Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Materials Science and Technology (35^{th}cycle)

Multiscale Molecular Modelling and Advanced Simulations of Supramolecular Systems

By

Luigi Leanza

******

Supervisor(s):
Prof. G. M. Pavan, Supervisor

Doctoral Examination Committee:
Prof. F. Pietrucci, Sorbonne Université
Prof. M. Salvalaglio, University College London
Prof. R. Pisano, Politecnico di Torino
Prof. G. Ragazzon, Université de Strasbourg
Prof. P. Ugliengo, Università di Torino

Politecnico di Torino
2023

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Luigi Leanza
2023

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

"Every step I have taken in my life has led me now, here."

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Giovanni M. Pavan, for his guidance and support throughout my PhD journey. His mentorship and expertise have been instrumental in inspiring my work and my academic growth, and contributed to the success of this Thesis.

I would also like to extend my thanks to the reviewers of my Thesis, Professor Fabio Pietrucci and Professor Matteo Salvalaglio, for their feedback and constructive criticism, which helped me to refine and improve the quality of my work.

I am also deeply grateful to Professor Salvalaglio for hosting me during my stay. His hospitality and mentorship have been an invaluable source of knowledge, inspiration, and motivation for me, and have enriched my research experience.

I would also like to acknowledge the support of my colleagues, whose collaboration and shared enthusiasm have created a stimulating research environment.

I thank my family for always supporting and encouraging me. Their love, understanding, and sacrifices have been a constant source of motivation and strength.

Finally, I want to express my deepest thanks to my partner for her unwavering support and belief in my abilities, especially during difficult times. Her presence has provided me with strength, encouragement, and motivation, and this accomplishment would not have been possible without her support.

Abstract

Many biological systems are *de facto* supramolecular materials that possess unique intriguing properties, such as the ability to self-heal, dynamically reconfigure their structure, respond to stimuli, and perform complex functions. These systems are composed of molecular subcomponents that interact and self-assemble, interconnecting with each other through noncovalent interactions. Designing artificial supramolecular systems that possess similar properties and self-assembling principles to those found in nature could lead to the development of a new generation of bioinspired materials that could revolutionize various fields, from biomedical and environmental, to energy and electronics. However, this requires gaining a detailed comprehension of the main molecular factors that control the structure and properties of these supramolecular systems, which is very challenging to attain. Molecular models and computer simulations are fundamental in this field. These tools permit us to investigate supramolecular systems from a privileged point of view and at an unprecedentedly high resolution. In this way, it is possible to reach a molecular-level comprehension of the behaviour and properties of supramolecular systems, which offers unique opportunities toward the design of new materials with bioinspired properties. This PhD Thesis focuses on the computational modelling study of a variety of supramolecular systems via the development of multiscale (atomistic and coarse-grained) molecular models and the use of classical molecular dynamics and advanced molecular simulation approaches (enhanced sampling techniques such as Metadynamics and Umbrella Sampling). Firstly, the chain-capping and supramolecular polymerization of porphyrin-based monomers is investigated. Through the extensive use of Metadynamics simulations and coarse-grained models, we gain a deeper understanding of the interaction between monomers and reveal hidden processes involved in the monomer exchange phenomena. The interaction between isolated nanotubes in solution is then assessed, by employing multiscale modelling. Our approach helps to identify the primary driving forces that lead to the formation

of intertwined supramolecular structures, providing a link between experimental observations and the molecular factors that drive the observed phenomena. We then extend the investigation to the interaction of a Benzene-1,3,5-tricarboxamide-based supramolecular polymer (BTA-OEG_4) with biological structures under physiological conditions to evaluate their potential as biomaterials. Our simulations reveal that the dynamic supramolecular nature of these fibres allows the interaction between BTA-OEG_4 and BSA, and elucidate the mechanism of monomer exchange in this system during the complexation. Finally, we show a flexible computational protocol that allows to reconstruct the thermodynamics and kinetics of different types of mechanically interlocked molecules (MIMs) at atomistic resolution. Though classical molecular dynamics and Metadynamics simulations, we gain deep insights into the free-energy characterization of MIMs and reveal the delicate balance between various intermolecular interactions (*e.g.* H-bonding, solvent effects, conformational entropy) that determine their behaviour. We also demonstrate the impact of changing environmental variables (such as the solvent type) on the dynamical behaviour of the MIMs. Our approach also includes a free-energy decomposition into enthalpy and entropy contributions, and detailed structural and dynamical analyses that provide unprecedented insights into the mechanisms and key transitions ruling the intermolecular movements between metastable states.

Overall, this PhD Thesis makes a significant contribution to the field of supramolecular systems, providing a detailed molecular-level understanding of fundamental processes involved in the self-assembly and dynamic behaviour of supramolecular polymers, and resolving the slow dynamics of rotaxanes and molecular shuttles with sub-molecular resolution. This work demonstrates the power and potential of molecular modelling in this area, paving the way for the rational design of artificial supramolecular systems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Self-assembly and supramolecular systems

Self-assembly is a fascinating phenomenon that has captivated researchers in the last decades[1, 2]. From the molecular level (*e.g.* the formation of crystals) to the macroscopic scale (*e.g.* the formation of cell membranes), self-assembly is ubiquitous in nature[1–6]. At its heart, self-assembly is the spontaneous organization of individual components into larger, more complex structures, that occurs without any external intervention[3, 4, 7, 8]. The key to self-assembly is the presence of specific interactions between the components that allow them to recognize and bind to one another in a organized manner[1]. These interactions are non-covalent forces (electrostatic, van der Waals interactions, hydrogen-bonding, aromatic stacking, metal coordination) that are relatively weak (2-50 kJ mol^{-1}) when compared to covalent bonds (100-400 kJ mol^{-1}), but many of these weak interactions together can generate a final highly stable assembly (Fig. 1.1).

Supramolecular chemistry has greatly benefited from the study of the self-assembly processes exhibited by nature. Nature, in fact, uses self-assembly to construct supramolecular structures that possess the ability to communicate with each other, respond to stimuli, convert energy from one form to another, control movement at the molecular level, and ultimately perform complex functions that are at the basis of many living materials (Fig. 1.2, top). In recent years, scientists have been inspired to use self-assembly as a promising strategy for designing and synthesizing supramolecular materials[1, 5, 9–12]. By controlling the interactions between

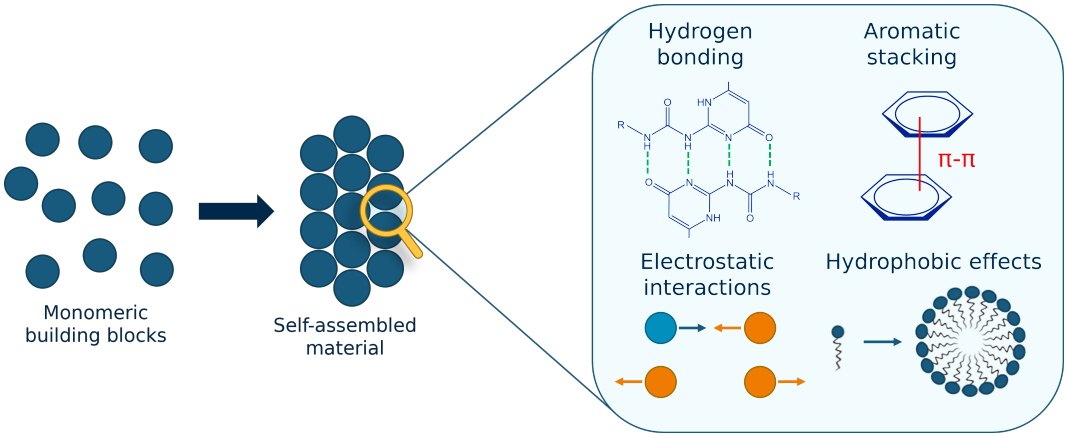

Fig. 1.1 Conceptual scheme of self-assembly: individual monomeric building blocks spontaneously assemble in larger aggregates, held together by non-covalent interactions.

the components, researchers have been able to create a wide range of structures, including nanoparticles[13], monolayers[14], vesicles[15], and even more complex architectures[16–18] (Fig. 1.2, bottom).

Two classes of supramolecular systems that have received particular attention in this context – and that will be the focus of this Thesis – are *supramolecular polymers* and *molecular machines*.

Supramolecular polymers[19] are formed through the self-assembly of monomers, which interact with each other via non-covalent forces to form long chains (Fig. 1.3). Supramolecular polymers have many potential applications, due to their ability to be tailored to have specific properties[20]. They can be controlled in the chemical nature of the building blocks[21], the strength of intra- and inter-molecular interactions[22], shape and size[23], or to have specific functional groups that can interact with other molecules[24]. This makes them potentially useful in a wide range of fields, including drug delivery[25], tissue engineering[26], and catalysis[27].

Molecular machines[28] are a class of molecules that are capable of performing mechanical work at the molecular level. These molecules are typically made of small organic molecules that are assembled in a specific way to perform a range of functions, such as the ability to move[29], respond to stimuli[30], and sense their environment[31]. Molecular machines have the potential to be used in a wide range of applications, including medicine[32], material science[33], and energy production[34] (see Fig. 1.4).

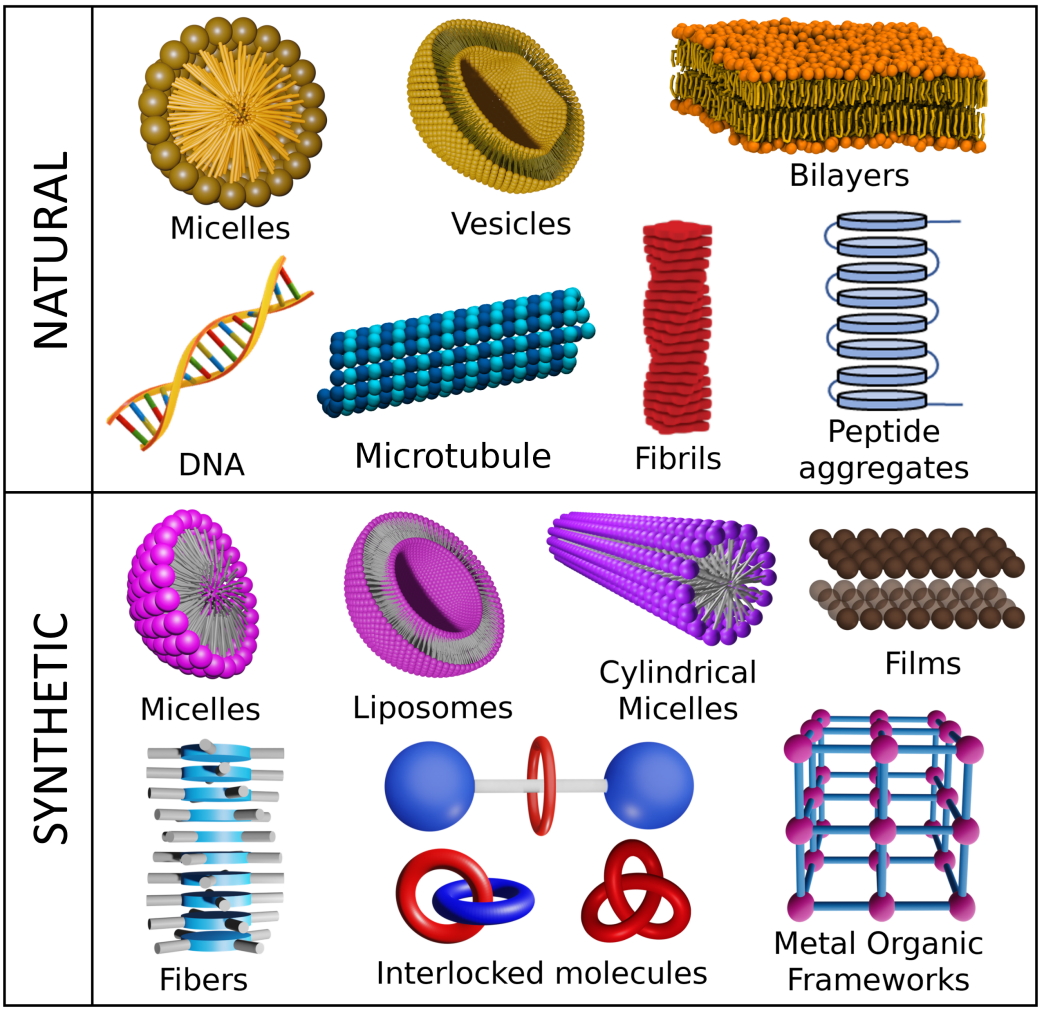

Fig. 1.2 Examples of natural (top) and synthetic (bottom) supramolecular systems.

The two classes of molecules that have been introduced in this section will be discussed in more detail in the following paragraphs. We will examine the unique properties and characteristics of some examples of natural supramolecular systems and then turn our attention to their artificial counterparts. We will also discuss their applications in various fields, exploring how these molecules are used in a wide range of contexts. Whether natural or artificial, these molecules play important roles in many areas of science and technology, and understanding their properties and characteristics is crucial to fully harness their potential.

1.2 Supramolecular polymers

Supramolecular polymers represent a promising class of self-assembly systems for material design and synthesis due to their unique properties and versatility. The reversible nature of the non-covalent bonds that hold monomers together allows for tunable mechanical and physical properties, making them suitable for a wide range of applications. The reversibility of the bonds also allows for easy modifications and recycling of these materials, when compared to traditional polymers, making them highly sustainable and environmentally friendly. There is no surprise, then, that there is an ever-increasing interest in supramolecular polymers, and researchers and industries continue to explore their potential.

1.2.1 Natural supramolecular polymers

Nature is a source of numerous examples of supramolecular polymers. A variety of biological structures exhibit, in fact, highly-organized and well-defined polymeric architectures, that are the results of intricate specific non-covalent interactions at the molecular level.
Microtubules (MTs) are a notable example of supramolecular polymers that can be found in nature. MTs are fundamental components of the cytoskeleton, as they provide structural support and help maintain cell shape. They are also involved in the movement of materials within cells, achieved with motor proteins (kinesin and dynein) which transport cargoes along the MTs[35].
MTs are long and thin tubes composed of subunits called *tubulin*[36] and assemble in a hierarchical structure. The basic unit of a MT is a *heterodimer* consisting of two different subunits, α- and β- tubulin, which are held together by non-covalent interactions, hydrophobic and polar interactions, stabilized by salt-bridges, and shows shape-complementarity interfaces[37]. The heterodimers then assemble into protofilaments, which are linear chains of alternating α- and β- tubulins. The MT is typically composed of 13 protofilaments, that are arranged to form a hollow cylinder and held together via electrostatic interactions. This arrangement gives a high rigidity to the structure. MTs can be highly dynamic, rapidly assembling and disassembling in response to various signalling pathways and environmental conditions.
In a sense, *DNA* is another example of natural supramolecular polymer, that contains the genetic information necessary for the development and functioning of all living

organisms. DNA is composed of two chains (strands), made up of *nucleotides* (adenine (A), thymine (T), cytosine (C), or guanine (C)), which form a double helix interacting via hydrogen bonds between complementary nucleotides, reinforced by $\pi - \pi$ stacking between the nitrogenous bases[38]. The double-helix formation – but also many other self-assembled processes – displays positive *cooperativity*, which is the phenomenon where monomer addition becomes more favorable as the chain grows[39, 40]. The stability of the double-stranded DNA is enhanced by this phenomenon, meanwhile the reversible nature of the hydrogen bonds between the two filaments is responsible for DNA replication and transcription[41], as these bonds can be easily disrupted and re-formed during these processes.
Another example of natural supramolecular polymers are *amyloid fibrils*, which are aggregates of proteins that are involved in a variety of diseases, including Alzheimer's and Parkinson's disease. Amyloid fibrils are characterized by parallel, highly ordered structures of polypeptide chains with a cross-β conformation, in which the polypeptide chains are oriented perpendicular to the direction of fibre propagation. In amyloid fibrils, the polypeptide chains interact through hydrogen bonding between the backbone amide groups of adjacent chains, forming extended β-sheets[42]. These β-sheets are the primary structural element of the amyloid fibrils, and further self-assemble into filaments and multiple filaments (fibrils). The assembly of fibrils involves various non-covalent intermolecular interactions, such as van der Waals interactions, hydrophobic effects, $\pi - \pi$ interactions[43]. The excellent mechanical properties of amyloid fibrils[44] – given by the hierarchical structure of fibrils and the high number of interactions – and biocompatibility, can be exploited, in principle, to design new functional materials.

The unique and extraordinary properties of many natural structures, including those explored in this Thesis, are due to their supramolecular nature. These properties are often the result of the specific arrangement and interactions of the individual molecules within the system. In an effort to replicate these properties on synthetic materials, scientists are constantly researching and exploring the complexity of supramolecular polymers. As the understanding of these systems advances, we can expect to see an increasing number of new materials designed to mimic supramolecular biopolymers, leading to a new generation of functional materials.

1.2.2 Artificial supramolecular polymers

With the advancement of supramolecular chemistry[45], the idea of replicating the principles employed by many living systems, described in the previous paragraph, on synthetic materials has emerged as an extraordinary innovation. In this context, the synthesis of the first supramolecular polymers was achieved[46].
Supramolecular polymers can be classified based on different principles, such as the type of monomers, the nature of the non-covalent interactions, or the mechanism of formation[19–21]. However, for the purpose of this Thesis, I will focus on ordered 1-dimensional structures. The key feature of this class of supramolecular polymers is the ability to finely tune the strength and the directionality of the interactions among the monomeric units, leading to materials with different features. Stronger interactions lead to the formation of static materials, while lower interactions yield highly dynamic materials. Within the intermediate range lies the opportunity to engineer materials with different properties such as adaptability, responsiveness, and self-repair[20].
The first example of a "mechanically-robust" material based on supramolecular polymers was reported by Meijer and co-workers[47] in 1997. In their study, they used the *quadrupole hydrogen-bonding 2-ureido-4[1H]-pyrimidinone (UPy)* unit to create a material with mechanical properties typically only seen in covalently-bonded macromolecules. A series of hydrogels were synthesized from UPy moieties, used to end-functionalize polyethylene-glycols (PEGs), also from the group of Meijer. By adjusting the supramolecular interactions through molecular design, it was possible to control rheological and material properties. These hydrogels showed promising results to support kidney regeneration[48] and reduce infarct scar[49].
This marked a significant breakthrough in the field, as it led to a drastic expansion in the use of supramolecular polymers for the development of novel materials[50–54].
Benzene-1,3,5-tricarboxamide (BTA) is another motif that has attracted significant interest in recent years in the field of supramolecular chemistry. This monomer consists of a benzene core, decorated with three N-centred or C=O-centred amides, which are capable of self-assembling in one-dimensional supramolecular polymers via three-fold hydrogen bonding between the amides (Fig. 1.3b). While several synthetic approaches towards BTAs have been proposed, they are beyond the scope of this introduction, and the reader is referred to Ref.[55] and related works.
Hydrogen bonding is the primary interaction driving the formation of BTA assem-

blies, but their stability is further enhanced by a combination of $\pi - \pi$ interactions and hydrophobic effects. Quantum calculations revealed that the cooperative growth of these assemblies is driven by dipole-dipole interactions and polarization effects[40]. The versatility of BTA-based supramolecular materials and their physical properties allow for a diverse range of applications, from energy storage to drug delivery. The ferroelectric switching exhibited by alkyl-substituted BTAs and thin films has been extensively investigated[56–58]. In a recent development by Akutagawa[59], a useful method for achieving low-voltage switching through the introduction of chiral alkyl chains was proposed, making chiral BTA derivatives ideal candidates for high-performance non-volatile memory devices. Branched lateral chains can be used in alkane solvents to form organogels, which have a variety of applications in industries, including the production of insulating layers for electric cables[60]. In water, the presence of large hydrophobic spacers that shield the amide units can lead to the formation of hydrogels. These hydrogels have the potential to be used as sensors, scaffolds for tissue engineering, molecular wires for electronic devices, but also for other applications[61, 62]. Recently, the potential of BTA-based supramolecular polymers for biomedical applications has been shown, in particular as magnetic resonance imaging contrast-agent[63], or as a drug delivery system[64]. Finally, it is worth mentioning that the ability of several BTAs to coordinate metal ions can drive benefits in the synthesis of new metal-organic framework materials[65].

Perhaps the most iconic example of how nature inspires scientists in supramolecular chemistry are poprhyrin-based supramolecular polymers. Porphyrins are aromatic molecules that contain a central metal atom surrounded by a ring of four nitrogen atoms (Fig. 1.3c). They can be found naturally in a variety of organisms, including plants, bacteria, and animals, where they play important roles in processes such as photosynthesis and oxygen transport. Scientists have been inspired by their high stability, excellent conductivity, and ability to absorb and emit light, and have worked to emulate these characteristics for the development of synthetic materials[66–68]. The strong ability of porphyrin cores of acting as electron-donors and their extensive π-conjugated systems, have made porphyrin-based compounds widely utilized in photovoltaics for the construction of organic solar cells. Porphyrins can be used as a part of dye-sensitized solar cells (DSSC)[69–72]. In DSSC, the porphyrin dye is applied to a layer of titanium dioxide nanoparticles and absorbs sunlight. The TiO_2 covered with the dye is immersed in an electrolyte solution and sandwiched between an anode and a cathode. When the dye absorbs sunlight, the electrons in the

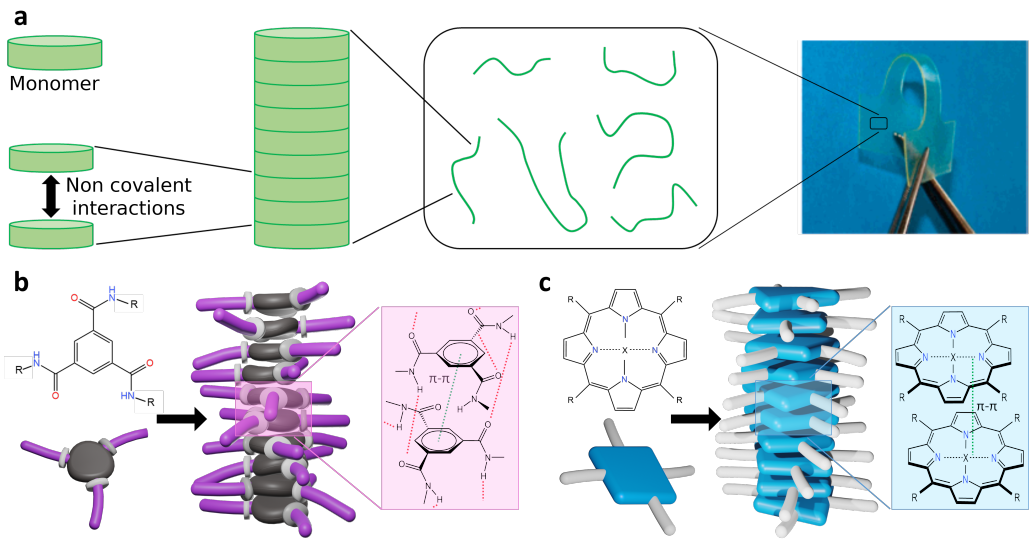

Fig. 1.3 Conceptual scheme of supramolecular polymers. (a) Monomers interacting via non-covalent interactions, assemble to form long fibres in solution. On the right, a supramolecular material based on the quadruple hydrogen bonded ureidopyrimidinone motif, adapted with permission from Ref.[20] (b) BTA- (left) and porphyrin-based (right) supramolecular polymers. Primary non-covalent interactions are showed in the zoomed coloured boxes.

dye are excited to the TiO_2. The excited electrons can flow through the electrolyte solution from the anode to the cathode, generating an electric current. Porphyrins and porphyrin-based polymers have shown excellent potential in the field of medicine, particularly for cancer treatment[73–75]. One of the main mechanisms of action is known as photodynamic therapy (PDT)[76, 77]. In PDT, porphyrins are activated by light, which causes them to react with oxygen to produce reactive oxygen species (ROS). These ROS are highly cytotoxic substances, making them an effective tool in the treatment of various types of cancer. Porphyrins can also absorb near-infrared radiations to create hyperthermia, which can kill tumour cells. This mechanism of action is known as photothermal therapy (PTT)[78, 79]. In addition to their use in PDT and PTT, porphyrins can also be used as a contrast agent in MRI, after chelation with manganese(III)[80] or gadolinium(III)[81]. The porphyrin chelation with radioactive isotopes, such as ^{64}Cu and ^{99m}Tc, also allow for their use in positron emission tomography (PET)[82] and single photon emission computed tomography (SPECT)[83].

The last category of systems we will discuss in this chapter is organic nanotubes (ONTs). These structures are made of organic molecules that can be either covalently

or non-covalently bonded that assemble into a tubular shape. ONTs have been found to exhibit extraordinary properties, such as high mechanical strength, thermal stability, and electrical conductivity. These properties make ONTs attractive for application in energy storage, electronics, medicine[84]. However, in the context of this Thesis, we will be focusing on the superassemblies formed by ONTs. ONTs can interact with each other through non-covalent forces into a variety of superassemblies. The group of Shinkai reported the formation of hollow spherical cages from the non-covalent self-assembly of carbon nanotubes[85]. In 2014, Stoddart and colleagues have successfully demonstrated that NDI-Δ units can be self-assembled into supramolecular nanotubes that intertwine in (E)-1,2-dichloroethene to form an organogel[86]. Single-walled carbon nanotubes (SWCNTs) can self-organize into "ropes", held together via van der Waals forces, $\pi - \pi$, or hydrophobic effects. The mechanical properties of the ropes can vary depending on the number of units, the strength of interactions, and the overall geometries[87]. These superassemblies can be considered as supramolecular systems, where the individual unit is the single organic nanotube. The resulting structures display unique physical and chemical properties that are distinct from their constituent monomers.

1.3 Molecular machines

Molecular machines have emerged as a relatively recent and impactful area of research, demonstrating enormous potential within the scientific community. These nanoscale machines are composed of individual components that interact via non-covalent forces, and can move with respect to each other generating motion when stimulated by energy or external factors. The importance of molecular machines is exemplified by the prestigious Nobel Prize in Chemistry awarded in 2016 to Jean-Pierre Sauvage, Sir Fraser Stoddart, and Bernard L. Feringa, recognizing their groundbreaking contributions in this field, and making it a central focus of scientific research. While reproducing the complexity of natural machines may be difficult, at least in the short term, very promising results are being achieved through the use of mechanical bonds, allowing for high stability, but also controllable large-amplitude motion of the constituent parts of these molecules.

1.3.1 Natural molecular machines

Molecular machines are ubiquitous in nature. They transform energy to perform specific cellular activities that are essential for life.

Myosin, kinesin, and dynein are a class of molecules known as *motor proteins*[88], which are essential for many biological processes. These motor proteins move along one-dimensional "tracks", actin filaments for myosin and MTs for kinesin and dynein, to perform their functions. *Myosin* is well-known for being a crucial component for muscle contraction. Myosin contains a motor domain that binds the actin filaments and hydrolyses adenosine triphosphate (ATP). The energy released by ATP hydrolysis causes conformational changes in the motor domain of myosin. As a result of such conformational changes, the myosin "steps" along the actin filaments toward one end. The myosin pulls on the actin filaments as they move along them, causing the actin and myosin filaments to slide, resulting in muscle contraction[89].

Kinesin and *dynein* are responsible for transporting various types of cargo within cells. Kinesin typically moves cargo toward the plus end of MT, while dynein moves cargo toward the minus end. Both motor proteins have a motor domain that allows them to bind MT and hydrolyse ATP, providing energy for movement. The structure of dynein differs significantly from kinesin, with the former having a hexameric ring of AAA-ATPase domain[90, 91]. Kinesin moves along the MT by swinging its motor domain from one tubulin subunit to another via a "walking" mechanism ("hand-over-hand" and "inchworm" mechanisms have been proposed)[90]. The mechanism proposed for dynein suggests that the conformational changes of the hexameric ring caused by ATP hydrolysis is transmitted to a stalk, which "swings" and moves the dynein along the MT[91].

As other examples, *ATP synthases* are enzyme complexes responsible for synthesizing ATP and are found in the membranes of mitochondria and chloroplasts in eukaryotic cells, and in the plasma membrane in some prokaryotes. The structure of ATP synthases consists of two domains, called F_0 and F_1. F_0 is embedded in the membrane and contains a proton-conducting channel, that pumps ion across the membranes, and a rotor; F_1 is found on the surface of the membrane and contains the catalytic subunits (α, β, γ, δ, ε) that synthesize ATP[92, 93]. For ATP synthesis, ATP synthases use a rotary motion. As the protons flow through the F_0 subunit, it rotates and transmits the rotation to the F_1 domain. The rotation causes the β-subunits of F_1 to change their conformation and bind adenosine diphosphate (ADP)

and inorganic phosphate (P_i), resulting in the formation of three ATP molecules after a full rotation[94].

The primary difference between the two classes of molecules discussed here is how they perform the motion: one is translational, while the other is rotational. Although replicating biological molecular machines is a prohibitive challenge, researchers hope that learning more about their mechanism of motion will aid in the design and construction of artificial molecular machines capable of performing specific tasks with controlled motion[28].

1.3.2 Artificial molecular machines

Numerous artificial molecular machines have been created over the past few decades[95]. One notable example is the creation of molecular switches[96, 97], which exhibit reversible shifting between two or more stable states in response to external stimuli[98]. These stimuli typically induce ring-opening and ring-closing steps, cis-trans isomerizations, and intramolecular proton-transfer reactions. Molecular switches have been used as components in sensors or logic gates[99, 100]. Molecular motors are another important class of artificial molecular machines, which are molecules that can perform continuous rotary motion, under the influence of certain input[101]. In particular, it is worth highlighting the work of Ben Feringa on the design and synthesis of the first light-driven rotary molecular motors[102].
While there are many other different types of artificial molecular machines (molecular tweezers, hinges, nanocars), the use of mechanically interlocked molecules[103] has greatly improved the development of such devices. Therefore, the primary focus of this Thesis is on the use of interlocked molecules as components for artificial molecular machines. Mechanically interlocked molecules (MIMs) are molecules that are linked together as a consequence of their topology, but they cannot be separated without breaking covalent bonds[103]. The two classic examples of this class of molecules are *rotaxanes* and *catenanes*.
A *rotaxane* is composed of a linear molecule, called axle, with a bulky group at each end, and one or more cyclic components, called macrocycles, threaded onto the axle and interlocked between the two bulky groups[104]. This design allows the macrocycles to move or rotate with respect to the axis while preventing dethreading. The resulting configuration is known as a *molecular shuttle* if a macrocycle can shuttle between distinct binding sites along the dumbbell. The binding sites may

be identical or different. Under the influence of various stimuli, such as light, pH, solvents, or ions[28, 105], the ring can move between the binding sites (Fig. 1.4, top panel).

