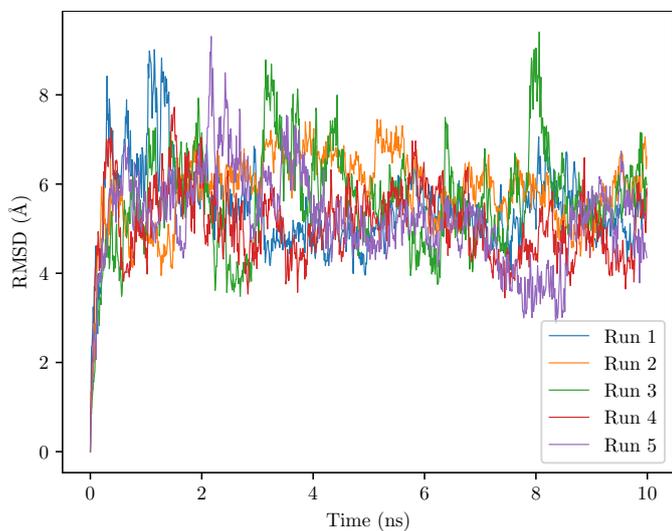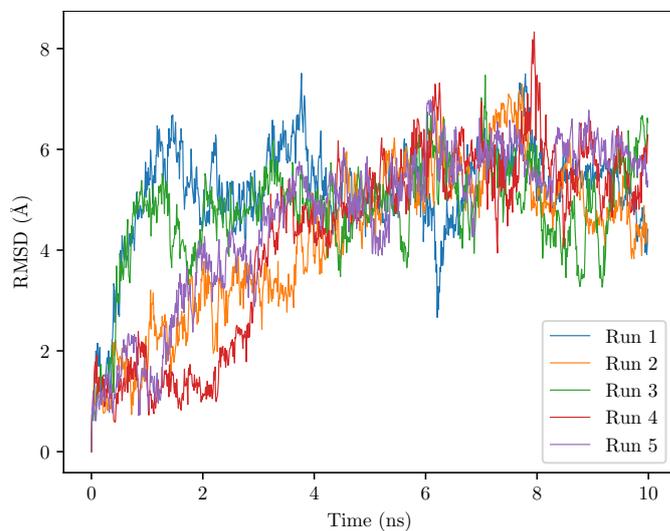**Sequence S_2**
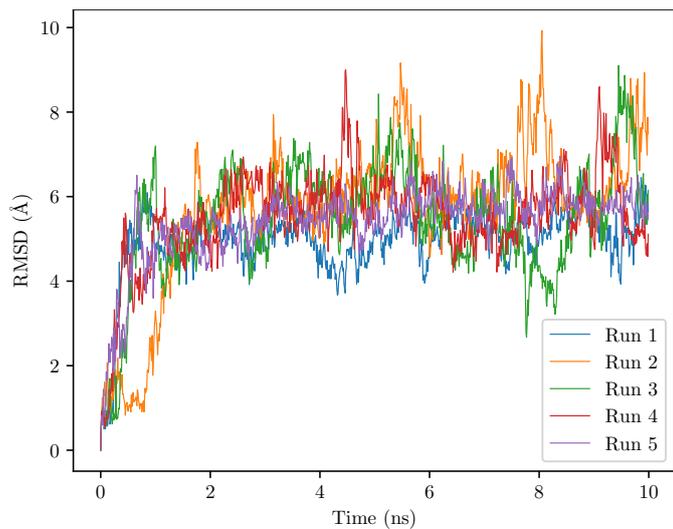
Sequence S_3