Rotaxanes have been studied as a potential tool for the delivery of anticancer drugs. Papot, Leigh, and coworkers[32], reported the synthesis of a molecular device, in which the ring component encapsulates and protects the drug from degradation by the body's enzymes. When the rotaxane reaches the target, the release of the drug is then triggered by a specific sequence of enzymatic activations. This method has the potential to increase the effectiveness of drugs and reduce side effects by delivering the drugs directly to cancer cells and in a controlled manner.

Rotaxane-based molecules have also been proposed as components of artificial molecular muscles[106]. "Daisy chains" are excellent candidates for artificial muscles. A daisy-chains rotaxane is a type of rotaxane that is made up of single rod-like molecules that are covalently linked to macrocycles and have a bulky group at one end and is cross-threaded to form cyclic or acyclic polymers. Each monomer can slide on top of the other, until the stopper is reached, imitating the motion of myosin and actin (Fig. 1.4, bottom panel). In the pioneering work of Sauvage[107, 108], the motion of the macrocycle is triggered by metal exchange reaction, where the two interlocked monomers can interact with two metal centres. In recent years, new architectures[109–112] or motion-triggering stimuli[106] for the design of muscle-like compounds have been investigated. These studies have enabled the creation of materials that can produce motion in response to stimuli such as pH, solvent, redox, and chemical switching, with very promising results[113].

Since the synthesis of the first rotaxane catalyst in 2004 by Tachibana and coworkers[114] for benzoin condensation, a significant amount of research has also focused on using these compounds to catalyze a variety of chemical reactions[115]. For example, rotaxane catalysts have been used for the formation of ester bonds[116] and the hydrogenation of enamides[117] or acid esters[118]. The idea behind using rotaxanes as catalysts is to take advantage of the macrocycle's mobility, which allows it to move along the thread of the molecule, to expose the catalytic sites to the reactants. This mobility can also be triggered by external stimuli[119–122], allowing for "on-demand" catalysis.

A *catenane* is a molecule that is composed of two or more interlocked macrocyclic components[104]. The control of the motion between the subcomponents of a catenane can be achieved by introducing recognition sites at macrocycles, that

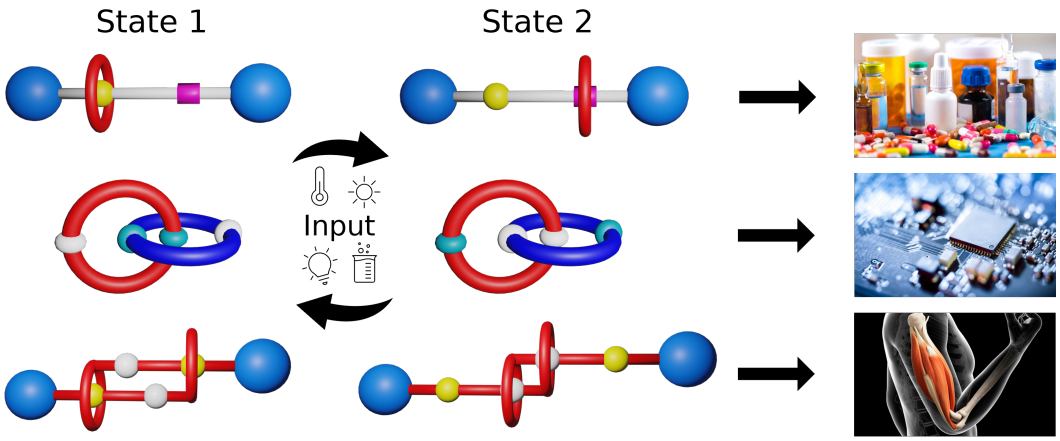

Fig. 1.4 Schematic representation of artificial molecular machines: Rotaxanes (top), catenanes (middle) and daisy chain (bottom) switch between different states (State 1 and State 2, in this scheme) under the influence of various stimuli. On the right, some potential applications of MIMs.

allow the rings to switch between specific states in a predetermined manner (Fig. 1.4, central panel). It is still uncertain whether the field of catenanes will lead to practical applications, at least in the near future. However, the ability to obtain controllable switches in a relatively easy way makes them very appealing as components for molecular rotary motors. The movement of the macrocycles in catenane rotary motors can be controlled through changes in the binding affinity between the macrocycles and the stations on the larger ring in which they are interlocked in response to external stimuli[101, 123, 124]. By carefully controlling a series of complexation/decomplexation reactions, it is possible to achieve directional rotation of the motor. Leigh and colleagues reported a four-station molecular machine that uses light, heat, or chemical stimuli to achieve the unidirectional rotation of two small rings[125]. Another example, also from Leigh and coworkers[126], is the synthesis of a reversible rotary molecular motor, in which the sense of rotation is only governed by the order in which the reactions are performed. The movement of the macrocycles can also be powered by DNA strands, via molecular recognition[127]. Catenane-based molecular motors have the potential to greatly benefit molecular electronics by enabling the creation of new types of molecular electronic devices[128, 129] that are more efficient, smaller, and more flexible than traditional electronic devices. One way that catenanes can be utilized in molecular electronics is by interfacing them with surfaces or solid supports[130].

1.4 Experimental characterization of self-assembled systems and limits

The characterization of supramolecular polymers and molecular machines is essential to understand their properties and behaviour, as well as for developing dynamic materials for new applications. A variety of experimental techniques has been used to study these materials, each with its own unique advantages and limitations. However, it is often necessary to combine multiple techniques to obtain reliable and consistent results and a satisfactory understanding of the behaviour of such systems.

Microscopy imaging can provide insights into the physical structure and organization of these systems (see Fig. 1.5 for some representative examples). Scanning tunnelling microscopy (STM) is able to image conducting and semiconducting surfaces at the atomic level, and has been used to study various supramolecular systems. As a notable example, STM revealed the double-layered structure of a cyclodextrin-based supramolecular polymer on a HOPG[1] surface[131]. A specialized variant of STM, ultrahigh-vacuum low-temperature STM, can provide even higher resolution images by exploiting the ultra-high vacuum conditions (avoiding contamination and minimizing noise) and low-temperature (reducing thermal motion). This has been recently applied for the surface characterization of a supramolecular hexagonal grid[132], and to visualize the structure of single metallo-supramolecular chains[133], but the extreme conditions of use that it requires can damage the sample.

Electron microscopy, including scanning electron microscopy (SEM) and transmission electron microscopy (TEM), are commonly used for characterizing supramolecular aggregates[134–136]. Recently, TEM has been used to study different strategies of self-assembly of DNA-catenanes[137], and to investigate the ring motility in DNA-based[138] rotaxanes. However, SEM and TEM samples are required to be completely dry, and therefore their images are not representative of these systems in solution.

Stochastic Optical Reconstruction Microscopy (STORM) is a super-resolution optical microscopy technique that can provide accurate images of self-assembled systems, and it has been used to study the exchange pathway in BTA-based supramolecular polymers[139] (Fig. 1.5a). Nonetheless, this method requires the use of fluorescent tags which may alter the structure. Furthermore, the typical frequencies of data acquisition are often insufficient to track molecular processes.

[1] Highly oriented pyrolytic graphite

Small-angle X-Ray scattering (SAXS) and small-angle neutron scattering (SANS) are powerful tools for the structural characterization of self-assembled materials[140–143]. SAS techniques can detect average properties of materials such as shape, size, and molecular weight. However, combined SAXS and SANS analysis can offer more insight not only into the morphology of the assembly, but also into its internal structure. As a representative example, the group of Yagai used combined SANS and SAXS analysis to demonstrate the self-assembly of hexameric hydrogen-bonded supermacrocycles in toroidal supramolecular polymers in solution[144]. SAXS and SANS analysis have also been applied to mechanically interlocked molecules to demonstrate the formation of rotaxanes in the absence of crystal structures[145] and to probe the dynamic of daisy-chain rotaxanes[146]. Despite their advantages, the low resolution (typically between 1 and 100 nm) limits their applicability.

Atomic force microscopy (AFM) is a high-resolution technique with a resolution of fractions of a nanometre. AFM is one of the most widely employed techniques to study the morphology of supramolecular surfaces[147], as, in contrast to STM, it can be applied also on non-conducting surfaces. The group of Zijlstra used AFM images[148], combined with fluorescence spectroscopy, to characterize the structure of 1,3,5-BTA supramolecular polymers, allowing for the measurement of height and width of aggregates. In addition, AFM can also measure intra- and inter-molecular forces with pN resolution, allowing for the determination of mechanical properties in polymers and supramolecular systems[147, 149], but it is often not applicable due to the fragile nature of these systems. Another limitation of AFM is, as for TEM and SEM, that the samples are investigated in dry-state.

In addition to the structural characterization, it is also crucial to determine the mechanism of self-assembly or the dynamics of supramolecular polymers and molecular machines. As in the case of the structural characterization, there are several techniques that can be used. AFM can also be utilized to follow the self-assembly of supramolecular polymers in real time[150, 151] (Fig. 1.5b). Spectroscopy techniques, such as ultraviolet-visible (UV-vis) spectroscopy, fluorescence spectroscopy, and circular dichroism (CD), have proven to be excellent techniques to get insights into the details of supramolecular polymerization and molecular motion in MIMs. Meijer and coworkers used a combination of CD and UV-vis spectroscopy to investigate the supramolecular polymerization of porphyrin-based monomers[152, 153]. Induced-CD is also a well-established method for studying cyclodextrin-based supramolecular complexes[154–156]. UV-vis and fluorescence spectroscopy have also been used

to study the interactions between monomers in hexameric supermacrocycles[157]. However, these techniques have some limitations, as they have low resolution and typically require labelling. Nuclear magnetic resonance (NMR), including a variety of 1-D and 2-D techniques, has been successfully applied for the structural characterization of supramolecular systems and molecular machines[158–161]. Furthermore, NMR spectroscopy has also been proven to be very effective for gaining understanding of the kinetic properties of a material. This powerful analytical technique allows for the determination of exchange rates between sites, by monitoring the NMR spectra at different time and fitting the data to a kinetic model, making it a widely used techniques for characterizing rotaxane[158, 162, 163] (Fig. 1.5d). It should be noted that NMR requires expensive deuterated solvents, which could interfere with the analysis and affect the free-energy profile.

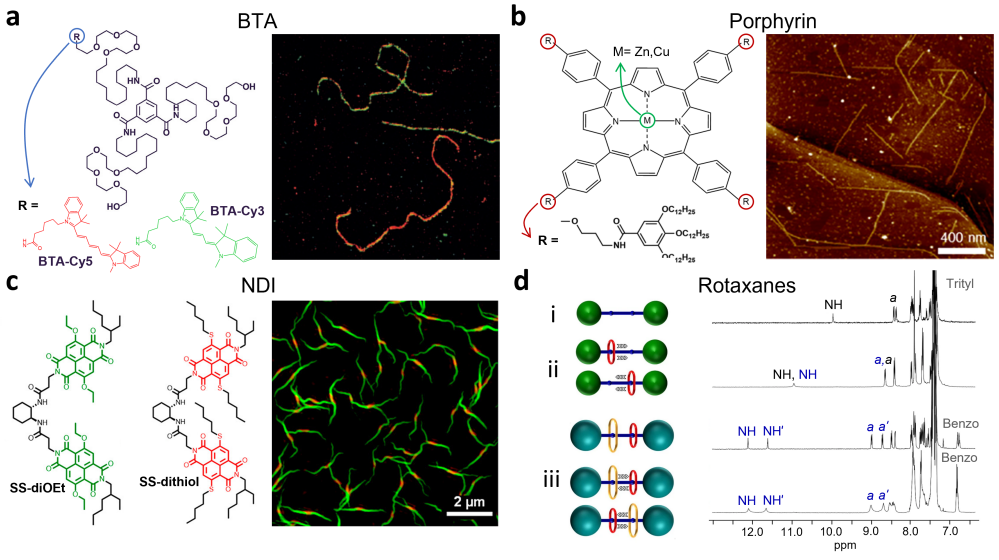

Fig. 1.5 Experimental characterization of supramolecular systems. Four panels showing (a) Structure of BTA, BTA-Cy5, BTA-Cy3 (left) and STORM imaging of BTAs after 1 hour of mixing (right). Reproduced with permission from Ref.[139]. (b) Structure of Porphyrin-based monomers (left) and AFM image of seeded-growth of SP_{4Zn} (right). Reproduced with permission from Ref.[150]. Copyright 2018 American Chemical Society. (c) Structure of two core-substituted NDI monomers SS-diOEt and SS-dithiol (left) and SIM microscopy images of annealing-induced mixing of homopolymers (right). Reproduced with permission from Ref.[164]. Copyright 2018 American Chemical Society. (d) On the left, schematic models of naked axle (i), [2]Rotaxane (ii) and [3]Rotaxane (iii); on the right, corresponding ^1HNMR spectra $CDCl_3$ at 298K. Reproduced with permission from Ref.[163].

Although experimental techniques have undergone significant developments in the recent years, that have made them more advanced, easy to use and less sensitive to contamination, obtaining a comprehensive understanding of supramolecular polymers and molecular machines remains challenging. The complexity of these systems, their sensitivity to the external environment, and the need for high-resolution characterization (both in space and time), make it difficult to fully characterize their properties and behaviour.

1.5 The potential of multiscale molecular modelling

In the recent years, computational studies have offered a fundamental support in the field of material science and engineering, providing a detailed understanding of the properties and dynamics of self-assembled materials, that complement experimental data.
The reliability of *in silico* methods in the examination of natural self-assembled materials has been established through a number of studies[165–170]. Self-assembly processes can occur over a range of timescales, from nanoseconds (formation of metal organic frameworks[171]) to minutes or even hours (slow protein folding[172]), and can involve systems spanning from nanometres (nanoparticles) to microns or millimetres (large polymers). Due to this complexity, multiscale molecular modelling approaches are often required to study self-assembly processes and materials. Multiscale molecular modelling allows for simulating self-assembled systems at different levels of detail, ranging from the quantum, atomistic, coarse-grained, to the meso scales. Each of these techniques captures relevant physical and chemical phenomena that occur at specific timescales (Fig. 1.6). By combining different modelling approaches, it is possible to explore a vast range of phenomena that occur across different time and spatial scales, such as local motion[173, 174], protein folding[175, 176], and crystal formation[177, 178].

The application of computational methodologies in the field of supramolecular chemistry is a rapidly developing area of research. One area of focus is dedicated to gaining a deep understanding of the structure, the key interactions between monomers, and fundamental factors controlling the supramolecular assemblies, typically when the initial structures are already available (from, *e.g.* X-ray, database. See Fig. 1.7a-b).

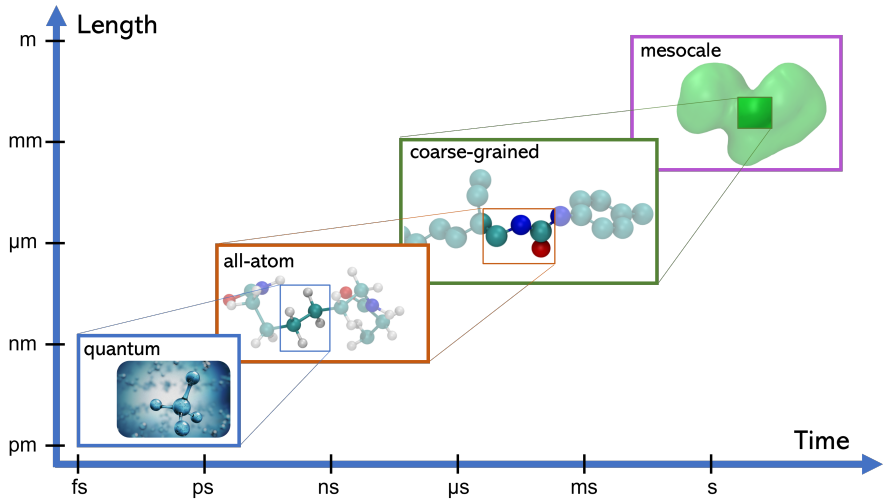

Fig. 1.6 Multiscale molecular modelling: characteristic timescales and space scales that can be explored by each method.

In this context, quantum mechanical calculations and *ab initio* simulations guarantee the highest level of resolution, but the high computational cost limits their applications to simple aggregates[179, 180].

All-atom (AA) molecular dynamics (MD) allow simulating larger systems, and have been recently employed to explore a variety of supramolecular materials[181, 182]. In a remarkable example, molecular mechanics (MM) and MD have been recently employed for in-depth characterization of the helical arrangement in C3 symmetric BTA-based compound[183], unequivocally revealing the preference for one helical twist over the opposite. Combined molecular mechanics (MM) and all-atom MD simulations allowed for a structural investigation of bisurea-based supramolecular polymers[184]. The in-depth analysis of the energy landscape identified three distinct structures similar in terms of stability (straight, helical and zigzag), all stabilized by hydrogen bonds. Similar bisurea-based tubular structures have been simulated in different organic solvents[185], suggesting that enthalpic stabilization is responsible for the enhanced stability of these assemblies in toluene.

The investigation of supramolecular polymers in polar solvents using molecular simulations is particularly useful, since the experimental characterization is hindered by the increased complexity of water-soluble monomers and hydrophobic effects. Pavan and coworkers have extensively used atomistic MD simulations to study both chiral and achiral BTA-based supramolecular polymers in water[186] (Fig. 1.7b), showing that the fibres fold under the influence of hydrophobic effects and that

small changes in the monomeric building block can lead to significant changes in the structure and dynamics of the polymer. Additionally, the cooperative self-assembly of 1,3,5-benzenetricarboxyester (BTE) – a derivative of BTA that lacks the ability to form hydrogen bonds (the amide group is replaced by an ester group) – was explored[187]. The findings of this study highlighted hydrogen bonds as the key interaction leading to the fibre growth, and further proved the link between the structure of the monomer and the dynamic and the properties of the polymer.
A second area focuses on the spontaneous aggregation of randomly dispersed monomers in solution (Fig. 1.7c-d), with a number of examples in literature describing the process with atomistic resolution[188–190].
However, observing the formation of longer aggregates is simply prohibitive within the timescale accessible to atomistic classical MD simulations. To explore larger ordered aggregates while retaining the atomistic resolution, enhanced sampling approaches can be used. In this area, Chami and Wilson[191] carried out multiple steered molecular dynamics (SMD) in aqueous solution to calculate the binding free-energy in stacks of two, three, and eight monomers, proposing a isodesmic aggregation mechanism. Bejagam and coworkers[192] used Adaptive Bias Force (ABF)[193] to demonstrate the cooperative nature of the self-assembly of BTA in organic solvent. Replica exchange methods have been effectively employed to study the supramolecular polymerization of BTA in explicit methylcyclohexane, successfully reproducing the helical H-bonding pattern of the polymer[194].
These techniques enable the exploration of longer timescales while maintaining atomistic details, but the associated computational cost remains extremely high due to the large size of the systems under investigation. One way to simplify the description of complex self-assembly processes is to use a coarse-grained (CG) representation. By replacing groups of atoms with single pseudo-atoms, the CG description drastically reduces the number of simulated particles and enables the use of larger integration time steps. This significantly decreases computational costs while preserving essential system features, facilitating the exploration of larger length and timescales (see Section 2.4, in the next Chapter).
In 2015, the group of Balasubramanian published a CG force field model capable of modelling the self-assembly of BTA in nonane in one-dimensional stacks, introducing a dipole point[195]. The non-bonded interactions between the CG beads were carefully set through radial distribution function (RDF) matching, and then further optimized to accurately reproduce dimerization and solvation free-energy

profiles. Following a similar procedure, Bochicchio and Pavan developed a CG model of BTA based on the MARTINI force field, with the notable advantage of transferability[196]. These seminal works paves the way for the future control of the structure-property relationships of similar BTA-based assemblies.
The combination of the aforementioned tools opens up a plethora of new possibilities for investigating the intrinsic dynamic properties of supramolecular polymers. By combining AA, CG, and advanced simulation techniques (Metadynamics, infrequent Metadynamics), Bochicchio et al. were able to achieve a molecular-level description of the mechanism, pathway, and kinetics of monomer exchange in various solvents, that has never been achieved so far[197]. Their findings revealed that the monomer exchange starts from the defects along the fibres and progresses in a step-wise manner. The detachment of the monomer from the stack to the surface was observed to occur in the μs timescale, while its diffusion in water takes place in the ms timescale[197]. From an higher perspective, implicit coarse-grained models and simulations have demonstrated that supramolecular polymers can be viewed as complex molecular systems, allowing us to study the way they communicate and exchange molecular fragments with each other, and with the surrounding environment[198] (Fig. 1.7d).

The potential of molecular modelling in the study of molecular machines is a noteworthy aspect of this field. The delicate nature of molecular machines, in which small changes in the surrounding environment, such as solvent composition, temperature, pH, and external stimuli, can have a significant influence on the dynamics of their subcomponents, requires precise characterization. Some experimental techniques that are used to study molecular machines operate under specific (sometimes extreme) conditions, such as dry environments or low temperatures. In this context, molecular modelling serves as a "virtual microscope" to study and characterize molecular machines behaviour in various environments at an atomic level, by simply varying some simulation input parameters, overcoming the limitations of traditional experimental techniques.
Density Functional Theory (DFT) calculation have been applied in a number of studies to understand the photo- or thermo-isomerization of molecular motors[200–202]. MD simulations have been employed in various studies to predict the structures and properties of several interfaces, like air/water[203], or Au(111) surface[204], and compare the results with experimental data. Zazza et al. combined MD and DFT to examinate hydrogen-bond network and the conformational flexibility of a neutral redox-responsive bistable rotaxane[205].

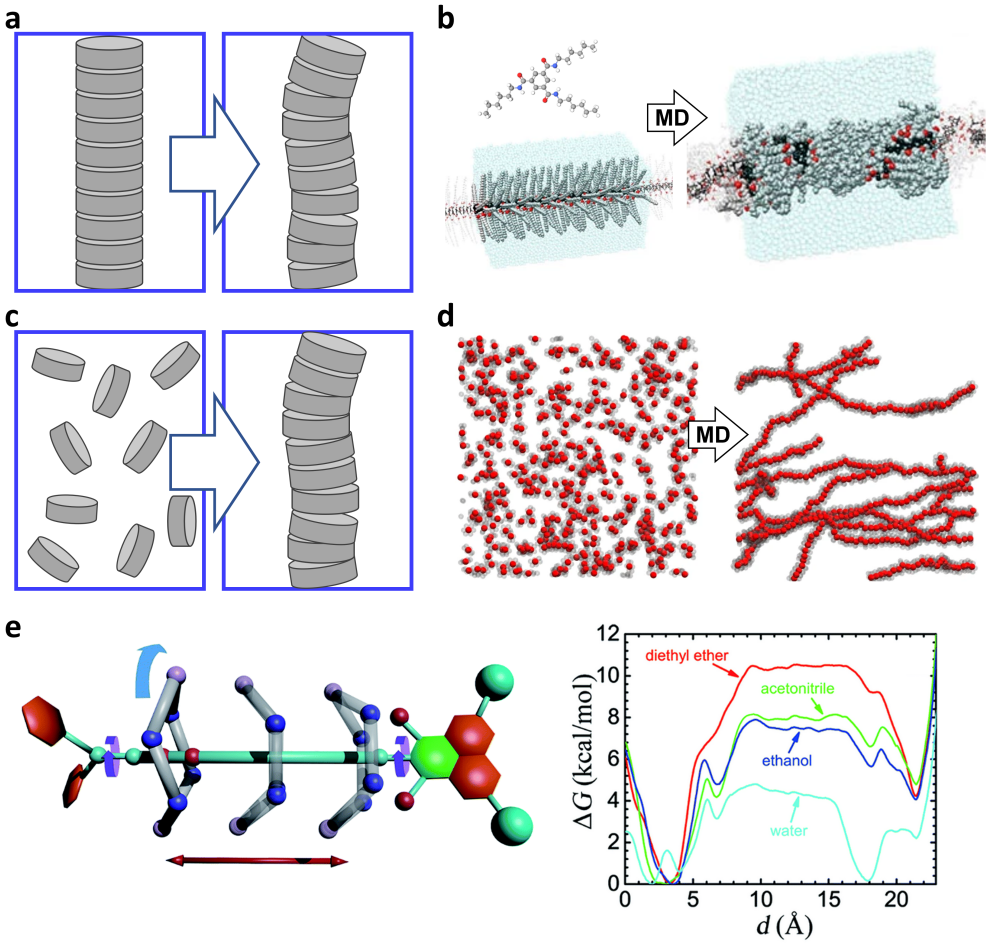

Fig. 1.7 (a) "Top-down" approach for studying supramolecular polymers, starting from preassembled stacks. (b) MD simulations of BTA-based supramolecular polymers in water. Reproduced with permission from Ref.[186]. (c) "Bottom-up" approach for studying self-assembly from free randomly dispersed monomers. (d) Self-assembly simulations of 500 CG particles modelling the behaviour of complex supramolecular systems. Reproduced with permission from Ref.[198]. The schematic views in panel (a) and (c) are adapted with permission from Ref.[181]. (e) On the left, schematic representation of the shuttling of rotaxanes. On the right, free-energy of the translational movement of the macrocycle, in various solvents. Reproduced with permission from Ref.[199].

On the other hand, the study of the molecular motion in self-assembled molecular machines is particularly challenging, due to the slow dynamics of these systems. In fact, the shuttling between binding stations is typically a rare event, occurring at a timescale that exceeds the classical AA simulations capability. This highlights the need for advanced computational methods that can capture the slow dynamics of

molecular machines and provide a deeper understanding of their behaviour and mechanism of action. The extensive work done by Cai and coworkers[199, 206–210] is an excellent example of how our understanding of molecular machines can benefit from enhanced sampling methods. By utilizing MD and ABF methods, they were able to effectively study the influence of various environmental factors such as solvent, pH, and ionic strength on the shuttling of molecular machines. In a notable example, they explored the motion of an amide-based rotaxane in different solvents and decoupled the shuttling process into the possible movements. Free-energy calculations and decomposition in physical contributions allowed them to gain insights into the driving forces responsible for the shuttling in the various solvents[199] (Fig. 1.7e). Furthermore, the team investigated the influence of the structure on shuttling rates by computing the free-energy profiles at different lengths of the axle[208]. The findings of the study suggest that the free-energy barrier increases with longer linkers, up to a certain length.

The aforementioned studies provide a strong foundation for the field of molecular simulations of supramolecular systems, and they represent the current state of the art, while still non-exhaustive due to length limitations. Despite these advances, a true multiscale understanding of the dynamical behaviour of these systems, spanning multiple space and timescales and linking the atomistic details of the monomers' interactions with the complex behaviour of their assemblies, is still lacking. This now brings us to the aim of the research performed during my PhD, which is described herein.

1.6 Aim of the Thesis

The aim of my PhD Thesis is to address the existing knowledge gap in the simulation of supramolecular systems by employing advanced multiscale modelling techniques for high-resolution, long-timescale simulations. While not developing new methodologies, my research applies existing cutting-edge methods to study new and unexplored problems in supramolecular systems, with a particular focus on their dynamic behaviour. Understanding the dynamic aspects of these systems with a submolecular resolution is fundamental for supporting the design of materials that exhibit controlled and predictable behaviour. However, this knowledge is still

lacking in the current state of research. My goal is to bridge this gap by providing unique insights into the mechanisms that regulate the dynamics of supramolecular systems, thereby enhancing our overall comprehension of these systems. By utilizing enhanced sampling techniques, my work aims to overcome the slow dynamics, timescale limitations, and computational costs associated with traditional molecular simulations in this field. Additionally, through enhanced sampling methods, I am able to obtain energetics and kinetics information, that can be used to identify rate-limiting steps in the explored processes and eventually aid in the design of strategies to overcome them. To simplify the description of complex molecular systems while preserving their essential features, I will develop coarse-grained models. These models will help to overcome the computational cost of high-resolution simulations and enable the investigation of systems that are too complex for atomistic models. Furthermore, my research seeks to complement experimental findings by providing insights into molecular interactions and dynamic behaviour at a level of detail that is not accessible by experiments. Therefore, I will work closely with experimentalists to ensure that my simulations can provide valuable insights into the systems under investigation, highlighting the mutually beneficial relationship between experimental and computational methods. Overall, by applying existing advanced multiscale modelling methodologies to study supramolecular systems at an unprecedented resolution, my PhD Thesis aims to advance our understanding of supramolecular systems and contribute to the design of more efficient and effective materials in the future.

Chapter 2

Methods

Molecular modelling and computer simulation have proven to be fundamental tools for understanding the behaviour of molecular systems. They allow for predicting structural, thermodynamic and dynamic properties of molecules, providing an understanding of complex molecular systems at the atomistic level. The ability to simulate the behaviour of molecules with thousands of atoms *in silico* has opened up new opportunities for research and innovation in a wide range of fields, including drug discovery, protein folding, and material science.

The aim of this chapter is to provide a comprehensive overview of the theoretical background behind molecular modelling and simulations, which will serve as the foundation for the research presented in this PhD thesis.[1]

I will start from a few relevant basic concepts of statistical mechanics[212], and then provide an overview of molecular mechanics and force fields[211], including both atomistic and coarse-grained representations. I will then discuss the molecular dynamics methods and finally, the enhanced sampling techniques I employed in my Thesis, such as Umbrella Sampling and Metadynamics, which are used to overcome the limitations of traditional simulations and obtain a more complete picture of the system under study.

[1] Parts of this section are adapted from Ref.[211]. The mathematical notation is adapted from Ref.[212].

2.1 Theoretical Background

Let's consider a system of N particles in three dimensions. The state of the system is determined by Newton's second law. This law completely determines the full set of positions $\{\mathbf{r}_1(t),...,\mathbf{r}_N(t)\}$ and momenta $\{\mathbf{p}_1(t),...,\mathbf{p}_N(t)\}$ of each particle in the system at any time t. It is, therefore, possible to fully characterize an N-particle system in three dimensions by specifying 6N numbers at any instant of time, three coordinates and three momenta per particle. These 6N numbers, that define the *microscopic state* of the system at a given time, represent a point in the 6N-dimensional space of all the possible degrees of freedom of the system, called *phase space*. The evolution of the system in time generates a *trajectory*. The associated total energy, expressed as a function of the positions and momenta, is given by the *Hamiltonian H*:

$$H(\mathbf{r}_1,...,\mathbf{r}_N,\mathbf{p}_1,...,\mathbf{p}_N) = K(\mathbf{p}_1,...,\mathbf{p}_N) + U(\mathbf{r}_1,...,\mathbf{r}_N) \tag{2.1}$$

where U is the potential energy, and K is the kinetic energy of the system, given by:

$$K(\mathbf{p}_1,...,\mathbf{p}_N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} \tag{2.2}$$

Using the Hamiltonian formulation of classical mechanics, the equations of motion can be obtained from the Hamiltonian using the following relationships:

$$\dot{\mathbf{r}}_i = \frac{\partial H}{\partial \mathbf{p}_i}, \qquad \dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{r}_i} \tag{2.3}$$

Hamilton's equations of motions are completely equivalent to Newton's second law of motion and describe the trajectory in the phase space of a system subject to initial conditions.

Statistical mechanics is based on the concept of *statistical ensemble*. A statistical ensemble is the collection of all individual microscopic configurations that share the same thermodynamic properties (such as temperature, volume, total energy). The most important statistical ensembles are:

- The Micro-Canonical ensemble (NVE), it is the collection of microstates with constant number of particles, volume and total energy, and it corresponds to an isolated system where energy is conserved;
- The Canonical ensemble (NVT), it is the collection of microstates with constant number of particles, volume and temperature, and it corresponds to a closed system in thermal equilibrium with a heath bath;
- The Isothermal-Isobaric Ensemble (NPT), it is the collection of microstates with constant number of particles, pressure and temperature, and it corresponds to a closed system in contact with a heat bath and subject to external pressure;
- The Gran-Canonical Ensemble (μVT), it is the collection of microstates with constant volume, temperature and chemical potential; the energy and the number of particles can change, hence it corresponds to an open system.

The introduction of *ensembles* by Gibbs implies that it is not necessary to know the precise motion of every particle in a system to know their properties, but it is sufficient to average over a large number of different microscopic configuration of the ensemble. Once the ensemble is defined, the average value of the observable $A(\mathbf{r}, \mathbf{p})$ is then obtained performing an ensemble average:

$$\langle A \rangle = \iint A(\mathbf{r}, \mathbf{p}) P(\mathbf{r}, \mathbf{p}) d\mathbf{r} d\mathbf{p} \tag{2.4}$$

The quantity $P(\mathbf{r}, \mathbf{p})$ is the probability density of the ensemble. In the canonical ensemble, the probability density is the Boltzmann distribution:

$$P(\mathbf{r}, \mathbf{p}) = \frac{1}{Z} e^{-\beta H(\mathbf{r}, \mathbf{p})} \tag{2.5}$$

where $\beta = 1/k_B T$, k_B is the Boltzmann's constant and T is the temperature. The quantity Z is known as *partition function*, and it is a measure of the number of accessible microscopic states in a given ensemble, and it is different in each ensemble. For the canonical ensemble, it is:

$$Z = \iint e^{-\beta H(\mathbf{r}, \mathbf{p})} d\mathbf{r} d\mathbf{p} \tag{2.6}$$

The ergodic hypothesis, which is at the essence of molecular dynamics, lies in the idea that, given a sufficiently long time, the dynamics of a system will visit all the

phase space with equal probability. This allows to replace the ensemble average, which represents the average behavior of a large number of identical systems, with the time average obtained from a single trajectory of the system over an extended simulation time:

$$\langle A \rangle = A_{time} \tag{2.7}$$

The average value of an observable A over a trajectory is given by the *time average*:

$$A_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} A\left(\mathbf{r}(t), \mathbf{p}(t)\right) dt \tag{2.8}$$

In molecular simulations, discrete time steps are used, so the equivalence becomes:

$$\langle A \rangle = \frac{1}{M} \sum_{n=1}^{M} A(\mathbf{x}(t_n)) \tag{2.9}$$

where M is the number of time steps and t_n is the simulation time.
The region of phase space that is explored in simulation time is the integration domain. This region cannot be infinite, due to limited computational resources, and it is therefore necessary to choose an appropriate integration domain that is relevant to our research question and can be compared with experimental data. Depending on the systems we are studying, different integration domains may be required to obtain equilibrium values. For example, if we are studying a simple gas of atoms, nanoseconds of simulation time may be sufficient to reach equilibrium. If we are studying a complex process, it may take microseconds or more to reach equilibrium, because the process has many and slow degrees of freedom. However, what is important to compare with experiments is not always the equilibrium value. Sometimes we may be interested in the properties of the system before it reaches equilibrium, or in the dynamics of how it approaches equilibrium. For example, if we are studying a polymerization process, we may want to know the mechanism of formation, which is not captured by equilibrium values. Or, if we are studying the motion of molecular machines, we may want to know how fast or slow the transition between metastable states occurs, or what the effects of external factors are. Therefore, when we choose the integration domain for our simulations, we need to consider both the physical characteristics of our system and the experimental data that we want to compare with.
The finite timescale that can be explored has practical implications for both the accuracy and interpretation of MD simulations. One implication is that finite sampling

timescales may introduce errors or biases in the estimation of thermodynamic and kinetic properties of molecular systems, such as free energies, entropies, reaction rates, etc. Another implication is that finite sampling timescales may limit the applicability and usefulness of MD simulations for certain problems or systems that require long timescales to be observed or predicted. To overcome these challenges, various enhanced sampling methods have been developed to accelerate the exploration of phase space and thus extend the accessible timescales in MD simulations, that will be discussed in Sec. 2.6.

2.2 Molecular Mechanics and Force Fields

Molecular modelling includes all the theoretical and computational methods that are used to model molecular systems, simulate their behaviour and study their properties by solving the equations of quantum and classical physics. As computer power is continuously increasing, it is possible to model and simulate even large and complex molecular systems (*e.g.* proteins, nucleic, polymers) with high details. *Molecular mechanics* is one of the methods in molecular modelling, and it uses the law of classical mechanics to describe molecular systems. The main assumption to move from a quantum to a classical description is the Born-Oppenheimer approximation[213]. This approximation allows to separate the motion of the atomic nuclei and the motion of the electrons in a molecule. This is based on the fact that the nuclei are much heavier and slower than the electrons, and therefore their positions can be considered as fixed while the electrons move around them. This way, we can write the potential energy surface as a function of only the nuclear coordinates, and ignore the electron motion. In molecular mechanics, molecules are described as charged points (the atoms), which are connected by covalent bonds and interact with other particles via non-bonded interactions.

The potential energy function U in Eq. 2.1, can be described using a general functional form, that is function of the atomic coordinates $\mathbf{r}^N$. Let V this functional form, also known as *force field*, it can be expressed as:

$$V = V_{bonded} + V_{non-bonded} \tag{2.10}$$

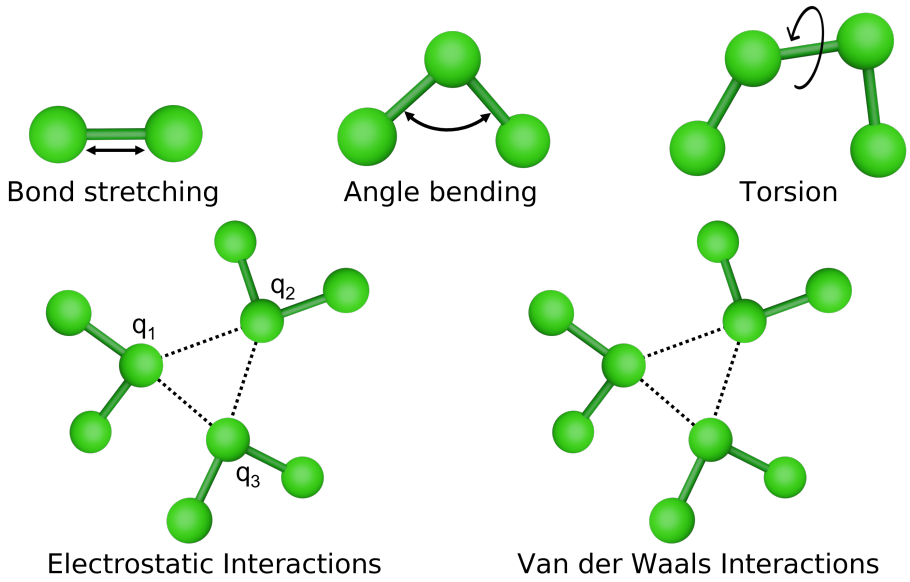

Fig. 2.1 Schematic representation of the main contributions in molecular mechanics force fields.

where the two term are given by the following summations:

$$V_{bonded} = V_{bonds} + V_{angles} + V_{dihedrals} \tag{2.11}$$

$$V_{non-bonded} = V_{electrostatics} + V_{vanderWaals} \tag{2.12}$$

Each term in the above equations (see Fig. 2.1 for a schematic representation of the terms) can be modelled in different ways and using different functional forms and parameter sets. Many of the molecular modelling force fields available in the literature use a functional form that describes how the energy changes as a deviation from reference values of bonds or angles, as a consequence of bond rotation, or due to the interaction between non-bonded parts of the system:

$$V\left(\mathbf{r}^N\right) = \sum_{bonds} \frac{k_{b_i}}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angles} \frac{k_{\theta_i}}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{dihedrals} \frac{V_{n_i}}{2}\left(1 + cos\left(n_i\omega_i - \gamma_i\right)\right) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \tag{2.13}$$

The first term describes the interaction between pairs of bonded atoms, modelled

with a harmonic potential, where k_l is the stretching constant of the bond, $l_{i,0}$ is the reference bond length (the value of the bond length when all the other terms in the force field are set to zero), and l_i is the bond length. The second term describes the deviation of angles from their reference values using again a harmonic potential, where k_θ is the force constant, $\theta_{i,0}$ is the reference angle value, and θ_i is the angle value. The third term is a torsional potential and describes how the energy profiles changes due to rotations about bonds: V_{n_i} is related to the barriers to rotation, n_i is the multiplicity (the number of minimum points in the function), ω_i is the torsion angle, and γ_i is the phase factor. The last term describes the non-bonded term and, in a simple force field, it is usually modelled using electrostatic and van der Waals interactions.
Using internal coordinates, such as bonds, angles, dihedrals, rather than absolute atomic positions offers some advantages. For example, internal coordinates leads to a more intuitive potential energy function, directly associated with changes in the structural features of the molecule and offers a more clear and intuitive physical interpretation of each term. This choice also facilitates a systematic parametrization process, thereby enhancing the transferability of the force field (the importance of force fields transferability will be further discussed in Sec. 2.3).

Electrostatic Interactions

The electrostatic interaction between two atoms is calculated using Coulomb's law:

$$V_{electrostatic} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{2.14}$$

where q are the partial atomic charges assigned to each atom, r_{ij} is the distance between atoms i and j, and ε_0 is the vacuum permittivity. Electrostatic interaction is a long-range interaction, as it decays as the inverse power of the distance. In order to efficiently compute the electrostatic interactions between all pairs of atoms in periodic systems, without introducing serious inaccuracies (*e.g.* using simple truncation), advanced approaches like the Ewald summation[214], fast multipole methods[215], or reaction field[216] are typically used.

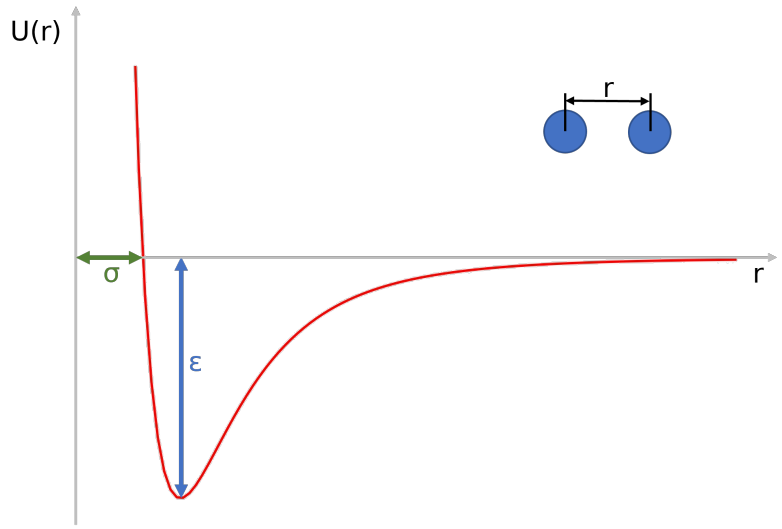

Fig. 2.2 The Lennard-Jones 12-6 potential used to describe van der Waals interactions between two interacting particles, as a function of the distance r.

Van der Waals Interactions

The van der Waals potential is usually described using the Lennard-Jones 12-6 pair potential:

$$V_{LJ} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.15}$$

The Lennard-Jones 12-6 function (displayed in Fig. 2.2) includes a repulsive part (that decreases as $1/r^{12}$), that models the Pauli repulsion between the electronic clouds of two atoms at short distances, and an attractive term (that decreases as $1/r^{6}$, that models the long-range attraction. The Lennard-Jones potential has the advantage of containing only two parameters: σ_{ij} is the collision diameter, the distance at which the interaction between two particles is zero, and ε_{ij} is the depth of the potential well.

The ε and σ values are usually given for pure species. The cross-interaction parameters σ_{ij} and ε_{ij} parameters can be obtained using *mixing rules*. The most used combination rule is the one of Lorentz-Berthelot[217], that calculates σ_{ij} as the arithmetic mean and ε_{ij} as the geometric mean of the values for the pure species:

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_{ii} + \sigma_{jj}\right), \qquad \varepsilon_{ij} = \sqrt{\varepsilon_{ii} + \varepsilon_{jj}} \tag{2.16}$$

The computational cost of such non-bonded interactions is significantly expensive, as the number of the interactions increases as the square of the simulated particles. Due to the short-range nature of this interaction, that decays rapidly at relatively short interatomic distances, the potential can be truncated, or truncated and shifted, and calculated up to a certain cutoff distance. Beyond the cutoff, the potential can be ignored or taken into account adding a tail-correction potential[218].

Despite the simplicity of the functional forms, force fields can give excellent results in predicting certain properties. Another key feature of classical force fields is their transferability, that is, the set of parameters developed for one molecule can also be applied to similar molecules, provided that care was taken in the parametrization, as discussed in the next section.

2.3 Force Field Parametrization

It seems quite clear that the development and parametrization of a classical force field for non-standard molecules is not always a trivial task. The first step for parametrizing a new molecule is to build the molecule and assign the *atom type*. The atom type definition is fundamental for a force field, because it implicitly contains information about the hybridization state and the neighbouring environment.
The second aspect to consider when parametrizing a force field is the charge distribution in a molecule. Many of the molecular force fields restrict the charge distribution to the atom centres, also referred to as *partial atomic charges*. Several approaches can be employed for calculating partial atomic charges. The electrostatic potential (ESP) derived charges, are obtained fitting the charges to reproduce the quantum mechanical potential electrostatic at a large number of grid point around the molecule. For larger systems, quantum mechanical (QM) calculations are extremely expensive and time-consuming. Among various semi-empirical methods, the Austin Model 1 (AM1)[219] is the one that better reproduces the charges derived from *ab initio* calculations. There are many resources to derive the missing parameters. For biological systems, the parameters can be derived from experimental data (X-Ray, neutron diffraction) or from large databases (PDB, CSD, CCDC). When experimental data are not available (*e.g.* for non-standard residues or new molecules), force fields are parametrized using QM calculations.

It should be kept in mind that a force field is designed to predict certain properties for a certain class of molecules and may therefore perform poorly or leads to erroneous results in predicting other properties for different molecules. For example, CHARMM[29], GROMOS[220] or AMBER[221] are very popular force fields designed specifically for simulations of biomolecules (proteins, lipids, or nucleic acids), while OPLS[222] has been optimized to reproduce experimental properties of liquids, such as density or heat of vaporization. Thus, transferability is a crucial feature for the development of a force field, because it allows the same set of parameters to be used for different molecules. In fact, a force field may fail in predictability because some parameters of the studied molecules are missing. For this purpose, in 2004 Kolmann et Al. developed a generalized version of the AMBER force field (GAFF)[223], which incorporates parameters missing in traditional AMBER for most of the organic molecules.

2.3.1 Generalized AMBER Force Field (GAFF)

GAFF is probably the most comprehensive, versatile and powerful force field available for material science. In this thesis, GAFF was widely used to parametrize all the molecular systems, unless otherwise specified. The functional form of GAFF is similar to other AMBER force fields:

$$V\left(r^N\right) = \sum_{bonds} k_{b_i}\left(l_i - l_{i,0}\right)^2 + \sum_{angles} k_{\theta_i}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{dihedrals} \frac{V_{n_i}}{2}\left(1 + cos\left(n_i\omega_i - \gamma_i\right)\right) + \sum_{i<j}^{N}\left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}}\right] \tag{2.17}$$

The default scheme in GAFF for calculating partial atomic charges is the restrained electrostatic potential (RESP)[224] algorithm, a modified version of the ESP method that applies restraints to non-hydrogen atoms. The RESP charges derivation requires *ab initio* calculations, that are performed using the Hartree-Fock (HF/6-31G*) level of theory. However, since the *ab initio* calculation may be expensive, AM1-BCC charge scheme[225] can be applied. AM1-BCC approach is much cheaper and uses the semi-empirical AM1 method with a bond charge correction to obtain results compatible with RESP charges.
Bonds and angles are modelled using harmonic potentials. The sources for the

reference bond lengths and angles in GAFF are: 1) the AMBER database, 2) *ab initio* calculations (MP/6-31G* basis sets), and 3) crystal structures. For torsional angles parametrization, the basic idea in GAFF is to perform a torsional angle scanning and find the set of parameters that best fit the torsional profile. Unless special cases (such as O-C-C-O torsional angle), the energy profile in AMBER depends only on the atom type of the central bond.

2.4 Coarse-grained Force Fields

Atomistic molecular modelling is routinely used in material science because it describes the behaviour of systems with a resolution unattainable by any experiments. However, the number of non-bonded interactions is proportional to the square of the number of particles in the system, and the computational effort required to simulate such system in relevant time scales is often prohibitive. A solution that allows to simulate larger complex systems for longer timescales is to reduce the number of interacting particles used to represent the molecules, using a simplified *coarse-grained* representations.

In a coarse-grained (CG) model, a group of atoms is modelled using a "pseudo-atom". As an example, the CG-reduction of a DPPC molecule is showed in Fig. 2.3. In general, a CG force field can be described by a similar formula to the all-atom force field (see Eq. 2.13), considering that some atoms are removed and their degrees of freedom are averaged out. A relevant consideration about coarse-graining modelling is that the free-energy profile is smoother than the atomistic one. This results in a faster exploration of the phase space (because the system avoid remaining trapped in local minima) and good estimation of free-energy differences, but also in inaccuracies of the thermodynamic properties: the coarse-graining of some degrees of freedom affects the entropy of the simulated system, which is compensated by reduced enthalpic terms[226]. As such, the time scale from a CG simulation should be interpreted with caution, and speed-up factors may need to be used.

A common strategy to develop force fields for CG models is using all-atom simulations as reference and find corresponding CG parameters that match some atomistic features. Different approaches can be employed to this end. Iterative Boltzmann inversion (IBI)[227] and inverse Monte Carlo (IMC)[228] use a target all-atom radial distribution function (RDF) and the CG pair potential is iteratively uploaded until it

reproduces the target RDF. The Force matching method[229] minimizes the mean square error between the forces acting on a CG pseudo-atom and the atomistic forces acting on the corresponding group of atoms. Other approaches for force field parametrization, such as relative entropy minimization (REM)[230] and conditional reversible work (CRW)[231], have been widely employed in the literature to study lipids, biomolecules and polymers. Several software tools, such as VOTCA[232], BOCS[233], Magic[234], PyCGTOOL[235], implement these approaches for tuning bonded interactions. With the advent of machine learning, the development and accuracy of CG models has been significantly accelerated. Within this context, our research group has developed a framework called *SwarmCG*[236–238], that is based on fuzzy self-tuning Particle-Swarm-Optimization (FST-PSO)[239]. This framework is designed to optimize both bonded and non-bonded parameters, by incorporating a combination of top-down (experimental data) and bottom-up (AA-MD simulations) reference information.

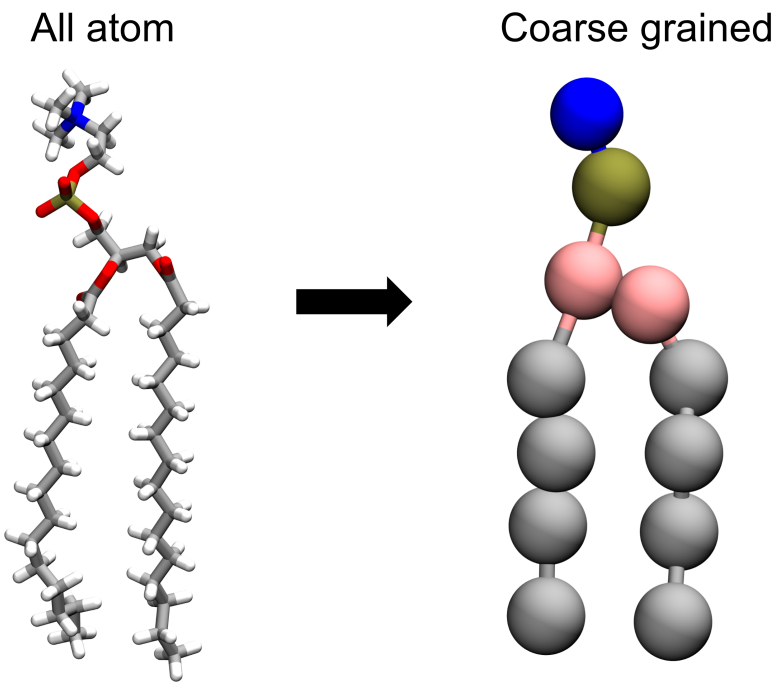

Fig. 2.3 Schematic representation of coarse-grained modelling of DPPC. On the left, the atomistic model used as reference. On the right, the reduced coarse-grained model.

2.4.1 Martini Force Field

MARTINI[240] is probably the most used CG force field. It was originally designed to model lipid membranes, and then extended to a variety of molecules, such as proteins, nucleic acids and polymers. The MARTINI force field is based on a 4 to 1 mapping, which means that up to four heavy atoms are modelled by a single MARTINI bead (all the hydrogens are neglected). Four bead types are defined, depending on the chemical nature of the atomic group: polar (P), apolar (C), non-polar (N) and charged (Q). The MARTINI beads are either divided based on their hydrogen-bonding capabilities (donor, acceptor, both, none), and based on the polarity (from 1, low polarity, to 5, high polarity). More accurate descriptions (more bead types, different sizes, improved interactions) were further introduced in MARTINI 2[241] and MARTINI 3[242] to better capture specific interactions and extend the applications. The MARTINI force field can be typically described by:

$$V_{MARTINI} = V_{bonds} + V_{angles} + V_{LJ} + V_{ele} \tag{2.18}$$

In this representation, non-bonded interactions between interacting particles are described by a Lennard-Jones (12-6) potential, where the minimum distance of approach σ_{ij} is assumed 0.47 nm, for regular beads, or 0.34 nm and 0.41 nm, for tiny and small beads, respectively, and the strength of interaction ε_{ij} depends on the type of the interacting particles and is defined for all possible pairs. The non-bonded parameters were optimized to reproduce some thermodynamic properties, such as the free-energy of vaporization, hydration and partitioning between water and some organic phases. Charged groups (Q types) interact via electrostatic Coulomb potential. Bonded parameters are modelled using simple harmonic potentials, based on atomistic geometries. Equation 2.18 can be therefore written as:

$$\begin{aligned} V(r) = &\sum_{bonds} \frac{k_b}{2}(r-r_0)^2 + \sum_{angles} \frac{k_\theta}{2}(cos\theta - cos\theta_0)^2 + \\ &+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r} \right) \end{aligned} \tag{2.19}$$

The particle types assignment based on chemical similarity and the use of standard interactions make the reduction from an atomistic to a CG representation straightforward and intuitive.

The accuracy of MARTINI compared to all-atom force fields depends on the property and system being studied. In general, MARTINI can accurately replicate the structural and thermodynamic properties of biomolecules, but it may have inaccuracies in capturing the kinetics. For example, MARTINI can reproduce the membrane thickness, area per lipid, and bending modulus of lipid bilayers with good agreement with experiments and all-atom simulations[241]. MARTINI can also reproduce the conformational flexibility of proteins with reasonable accuracy[243, 244]. However, MARTINI may overestimate the diffusion coefficients of molecules in solution. The comparison of diffusion constants of alkanes and lipids[240] indicates MARTINI is 3-8 times larger than all-atom force fields. A standard conversion factor of 4 can be applied when interpreting the results from MARTINI models, which is the speed-up factor of water dynamics when compared to real water. The overall dynamics of other processes, such as water permeation across a membrane, lipid aggregation into bilayers or vesicles, and self-diffusion of lipids, transmembrane peptides, and proteins, are accelerated by the same order of magnitude[245]. It is important to note that the speed-up factors may vary in other systems, and care should be taken when interpreting the time scale of MARTINI simulations. Nonetheless, despite the observed difference in diffusion constants, MARTINI models – and CG models in general – are highly valuable for studying the dynamics of large systems, enabling comparisons between different dynamics and providing insights into system behavior.

2.5 Molecular Dynamics

Molecular Dynamics (MD) is a computer simulation method that can predict the motion of atoms and molecules by solving the equations of motion, and the generated trajectories are used to calculate average properties of systems.

2.5.1 Periodic Boundary Conditions

In order to simulate a realistic system using a relatively small number of particles *periodic boundary conditions (PBC)* are often used in computer simulations. PBC are used to approximate an infinite system by repeating a small part called a *unit cell*. The unit cell has images or copies that surround it in all directions, forming a

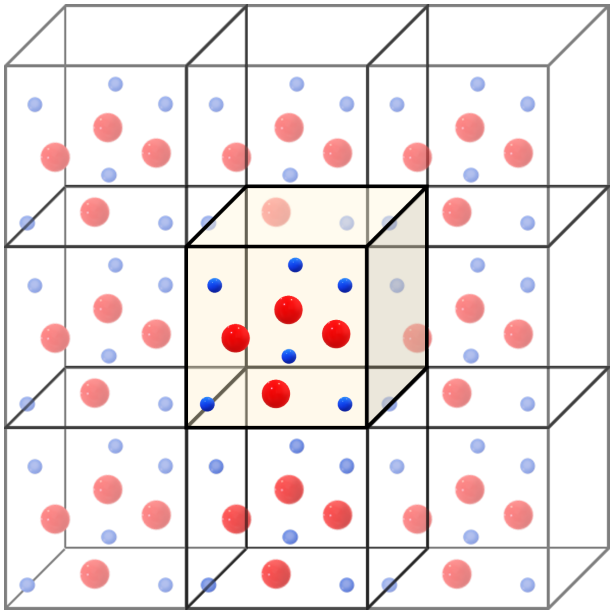

Fig. 2.4 Schematic representation of three-dimensional periodic boundary conditions (PBC).

periodic lattice (see Fig. 2.4). If, during the simulation, a particle leaves the box it is replaced by an image particle that enters from the opposite side. A method to apply PBC in molecular dynamics simulations is the *minimum image convention*. The minimum image convention states that each particle in the unit cell interacts only with the closest image of the other particles in the system. Consequently, the size of the box has to be large enough to avoid multiple interactions between particles and their images. This also means that the cut-off radius for the non-bonded interactions, such as Lennard-Jones or Coulomb, cannot exceed half the shortest box vector.