Sequence D_1
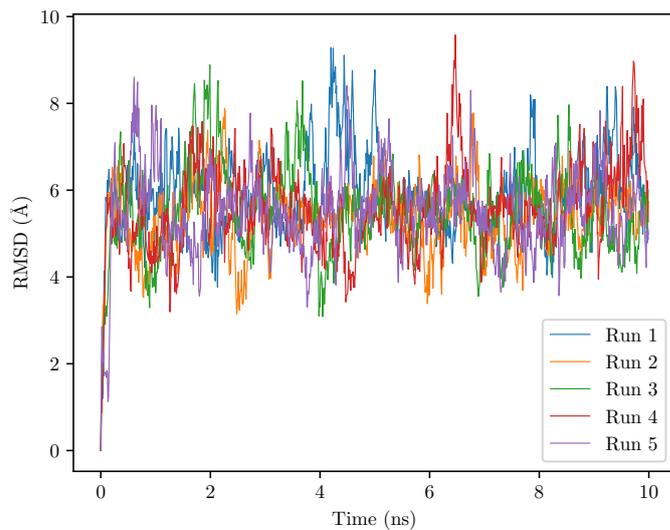
Sequence D_2

Sequence D_3

Sequence Z_1
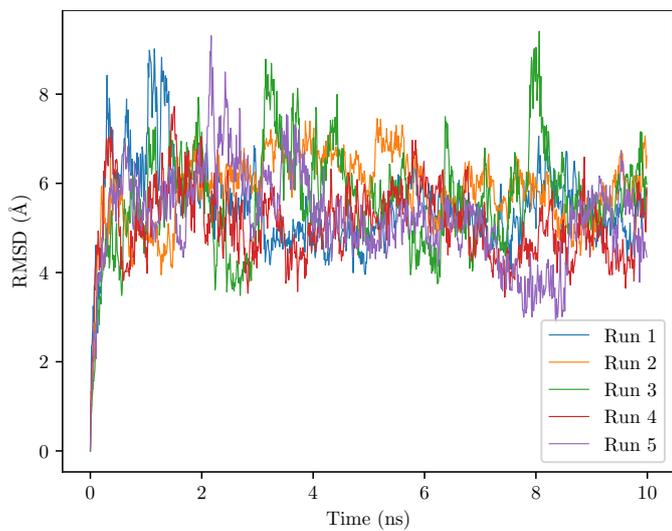
Sequence Z_2

Sequence Z_3
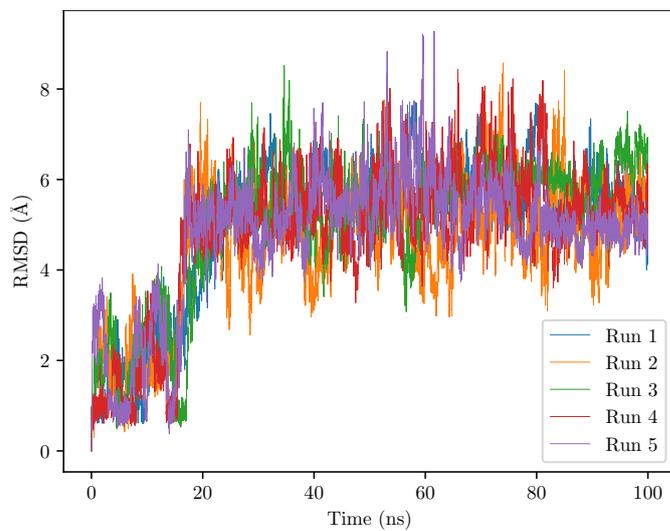
List of abbreviations

3D	Three dimension
4-CaPh	4-carboxyphenyl
AFM	Atomic Force Microscopy
AM1	Austin Model 1 semi-empirical method
AMBER	Assisted Model Building with Energy Refinement
AS	Annealing Simulation
BCC	Bond Charge Corrections
BO	Born-Oppenheimer
cc	Cis-cisoid
CCW	Counterclockwise
CD	Circular Dichroism
CHARMM	Chemistry at Harvard Macromolecular Mechanics
ct	Cis-transoid
CW	Clockwise
D3-BJ	Grimmes' dispersion correction with Becke-Johnson dumping
def2-TZVPP	Karlsruhe basis set split valence with polarization functions on heavy atoms
DFT	Density Functional Theory
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DP	Degree of polymerization
ECN	Effective Coordination Number
FF	Force Field
GAFF	General Amber Force Field
GROMACS	GROningen MAchine for Chemical Simulations
H-Bond	Hydrogen Bond
H-K	Hohenberg-Kohn
HIE	Helical Index Error
HPC	High Power Computer
ICD	Induced Circular Dichroism
IHP	Induced Helical Polymer
KIT	Karlsruhe Institute of Technology
KS-DFT	Kohn-Sham Density Functional Theory
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
MD	Molecular Dynamics
MS	Mass Spectroscopy
NMR	Nuclear Magnetic Resonance
NNN	Number of nearest neighbors
OE	Oligomer Encoder
OPLS	Optimized Potentials for Liquid Simulations
PES	Potential energy surface
PFP	Pentafluorophenyl
PMMA	Poly methylmetacrilate
PPA	Poly phenylacetylene
RESP	Restrained electrostatic potential
RMSD	Root Mean Square Error
SDO	Sequence-defined Oligomer
SIMONA	Simulation of Molecular and Nanoscale Systems
SMILES	Simplified Molecular Input Line Entry System
STM	Scanning Tunneling Microscopy
VASP	Vienna Ab initio Simulation Package
vdW	van der Waals
WaNo	Workflow-active Node
XDR	X-ray diffraction
XML	Extensible Markup Language

Eidesstattliche Erklärung

Hiermit versichere ich eidesstattlich, dass ich die hier vorgelegte Dissertation selbständig angefertigt habe und keine anderen Quellen und Hilfsmittel genutzt habe als die hier angegebenen. Die wörtlich und inhaltlich übernommenen Stellen wurden als solche kenntlich gemacht. Die Regeln zur Sicherung guter wissenschaftlicher Praxis des Karlsruher Instituts für Technologie (KIT) in der gültigen Fassung wurden beachtet und Primardaten gemäß Abs. A(6) gesichert. Die elektronische Version der Arbeit stimmt mit der schriftlichen überein. Die Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde zur Erlangung eines akademischen Grades vorgelegt.

Karlsruhe, 02.11.2022

Montserrat Peñaloza-Amion