2.5.2 Integrating the Equations of Motions

MD solves equations of motion numerically and generates trajectories for a system of N interacting particles. According to the Newton's second law, the acceleration of a particle of mass m_i along the position $\mathbf{r}$ is:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} \tag{2.20}$$

For a general N-particle system, the force $\mathbf{F}_i$ acting on the particle i is a function of the position of all the N particles. The solution of the equations of motion cannot

be found analytically, but requires a numerical integration scheme. The principle of integration schemes is that the integration is decomposed in small stages, each separated in time by a fixed time discretization parameter Δt, called *time step*. The total force on each particle at time t is calculated as the sum of the interactions with the other particles and, from the force, the acceleration, velocity and position at time t are used to obtain velocity and position at time $t+\Delta t$. The choice of the time step is a crucial step in MD, as it determines the accuracy and efficiency of the numerical integration of the equation of motions. The time step should be small enough to capture the fastest motions and vibrations in the system, but not too small to waste computational resources and limit the accessible timescales. A simple guideline is to choose a time step that is one-tenth of the period of the fastest motion in the system. For example, for an atomistic system with C-H bonds, which have a period time of 10 fs, a time step of 1 fs is recommended[211]. One possible strategy to increase the integration time step without affecting significantly the accuracy of the simulation is to use constraint algorithm, that keep some internal coordinates fixed at their equilibrium value during the simulation, typically the highest frequency motion present in the system. LINCS[246] and SHAKE[247] are two commonly used algorithm to apply constraints. The simplified description of CG force fields neglets some degrees of freedom of the atomistic representation, thus allowing for larger time steps. Time steps of 20-30 fs can be used for MARTINI force field. Using time steps lower than 10 fs is not efficient, while time steps of 40 fs and larger may introduce artifacts or cause numerical instabilities. There are different integration methods, for example the Verlet algorithm[248], the velocity Verlet algorithm[249] and the leap-frog algorithm[250]. All of them use Taylor expansions to approximate the position of a particle at time $t+\Delta t$:

$$\mathbf{r}_i(t+\Delta t) \approx \mathbf{r}_i(t) + \Delta t \dot{\mathbf{r}}_i(t) + \frac{1}{2}\Delta t^2 \ddot{r}_i(t) \tag{2.21}$$

Since $\dot{\mathbf{r}}_i(t) = \mathbf{v}_i(t)$ and $\ddot{\mathbf{r}}_i = \mathbf{F}_i(t)/m_i$, Eq. 2.21 can be written as:

$$\mathbf{r}_i(t+\Delta t) \approx \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2m_i}\mathbf{F}_i(t) \tag{2.22}$$

The Leap-Frog algorithm

The default MD integrator implemented in the MD engine used in this Thesis is the *leap-frog* algorithm. The leap-frog algorithm uses the following relationships:

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i\left(t+\frac{1}{2}\Delta t\right) \tag{2.23}$$

$$\mathbf{v}_i\left(t+\frac{1}{2}\Delta t\right) = \mathbf{v}_i\left(t-\frac{1}{2}\Delta t\right) + \frac{\Delta t}{m_i}\mathbf{F}_i(t) \tag{2.24}$$

The algorithm starts from velocities at time $t - \frac{1}{2}\Delta$t and accelerations at time t – determined from the forces $\mathbf{F}_i(t)$ – to calculate velocities at time $t + \frac{1}{2}\Delta$t; Then, from the velocities just obtained and positions at time t, positions at time $t + \Delta$t are calculated. The velocity at time t is calculated as:

$$\mathbf{v}_i(t) = \frac{1}{2}\left[\mathbf{v}_i\left(t+\frac{1}{2}\Delta t\right) + \mathbf{v}_i\left(t-\frac{1}{2}\Delta t\right)\right] \tag{2.25}$$

MD Algorithm

The first step to perform an MD calculation is setting up the initial conditions of the system. Starting from the coordinates of the input structure (that can be taken from experimental structures, databases, or properly prepared with specific software packages), the initial velocities of each particle are generated from a Maxwell-Boltzmann distribution at a given temperature. The algorithm continues calculating the forces acting on each atom and the time integrator updates the positions and velocities of the atoms. The last two steps are repeated for a certain number of steps until an equilibrium state is reached. Once the equilibrium is established, the thermodynamical properties (*e.g.* temperature, pressure, energy) can be calculated as time averages.

2.5.3 Stochastic Dynamics

In many molecular systems, we are mainly interested in the behaviour of the solute. Additionally, the solvent's degrees of freedom are the most computationally expensive part. A widely employed approach to tackle these problems is the incorporation

of the solvent effects into the MD of the solute particles. *Stochastic Dynamics* relies on *Langevin equations* and considers the forces acting on a particle as the result of three contributions. The first contribution results from the interaction between the reference particle and all the other particles in the simulation box. This force $\mathbf{F}_i$ is a function of the relative positions of the particles, and it is modelled by a *potential of mean force (PMF)*. The frictional term is the term that describes the drag force that a particle experiences when moving in a fluid. This frictional force $\mathbf{F}_{frictional}$ is proportional to the velocity $\mathbf{v}$ of the particle and the friction coefficient ξ:

$$\mathbf{F}_{frictional} = -\xi \mathbf{v} \tag{2.26}$$

The friction coefficient is proportional to the friction constant γ by the relationship $\gamma = \xi / m$. The inverse of the friction coefficient has the physical meaning of a velocity relaxation time. The last term, indicated as $\mathbf{R}$, comes from occasional impacts of the particles with molecules of the surrounding medium. The Langevin equation for the particle i can be therefore written as:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i(\mathbf{r}) - \gamma_i \frac{d\mathbf{r}_i}{dt} m_i + \mathbf{R}_i \tag{2.27}$$

The second fluctuation-dissipation theorem[251] establishes the fundamental relationship between the friction term and the random force in the Langevin equations. At thermal equilibrium, the strength of the fluctuations is proportional to the magnitude of the friction acting on the system[252].

Some assumptions are made about the terms in Eq. 2.27 in different simulation methods. $\mathbf{F}_i$ is assumed constant over the time step of integration, γ_i is independent of time and position and the random force $\mathbf{R}_i$ is uncorrelated with positions, velocities, and forces acting on particles.

Various integration methods are used to integrate Langevin equations of motion[253–255]. In GROMACS, the friction and the noise terms are applied in an impulse fashion[256].

The main advantage of stochastic dynamics is that the computational time is drastically reduced, not only because of the smaller number of particles that are simulated but also because longer time steps can be used. In fact, in this Thesis, I employed stochastic dynamics as a method for simulating coarse-grained molecular systems. In this context, many fast degrees of freedom that are not essential for

long-time scale dynamics are eliminated. This increase the numerical stability of the integration scheme, allowing larger time steps to be used, similar to what was discussed in Sec. 2.5.2. Generally speaking, the larger the friction coefficient, the smaller the time step that can be used in stochastic dynamics. This is because a large friction coefficient implies a strong damping effect on the system, which reduces its characteristic time scale. If the time step is too large compared to the characteristic time scale, the simulation may become inaccurate or unstable. Therefore, one needs to choose a suitable time step that can capture the essential features of the system dynamics under a given friction coefficient.

2.5.4 Thermostats and Barostats

Molecular Dynamics simulations are typically performed in NVT and NPT ensembles, to compare simulations with experimental results.
The temperature is kept constant by using *thermostats*. Several temperature-coupling algorithms are implemented within the MD engines, such as Berendsen[257], Nose-Hoover[258, 259], Andersen[260] and velocity-rescaling[261]. All the simulations performed in this Thesis employed the velocity-rescaling algorithm, a modified version of the Berendsen thermostat, which corrects the deviation of the system temperature from the reference, but also provides the correct kinetic energy distribution.
Similarly, the system can be coupled with a *barostat* to maintain a constant pressure. The Berendsen barostat[257] scales the coordinates of the system and box vectors every step but does not generate the correct NPT ensemble. The Parrinello-Rahman barostat[262, 263] gives the true NPT ensemble, but it can be unstable if the system pressure is far from equilibrium.

2.6 Enhanced Sampling and Free Energy Calculations

MD simulations (atomistic and CG) have proven to be very efficient tools for obtaining information on a wide range of systems. Much progress has been made in this field, mainly due to improvements in hardware, which have made it possible to simulate larger systems for a longer time. However, the MD-accessible timescale

– and thus, its application – is still limited. This limitation is given by the small time step that can be employed, which must be smaller than the fast degrees of freedom of the systems to ensure stability and accuracy of the simulations. In fact, many interesting systems in biology, chemistry, or material science are characterized by various metastable states separated by large barriers, that cannot be sampled with standard simulations. To overcome this limitation, many enhanced sampling techniques have been proposed – see Refs.[264–268] for some reviews – including methods[269–274] that add an external potential to let the system escape from local minima and explore different regions in the FES, such as Umbrella Sampling[275] and Metadynamics[276]. These methods require first to choose a limited number of degrees of freedom, often referred to as collective variables (CVs), that are functions of the coordinates of the system $\mathbf{r}$. If we denote $\mathbf{s}=(s_1,...,s_d)$ the vector of d CVs, with $d \ll 3N$, and $\xi = (\xi_1,...,\xi_d)$ the functions mapping the high-dimensional $\mathbf{r}$ in the CVs space, the probability distribution P($\mathbf{s}$) can be obtained by integrating the Boltzmann probability P($\mathbf{r}$) on all the other degrees of freedom at constant $\mathbf{s}$:

$$P(\mathbf{s}) = \int P(\mathbf{r})\delta(\mathbf{s}-\xi(\mathbf{r}))\,d\mathbf{r} \tag{2.28}$$

Once P($\mathbf{s}$) is known, the free-energy as a function of $\mathbf{s}$ is:

$$F(\mathbf{s}) = -\beta^{-1} lnP(\mathbf{s}) \tag{2.29}$$

Normally, F($\mathbf{s}$) is defined up to a constant. The constant can be ignored, as often we are interested in free-energy differences between two metastable states.
Being the two macrostates A and B, $P_A = Z_A/Z$ and $P_B = Z_B/Z$, the free-energy difference between A and B is:

$$\Delta F_{A,B} = F_A - F_B = -\beta^{-1} ln\frac{Z_A}{Z_B} = -\beta^{-1} ln\frac{P_A}{P_B} \tag{2.30}$$

However, for many relevant applications (chemical reactions, protein folding, ligand binding), it is prohibitively expensive to observe even a single transition, let alone sample many to collect meaningful statistics. To overcome this limitation, and if prior knowledge of the free-energy surface is available, the exploration can be accelerated adding a potential to the MD simulations $\widetilde{V}(\mathbf{r}) = V(\mathbf{r}) + V^{bias}(\xi(\mathbf{r}))$ and the associated probability distribution becomes:

$$\widetilde{P}(\mathbf{s}) = C \int d\mathbf{r}\ \delta(\mathbf{s} - \xi(\mathbf{r}))\widetilde{P}(\mathbf{r}) = \frac{e^{-\beta\left[F(\mathbf{s})+V^{bias}(\mathbf{s})\right]}}{\int d\mathbf{s}\ e^{-\beta\left[F(\mathbf{s})+V^{bias}(\mathbf{s})\right]}} \tag{2.31}$$

where $\widetilde{P}(\mathbf{r}) = e^{-\beta\left[V(\mathbf{r})+V^{bias}(\xi(\mathbf{r}))\right]}/\widetilde{Z}$ is the biased Boltzmann distribution, $\widetilde{Z} = \int d\mathbf{r} e^{-\beta\left[V(\mathbf{r})+V^{bias}(\xi(\mathbf{r}))\right]}$ is the biased partition function, and C is a normalization constant.

Therefore, a flat probability distribution can be obtained if $V^{bias}(\mathbf{s}) \approx -F(\mathbf{s})$. In this scenario, the simulation diffuses along the landscape and the unbiased free-energy can be obtained up to a constant as: $F(\mathbf{s}) = -V^{bias}(\mathbf{s}) - \beta^{-1} ln(\widetilde{P}(\mathbf{s}))$.

Collective Variables

Before proceeding with the discussion of the enhanced sampling methods employed in this Thesis, it is useful to discuss the concept of *collective variable* and the key role it plays for the success of the methodology.
The major disadvantage of methods based on the application of an external potential over a few selected degrees of freedom (such as Umbrella Sampling and Metadynamics and its variants) is that a poor choice of collective variables can generate an incorrect estimate of the free-energy or fail to accelerate the dynamics, with no improvement in the sampling.
There are some requirements the CVs must satisfy. The first requirement is that the CV must have a different value in each of the metastable state and the transition state between them. If this requirement is not met, the bias added to one state will be indistinguishable from the bias added to any other state with the same value of CV. The second requirement is that the biased CVs should include all the "slow" degrees of freedom, that are not sampled by classic MD. Indeed, neglecting some slow modes can result in the system remaining stuck in local minima or incorrectly estimating barriers[277]. The last requirement is that the number of CVs is not too large, because a multidimensional space requires more computational time to be explored, and it becomes difficult to converge.

2.6.1 Umbrella Sampling

In Umbrella sampling (US), a series of independent simulations are restrained along the selected CV applying a bias potential. Typically, the bias applied to the system is a harmonic potential, with the equilibrium value chosen in a set of intermediate points of the CV space:

$$W_k(s) = \frac{K}{2}(s - s_k)^2 \tag{2.32}$$

where W_k is the bias potential of the k-th window, K is the force constant, and s_k is the reference value of s at the window k.
When performing US, care must be taken to the choice of K, the strength of the harmonic bias. Typically, the force constant K has to be large enough to sample the desired subset of the CV space, but not too large to ensure sufficient overlap between the distribution of adjacent windows.

Weighted Histogram Analysis Method (WHAM)

After a proper sampling of each window is achieved, all the data from each simulation must be combined to obtain the unbiased probability distribution and to estimate free-energies. The *weighted histogram analysis method* (WHAM)[278] is probably one of the most popular methods to calculate the free-energy profile from US simulations. WHAM calculates the unbiased probability distribution as the weighted average of the distribution generated from each window:

$$P^U(s) = \sum_{k=1}^{N} w_k(s) P_k^U(s) \tag{2.33}$$

where N is the total number of windows, and w_k are the weights and must satisfy the condition $\sum w_k = 1$. The coefficients w_k are determined minimizing the statistical error σ in the unbiased distribution:

$$\frac{\partial \sigma^2\left(P^U\right)}{\partial w_k} = 0 \tag{2.34}$$

Assuming that the error function is the same in all the N windows, and that the sampling is adequate in each window, the weights are given by:

$$w_k(s) = \frac{n_k e^{-\beta W_k(s)+\beta F_k}}{\sum_{j=1}^{N} n_j e^{-\beta W_j(s)+\beta F_j}} \tag{2.35}$$

where n_k are the number of samples from the k-th window. The F_k are calculated as:

$$e^{-\beta F_k} = \int P^U(s) e^{[-\beta W_k(s)]} ds \tag{2.36}$$

Eqs. 2.34 and 2.36 are coupled because $w_k(s)$ enters the calculation of F_k and *vice versa*. The two equations must be iterated, usually starting from an initial guess for F_k, until a self-consistent solution is reached.

A complete representation of the US scheme and reweight is showed in Fig. 2.5. One possible strategy for assessing the convergence of the Potential of Mean Force (PMF) – that is, the free energy surface along a given reaction coordinate or collective variable – generated from WHAM is to compute the PMF at different time intervals and compare the profiles. If the profiles are similar over time, it indicates convergence. However, this condition is insufficient for more complex systems[280]. Therefore, it is advisable to perform an error analysis to estimate the uncertainty and reliability of the PMF. The error can be calculated with different methods, such as bootstrap[281] or block average[282]. Bootstrap is a method that resamples the dataset generating a new "bootstrapped" trajectory for each umbrella histogram and yields to a new histogram of the bootstrapped trajectory. WHAM is then applied to the bootstrapped histograms to compute a new PMF. The procedure is repeated several times. The standard deviation of the PMF values from different resampled data sets gives an estimate of the error[283]. Block average is a method that divides the data of the collective variable x in each window i into blocks of different sizes, and the variance of the x is estimated as the variance of these averages. From the variance of x in each window, the variance of the PMF *G(x)* is given by[284]:

$$var[G(x)] \approx (K\Delta r)^2 \cdot \sum_{i=1}^{(x-r_0)/\Delta r} var(\bar{x}_i) \tag{2.37}$$

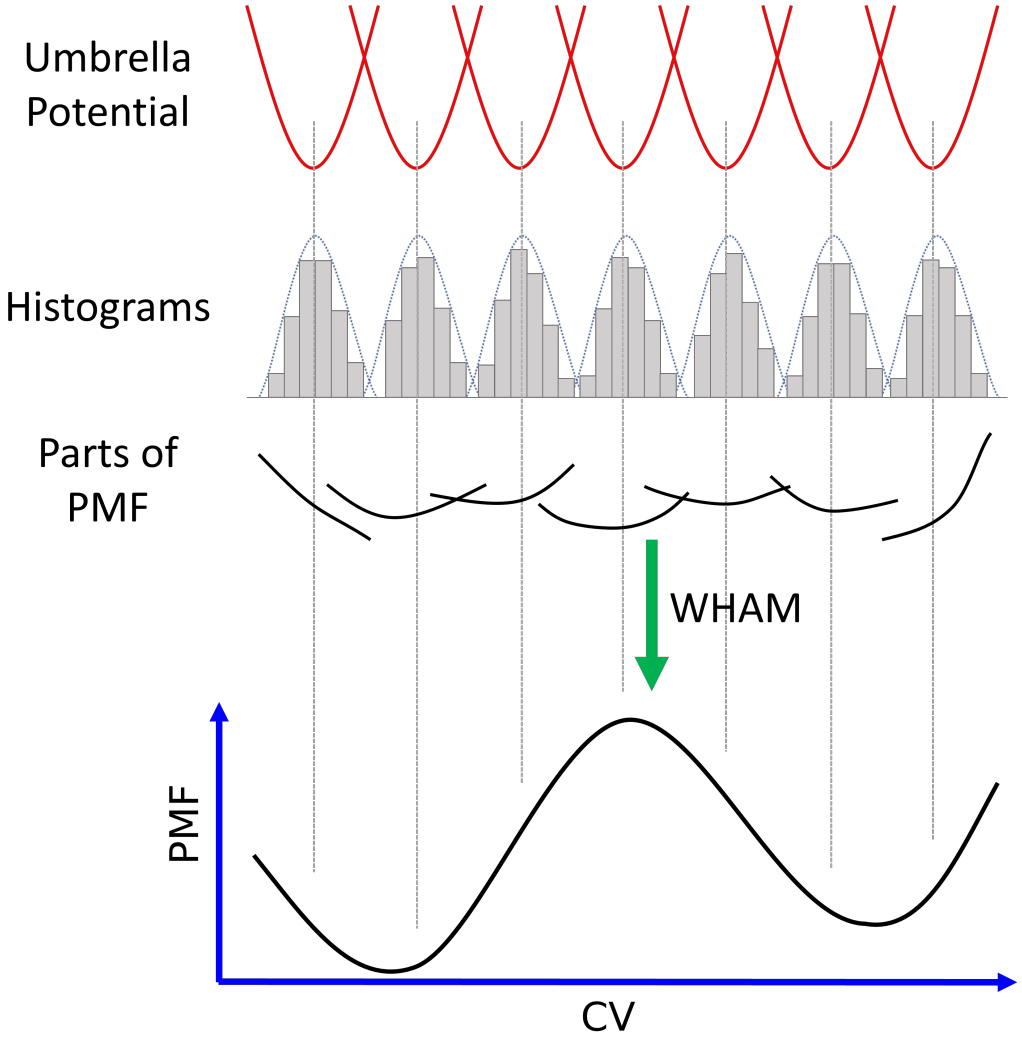

Fig. 2.5 Schematic representation of Umbrella Sampling (US). The CV space is divided in multiple windows, and each window is sampled independently (top row). From each independent simulation window, one build the histogram of the configurations (second row) and, subsequentially, the potential of mean force (PMF) (third row). WHAM algorithm combines all the data from each simulation, to obtain the global free-energy surface. Adapted with permission from Ref.[279].

2.6.2 Metadynamics

In Metadynamics[276], the MD potential is combined with a history-dependent term. Namely, the time-dependent bias, in the form of Gaussian function, is deposited in the CV space with a user-defined frequency following the system sampling. The Metadynamics bias at time t can be written as:

$$V_G(\mathbf{s},t) = \int_0^t dt' \omega exp\left(-\sum_{i=1}^{N} \frac{\left(s_i - s_i(t')\right)}{2\sigma_i^2}\right) \tag{2.38}$$

where σ_i is the width of the gaussians of i-th CV, ω is an energy rate, and it can be expressed in terms of the gaussian height W and the deposition stride τ_G as:

$$\omega = \frac{W}{\tau_G} \tag{2.39}$$

This potential gradually deposits a bias potential in the CV space, ideally obtaining the convergence condition, *i.e.,* allowing the system freely diffuse among metastable states (Fig. 2.6). In this way, the sum of the underlying FES and the deposited potential will return a uniform landscape and the inverse of the deposited potential can be considered an estimation of F(s).
The strongest point of this technique is its twofold goal: first, it allows to "escaping free-energy minima", enhancing the sampling of the simulated system in a small computational time. Secondly, at convergence, it returns a full free-energy landscape in the chosen CVs.

However, Metadynamics also has some limitations. Firstly, it is not straightforward to understand when a Metadynamics simulation converges, because the continuous bias deposition can create artefacts on the landscape. Moreover, the system can be pushed in physically unrealistic (or at least irrelevant) regions, under the effect of the bias. In order to tackle these issues, *Well-Tempered Metadynamics* have been proposed.

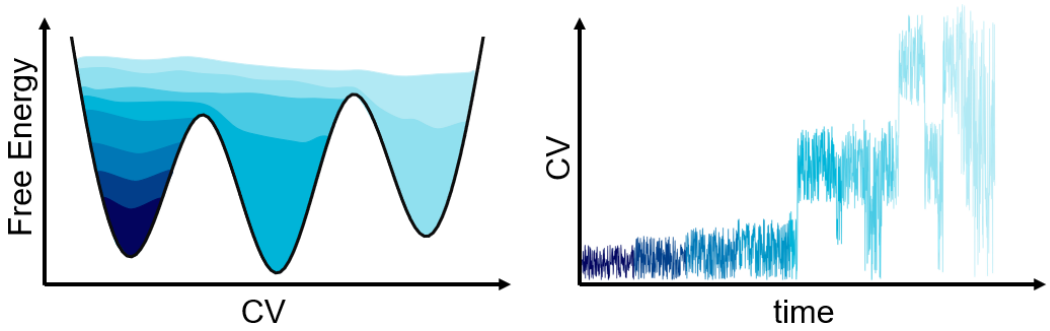

Fig. 2.6 Schematic representation of Metadynamics; On the left, the free-energy landscape (black line) as a function of a collective variable (CV). The blue scale filled area represents the bias potential added to the CV space at different times. On the right, the trajectory of the system in the CV space.

Well-Tempered Metadynamics

In Well-Tempered Metadynamics[285] (WT-MetaD), the bias deposition rate decreases with time. The new expression of the Metadynamics potential can be written:

$$V(\mathbf{s},t) = k_B \Delta T \, ln\left(1 + \frac{\omega N(\mathbf{s},t)}{k_B \Delta T}\right) \tag{2.40}$$

where $N(\mathbf{s},t)$ is the histogram of the variable $\mathbf{s}$ sampled in the biased simulation, ΔT is a temperature. The change in the deposition rate is implemented by rescaling the height of each Gaussian to:

$$W = \omega \tau_G e^{-\frac{V_G(s,t)}{k_B \Delta T}} \tag{2.41}$$

Thus, the height of the Gaussian deposited at a given point of the CVs space is inversely proportional to the time already spent at that point.
In the long-time limit, WT-MetaD does not converge to the negative of the free-energy $-F(S)$, like standard Metadynamics, but rather to:

$$V(\mathbf{s}, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(\mathbf{s}) + C \tag{2.42}$$

and samples the distribution $P(\mathbf{s}) \propto e^{-\frac{F(\mathbf{s})}{(T+\Delta T)}}$.
The limit case of $\Delta T \to 0$ corresponds to classical MD, while for $\Delta T \to \infty$ standard

Metadynamics is recovered. In intermediate cases, ΔT controls the effective temperature at which the CV is sampled. Thus, ΔT is a key input parameter to be tuned, as it increases the barrier that can be crossed, but also limits the exploration of the FES to the physically relevant regions of the CVs space. It can be expressed in terms of the *bias factor* $\gamma = \frac{T+\Delta T}{T}$.

To assess the convergence of Metadynamics simulations, some preliminary steps can be followed. Firstly, the system should diffuse rapidly in the entire space of the CV. Secondly, one can compute the free energy surface as a function of simulation time. At convergence, the free energy profiles at different times should be similar. Another condition to monitor the convergence of simulation is the observation of the time-dependent function *c(t)*, which is an estimator of the reversible work done by the Metadynamics bias and will be detailed in the next paragraph, and which should grow logarithmically in the regime of convergence. If the three qualitative conditions are satisfied, one can perform block-analysis to have quantitative signs of convergence. Firstly, it is advisable to discard the initial part of the simulation. After that, the simulation trajectory is divided into blocks and the histogram of the CVs is computed in each block, as well as its error. Then, the error on the histogram is converted into an error on the free-energy estimator. By visualizing the error as a function of the block size, the error should increase until it reaches a plateau when the dimension of the block exceeds the correlation time between data points. If a plateau is observed, the convergence is reached.

Despite its undoubted advantages in boosting convergence and exploring the phase space, it is also important to consider WT-MetaD limitations. These limitations include the need for some *a priori* knowledge of the free-energy surface, in particular regarding the depth of the free-energy wells. This knowledge is important as it directly influences the choice of the bias factor. Another limitation – which is common to every CVs-based method – is that the identification of a low number of CVs to effectively explore complex events is typically difficult, making the application of WT-MetaD not always effective or easy to handle, particularly for complex systems. Thus, the careful consideration of these limitations is essential when utilizing WT-MetaD for molecular simulations.
Lastly, we note that the introduction of the bias potential changes the probability distribution. For the biased variables, the unbiased probability distribution can be easily recovered from the sampled distribution P(**s**). It is not so straightforward

for all the others degrees of freedom. Several techniques have been proposed to reconstruct the free-energy also as a function of non-biased variables[286–288].

Reweighting Metadynamics Bias

A common way – and the most used in this Thesis – to obtain the unbiased probability distribution from a biased simulation was introduced by Tiwary and Parrinello[288] and takes into account the time-dependence of the bias potential. The time-dependent constant $c(t)$ is the logarithm of the ratio of the unbiased and the biased partition functions:

$$c(t) = \frac{1}{\beta}\frac{Z}{\widetilde{Z}(t)} = \frac{1}{\beta} log \left(\frac{\int d\mathbf{s} e^{-\beta F(\mathbf{s})}}{\int d\mathbf{s} e^{-\beta \left(F(\mathbf{s}) + V(\mathbf{s},t) \right)}} \right) \tag{2.43}$$

or, directly from the WT-MetaD run, it is estimated using the formula:

$$c(t) = \frac{1}{\beta}\frac{Z}{\widetilde{Z}(t)} = \frac{1}{\beta} log \left(\frac{\int exp \left[\frac{\gamma}{\gamma - 1} \beta V(\mathbf{s},t) \right] d\mathbf{s}}{\int exp \left[\frac{1}{\gamma - 1} \beta V(\mathbf{s},t) \right] d\mathbf{s}} \right) \tag{2.44}$$

The average of any observable O that depends on the atomic positions will be:

$$\langle O(\mathbf{r}) \rangle = \frac{\langle O(\mathbf{r}) e^{\beta [V(\mathbf{s},t) - c(t)]} \rangle_{MetaD}}{\langle e^{\beta [V(\mathbf{s},t) - c(t)]} \rangle_{MetaD}} \tag{2.45}$$

Infrequent Metadynamics

The WT-MetaD framework has been extended to obtain kinetics for rare events, introducing the *infrequent* Metadynamics[289].
The rate of the rare event of escaping from state A to state B is given by:

$$k = \omega \kappa \frac{Z_{TS}}{Z_A} \tag{2.46}$$

where Z_{TS} and Z_A are the partition functions of the transition state and of the state A, respectively, ω is a normalization constant that depends on the temperature and mass of the reaction coordinates, and κ is the transmission coefficient, which takes into account recrossing events and is assumed to be equal to 1.

In practice, the transition state is hardly visited by MD trajectories, thus it is not trivial to have a converged value of Z_{TS}.
Now, assuming that (i) the time taken to cross the barrier is much smaller than the time spent in each basin, and (ii) one have some CVs that are able to distinguish between the two free-energy basins, one can perform a Metadynamics run to gradually enhance the probability to observe a transition. The kinetic rate from the biased simulation can be written as:

$$k^* = \omega\kappa\frac{Z_{TS}}{Z_A^*} \tag{2.47}$$

If no bias is added to the transition state, the quantities κ and Z_{TS} are the same as in Equation 2.47. The ratio of k^* and k is called *acceleration factor* α:

$$\alpha(t) \approx \frac{Z_A}{Z_A^*} = \langle e^{\beta V(\mathbf{s},t)} \rangle_{MetaD} \tag{2.48}$$

The angular brackets $\langle ... \rangle_{MetaD}$ indicate an average over the Metadynamics run and V(**s**, t) is the Metadynamics bias.
Assuming that conditions (i) and (ii) are met, another crucial condition for the method's success is to prevent depositing bias over the transition state. This can be achieved by reducing the deposition rate between two consecutive Gaussians.
The real (unbiased) transition time t_t is obtained as:[268]

$$t_t = \sum_i^{n_{MD}} dt e^{\beta V_{t_i}(\mathbf{s}(t_i),t_i)} \tag{2.49}$$

where n_{MD} are the total MD steps, $t_i = i\ dt$ is the MD time at the i-th time step and $V_{t_i}(\mathbf{s}(t_i),t_i)$ is the Metadynamics bias at time $t = t_i$.
By performing multiple simulations, one can build an empirical distribution of escape times. The reliability of the dynamics reconstructed from Metadynamics can be assessed[290] by performing a statistical test that evaluate how well the computed distribution fit with the ideal Poisson distribution, expected for rare events:

$$P_{n\geq 1} = 1 - P_0 = 1 - e^{-\frac{t}{\tau}} \tag{2.50}$$

where τ is the characteristic timescale of the transition.
In this Thesis, to quantitatively compare the theoretical and empirical distribution, the Kolmogorov-Smirnov (KS) test[291] has been performed.
The barrier associated to the transition $\Delta G^{\ddagger}$ can be calculated from the Eyring

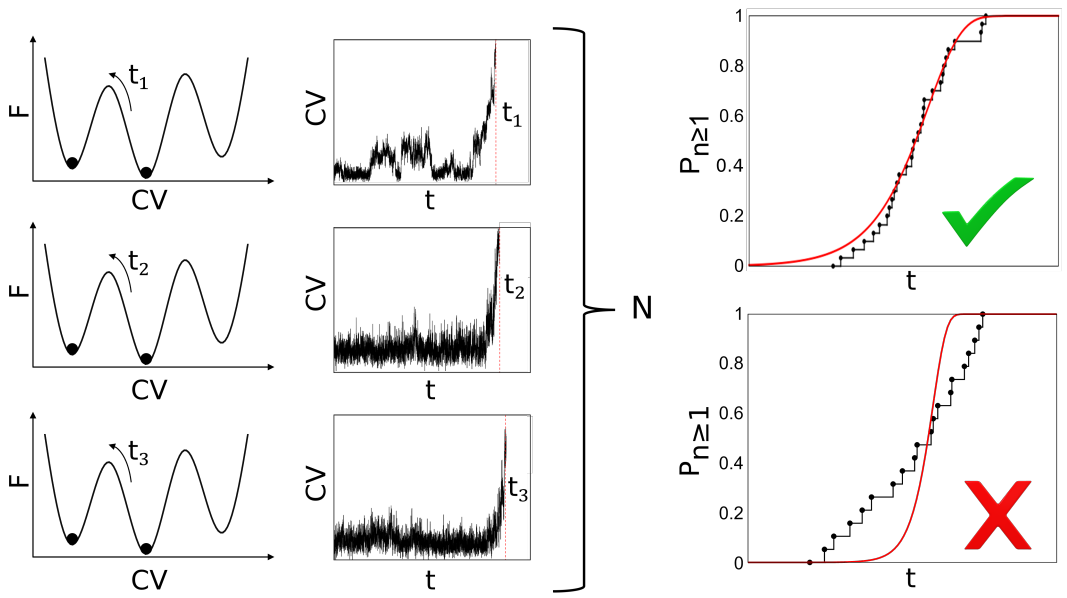

Fig. 2.7 Schematic description of the infrequent Metadynamics approach: N single-transition replicas are run in parallel; the transition time collected from each replica (t_1, t_2, ..., t_N) is used to build an empirical cumulative distribution function (on the right, a black step identifies each time) that is fitted with a Poissonian distribution function (red curves). Statistical tests (*e.g.* Kolmogorov-Smirnov test) must be performed to assess the reliability of the empirical distribution.

equation, using the constant rate k^* calculated at the previous step:

$$k^* = \frac{\kappa k_B T}{h} e^{\frac{-\Delta G^{\ddagger}}{RT}} \tag{2.51}$$

where κ is the transmission coefficient (equal to one in the no-recrossing assumption of transition state theory), k_B is Boltzmann's constant and h is Planck's constant. A schematic representation of the infrequent Metadynamics principles is showed in Fig. 2.7.

2.7 Analysis of MD simulations

2.7.1 Mapping auxiliary variables in CV space

In this Thesis, I used the approach described by Gimondi et Al.[292] to build maps of auxiliary variables in CV space (*e.g.* number of hydrogen bonds, SASA, enthalpy

and entropy). Adopting the notation of the reference paper, **s** is the set of CVs where the system is mapped and $\bar{s}$ is the auxiliary variable. Each point in **s** represents an ensemble of degenerate configurations. One can define a local probability density for $\bar{s}$, for any values of **s**, namely the conditional probability density of $\bar{s}$ under a constraint on the value of **s**:

$$p(\bar{s}|\mathbf{s}) = \frac{\int e^{-\beta F(\mathbf{s}|\bar{s})} \delta(\mathbf{s}-\mathbf{s}')d\mathbf{s}'}{\iint e^{-\beta F(\mathbf{s}|\bar{s})} \delta(\mathbf{s}-\mathbf{s}')d\bar{s}d\mathbf{s}} \tag{2.52}$$

From the probability density, $p(\bar{s}|\mathbf{s})$ one can compute the ensemble average of $\bar{s}$ over the ensemble of configurations degenerates in **s**:

$$\langle \bar{s} \rangle_{\mathbf{s}} = \int \bar{s} p(\bar{s}|\mathbf{s}) d\bar{s} \tag{2.53}$$

2.7.2 Enthalpic and Entropic Contribution to the Free Energy

The WT-MetaD simulations allow us to compute the projection of the free-energy surface on the space of the collective variables. The Gibbs free-energy can be written as the sum of the enthalpic and the entropic part:

$$\Delta G(\mathbf{s}) = \Delta H(\mathbf{s}) - T\Delta S(\mathbf{s}) \tag{2.54}$$

The approach proposed in Ref.[292] and recently applied for studying the early stage of nucleation of metal-organic frameworks[293] was employed here for a systematic breakdown of the free-energy surface. The enthalpic contribution to the free-energy is:

$$\Delta H(\mathbf{s}) = \Delta U(\mathbf{s}) + P\Delta V(\mathbf{s}) \tag{2.55}$$

where P is the pressure and V(**s**) is the volume, and U(**s**) is the sum of the ensemble averages of the potential and kinetic energy:

$$U(\mathbf{s}) = \langle E^P \rangle_{\mathbf{s}} + \langle E^K \rangle_{\mathbf{s}} \tag{2.56}$$

At this point, it is worth noting that for negligible variations of the volume mapped on the CV space, the term PΔV(**s**) is constant, and the enthalpy reduces to the ensemble

average of the internal energy on **s** U(**s**). Moreover, at constant temperature, the kinetic energy does not depend on **s**, and the internal energy reduces to the potential energy of the system $\Delta E_{P(\mathbf{s})}$.
For systems immersed in explicit solvent, as for the systems analyzed in Chapter 6, the potential energy of the system can be further decomposed as follows:

$$\Delta E_P(\mathbf{s}) = \Delta\langle E_P^{Solute}\rangle_{\mathbf{s}} + \Delta\langle E_P^{Solvent}\rangle_{\mathbf{s}} + \Delta\langle E_P^{Solute-Solvent}\rangle_{\mathbf{s}} \tag{2.57}$$

The Eq. 2.57 can be reduced, noting that the term $\Delta\langle E_P^{Solvent}\rangle_{\mathbf{s}}$ is independent from the value of the CV, **s**.
In conclusion, the enthalpy mapped on the CV space reduces to the ensemble average of the potential energy of the system along the CV, and the entropy can be obtained by difference.

2.8 Software

The GROMACS simulation package[294, 295] was used to run all the simulations in this Thesis. The PLUMED plugin[296, 297] was used to bias MD simulations and perform analysis on trajectories. Non-standard molecules were created in Avogadro[298] and then parametrized with the AMBER package[299]. *Ab initio* calculations for charges derivation were performed using Gaussian software[300] (RESP charges) or *am1bcc* module[301] in AMBER (BCC charges). VMD[302] was used to visualize trajectories and render snapshots. Blender software[303] was used to model and render 3D models.

Chapter 3

Controlling the length of porphyrin supramolecular polymers

This work has been carried out in collaboration with the experimental group of Prof. E. W. Meijer, at the Eindhoven University of Technology (Netherlands). This section will focus on the computational aspects and the role of molecular modelling in supporting the experiments. The Introduction is largely taken from the postprint version of the article published on *Nature Communications*. The experimental results are briefly summarized at the beginning of the "Results" section. Details regarding the experimental techniques are available in the full published paper[304].

Full bibliographic reference: Weyandt, E., **Leanza, L.**, Capelli, R., G. M. Pavan, G. Vantomme, E. W. Meijer. Controlling the length of porphyrin supramolecular polymers via coupled equilibria and dilution-induced supramolecular polymerization. *Nature Communications* 13, 248 (2022). DOI:10.1038/s41467-021-27831-2[1]

In this chapter, we will be focusing on porphyrin-based polymers, an important class of supramolecular polymers, whose importance and application have been explored in Chapter 1. Specifically, we will examine their behaviour in multi-component systems, which refer to systems that involve two or more types of molecules or species.

[1]My contribution to this article, as the first computational author, has been the development of all the atomistic and coarse-grained models, performing all MD simulations and analysis and contributing to the interpretation of the results and in the writing of the manuscript.

Multi-component systems often display convoluted behaviour, pathway complexity and coupled equilibria. In recent years, researchers have explored ways to control complex systems by manipulating the subtle balances of interaction energies between the individual components and thereby shifting the equilibrium between different aggregate states. By adding a monotopic Mn^{3+}-porphyrin monomer, Zn^{2+}-porphyrin-based supramolecular polymers exhibit enantioselective chain capping and dilution-induced supramolecular polymerization. When mixing the zinc and manganese centred monomers, the Mn^{3+}-porphyrins act as chain-cappers for Zn^{2+}-porphyrin supramolecular polymers, effectively hindering growth of the copolymer and reducing the length.
Through the use of advanced simulation techniques (i.e. Metadynamics) and coarse-grained modelling we gained insights into the intricate supramolecular interactions and dynamics involved in co-assembly and chain-capping events, as well as possible exchange pathways in capper-release from chain ends, which is essential for recovering supramolecular polymerization.
Our work also demonstrates that integrating experimental and theoretical methods provides a comprehensive understanding of complex supramolecular systems, offering a valuable framework for future research in the development of novel functional materials based on supramolecular interactions.

3.1 Introduction

Porphyrin-based polymers have recently become the focus of intense research due to their remarkable photochemical properties, but also to their intricate and complex assembly behaviour[305]. The extended π-systems surrounding the core lead to high levels of absorptions in the ultra-violet and visible region, but also to a strong tendency to aggregate through π-stacking and solvophobic interactions[306]. The assembly behaviour is particularly sensitive to small structural changes, such as changes in substituents, porphyrin core structure, and the addition of metal ions[153]. For example, *C*-centred monomers tend to form highly cooperative supramolecular polymers, while *N*-centred monomers only assemble into short, non-helical J-aggregates due to the high rotation barrier around the amide groups hindering the formation of hydrogen bonds. The complex and dynamic nature of porphyrins often result in pathway complexity, with monomers assembling into multiple types of

aggregates via competing pathways and mechanisms[152, 307]. By controlling the assembly equilibria under kinetic or thermodynamic conditions, interesting phenomena have been discovered. Sugiyasu and colleagues utilized the competition between different aggregate states in a living supramolecular polymerization to produce fibres with controlled lengths and narrow dispersities[308–313]. Similar techniques have been used for seeded growth of supramolecular polymers in one or two dimensions and the preparation of supramolecular block copolymers and polymorphs through kinetic control[150, 314–316].

The complexity of porphyrin assembly behaviour increases in multi-component systems, where the interplay between multiple components can give rise to a wider range of assembly possibilities, under varying conditions such as temperature, composition, solvent, and concentration. Many studies have demonstrated the selectivity of one assembly pathway over another through small changes in solvent composition[317–320]. The Aida group showed the thermally bisignate polymerization of porphyrin monomers by tuning the interactions between an alcohol and the monomers across a wide temperature range[321, 322]. The Meijer group demonstrated the dilution-induced self-assembly of a monomer in the presence of pyridine by manipulating the coupled equilibria between complexed monomer and polymer formation across a range of concentrations[323].

Enantioselective interactions have been found to reduce the level of complexity in these multi-component systems by introducing specificity in the interactions of the components, which can be helpful to isolate effects in the aggregation pathways. Chiral recognition and exchange of chiral information are essential in both biological systems and in chemical catalysis[324–327], host-guest complexes[328–330] and supramolecular systems[331–336]. In the supramolecular polymerizations of homochiral monomers, either *M*- or *P*-helical fibres are formed. With heterochiral monomer mixtures, the monomers can intercalate into a stack of majority-preferred helicity[337, 338], form alternating heterochiral polymers[339], or the monomers narcissistically self-sort if the mismatch penalty for co-aggregate formation is too high. Nakashima et al. reported a method to control aggregate length and morphology by balancing the enantiomeric excess (ee) of the components[340]. George et al. demonstrated chirality-driven self-sorting and stereoselective polymerizations with Naphthalene diimide (NDI) monomers[331, 332, 341, 342]. Meijer's group reported self-sorting of zinc-porphyrin monomers into homochiral stacks and selective depolymerization by adding a Lewis base[343], proposing that the chiral

discrimination exhibited in self-sorting systems can control aggregate microstructure in multi-component systems.
As shown in Chapter 1, molecular modelling has become essential for understanding the underlying mechanisms and rationalizing experimental results of narcisistically self-sorted supramolecular polymers. Detailed molecular dynamics simulations were used to explore the pathway selectivity in NDI assemblies, and to estimate the relative rates of monomer exchange dynamics[164, 316].
In the field of porphyrin-based supramolecular polymers, molecular modelling has been used to study polymerization and the impact of the organic catalyst DMAP on the depolymerization process[150].
In this study we focus on a zinc-centred porphyrin-based monomer (**S-Zn**), which in apolar solvents such as methylcyclohexane (MCH) forms highly cooperative, helical supramolecular polymers through fourfold hydrogen bonding interactions. The **S-Zn** monomers exhibit pathway complexity and form next to helical H-aggregates also non-helical weakly coupled J-aggregates via an isodesmic mechanism. The use of monotopic manganese^{III} porphyrin monomers (**S/R-Mn**) with an axially bound chloride ion causes chain-capping for zinc centred porphyrin monomers (Fig. 3.1). By adding homo- or heterochiral manganese chain-cappers we enantioselectively control the capping of H- or J-aggregates and reduce the length of the supramolecular polymers. Because of coupled equilibria, H-aggregates can be regained from depolymerized mixtures of manganese and zinc monomers by simply reducing the overall concentration of all components through dilution-induced supramolecular polymerization.
The aim of this work was to elucidate the complex interplay of interactions between monomers, contributing to obtain a deeper understanding of supramolecular aggregation in multi-component systems, by integrating computational simulations with experimental data.

3.2 Results and Discussion

3.2.1 Experimental section

To provide a clear context for the theoretical calculations, we have summarized the main experimental results, obtained by the group of Prof. E. W. Meijer, as follows.

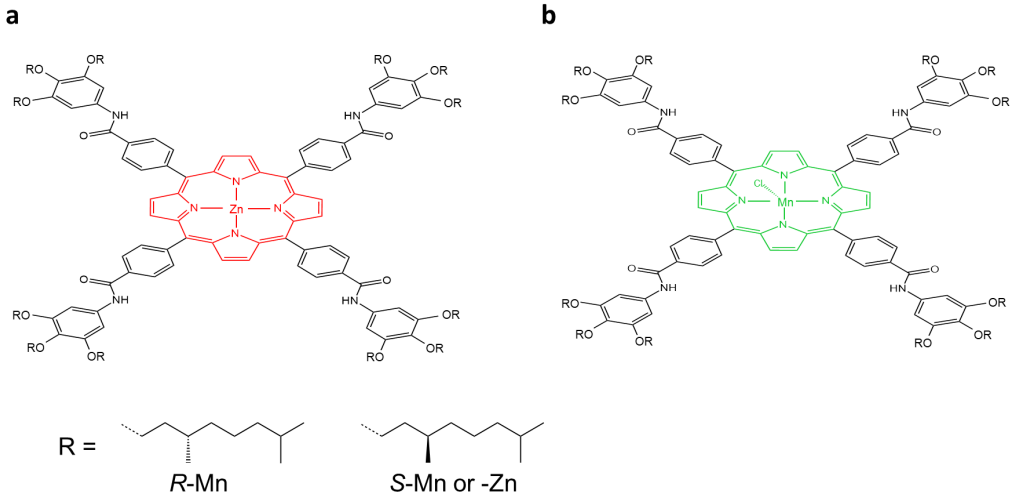

Fig. 3.1 Molecular structures of **S-Zn** (left) and **R/S-Mn** (right). Adapted from Ref.[304].

Supramolecular homopolymerization. The supramolecular polymerization and optical properties of **S-Zn** are known from previous studies[319, 344]. H-aggregates exhibits an adsorption band at λ=392 mm, while the J-type aggregate has its maximum absorption band at 425 nm and. Due to the axially bound chloride, **S/R-Mn** can only be monomeric or dimeric. In diffusion ordered spectroscopy (DOSY) NMR experiments we find short species with a diffusion coefficient of 2.32 $\cdot 10^{-10} m^2 \cdot s^{-1}$, corresponding to a diameter of around 22.7 Å. Macroscopically, the formation of small monomeric or dimeric species for **S-Mn** is evidenced by excellent solubility and low viscosity in MCH. For **S-Zn**, the long supramolecular fibres diffuse too slowly to be measured with DOSY-NMR. Fourier-transform (FT)-IR measurements of 2.0 mM solutions of **S-Zn** and **S-Mn** indicate hydrogen-bonded organization for **S-Zn** and weak hydrogen-bonding for **S-Mn**.

Enantioselective chain-capping of porphyrin stacks. When two monomers **S-Zn** and **S/R-Mn** are mixed (the molecular structure of the monomers is displayed in Fig. 3.1), there are several possibilities for co-aggregate formation: (1) no interaction and both monomers self-sort; (2) homochiral interaction, only **S-Mn** is a chain-capper for **S-Zn** polymers; (3) heterochiral interaction, both **S-** and **R-Mn** are chain-cappers for **S-Zn** polymers.
We prepared mixed solutions of **S-Zn** with either homochiral **S-Mn** or heterochiral **R-Mn**. For the homochiral monomer/chain-capper pair, a decrease in CD of the H-aggregate is observed when increasing the amounts of **S-Mn** to a 1:1 ratio. When the

length of the H-aggregates decreases through the interaction with the chain-cappers, the H-aggregates become destabilized and convert into J-aggregates as the interaction energy of the aggregates decreases with decreasing length. For the heterochiral pair, these effects are less pronounced, but a slight increase in J-aggregates is observed, when **R-Mn** is added. Double capping of the oligomers with S/R-Mn on both ends is also possible, but the experiments do not indicate one or the other. These results seem to support that only **S-Mn** can act as a chain-capper for both H- and J-aggregates of **S-Zn**, while **R-Mn** only interacts with the achiral J-aggregates.

Dilution-induced supramolecular polymerization. AFM experiments support the evidence that **S-Mn** causes the depolymerization and transformation of H- into J-aggregates of **S-Zn**. However, repeating the AFM experiments at lower concentration the reformation of H-aggregates from the J-aggregates is observed. For heterochiral mixtures, this phenomenon is less pronounced. This counterintuitive repolymerization with decreased total concentration is a consequence of coupled equilibria and pathway complexity.
By lowering the concentration and shifting of the equilibria, the **S-Mn/S-Zn** interaction becomes less favourable than **S-Zn** homo-interaction. This causes a release of **S-Mn** chain-capper from the chain ends and thereby reformation of the supramolecular polymers.

3.2.2 Computational section

The computational simulations allowed us to elucidate the underlying mechanisms driving the experimental observations, providing a more complete picture of the complex supramolecular interactions and dynamics involved in the co-assembly and chain-capping processes.

Coarse-grained molecular modelling of Zn and Mn monomer exchange

Our initial focus is investigating the interactions between monomers. To this aim, we initially built an all-atom (AA) model of the monomers (computational details are available in the "Methods" section of this Chapter). However, the size and the timescales involved in the self-assembly and dynamic behaviour of these supramolecular systems exceed the possibilities for AA models. As recently done for similar structures[150, 196], we thus developed coarse-grained (CG) models for the

monomer (Fig. 3.2a). The limited resolution of these CG models (3-4 heavy atoms per CG-bead) does not allow distinguishing between S- or R- chirality. Nonetheless, such CG model can safely be used to qualitatively compare the dynamics of two (**Zn** and **Mn**) porphyrin-based supramolecular building blocks, and to shed light on the complex mechanism of monomer exchange and of interaction between comonomers of similar nature[164, 197, 316, 345]. The model for **Zn** monomers was based on previous work on zinc-centred porphyrin polymer fibres[150], and the non-bonded interaction were optimized in order to reproduce the dimerization free-energy of two atomistic porphyrins' cores, using a well-tested Metadynamics-based protocol[150, 164, 196, 316]. Similarly, we used the dimerization free-energy of two Mn-centred all-atom monomers as a reference to optimize the CG interactions. We used these two CG models to study and compare via WT-MetaD simulations the **Zn-Zn**, **Zn-Mn**, and **Mn-Mn** interactions. Our WT-MetaD simulations provided a dimerization free-energy ΔG for **Mn-Mn** cores of $\sim$18.0 kJ mol^{-1} (Fig. 3.2d). In previous work, the ΔG for **Zn-Zn** was found $\sim$45.2 kJ mol^{-1}[150]. The **Zn-Mn** core interactions are $\sim$18.8 kJ mol^{-1} and just slightly stronger than **Mn-Mn** core interactions (Fig. 3.2b). The difference observed in monomer interactions between the three pairs can be attributed to various factors. For instance, Mn larger atomic radius results in a greater distance between the two nuclei, weakening their interaction strength. Additionally, the presence of the chloride counterion appears to play a crucial role in influencing monomer interactions. The metal Mn and the positively charged atoms of the core are attracted to Cl^- ions, which induces structural changes, making the porphyrin molecule less planar. This geometric change results in a reduction of overlap between aromatic rings and alignment of two molecules that are crucial for $pi - pi$ stacking and metal-metal interactions.

Subsequently, CG models for a **Mn** dimer (Fig. 3.2c) and for a fibre composed of 40 **Zn** monomers (Fig. 3.2d) were pre-equilibrated in a periodic simulation box filled with explicit MCH solvent molecules, and we performed molecular dynamics (MD) simulations at a temperature of 300 K and a pressure of 1 atm. Both the **Mn** dimer and the **Zn** fibre resulted in stable structures in the timescale accessible by these CG-MD runs.

We employed infrequent WT-MetaD[289] simulations to obtain information on the relative characteristic kinetics for the events of **Mn** dimerization and **Zn** monomer exchange from the **Zn** fibre tip, using a similar approach to the one recently used to study the dynamics of other supramolecular polymers[164, 197, 316]. Activating

monomer exchange in the two systems, these simulations allowed us to retrieve the free-energy barriers involved in the exchange and the related characteristic timescales (t_{CG}) expected for these exchange phenomena. The **Mn** dimer breakage requires the system to cross an activation barrier of $\sim$51.8$\pm$0.8 kJ mol^{-1}. On the other hand, exchanging one monomer from the **Zn** fibre tip requires crossing a free-energy barrier of $\sim$71.5$\pm$2.0 kJ mol^{-1}. The transition probability distributions obtained from multiple infrequent WT-MetaD simulations allowed us to estimate the kinetics of monomer exchange and to compute the characteristic exchange timescale. While the exchange timescales estimated from these (approximated) CG models should be considered as qualitative, these are still useful to compare the dynamics of the two systems simulated at the same level of resolution[164, 197, 316, 345]. The characteristic timescale for monomer exchange is found in the order of $\sim 10^1$s for the **Mn** dimer. In comparison, monomer exchange from **Zn** fibre occurs on a characteristic timescale of $\sim 10^4$s, in agreement with the static nature of similar **Zn** porphyrin fibres reported recently[150]. These results indicate that exchanging a monomer with the solvent from a **Mn** dimer is three order of magnitude faster (or more probable) than exchanging a monomer out from the **Zn** fibre. The probability of finding free/disassembled **Mn** monomers in solution is much higher compared to that of finding disassembled **Zn** monomers. Additionally, once a **Mn** monomer is present in the solution, it is likely to bind onto a fibre tip, as the interaction between **Zn-Mn** is similar to that of **Mn-Mn**. Therefore, the chain-capping of the **Zn** fibres by the binding of **Mn** monomers is a likely event.

Co-assembly of Zn and Mn monomers

After examining the interactions between monomers, we proceeded to investigate the self-assembly of mixed species of **Zn** and **Mn** monomers in solutions. We simulated two different systems, (i) starting from free **Mn/Zn** monomers in solution that can freely interact and self-assemble or (ii) starting from pre-formed **Zn** fibres surrounded by free/disassembled **Mn** monomers (see Fig. 3.3). We compare the behaviours of both systems by means of CG-MD simulations in explicit solvent. Based on the results on the exchange dynamics of the respective aggregate types, we would expect both sequestration and chain-capping to occur for the first case, while for the second case chain-capping should be most common as the **Zn** fibres exchange monomers very slowly. After 5μs of CG-MD simulation time for the

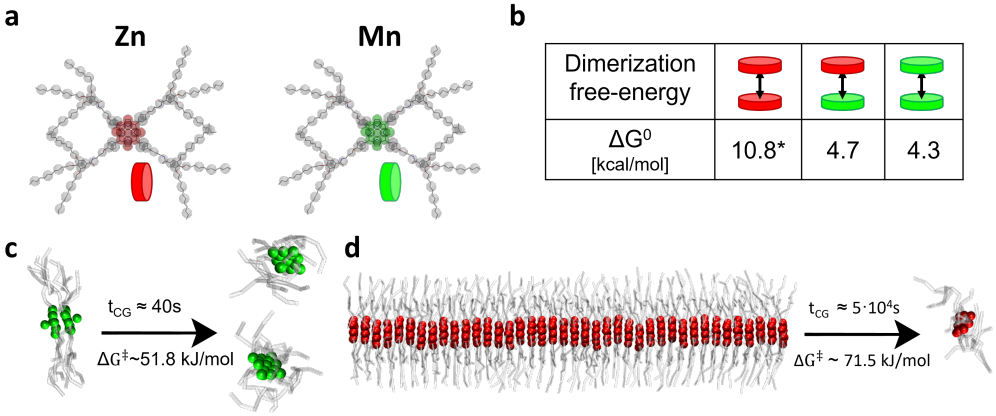

Fig. 3.2 Molecular modelling of porphyrin-based monomers. (a) All-atom (AA, full bonds) and coarse-grained (CG, transparent beads) models of **Zn** (left) and **Mn** (right); the cores of the monomers are represented by the coloured (red and green) disks; (b) Dimerization free-energies obtained from WT-MetaD simulations, in explicit CG MCH solvent; (c) and (d) The free-energy barrier and the CG-time of exchange from the dimer and the fibre tip obtained from infrequent WT-MetaD simulations. Adapted from Ref.[304].

first scenario, short aggregated species are observed as spontaneously appearing in the system (Fig. 3.3a, right panel): these are homo- or hetero-dimers of the two monomers, sandwich type **Mn-Zn-Mn** complexes or short, chain-capped stacks of **Zn**. The occurrence of sandwich-type species in the simulations supports the previously mentioned probability that also longer stacks of **S-Zn** could be capped by two **S/R-Mn** monomers.

For the second case, since the spontaneous chain-capping event could not be observed within the timescale accessible via classical/unbiased CG-MD simulations, for explorative purpose we used WT-MetaD to accelerate the sampling of the system. This allowed us to qualitatively observe that **Zn** chain-capping by **Mn** is indeed possible in the system. Interestingly, we could observe that over the course of this simulation the **Mn** monomer first starts to interact with the surface of the **Zn** stack (Fig. 3.3b, central panel). After some time, it finally binds and chain-caps the pre-existing supramolecular **Zn** fibre (Fig. 3.3b, right panel). Such a complex adsorbing-sliding-stacking mechanism is in line with what recently observed also for other types of supramolecular polymers[197, 345, 346]. We underline once again that this simulation was performed with an explorative purpose, aimed at offering preliminary insights into the possibility of **Zn** chain-capping by **Mn** monomers, since the convergence of such event with a single simulation cannot be guaranteed.

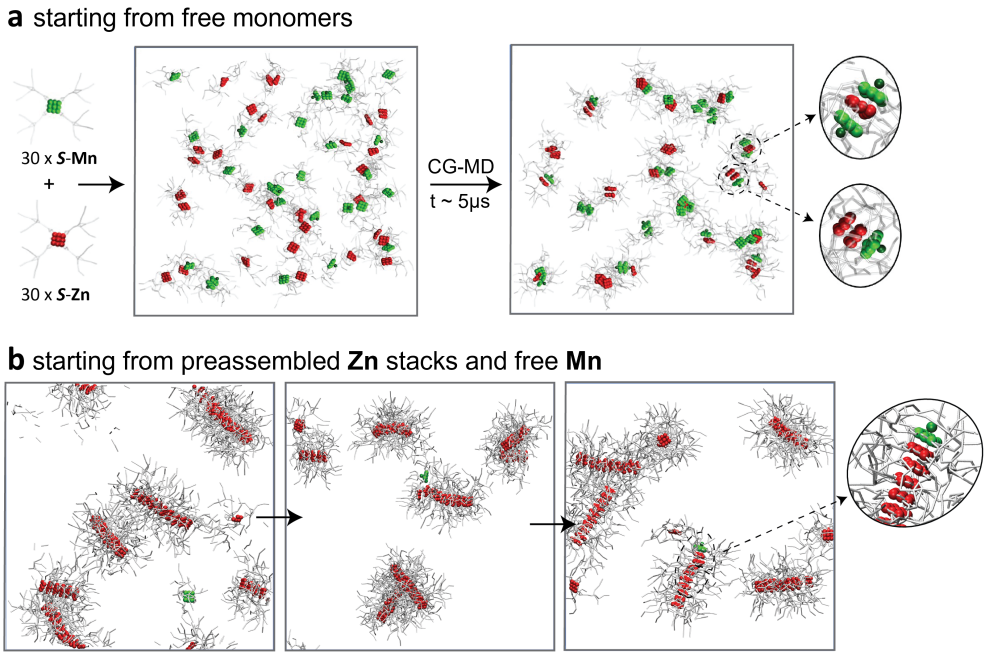

Fig. 3.3 Simulations starting from (a) free monomers or (b) preassembled Zn stacks and free Mn. The solvent is omitted for clarity. Adapted from Ref.[304].

Chain-capper release from Zn chain ends

To gain a deeper understanding of **Mn** monomers release from the chain ends, we performed WT-MetaD simulations on the exchange of a single **Mn** monomer from a **Zn** fibre. The binding and unbinding of a **Mn** monomer from the tip of a fibre composed of 30 **Zn** monomers were enhanced through the addition of Metadynamics potential, to efficiently sample the chain-capper release and to obtain insights into the most probable exchange pathways (Fig. 3.4a). Additional details about the WT-MetaD setup are provided in the Methods section of this Chapter. The FES obtained from WT-MetaD simulation and displayed in Fig. 3.4b shows a global minimum at **Mn-Zn** stacking distance ($\sim$0.5 nm in this CG model), and number of contacts between the **Mn-Zn** cores s equal to 5 (according to the switching function used to compute the number of contacts, 5 corresponds to the number of contacts between **Mn** monomer and **Zn** fibre in the perfect stacked configuration). State A identifies a system configuration where the **Mn** monomer is stacked onto the **Zn** fibre tip. A second broader local minimum (state B) is found at larger **Zn-Mn** core-core distance and reduced number of contacts ($\sim$10-40% of the stacked state A). State B

identifies the configurations where the **Mn** monomer is de-stacked from the fibre tip, and adsorbed on the arms of the **Zn** fibre. From state B, the **Mn** monomer can then jump to solution (state C).

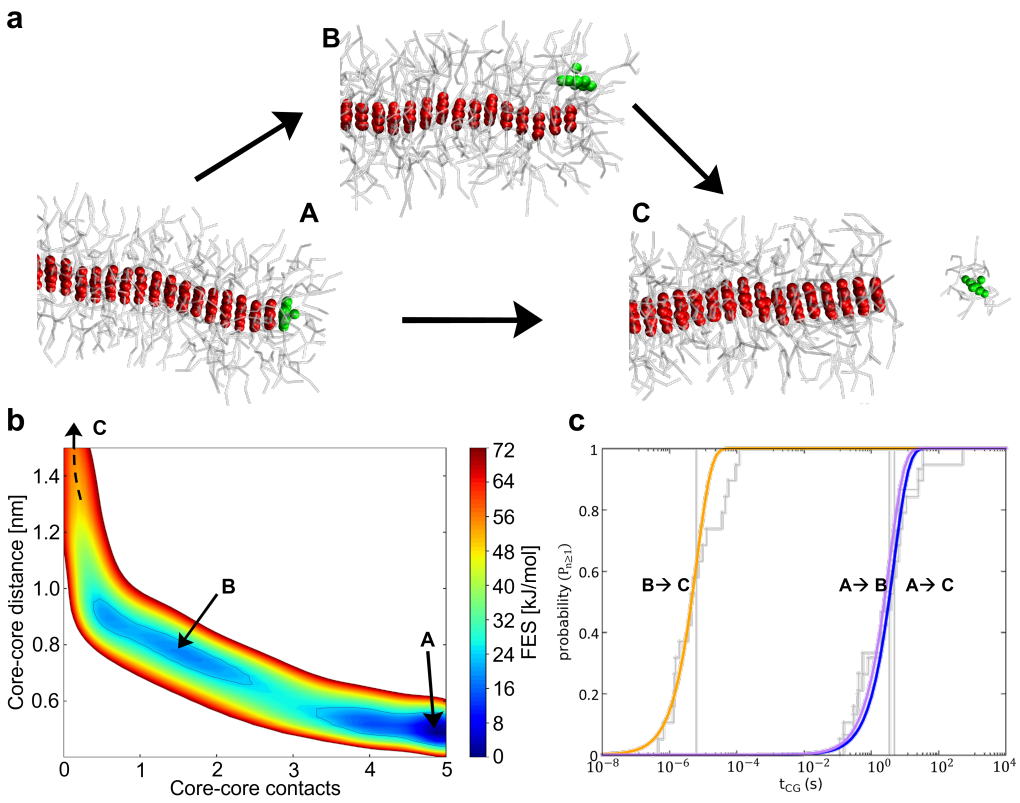

Fig. 3.4 (a) Two hypothesized mechanism of monomer-exchange: 1-step transition, from A to C, or 2-steps transitions, from A to B, and from B to C; (b) Free-energy surface (FES) obtained from WT-MetaD: the global minimum A corresponds to the stacked configuration; the wide cyan area corresponds to the progressive detachment of the **Mn** monomer from the tip; state C corresponds to the unbound-state; (c) Transition times distributions (gray vertical segments) and related Poissonian fits (coloured curves) allowing to estimate the characteristic timescales (coloured dots on the x axis) for the exchange transitions from A to B (violet), from B to C (orange), and from A to C (blue). Adapted from Ref.[304].

We then used multiple infrequent WT-MetaD simulations to kinetically characterize all the steps involved in the monomer exchange. In particular, we proposed two mechanisms of exchange: (i) 1-step transition, where the **Mn** monomer jumps from the fibre tip to the solution (A→ C), or (ii) a two-steps mechanism, where the **Mn** monomer first detaches itself from the tip and is adsorbed onto the lateral surface of **Zn** fibre (A→ B), and then jumps to solution (B→ C).

In the first case, the A→ C transition (Fig. 3.4c, blue line) time is found in the order of $\sim 10^0$s. This timescale is very similar to the one required to exchange a **Mn** monomer ($\sim 10^1$s), and consistent with the similar dimerization free-energies **Mn-Mn** and **Mn-Zn**.
In the second case, we observe that the A→ B transition is again very similar to the A→ C transition (Fig. 3.4c, purple line, $\sim 10^0$s). In contrast, B→ C transition is much faster (Fig. 3.4c, orange line, $\sim 10^{-6}$s). These findings convincingly suggest that breaking the strong directional interactions between the porphyrins core is the rate-limiting-step in exchanging monomers in solution, while jumping in solution is a much easier step. Conversely, in the chain-capping event, the **Mn** monomer first impacts on the lateral surface of the **Zn** fibre (most probable event, considering the aspect ratio of these fibres) and, after surfing along the surface, it eventually reaches the tip of the fibre.

3.3 Conclusions

In this study, we investigated the supramolecular polymerization and enantioselective chain-capping of porphyrin stacks in mixed solutions of **S-Zn** and **S/R-Mn**. The experimental results obtained from our collaborators showed that only **S-Mn** can act as a chain-capper for both H- and J-aggregates of **S-Zn**, while **R-Mn** only interacts with the achiral J-aggregates. The dilution-induced supramolecular polymerization was also observed and described.
The computational simulations, performed using the MARTINI force field-based coarse-grained molecular models, played a crucial role in complementing and deepening the understanding of the experimental results. The predictions of these models provided further insights into the supramolecular interactions between the monomers, helping us to quantify the relative stability of the different aggregate species and probe the mechanism of supramolecular polymerization. We could observe the underlaying dynamics of co-assembly and chain-capping events, revealing that multi-step processes (absorb-slide-stack) are likely involved in the exchange phenomena.
Overall, this study demonstrates the effectiveness of combining experimental and computational techniques in investigating supramolecular polymerization. The comprehensive view of the system obtained through the combination of experiments and simulations sheds light on the intricate supramolecular interactions and dynamics

and could have significant implications for the development of novel functional materials based on supramolecular interactions, following the aim of this Thesis.

3.4 Methods

3.4.1 Coarse-grained modelling of Zn and Mn

The coarse-grained (CG) model of **Zn** monomer core used here has been already published in a previous work The all-atom (AA) model of **Mn** monomer was built using General Amber Force Field (GAFF)[223] with RESP charges[224] computed at B3LYP level of theory with LANL2DZ/6-31+G** mixed basis set and used as a reference to tune the CG model of **Mn** porphyrin, based on the MARTINI force field[241] mapping. Due to the complexity of the systems and the fact that arms are identical between **Zn** and **Mn**, we focused on the study of the porphyrin cores, similarly to what has previously done in the literature with supramolecular polymers modelling[196, 316]. Two **Mn** porphyrin cores and a **Mn-Zn dimer** have been placed in two different pre-equilibrated simulation boxes of cyclohexane (CHX, equivalent to methylcyclohexane at CG level) and their dimerization free-energy have been estimated via Well-Tempered Metadynamics (WT-MetaD)[285]. The **Mn-Mn** interaction in solvent was found $\sim$18 kJ mol^{-1}, slightly lower than the interaction between **Zn** and **Mn** ($\sim$18.8 kJ mol^{-1}) and significantly lower than **Zn-Zn** ($\sim$45.2 kJ mol^{-1}, obtained from the literature[150]). Then, the non-bonded parameters of the **Mn** model were tuned in order to have a good agreement with the **Mn-Mn** dimerization energy obtained with AA simulations.

3.4.2 Simulation details

All the simulations have been performed with GROMACS[295] version 2018.6, patched with the PLUMED plugin version 2.5.0[296, 297] and conducted at 300K. In AA production runs we used leap-frog integrator with a time step of 2 fs, the v-rescale thermostat[261] with a coupling constant of 0.2 ps and Parrinello-Rahman barostat[262] with a coupling constant of 4 ps. Non-bonded interactions were treated with a cutoff distance of 0.9 nm. Long-range electrostatic interactions were evaluated using the Particle Mesh Ewald (PME) method[347]. In CG production runs we used

the leap-frog integrator with a time step of 10 fs, the v-rescale thermostat[261] with a time constant of 1 ps and Berendsen barostat[257] with a time constant of 4 ps. All the AA and CG WT-MetaD simulations for the estimation of dimerization free energies have been performed using the distance between metal atoms of the two monomers as collective variable. We used a bias factor of 25, a gaussian height of 0.4 kJ mol^{-1} and a deposition rate of 1 gaussian every ps and 5 ps in AA and CG, respectively.

3.4.3 Coarse-grained molecular modelling of Zn and Mn monomer exchange

The exchange of monomers out from an assembly (fibre or dimer) is a rare event at the timescale accessible to classical AA and CG simulations. We thus employed infrequent WT-MetaD[289] CG simulations activating the monomer exchange in the two systems. First, we built a **Mn** dimer and a **Zn** fibre composed of 40 monomers, and we equilibrated them in solvent. To obtain a transition probability distribution, we run 20 infrequent WT-MetaD simulations and fitted the empirical distribution with a Poisson distribution[290] (see Chapter 2, section 2.6.2). All the simulations were performed using two collective variables (CVs):

1. **Mn** dimer: the distance between the central beads of manganese (CV1) and the number of contacts between the two monomers (CV2).
2. **Zn** fibre: the distance between the central beads of zinc (CV1) and the number of contacts between the last monomer and the rest of the fibre (CV2).

The number of contacts s_{ij} was calculated using the following switching function:

$$s_{ij} = \frac{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^n}{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^m} \tag{3.1}$$

where r_0=0.3 nm, n=6, m=12, d_0=0.
We used a bias factor of 10, a gaussian height of 1 kJ mol^{-1} and a deposition rate of 1 gaussian every 200 ps. The obtained empirical distribution functions are displayed in Fig. 3.5. We have performed previous tests using either the number of contacts

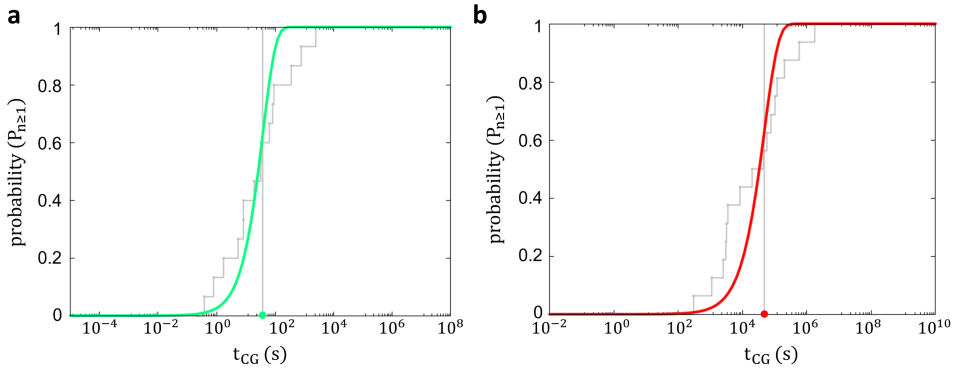

Fig. 3.5 Transition (monomer exchanges) times estimated from multiple infrequent WT-MetaD simulations (each transition time is identified by the vertical grey segments) and Poissonian probability distributions for the detachment of one **Mn** monomer from a **Mn** dimer (panel a) and of a **Zn** monomer out from a **Zn** fibre (panel b). From the Poissonian fits, it was possible to estimate the characteristic timescales for the monomer exchange events, identified by the coloured dots in the x-axis of the plots. Adapted from Ref.[304].

or the distance as the only collective variable to describe the unbinding event. We found that both variables give a similar order of magnitude for the unbinding rate, but the best fit with the Poissonian distribution is obtained using the combination of the two variables.

3.4.4 Co-assembly of Zn and Mn monomers

We performed two CG-MD simulations starting from different initial conformations, namely: 1) 30 **Mn** and 30 **Zn** free monomers, and 2) one free **Mn** monomer and six pre-stacked **Zn**-fibres consisting of 10 monomers each. The first system was simulated for 5μs. During the simulation, we observed the formation of short homo- or hetero-dimers, sandwich type **Mn-Zn-Mn** complexes or short, chain-capped stacks of **Zn**. In the second simulation, the spontaneous chain-capping event is not observable from the timescale explored by unbiased CG-MD simulations, thus we sped up the process employing WT-MetaD. This simulation was performed using two CVs: 1) the minimum distance between the **Mn** monomer and **Zn** monomers of the tip of the fibres, and 2) the minimum distance between **Mn** monomer and all the **Zn** monomers of the fibres. We used a bias factor of 50, a gaussian height of 1 kJ mol^{-1} and a deposition rate of 1 gaussian every 5 ps.

3.4.5 Chain-capper release from Zn chain ends

To clarify the interaction mechanism between **Mn**-based chain cappers and the **Zn**-fibres, we performed a CG WT-MetaD simulation. A **Mn** monomer was placed on the tip of a fibre formed by 30 **Zn** monomers, and pre-equilibrated. Two CVs were used in this CG WT-MetaD simulation: the distance between the central bead of the **Mn** monomer and the central bead of the last **Zn** monomer in the fibre (CV1), and the number of contacts between the core of the **Mn** monomer and the cores of **Zn** fibre (CV2). We used a bias factor of 25, a gaussian height of 1 kJ mol^{-1} and a deposition rate of 1 gaussian every 5 ps. Given the large conformational space that can be explored during the capping/decapping events, and the size of the system, we used a 1.5 nm cutoff for CV1.
To characterize all the exchange steps, we used multiple infrequent WT-MetaD simulations, using the same parameters of section 3.4.3.

Chapter 4

Multiscale molecular modelling of intertwining covalent organic nanotubes

This work has been carried out in collaboration with the experimental group of Prof. R. Banerjee, at IISER Kolkata (India). This section will focus on the computational aspects and the role of molecular modelling in supporting the experiments. The Introduction is largely taken from the postprint version of the article published on *Nature Chemistry*. The experimental results are briefly summarized at the beginning of the "Results" section. Details regarding the experimental techniques are available in the full published paper[348].

Full bibliographic reference: Koner, K., Karak, S., Kandambeth, S., Karak, S., Thomas, N., **Leanza, L.**, Perego, C., Pesce, L., Capelli, R., Moun, M., Bhakar, M., Ajithkumar, T. G., Pavan, G. M., Banerjee, R. *Nature Chemistry* 14, 507–514 (2022). DOI:10.1038/s41557-022-00908-1.[1]

In this Chapter, we extend our investigation to more complex hierarchical assemblies, based on covalent organic nanotubes. Unlike porphyrins, which assemble in linear supramolecular polymers – due to the highly directional interaction between porphyrins core – these covalent nanotubes interact with each other in solution via

[1]My contribution to this article, as the first computational author, has been the development of all the atomistic and coarse-grained models, performing all MD simulations and analysis and contributing to the interpretation of the results and in the writing of the manuscript.

non-covalent forces to form more complex superassemblies with intricate shapes and structures. This behaviour is similar to how two single long strands made of covalently bonded nucleotides interact with each other to form the characteristic double helix structure of DNA, which can further fold to form 2- and 3-D shapes (*e.g.* DNA origami).

Carbon nanotubes (CNTs) and synthetic organic nanotubes have demonstrated great potential in a range of applications, such as electronic devices, energy storage, catalysis, and biosensors. In this context, the exceptional properties exhibited by covalent organic nanotubes (CONTs) have led to an increasing interest for their use in nanomaterials research. The strong covalent bonds formed between carbon, nitrogen, and oxygen atoms result in the CONTs exhibiting high thermal and chemical stability. In this work, we focus on new covalently bonded porous organic nanotubes (CONTs), synthesized by Schiff base reaction. Morphological characterization of CONTs showed that, upon ultrasonication, they form intertwined structures that subsequently coil and form toroidal superstructures. Using multiscale molecular modelling and enhanced sampling methods, we investigated the impact of the solvent on the formation of intertwined assemblies and the strength of interaction under different conditions. Our computational results provide valuable insights into the role of solvophobic interactions in the formation of covalent organic superstructures, which can aid in the design of novel nanomaterials for various applications.

4.1 Introduction

Covalent bonds are well known for their strength, directionality and versatility[349]. These bonds impart exceptional thermal and mechanical properties to extended solids, such as diamond and silicon carbides[350], and their unique properties have been harnessed in various technological applications. Additionally, the diverse nature of covalent bonding enables the formation of a vast array of organic, bioorganic molecules, and polymers. Through dynamic covalent chemistry a wide variety of organic cages (zero-dimensional) and two- and three-dimensional covalent organic frameworks have been synthesized[351–353].
Carbon nanotubes and synthetic organic nanotubes in general are one-dimensional hollow tubular nanostructure attractive for applications in electronic devices, energy storage, catalysis, membrane separation and biosensors[354–356]. CNTs are

generally synthesized by rolling two-dimensional graphite sheets along their edges by various methods, including arc discharge, electrolysis, chemical vapour deposition, plasma torch and hydrothermal techniques[357–362]. These methods demand harsh reaction conditions and high temperatures. Furthermore, the incorporation of predesigned functionalities is difficult due to the insolubility of CNTs in common organic solvents.

Covalent organic nanotubes (CONTs) represent a promising class of one-dimensional structures that offer a solution to some of the challenges posed by CNTs. CONTs are organic nanotubes that are formed through the covalent bonding of organic building blocks that can be synthesized under milder conditions and are soluble in common organic solvents, enabling their functionalization and incorporation into various applications[84, 363, 364]. The strong covalent bonds formed between carbon, nitrogen, and oxygen atoms make CONTs highly stable, and impart unique mechanical, thermal and chemical properties to the materials.

One of the main challenges during bottom-up synthesis of covalent organic nanotube is controlling both the structure and morphology of self-assembled nanostructures. This is because even slight variations in the composition of building blocks can affect the nanoscopic assembly of the system, resulting in changes to the overall morphology of the structures. As a result, it can be challenging to systematically tune the size, shape, or porosity of the resulting structures[365, 366].

Molecular modelling can be a valuable tool in overcoming this challenge. It can elucidate the mechanism of assembly, the mechanism of interaction, and aid in the design of new structures. By simulating the behaviour of individual molecules, molecular modelling have been widely employed to obtain insights into the global properties and behaviour of organic materials[367–371]. Moreover, molecular modelling has shown great promise in predicting how different building blocks can be combined to produce porous organic nanostructures with specific properties, such as cages[372] and metal organic frameworks[373].

As discussed in Chapter 1, the extensive application of molecular simulations in this field is currently limited by their high computational costs, which is due to the large length and time scales involved. As was the case in Chapter 3, one possible approach to address this issue is to utilize coarse-grained (CG) models, as they can significantly reduce computational costs while still providing accurate results[374, 375].

Our collaborators introduced in this study a new design strategy to synthesize purely covalent porous organic nanotubes, using organic building block designed to

assemble into one-dimensional CONTs through dynamic covalent chemistry. The synthesized CONTs display remarkable chemical and thermal stability, due to the high strength and stability of covalent bonds. Furthermore, a detailed investigation of the morphological evolution of CONTs revealed that the isolated tubular morphologies go on to form intertwined structures, which then further aggregate to form toroidal superassemblies.
Through the use of multiscale molecular modelling approach combined with enhanced sampling techniques, we aimed to gain insights into the mechanism of intertwining of nanotubes, the complex interaction between isolated nanotubes in solution, and the effect of different environmental condition on their behaviour. Our simulations provided a detailed understanding of the interplay between intermolecular forces and geometrical constraints, which govern the self-assembly of the nanotubes and the formation of the intertwined structures.

4.2 Results and Discussion

4.2.1 Experimental section

In the present section, I will report the main experimental results, obtained by the group of Prof. R. Banerjee, regarding the structural characterization and morphology of CONTs to provide context for the theoretical calculations. More details about the design, synthesis and characterization of CONTs (CONT-1 and CONT-2) are available in the reference paper[348].

Morphology of nanotubes

Scanning electron microscopy (SEM) images of CONTs reveal uniform tubular morphology, with an average diameter of $\sim$5 nm. High-resolution transmission electron microscopy (HRTEM) and AFM identify the hollow tubular nature at the interior with a constant diameter of $\sim$5 nm throughout the entire length of the nanotube. Additionally, electron microscopy images show that the single-walled CONTs are intertwined, which might be due to their high length-to-width ratio (average, $\sim$300:1). The microscopy analysis further confirms that the individual units of the

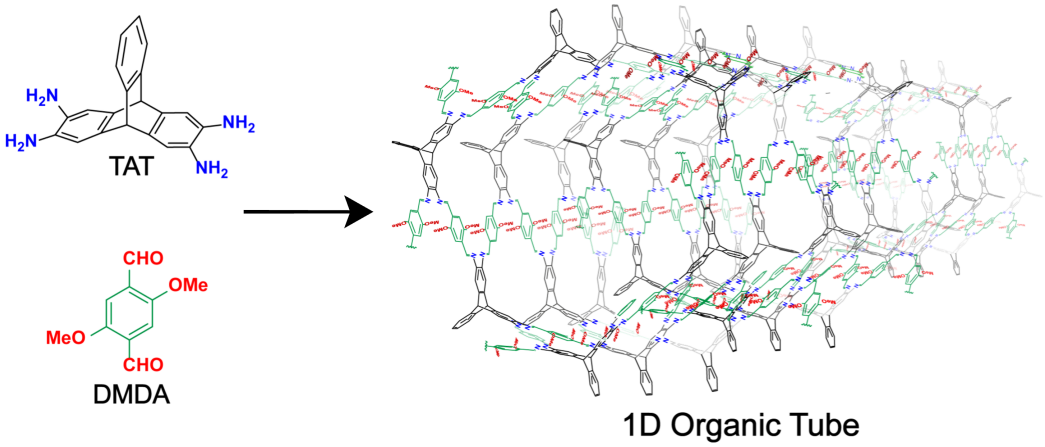

Fig. 4.1 Structure of the porous CONTs synthesized from TAT and from the linear dialdehydes DMDA. Adapted from Ref.[348].

intertwined nanotubes have diameters very similar (∼5 nm) to those of isolated single nanotubes. CONT-1 formation starts within 6h of reaction. These CONTs are 100-200 nm in length and ∼5 nm in diameter. After 12h, the CONTs grow up to 500-800 nm length with a diameter of ∼5 nm (average length-to-diameter ratio, 130:1). The intertwining starts after 24-30h when the length-to-diameter ratio of the CONTs increases substantially (average, ∼200:1).

The rapid increase in the length of CONTs induces high flexibility, promoting intertwined structures. After 36h of reaction, almost all CONTs become completely intertwined, and no notable morphological changes are observed. After 6h, two nanotubes uniformly intertwine, following a particular pattern where the twining pitch is 70±10 nm.

The nanotubes retain their morphology in a broad range of solvents of various polarities, as nitrogen adsorption isotherm confirmed. However, the width of intertwined CONT-1 varied with the nature of the solvents due to the solvent environment apparently affecting the interaction among the CONTs.

Interestingly, the intertwined CONTs further self-assemble to form a toroidal superstructure upon ultrasonication. THF is the best solvent to obtain the toroids in high yields, but their formation proceeds also in other solvents, such as xylene or dichlorobenzene. The toroid morphology is confirmed by combination of AFM, SEM and TEM. The toroid formation will not be investigated by molecular modelling in this study.

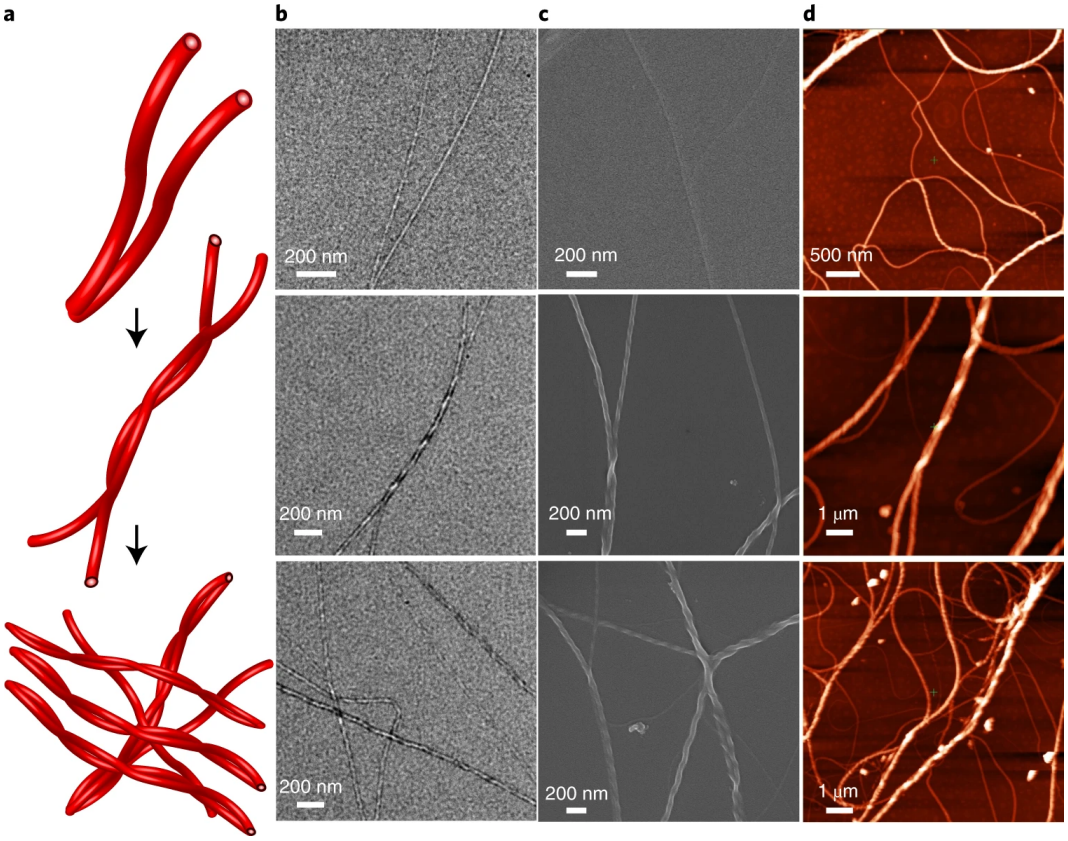

Fig. 4.2 (a) Graphical representation of the increasing intertwining of the nanotubes, from top to middle to bottom. (b)-(d) Characterization by HRTEM (b), SEM (c) and AFM (d) of each of the situations represented in a). In d) two flexible CONTs interconnect first at a single point (top); this interconnection leads to the formation of intertwined structures with a characteristic average pitch (middle); this in turn generates the assemblies shown in the bottom panel. Adapted from Ref.[348].

4.2.2 Computational section

To obtain a deeper insight into the molecular factors that drive CONT-1 intertwining in different solvents, we used multiscale molecular models to simulate the CONTs in different solvent conditions. We developed an all-atom (AA) model of CONT-1 composed of 16 layers (Fig. 4.3)

This AA model was simulated in explicit DCM, in THF, in water and in gas phase (that is, in the absence of solvent), at 20°C. The final snapshots of each simulation are displayed in Fig. 4.4, panel b. To quantitively estimate changes in the global shape of nanotubes we defined two characteristic angles: angle 1 and angle 2, which are displayed in Fig. 4.4a. The distributions of these two angles, as depicted in Fig.

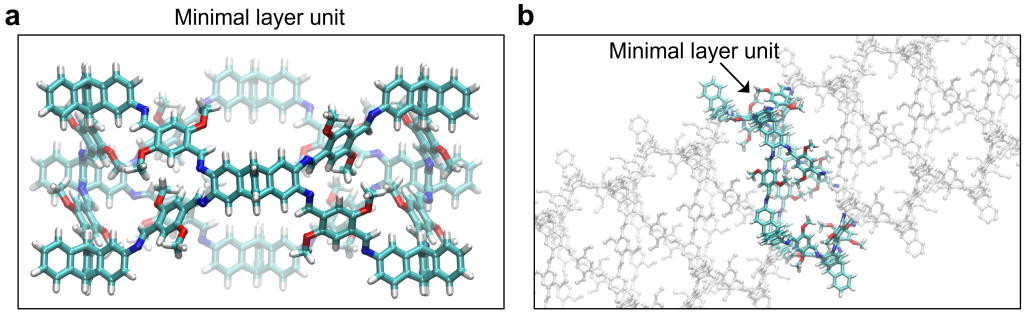

Fig. 4.3 (a) Atomistic molecular structures of the minimal layer unit; (b) the minimal unit (highlighted) is then replicated along the z-direction to form a 16-layer long nanotube used for the MD simulations. Adapted from Ref.[348].

4.4c and also clearly visible from the final snapshots in panel b, reveal that in DCM and THF the CONT-1 equilibrates to configurations that slightly deviate from the initial perfect one (Fig. 4.4b): angle 1 shifts from ∼106.1° (reference value in the ideal configuration) to ∼107.1° (in DCM) and ∼106.7° (in THF), while angle 2 changes from ∼45.3° (perfect tube) to ∼41.9° (DCM) and ∼43.7°(THF). On the contrary, in water, the tubules tend to collapse along the longitudinal axis due to strong solvophobic effects, causing angle 1 to change to ∼115.6° and angle 2 to ∼24.8°. A similar structural compression is also observed in the gas phase, with angle 1 shifting to ∼114.5° and angle 2 to ∼27.1°. In all cases, the diameter of the AA CONT models remains compatible with that estimated experimentally. We then used these AA models as a guideline to develop a minimalistic CG model that, while more approximated, allowed us to study the behaviour and interactions between the CONTs on a larger scale. In this CG model, each TAT unit in the CONT-1 structure is represented by a single CG particle, interconnected with the other neighbour TAT particles via harmonic bonds (Fig. 4.5). The CG particles interact with each other via a simple Lennard–Jones potential. The parameters of this CG model were initially optimized to obtain a behaviour consistent with that of the AA CONT-1 model in explicit DCM solvent. This allowed us to simulate with reasonable accuracy the behaviour of long CONT models composed of 500 TAT layers (as shown in Fig. 4.6a, left, with a tubule length of approximately 820 nm). We started by creating a system configuration with two parallel tubes, and then performed CG-MD simulations by varying the depth of the LJ potential, ε, acting between the CG beads. The ε value determines the strength of the interaction between the CG particles of the CONTs, allowing us to effectively simulate changes in the solvent conditions

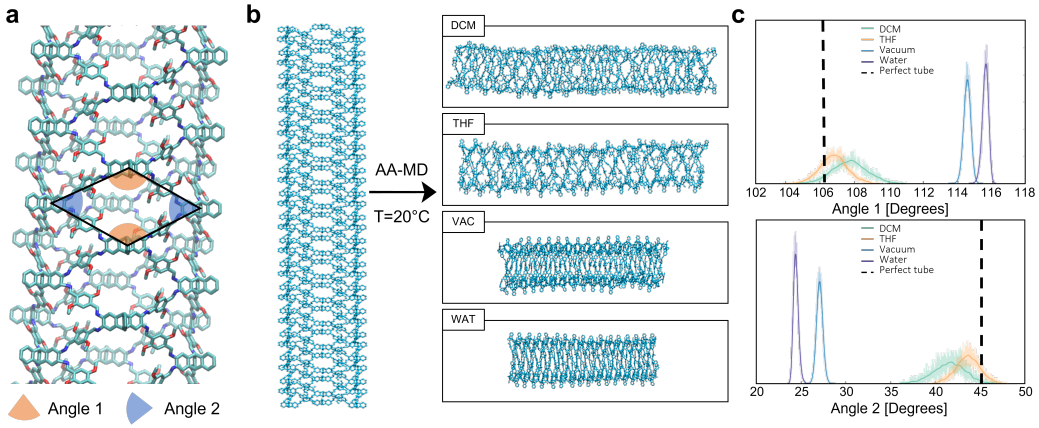

Fig. 4.4 (a) The two characteristic angles (angle 1 and angle 2), formed by the TAT moieties in the CONT structure are highlighted in orange and blue; (b) MD snapshots of the AA model of the initially perfect CONT-1 (left) and the equilibrated CONT-1 structures after 200 ns of AA-MD in DCM, THF, vacuum and water; (c) Distributions of the angles 1 and 2 in the structure of the equilibrated CONT-1 computed along the AA-MD in the different environments. Vertical black-dashed lines indicate the values of angle 1 and angle 2 in the initially perfect configuration of the tube. In water and vacuum the nanotube undergoes significant shrinkage, adopting a more compact configuration. Adapted from Ref.[348].

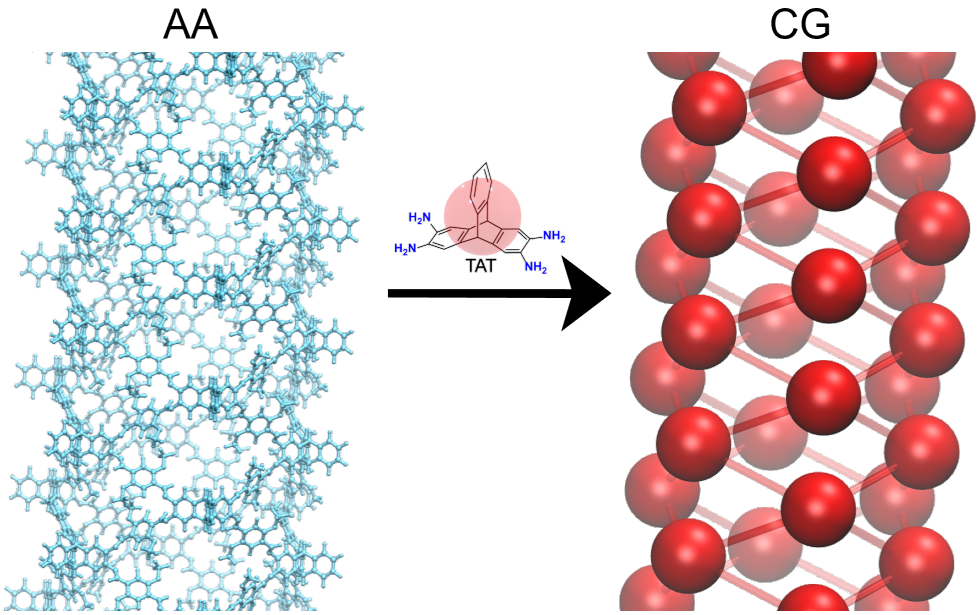

Fig. 4.5 Mapping of the AA structure of the model into the CG CONT model: each TAT unit is replaced by a CG bead. Adapted from Ref.[348].

by adjusting the nanotube-nanotube interaction. For ε values < 1 kJ mol^{-1} the two CONTs interacted only weakly and intermittently, and no intertwining was observed

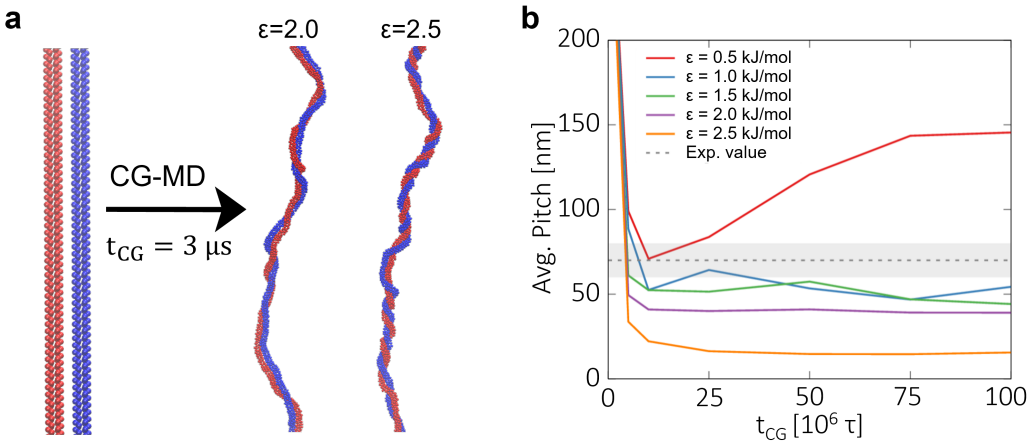

Fig. 4.6 Spontaneous intertwining of two CONTs during CG-MD simulations. (a) Starting from initially parallel CONTs (left), different average intertwining pitches are obtained during the CG-MD as a function of the interaction strength ε between the CG beads (the simulation time is expressed in CG-MD integration time steps unit, τ); (b) Average intertwining pitches as a function of CG-MD simulation time, measured for different values of ε. For ε=1 and ε=2, the nanotubes exhibit a similar behaviour to the nanotubes characterized experimentally. Adapted from Ref.[348].

during the CG-MD (weak solvophobicity). Instead, for $\varepsilon \geq 1$ kJ mol^{-1} we observed persistent interactions and intertwining of the two CONTs. In these cases (Fig. 4.6b), the CG-MD showed an average intertwining pitch consistent with that observed experimentally ($\sim 70 \pm 10$ nm). However, at $\varepsilon > 2.5$ kJ mol^{-1}, the formation of well-defined helices becomes less favoured, and the CONTs tend to interact further, generating tighter and less-defined hierarchical assemblies.

The interaction between two CONTs was qualitatively estimated through the use of Umbrella Sampling (US)[275] calculations. In these simulations, we utilized two shorter CONT-1 models, each composed of 6 TAT ring layers, placed parallel to each other in the simulation box. In every window of the US calculation, restraining potentials were utilized to maintain the parallel configuration of the CONTs, enabling interaction only through their lateral surfaces and preventing rotational movements. The resulting trajectories were combined through the weighted histogram analysis method (WHAM) to obtain the potential-of-mean-force profile (see, for example, the profile obtained in explicit DCM solvent in Fig. 4.7a.

Our findings support the idea that the interaction between adjacent nanotubes is primarily driven by solvophobic effects. As solvophobicity increases, the free-energy difference between the bound and unbound states also increases (as shown in Fig.

4.7b). The US calculations were repeated at the CG level, while varying ε values. Interestingly, the results indicate that values of 2 and 2.5 kJ mol^{-1} provided a CONT-CONT interaction that is consistent with the results obtained from the AA models in explicit DCM and THF solvents, respectively. In addition, the results highlight that the interactions between the CONTs in water (where solvophobic effects are stronger) or in the gas phase are compatible with higher ε values in the CG representation.

Although these analyses have a qualitative nature, as they refer to the ideal lateral interaction between reduced tubule portions, they provide valuable insights into the relative local interaction strength between the tubules in various solvents.

We calculated the CONT-CONT interaction energy, and decomposed it to electrostatic and van der Waals components in AA systems.

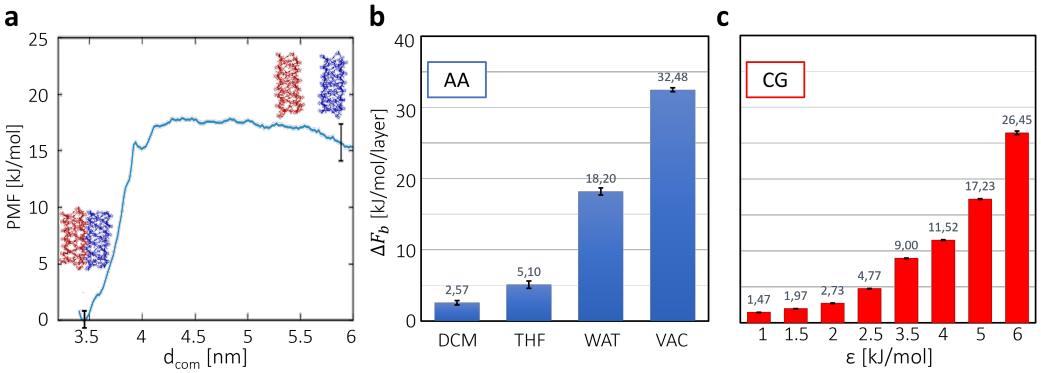

Fig. 4.7 (a) Potential of mean force (PMF) profile computed via Umbrella Sampling (US) for two AA CONTs (shown in the insets) immersed in DCM; (b) Lateral interaction free energies for the AA systems in different solvent conditions. The values are indicated per layer units; (c) Free energies of lateral binding for the CG models at different ε values, computed via US. For ε=1.5-2, the behaviour of CG models is similar to that of nanotubes immersed in DCM. For ε=2.5 the CG behaviour is compatible with the one observed in THF solvent. The effect in water and vacuum is modelled by using higher ε values (ε>5). Adapted from Ref.[348].

4.3 Conclusions

In this study, we aimed to investigate the mechanism of nanotube intertwining and the interaction between isolated nanotubes in solution, in order to complement the

experimental findings. Our multiscale molecular modelling approach, combined with enhanced sampling techniques, allowed us to gain valuable insights into the process. Specifically, our simulations revealed that the different behaviour of isolated nanotubes in solution is controlled by solvophobic effects, which can be attributed to the hydrophobicity of the building blocks. Solvophobic effects are also the primary driving forces for the intertwining of nanotubes. Through the use of CG models, we were able to simulate the large-scale folding of nanotubes and effectively reproduce the experimental conditions. Our results suggest that the model, despite its approximations, exhibits a good qualitative agreement with the experimental observations, and could provide some insights into the behavior of the system under different environmental conditions. Future research directions could include the study of toroid formation from intertwined structures, that has been observed experimentally. This could further elucidate the pivotal role of solvophobic interactions in the formation of complex covalent organic superstructures.

4.4 Methods

Computational setup

All simulations were performed with the GROMACS 2018.6 software[295] equipped with PLUMED 2.5.0[296, 297] for the US simulations and systems analysis.
The AA CONT-1 model was parameterized based on the General Amber Force-Field (GAFF)[223] using the Antechamber software[376]. The partial atomic charges for the tubule were computed using the Restrained Electrostatic Potential (RESP) method[224] on an elementary TAT ring layer (composed of 6 TAT and 12 DMDA units, which, replicated along the main tubule axis, generates the complete tubule structure) after geometry optimization performed using PM6 and Hartree-Fock (with the 6-31g* basis set) levels of theory and the Gaussian 16 software[300].
The AA tubule models used in the AA-MD simulations of Fig. 4.4 are composed of 16 TAT layers (plus 3 extra TAT units to saturate the tubule structure), corresponding to a tubule length of ~ 21 nm. This AA CONT-1 model was simulated in different environment conditions: in the absence of solvent (vacuum), in water, in DCM, or THF solvents. In all these AA-MD simulations, the nanotube was placed in a simulation box with periodic boundary conditions. To prevent the CONT interaction with its periodic images via tubule rotation, we imposed external wall potentials

acting on the centres of mass of the tube ends (terminal TAT ring layers) that restrained their position to the initial ones within a cylinder having a 0.2 nm radius on the XY planes. In this way, the AA tubule model is globally free to relax, and it is free to compress or enlarge along the z-direction, but it cannot rotate perpendicular to the major axis during AA-MD. Preliminary energy minimization of the systems was conducted via steepest-descent and short equilibration cycles with a 0.2 fs time step. The equilibration AA-MD simulations were 200 ns long for each system and were performed using a time step of 2 fs. The temperature was maintained at 20°C and the pressure at 1 atm by means of the v-rescale thermostat[261] and Berendsen barostat[257] using isotropic pressure scaling. The organic solvent molecules DCM and THF used in the AA-MD simulations were also parametrized compatibly with the GAFF force field, while simulations in water used the TIP3P water model[377].

Umbrella sampling[275] calculations were employed to qualitatively estimate the local energy of lateral interaction between two CONTs. For these simulations, two shorter CONT-1 models were used, each composed of 6 TAT ring layers positioned parallel to each other in a simulation box. For these analyses, AA-MD simulations were conducted in periodic boundary conditions, with the tubes restrained at a fixed distance by means of a harmonic potential acting on the distance between the two CONT-1 centres of mass. This was repeated for 66 runs, each with a different equilibrium distance between the tubes, ranging from 3.5 to 6.5 nm. In these runs, the CONTs were kept in the parallel configuration by using restraining potentials that maintained the centre of the nanotube ends within a cylinder of 0.9 nm of radius, in practice allowing the tubes to interact only via their lateral surfaces and preventing their rotation. The force constant of the harmonic potential restraining the distance between the tubes was 10^4 kJ mol^{-1}nm $^{-2}$. The resulting trajectories were combined through the WHAM[278] to obtain the PMF profile. From there, we extracted the tubule-tubule interaction energies that are shown in Fig. 4.7. The errors were computed using the procedure of Ref.[284].

Coarse-grained modelling

The centres of mass of the TAT molecules in the CONT structure were mapped to CG particles for building a minimalistic CG CONT-1 model, as illustrated in Fig. 4.5. Each CG-bead is connected to 4 other beads, consistently with the geometry of

the atomistic CONTs, by means of harmonic potentials of the form:

$$V_b = \frac{k_b}{2}(r - r_0)^2 \tag{4.1}$$

where r is the distance between two neighbouring CG-beads, k_b is the force coefficient, and r_0 is the equilibrium distance of the bond. Angular potentials were also applied to reproduce the structure and flexibility of the nanotubes. Two angles, angle 1 and angle 2 (Fig. 4.4a) were restrained by potentials of the following form:

$$V_{a,i} = \frac{k_{a,i}}{2}(cos\theta - cos\theta_{0,i})^2 \tag{4.2}$$

Where i indexes the angle type (1 or 2), and $k_{a,i}$ and $\theta_{0,i}$ are the force coefficient and equilibrium value associated with the i-th angle.
Non-neighbouring CG-beads interact with each other utilizing Lennard-Jones potentials that implement the steric hindrance, mutual attraction, and hydrophobic effects driving the interaction between the CONTs.

$$V_{LJ} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{4.3}$$

We distinguish two kinds of non-bonded interactions described by the equation above, those between beads belonging to different CONTs, with parameters ε and σ, that regulate the interaction between the tubes, and those between beads belonging to the same CONT, with parameters ε' and σ', that regulate the self-interactions within the tube. In all the CG models, the Lennard-Jones potential was truncated and shifted to 0 at a distance of r_{cut}= 3.54 nm.
The solvent is represented implicitly via the noise term in Langevin dynamics and the interaction between the assembly beads, which also encode the solvophilic/solvophobic behaviour of the nanotubes.
The different parameters of the potential terms were optimized to reproduce the behaviour of the AA model in DCM solvent. First, we tuned the bonded potential CG parameters (bond, angles) by comparing the AA-MD of a single, equilibrated 16-layer-CONT with that of an equally sized CG-CONT in order to tune the interactions of a single tube in solution. The length of these CG-MD runs was of 10^7 integration steps (τ).
To this end, we compared the AA and CG distributions of the bond distance and angles obtained and set the force constants and equilibrium values accordingly. (Fig.

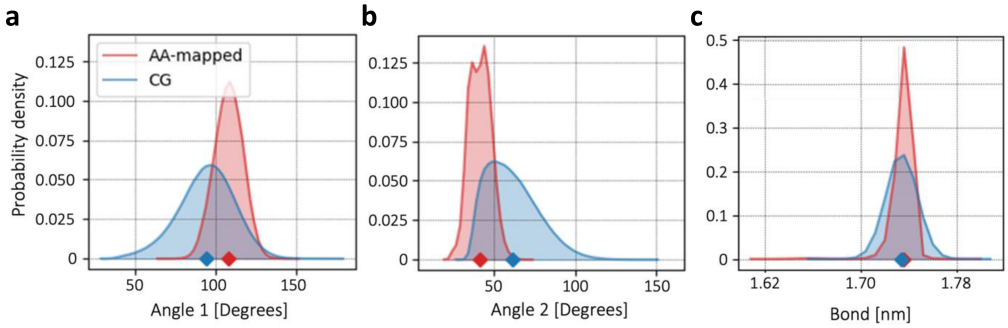

Fig. 4.8 (a) Distribution of angle 1 formed by the centres of mass of neighbouring TAT moieties (red area) and by the corresponding CG pseudo atoms in the CG model (blue area). (b) Distribution of angle 2 formed by the centres of mass of neighbouring TAT moieties (red area) and by the corresponding CG pseudo atoms in the CG model (blue area). Angle 1 and Angle 2 are showed in Fig. 4.4. (c) Distribution of the bond, namely the distance between the centres of mass of neighbouring TAT moieties (red area) and by the corresponding CG pseudo atoms in the CG model (blue area). Adapted from Ref.[348].

4.8) shows the comparison of the distributions obtained with AA-MD and CG-MD with the selected parametrization values, reported in Table 4.1. It is worth noting that in the chosen parametrization, the potential acting on angle 2 is inactive ($k_{a,2} = 0$, because angle 1 potential, together with the intra-tube excluded volume interactions, already enforce the distribution of angle 2 to a satisfactory agreement with the AA one. Due to the vast difference in resolution between the AA and CG descriptions, we looked for the best possible qualitative agreement, as the fine details of the CONT geometry, accessible with the AA description, are necessarily out-of-reach with the minimalistic CG model. In this phase, we also set the parameters ε' and σ', reported in Table 4.1. For monitoring and tuning the angular and bond distributions of the beads, we have used the analysis tools available in the *Swarm-CG* software[236].
We then proceeded tuning non-bonded interaction parameters ε and σ. The length scale σ was maintained at the fixed value of 1.51 nm, in order to avoid inter-penetration of the interacting CONTs. For the ε values, we repeated the US calculations systematically varying the ε values in the range between 1 and 6 kJ mol^{-1}. All parameters used in the CG model are reported in Table 4.1.

We performed the simulation of two long (500 TAT ring layers, corresponding to a tubule length of $\sim$820 nm) CONTs interacting with variable strength ε. Initial configurations containing two parallel tubes at a distance of 6.4 nm between each other were generated. After preliminary minimization via steepest-descent and short

V_b		$V_{a,i}$				V_{LJ}			
k_b $\left[\frac{kJ}{molnm^2}\right]$	r_0 [nm]	$k_{a,1}$ $\left[\frac{kJ}{mol}\right]$	$\theta_{0,1}$ [°]	$k_{a,2}$ $\left[\frac{kJ}{mol}\right]$	$\theta_{0,2}$ [°]	σ [nm]	ε $\left[\frac{kJ}{mol}\right]$	σ' [nm]	ε' $\left[\frac{kJ}{mol}\right]$
967.2	1.74	10.0	106.1	0	45.3	1.51	1.0-6.0	1.25	0.2

Table 4.1 Parameters of CG model of CONT-1

equilibration cycles with a 0.1τ integration step, each of these systems simulated for 10^8 steps of CG-MD simulation conducted in periodic boundary NVT (constant N: number of particles, V: box volume, and T: temperature during the run) conditions. All CG-MD simulations used a leapfrog stochastic dynamics integrator, with an inverse friction constant of $\tau_t = 25\tau$, regulating the solvent viscosity and thermalization of the system at T=20°C.

The umbrella sampling calculations to compare the tube-tube lateral interactions with the CG case were conducted mirroring the AA model's procedure. In particular, two CONT CG portions composed of 6 TAT ring layers each (same length of the AA analysis) were placed in a parallel configuration and maintained at a fixed distance via harmonic potential restraining the distance between the two centres of masses of the tubes. This was repeated for 40 runs, each with a different equilibrium distance between the tubes, ranging from 2.86 nm to 9.0 nm. Also, in this case, the CG tubules portions were allowed to approach each other only in parallel fashion and were prevented from rotating by using restraining potentials restraining the centres of mass of both ends within a cylinder of 2.89 nm. The force constant of the harmonic potential restraining the distance between the tubes was 967.2 kJ $mol^{-1}nm^{-2}$. The resulting trajectories were then combined using the weighted histogram analysis method to obtain the PMF profiles.

Chapter 5

Exploring the interaction of benzene-1,3,5-tricarboxamide supramolecular polymers with proteins

This work has been carried out in collaboration with the experimental group of Prof. E. W. Meijer, at the Eindhoven University of Technology (Netherlands). This section will focus on the computational aspects and the role of molecular modelling in supporting the experiments. The Introduction is largely taken from the postprint version of the article published on *Biomacromolecules*. The experimental results are briefly summarized at the beginning of the "Results" section. Details regarding the experimental techniques are available in the full published paper[378].

Full bibliographic reference: Varela-Aramburu, S., Morgese, G., Su, L., Schoenmakers, S. M. C., Perrone, M., **Leanza, L.**, Perego, C., Pavan, G. M., Palmans, A. R. A., Meijer, E. W. *Biomacromolecules* 21(10), 4105-4115 (2020).
DOI:10.1021/acs.biomac.0c00904[1]

After investigating the nature of interactions in supramolecular polymers – porphyrin-based and organic superassemblies – in this Chapter we will shift our focus

[1]In this article I equally contributed with M. Perrone in the development of the atomistic and coarse-grained models, performing all MD simulations and analysis.

to another crucial aspect of supramolecular polymers: their potential as biomaterials and their interactions with biological structures under physiological conditions.

The ability to mimic the fast dynamics that occur in natural processes is crucial for creating biomaterials. Synthetic biomaterials, particularly water-soluble supramolecular polymers, have demonstrated great potential in emulating the dynamic behaviour of natural systems. In particular, benzene-1,3,5-tricaboxamide (BTA)-based supramolecular polymers have shown to be highly dynamic due to the exchange of monomers within and between fibres. However, their suitability as biomaterials has not been fully explored. Experimental investigations assessed the interactions of BTA supramolecular polymers bearing either tetraethylene glycol or mannose units at the periphery with different biological entities. When BTA fibres were incubated with bovine serum albumin (BSA), the protein conformation was only affected by the fibres containing tetraethylene glycol at the periphery (BTA-OEG_4). Herein, we conducted coarse-grained molecular simulations to gain insights into the molecular interactions between BTA-OEG_4 and BSA. Our results complemented experimental evidences revealing that BSA interacted with BTA-OEG_4 fibres rather than with BTA-OEG_4 monomers present in solution or that may exchange out of the fibres. Our findings provide new insights into the potential of BTA-OEG_4 as a viable biomaterial, which could have various applications in drug delivery and tissue engineering.

5.1 Introduction

Although the field of biomaterials has made significant advances in recent years, mimicking nature remains an arduous task. This is primarily due to the highly dynamic behaviour of natural systems, which continuously evolve and adapt through non-covalent recognition motifs. In this context, there has been a significant effort from supramolecular chemists to replicate nature's unique dynamics using biomaterials assembled through non-covalent interactions. In particular, water-compatible supramolecular polymers have been thoroughly investigated as the building blocks for creating dynamic biomaterials[379, 380].
Supramolecular polymers offer a wide range of possibilities for the fabrication of one-dimensional (1D) fibres in water and buffers[43], making them ideal for mimicking the fibrillary components found in natural systems[381]. Host-guest interactions

have been utilized to create 1D supramolecular polymers through their hydrophobic nature[382, 383]. However, it is difficult to achieve complex modular host-guest biomaterials due to synthetic limitations. As an alternative, modular supramolecular building blocks based on π-conjugates have been chosen, utilizing hydrophobic interactions to drive their assembly. This approach allows for direct coupling to specific biofunctional moieties, simplifying the synthetic challenge[384, 385].
Nature-inspired peptide assemblies are another example of fibre-like materials that mimic 1D fibrillary natural entities. These peptide assemblies contain natural amino acids, which give them intrinsic biofunctionality. Peptide amphiphiles (PA) are a type of such peptide assemblies, made up of peptide sequences covalently linked to hydrophobic chains[386–388]. These PA materials have demonstrated promising properties as biofunctional materials[389]. Fascinating applications have utilized PA fibres, while understanding their structure-property relationship has been essential for advancements. To increase modularity and expand the bioapplications of supramolecular polymers, researchers have developed 1D materials based on ureido-pyrimidinone (UPy, see 1.2.2). The UPy moieties dimerize through quadruple hydrogen bonding, and the addition of urea groups enables lateral stacking into fibres held together by a combination of hydrogen bonding, hydrophobic interactions, and π-π stacking[48, 390]. In a relevant study, UPy fibres have been enriched with cationic moieties in dilute form to facilitate intracellular delivery of siRNA[391]. Another intriguing application is the incorporation of specific peptides to enable growth factor stabilization[392]. Another example that involves hydrogen bonding and hydrophobic effects is benzene-1,3,5-tricaboxamide (BTA)-based supramolecular polymers. In the 'Introduction' chapter of this thesis, BTA-based supramolecular polymers were discussed extensively, along with some potential applications in various fields that were explored. However, the deep understanding gained from fundamental studies on their assembly in water makes them highly promising materials also for bioapplications[64, 393, 394], which is the primary focus of this specific work.
BTAs 1D fibres are highly dynamic[139, 197, 345, 395] and can incorporate multiple functionalities in a modular fashion through copolymerization[312, 396]. Moreover, they can be decorated with peptides, charges, carbohydrates, and DNA, inducing cellular recognition, intracellular delivery, mimicking the glycocalyx, and recruiting proteins[64, 396–399]. Recently, hydrogels were obtained by increasing the concentration of the fibres[398, 400]. The dynamic and self-healing properties of these

hydrogels make them extremely powerful tools to emulate the extracellular matrix. Despite their potential, the use of these supramolecular materials as biomaterials, such as in tissue engineering, is limited by a lack of detailed understanding of their behaviour in physiological conditions, including their interaction with biologically relevant entities such as proteins and cells.
In this study, we explore the potential of water-soluble BTA derivatives as biomaterials. Specifically, we investigate the stability of BTA-OEG_4 when exposed to bovine serum albumin (BSA) at physiological concentrations using molecular modelling techniques. Our results not only support the experimental evidence but also offer significant insights into the potential of BTA-OEG_4 as a biomaterial.

5.2 Results and Discussions

5.2.1 Experimental section

In the present section, I will report the main experimental results regarding the interaction between BTA-OEG_4 and BSA, obtained by the group of E.W. Meijer, to provide context for the theoretical calculations that I conducted in this work. More details about the investigation of the interaction between different types of BTA-based monomers (BTA-Man and BTA-OEG_4-Man) and BSA, and between BTA-OEG_4 and FBS and cell media are available in the reference paper[378].

Since protein adsorption onto materials is depicted as the first event occurring in any biological system, the interaction of BTA fibres with proteins was first investigated by using BSA as model protein using fluorescence spectroscopy. Different amounts of BTA-OEG_4 fibres were incubated in physiological concentration, and the changes in tryptophan fluorescence emission were followed (Fig. 5.1b-d). The fluorescence quenching was overall low and only at the highest concentration of BTA-OEG_4, when the ratio BTA-OEG_4/BSA was close to 1:3, the fluorescence emission spectra overlapped with the one containing urea, which is a known protein denaturing agent. In order to investigate the mechanism of this interaction, the same experiment was repeated with more hydrophilic BTA-based supramolecular copolymers (see Fig. 5.1a). The tryptophan fluorescence showed almost no quenching by these two copolymers indicating a minor interaction between BSA and BTA copolymers bearing mannose moieties at the periphery. The role of mannose in

reducing the BTA interaction with proteins was further confirmed by lower Forster Resonance Energy Transfer (FRET). These experiments also showed that higher BTA concentrations induced an increase in interactions, corroborating the tryptophan fluorescence emission results (Fig. 5.1e).
CryoTEM and TIRF microscopy suggested that the interaction with BSA partially affects the BTA-OEG_4 microstructure by shortening the fibre length, but the fibre morphology is retained.
The stability of BTA-OEG_4 fibres and hydrogels in fetal bovine serum (FBS) and cell media and cytotoxicity were evaluated. The results, available in the reference paper[378], showed that BTA-based supramolecular biomaterials can be typically used for bioapplications.

5.2.2 Computational section

Experimental data has demonstrated that the strongest interaction occurred between BTA-OEG_4 and BSA. Therefore, in our computational investigation, we focused our molecular modelling on the interaction between these two compounds.

Coarse-grained modelling of the BSA/BTA-OEG_4 interaction in solution

Two mechanisms were initially hypothesized for the BTA-OEG_4/BSA interaction: either BTA-OEG_4 monomers were sequestered by BSA and trapped into the protein hydrophobic pockets (i.e. monomer exchange out of the fibres into the protein), or BSA was destabilized by BTA-OEG_4 (i.e. interaction of BSA with BTA-OEG4 fibres, rather than monomers). In the first case, the BSA/BTA-OEG_4 interaction would be governed by the dynamics of the monomer exchange out of the BTA-OEG_4 fibres, which determines how many disassembled monomers were present in the solution. In the second case, the destabilization would involve a macromolecular interaction between BSA and the whole BTA-OEG_4 fibres.
In order to have qualitatively insights into the most probable BSA/BTA-OEG_4 mechanism of interaction, we employed coarse-grained molecular dynamics (CG-MD) simulations. In particular, two in silico experiments were designed to assess the effect on the BSA in the two types of interactions, that is, BSA interaction with disassembled BTA-OEG_4 monomers in solution vs BSA interaction with assembled

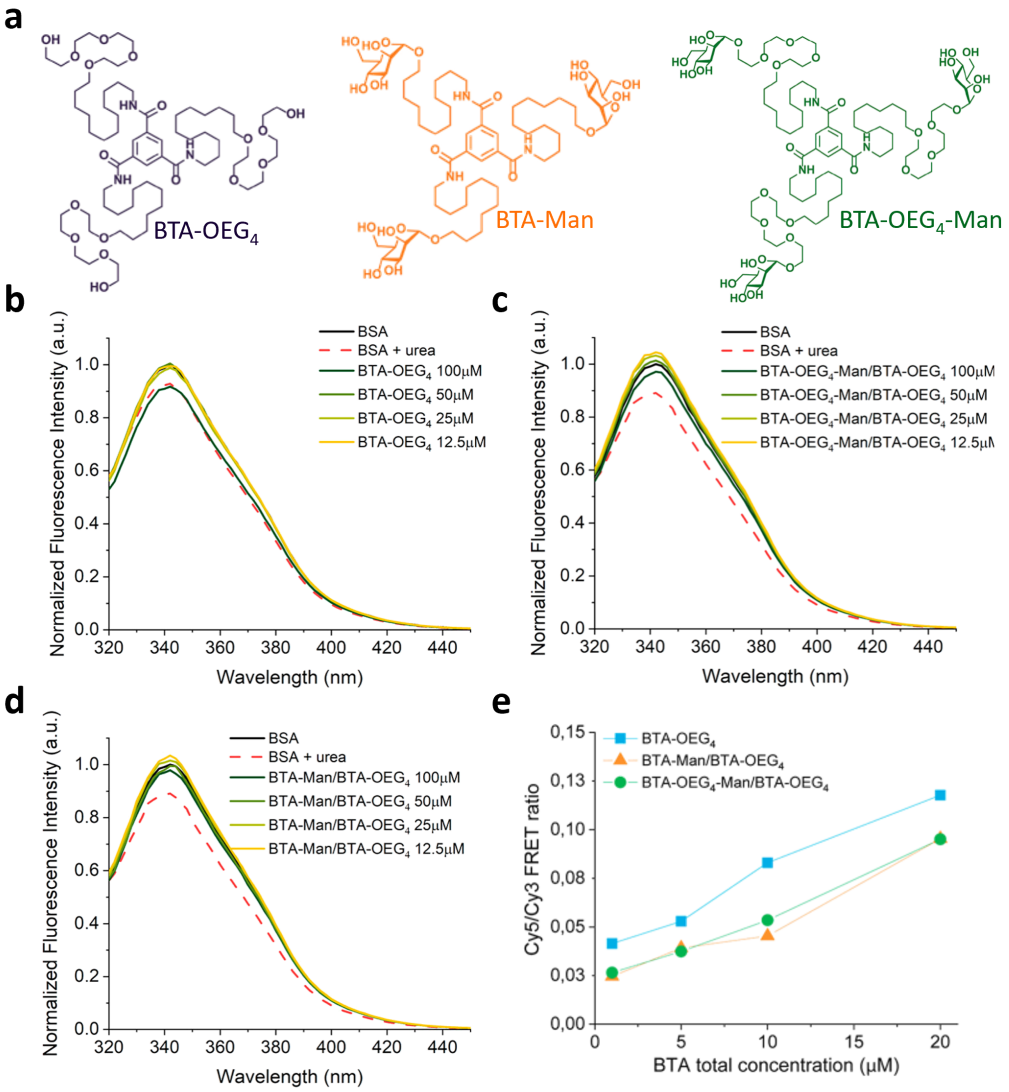

Fig. 5.1 (a) Chemical structure of water-compatible BTA monomers selected for this study; (b) BSA tryptophan fluorescence emission showed the interaction between BTA-OEG_4 and physiological concentration of BSA at 37°C. The introduction of mannose moieties at the fibre periphery reduced this interaction (c, d). (e) FRET experiments showed enhanced interaction between Cy3-labeled BTA fibres and Cy5-BSA increasing the fibres concentrations. Adapted from Ref.[378].

BTA/OEG_4 fibres. We thus simulated two cases where one BSA dimer interacts with 100 initially disassembled BTA monomers vs one pre-equilibrated BTA fibre composed of 100 assembled monomers in aqueous solution. The effect of the interaction with BTA-OEG_4 on the BSA structure in both cases was then compared.

In general, both CG-MD simulations showed that the protein had a strong interaction with BTA/OEG_4. In the first case, the monomers tended to distribute across the BSA surface, targeting specific hydrophobic patches (Fig. 5.2a). In the second situation, the BTA-OEG_4 fibre strongly attached to the protein surface and bent surrounding it (Fig. 5.2b). Analysis of the CG-MD trajectories indicated that the number of contacts between BSA and BTA/OEG_4 was higher in the case of free/disassembled BTA/OEG_4 monomers in solution (Fig. 5.2c). Nevertheless, the BSA 3D structure deviated more from its native conformation when interacting with the BTA-OEG_4 fibre (Fig. 5.2d).

Since CG models can be limited by the approximations in the CG scheme and by the use of the elastic network applied to the protein model (see the Methods section of the Chapter), we cross-checked these comparative simulations in presence and in absence of an elastic network potential (preserving the BSA structure). These additional simulations confirmed the trends observed in the data shown in Fig. 5.2d for both cases. This demonstrated that the BSA/fibre interaction entailed a strong effect on the protein secondary and tertiary structure. This phenomenon, which can be compatible with protein denaturation, is experimentally corroborated by the similar effect of urea and BTA-OEG_4 seen in the fluorescence spectroscopy. On the other hand, although the BSA interaction with free monomers was observed in the CG-MD (Fig. 5.2c: number of contacts with monomers was even higher than those with the BTA-OEG_4 fibre), the effect on the BSA's 3D structure was considerably smaller. Altogether, this indicated that the interaction with a full fibre had a stronger destabilization effect on the BSA structure than the interaction with the same number of individual BTA-OEG_4 monomers. Since the concentration of free monomers when the fibres are present is very low, BSA proteins interacted more likely with BTA-OEG_4 fibres than with disassembled BTA-OEG_4 monomers. Thus, these CG-MD simulations suggested that the tryptophan fluorescence quenching and the shortening of the fibres observed experimentally were caused by the interaction of BSA with the BTA-OEG_4 fibre rather than with the monomers. This was further confirmed by thermodynamic data extracted from the CG-MD simulations, which showed a more favourable interaction energy for BSA with BTA-OEG_4 fibres compared with the same number of disassembled BTA-OEG_4 monomers, with a difference in interaction energy of $\sim$2.2 kcal mol^{-1} per-monomer, see Table 5.1.

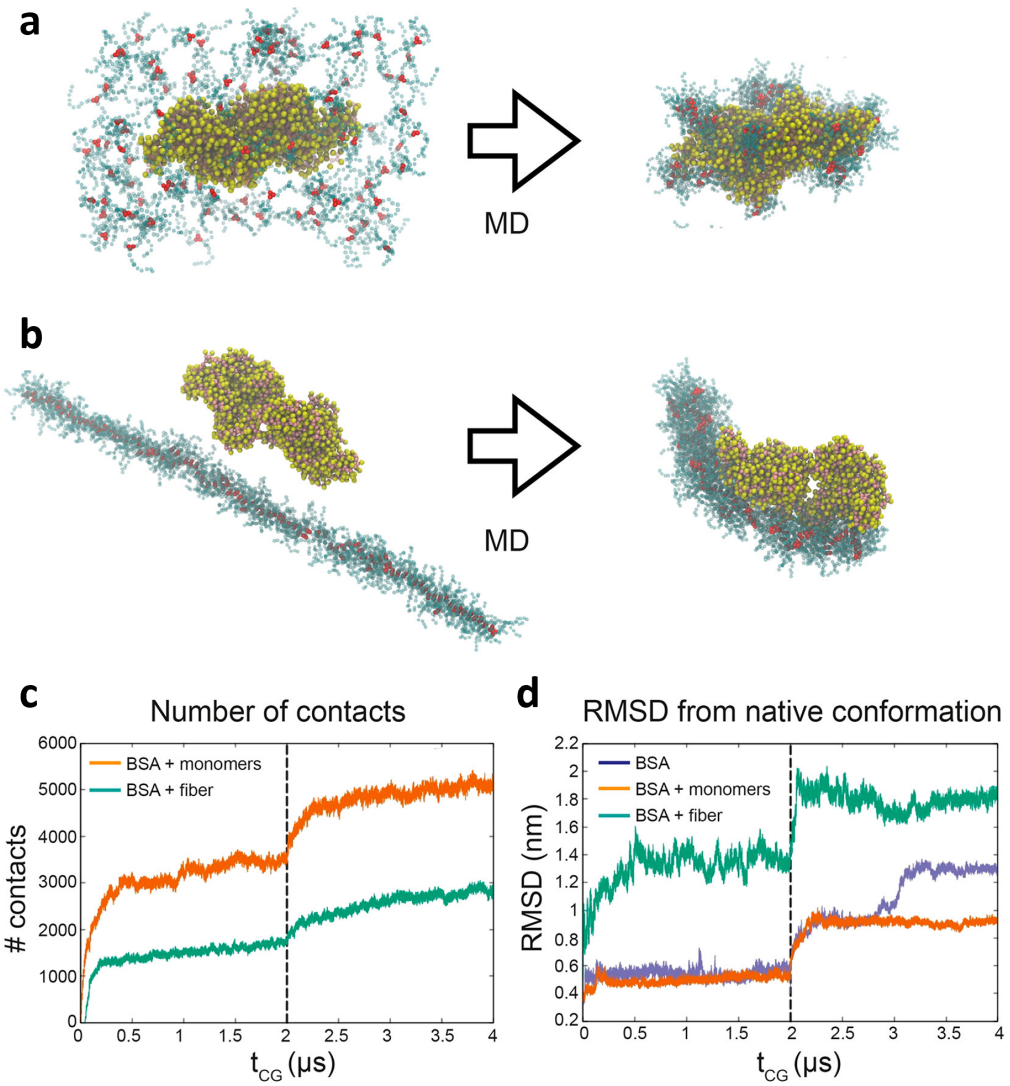

Fig. 5.2 (a) BSA dimer (yellow and pink beads) interacting with 100 disassembled BTA-OEG_4 monomers (red and blue beads) and (b) BSA dimer interacting with the BTA-OEG_4 fibre composed of 100 monomers. The snapshots on the left show the initial configuration, while those on the right are taken after 2μs of CG-MD simulations. (c) Evolution of the number of contacts between BSA and BTA-OEG_4 in the two systems (a) and (b). (d) Evolution of root-mean-square displacement (RMSD) of BSA conformation from the native structure, for the two systems (a) and (b) and for a control system with BSA in water. The first 2μs of CG-MD in (c) and (d) employed an elastic network potential, which was removed after 2μs (vertical dashed line). Adapted from Ref.[378].

Monomer exchange from the BTA-OEG_4 fibre to BSA

Additional evidence for stronger interaction of BSA with BTA-OEG_4 fibres rather than with BTA-OEG_4 monomers was obtained by CG Metadynamics (CG-MetaD)

	BSA + 100 BTA monomers	BSA+BTA fibre	Δ (fibre-monomers)
LJ energy per-BTA monomer [kcal/mol]	-2682.68 ± 0.14	-2684.79 ± 0.14	-2.11
Total potential energy per-BTA monomer [kcal/mol]	-2643.59 ± 0.06	-2645.75 ± 0.14	-2.17

Table 5.1 Values of the potential energy and the Lennard Jones (LJ) energy in the systems (per-monomer), computed on the equilibrated phase of 400 ns of the CG-MD simulations.

simulations. Starting from a BTA-OEG_4/BSA bound state (i.e. BTA-OEG_4 fibre bent and bound to the BSA surface), a CG-MetaD simulation was performed where one BTA-OEG_4 monomer at the interface was biased to exchange/diffuse from the BTA-OEG_4 fibre within the protein and back. During the CG-MetaD run, the monomer crossed multiple times back and forth from the fibre into the protein. From this CG-MetaD simulation, we computed the free-energy surface (FES) of the monomer exchange between fibre and protein. The FES showed that a BTA-OEG_4 monomer could exist in different states with favourable free-energy (Fig. 5.3). In the first case, the monomer could be stacked into the BTA-OEG_4 fibre backbone having negligible interaction with BSA (Fig. 5.3i). Another possibility could see the monomer stacked onto a defect of the BTA-OEG_4 fibre backbone at the interface between BSA and the BTA-OEG_4 fibre and interacting with both (Fig. 5.3ii). The monomer could also be placed at the interface between BSA and the BTA-OEG_4 fibre but adsorbed on the fibre surface (Fig. 5.3iii). In this case, the monomer interaction was stronger with BSA than with the fibre. Finally, the monomer could be sequestered by the protein, having no interaction with the fibre (Fig. 5.3iv). Remarkably, the FES demonstrated that the global free-energy minimum was achieved when the exchanging monomer was at the interface, interacting more with BSA (Fig. 5.3iii). The shape of the FES suggested that, once a BSA/BTA-fibre binding was completed, the diffusion of monomers into the protein was quite likely. From the global minimum, the monomer could indeed diffuse inside the BSA structure by crossing a free-energy barrier of just $\sim$5 kcal mol^{-1}, which is half of the free-energy ($\sim$10 kcal mol^{-1}) required to exchange a monomer out from the fibre into water. This means that BTA-OEG_4 monomers were more favourably exchanged directly from the fibre into the BSA structure via a direct surface/interface contact, rather than through a multistep

mechanism involving monomer exchange from fibre into water and from water into BSA. This further supported the hypothesis that the conformational change of BSA observed from the tryptophan fluorescence quenching experiment was also likely due to the interaction of BSA with BTA-OEG_4 fibres rather than with exchanging BTA-OEG_4 monomers present in solution.

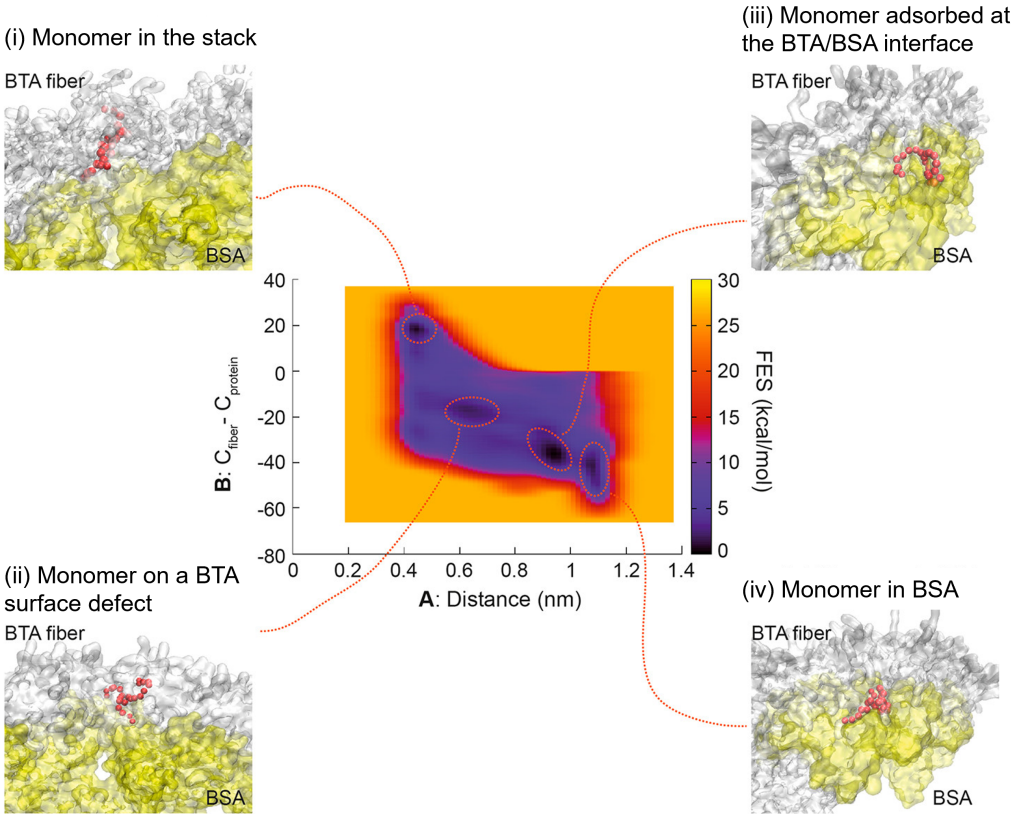

Fig. 5.3 In the centre, the free-energy surface (FES) estimated from WT-MetaD simulations as a function of the two variables *A* and *B* (see Methods section). Snapshots from the main free-energy minima are reported: (i) The monomer (red beads) is part of the backbone of the fibre; (ii) The monomer is in a defect configuration, interacting with the protein; (ii) The monomer is adsorbed at the interface between BSA and BTA-OEG_4 fibre; (iv) The monomer detaches from the fibre and is sequestered inside the protein. Adapted from Ref.[378].

5.3 Conclusions

In conclusion, the study investigated the suitability of water-soluble BTA supramolecular polymers as biomaterials. Three functional BTA materials, BTA-OEG_4, BTA-Man, and BTA-OEG_4-Man, were studied in the presence of different biological entities (BSA, FBS, and cells). The results demonstrated that only BTA-OEG_4 interacts with BSA, as observed through tryptophan fluorescence and FRET studies. CG-MD simulations corroborated this interaction and showed that the BSA conformation was affected by the interaction with BTA-OEG_4 fibres rather than with BTA-OEG_4 monomers. Additional CG-MetaD simulations showed that the exchange of monomers from the BTA-OEG4 fibres to BSA occurred most likely via a direct contact between the fibre and the BSA surface without the need of a monomer exchange to water as an intermediate step. In addition, the study also demonstrated high biocompatibility and low cytotoxicity of BTA-OEG_4 in the presence of cells, indicating their potential for use in biological applications. This work underscore the significance of molecular modelling in the development of novel biomaterials. The use of computational methods such as CG-MD and CG-MetaD allowed us to obtain valuable insights into the interaction between BTA-OEG_4 and BSA, which would have been difficult to obtain experimentally. Furthermore, this Chapter highlights the potential benefits of integrating experimental and computational methods in biomaterials research. This combination of approaches can provide a more comprehensive understanding of the behaviour of materials in biological systems, ultimately leading to the development of more efficient and effective biomaterials.

5.4 Methods

CG-MD Simulations of BTA-OEG_4-BSA Interaction–BSA Destabilization

A CG description of the system was chosen by building the molecular model based on the MARTINI force field[241], as it is suited to study the interactions between supramolecular polymers and biomolecules[181, 346]. The CG model of the BSA protein dimer was based on the PDB structure 4f5s[401], and the topology and interaction parameters were constructed using the martinize.py tool (github.com/cgmartini/martinize.py), employing the MARTINI force field extension

for proteins[402, 403]. The MARTINI force field has some limitations in modelling proteins accurately. The secondary, tertiary and quaternary structure of proteins are stabilized by various types of interactions, such as hydrophobic interactions, hydrogen bonds, polar interactions, disulfide bonds, etc. However, most CG models, including MARTINI, have difficulties in reproducing the directionality of H-bonds, which are essential for the formation and stability of protein structures. To overcome this limitation, one possible solution is to combine an elastic network potential with the MARTINI model[404, 243]. The elastic network potential introduces a set of restraints (harmonic forces) among non-bonded beads within a certain cutoff to preserve the correct native protein structure. For the BSA protein, we used a cutoff of 0.9 nm and a force constant of 500 kJ $mol^{-1}nm^{-2}$ for the elastic network bonds. However, the elastic network potential also introduces some bias and constraints that may affect the realistic dynamics of the protein (and consequently, its interaction with BTA-OEG_4). Therefore, we simulated BSA both with and without the elastic network. By comparing the results of both simulations, we can assess the impact of the elastic network on the protein behavior and structure, and evaluate any inconsistencies or discrepancies between the two models.

A MARTINI-based CG model, previously developed and optimized using the standard MARTINI parametrization procedure[196], was employed for BTA-OEG_4. This model includes a dipole to model the directionality of H-bonding of the BTA-based monomer. A BTA-OEG_4 fibre model was preliminarily equilibrated in explicit MARTINI water. This model was used to study the interactions of a BTA-OEG_4 fibre with BSA compared with interaction of dispersed/disassembled monomers.
The BSA+BTA-OEG_4 monomers system contained the protein dimer, 100 BTA-OEG_4 monomers, and 34 Na^+ ions to neutralize the protein charge. The BSA + BTA-OEG_4 fibre model contained the same number of proteins, BTA-OEG_4 monomers (arranged into a pre-equilibrated fibre), and neutralizing Na^+ ions as the BSA + BTA-OEG_4 monomer system. Both CG models were solvated in standard MARTINI CG water (W) beads.
We conducted 2μs of MD simulations starting from model configurations where the BTA-OEG_4 monomers or the BTA-OEG_4 fibre were placed in proximity to the BSA dimer. During these runs the interaction between BSA and the BTA-OEG_4 monomers or fibre reached the equilibrium in the CG-MD regime. From this point, the simulation box size was resized in order to have the same number of water

beads in the two models. An additional 0.5 μs of MD simulations were conducted to calculate the energies of the two systems and compare the two cases of BSA interacting (i) with BTA-OEG_4 monomers vs. (ii) with a BTA-OEG_4 fibre.
In all CG-MD runs, the boundaries of the simulation box were treated with Periodic Boundary Conditions to reduce finite size effects. To prevent the BSA from rotating and directly interacting with its periodic images, a restraining potential was imposed, acting on two residues (Asp363 and Lys312 of first and second BSA monomers, respectively). As verified by the comparison of two trajectories with and without the restraining potential, the effect of these restrains was negligible. The leapfrog algorithm was used to integrate the equations of motion, with a standard CG time step of 20 fs. The temperature of the systems was set at 293 K by means of the v-rescale thermostat method[261], with a coupling time of 2 ps, while the pressure was kept at 1 bar by means of Parrinello–Rahman barostat[262], with a time constant of 6 ps.

CG-MetaD Simulations–Mechanism of BSA–BTA Interaction

Metadynamics (MetaD)[276] was employed to efficiently explore the configurations and exchange of a BTA-OEG_4 monomer across the interface between the fibre and the BSA dimer. The applied bias acted along two collective variables A and B, relative to a preselected monomer **M**, initially interacting with both the fibre and the protein surface. A was defined as the minimum distance between the **M** core atoms (the benzene ring of BTA-OEG_4) and the backbone of the fibre. B was defined as $C_{fiber} - C_{protein}$, where C_{fiber} was the number of contacts between the core atoms of **M** and the backbone of the fibre (the other BTA cores) and $C_{protein}$ was the number of contacts between the core of **M** and the BSA atoms. The bias was constructed by depositing every 1000 CG-MD steps Gaussian kernels of height 0.2 kJ mol^{-1} and width of 0.05 (in the A direction) and 0.25 (in the B direction). All MD and MetaD simulations and analyses were performed with GROMACS 2018[294, 295], equipped with PLUMED 2.5[296].

Chapter 6

Into the dynamics of rotaxanes at atomistic resolution

The work presented in this Chapter is entirely computational and it has just been published on *Chemical Science*. Part of the analysis conducted in this work has been carried out at University College London during my 6-months period as visiting student. The Chapter is largely taken from the postprint version of the published article[405].

Full bibliographic reference: **Leanza, L.**, Perego, C., Pesce, L., Salvalaglio, M., von Delius, M., Pavan, G. M. *Chem. Sci.* (2023). DOI:10.1039/D3SC01593A

The last and also the main study of my PhD Thesis will examine the second class of supramolecular systems that was discussed in Chapter 1, molecular machines.

Mechanically-interlocked molecules (MIMs) are at the basis of artificial molecular machines and are attracting increasing interest for various applications, from catalysis to drug delivery and nanoelectronics. MIMs are composed of mechanically-interconnected molecular sub-parts that can move with respect to each other, imparting these systems dynamical behaviours and interesting stimuli-responsive properties. Among them, extensively studied examples are rotaxanes and molecular shuttles. The rational design of MIMs with desired functionalities requires studying their dynamics "in action" at sub-molecular resolution and on relevant timescales, which is challenging experimentally and computationally.
In this work, we combine molecular dynamics and Metadynamics simulations to reconstruct the thermodynamics and kinetics of different types of MIMs at atom-

istic resolution under different conditions. As representative case studies, we use rotaxanes and molecular shuttles substantially differing in structure, architecture, and dynamical behavior. Our computational approach provides results in excellent agreement with available experimental evidences and a direct demonstration of the critical effect of the solvent on the dynamics of the MIMs. At the same time, our simulations unveil key factors controlling the dynamics of these systems, providing submolecular-level insights into the mechanisms and kinetics of shuttling. Reconstruction of the free-energy profiles from the simulations reveals details of the conformations of macrocycles on the binding site that are difficult to access via routine experiments and precious for understanding the MIMs' behavior, while their decomposition in enthalpic and entropic contributions unveils the mechanisms and key transitions ruling the intermolecular movements between metastable states within them. The computational framework presented herein is flexible and can be used, in principle, to study a variety of mechanically-interlocked systems, and to explore how to control their dynamics.

6.1 Introduction

Biomolecular machines and motors are at the basis of many biological processes, such as DNA replication[406], muscle contraction[407–410] and ATP synthesis[411–413].

For the last decades, one of the biggest challenges for chemists and physicists has been to design and synthesize artificial molecular machines (AMMs) with controllable movements, mimicking nature's technology[28, 95, 414]. Within this context, the pivotal contribution provided by the use of mechanically interlocked molecules (MIMs)[103, 415–417], such as catenanes[418] and rotaxanes[419], has been carefully addressed in Chapter 1.3.2. To recapitulate briefly, catenanes consist of two or more entangled rings forming a mechanically interlocked chain, while rotaxanes and shuttles typically consist of one or more macrocyclic rings mechanically interlocked around a dumbbell-shaped molecule (the axle) with two bulky groups at both ends (stoppers). In rotaxanes, the macrocycles can move back and forth along the axle, and we take the liberty to describe this translational movement from now on as "shuttling". The same type of linear movement occurs also in the so-called "molecular shuttles", while the fundamental difference of a shuttle (as opposed to a simple rotax-

ane) is that there are at least two distinct binding sites for the ring within the thread, such that there is not one single global thermodynamic minimum for the stochastic location of the ring (but at least two local minima). If properly designed, one can achieve control over the relative motion of subcomponents, making rotaxane-based architectures the perfect candidate components for engineering AMMs.

As a significant development, Green et al.[420] designed a monolayer of bistable molecular switches as a storage element, relating the relaxation kinetics with the memory retention time of the device. Significant breakthroughs were also achieved in catalysis, exploiting the threading of the macrocycle along the thread to expose/hide organocatalytic groups[120, 421–423]. Similarly, the disassembly of a biocompatible [2]rotaxane can trigger the release of anticancer drugs[32]. Designing MIMs with controllable dynamical properties could benefit from a deeper understanding of the factors controlling their structure and dynamics. In particular, reaching a molecular understanding of rotaxane dynamics in solution and unraveling the details of the mechanisms (thermodynamics and kinetics) underpinning their behavior is crucial for designing MIMs with desired properties. Experimental techniques based, for example, on nuclear magnetic resonance, coalescence methods or cyclic voltammetry provide good estimates of the dynamics of rotaxanes and molecular shuttling. For example, the influence of the axle length[161, 424], the conformational flexibility of spacers[425, 426] and of different environments[427, 428] on rate constants were extensively investigated.

However, the limited resolution of kinetic experiments and the high sensitivity of such architectures to numerous factors (structure of subcomponents, solvent composition, pH) hinder the complete characterization of MIMs. In this regard, computational methods, such as Molecular Dynamics (MD), can be extremely helpful as they can provide a detailed viewpoint on the atomistic/molecular mechanisms that regulate the dynamics of MIMs[203–205, 429–438]. In particular, a significant contribution was made by Cai and coworkers[199, 206–210, 439], on the investigation of how the solvent and the surrounding environment modulate the thermodynamics of various rotaxanes.

Nevertheless, the rare-event dynamics that govern shuttling in MIMs exceed the capability of standard MD simulations. To this aim, as discussed in the previous Chapters, enhanced sampling techniques (*e.g.* Metadynamics[276], adaptive bias force[269]) were recently proven to be efficient in investigating rare events with atomistic/sub-molecular resolution, enabling the simulation of the relevant timescales

over which these rare events unfold[197, 289, 290, 304, 440–444]. For example, Metadynamics simulations have been recently used to investigate the dynamics of monomer exchange in supramolecular polymers[197, 304, 444] or the innate guest exchange dynamics in-and-out the cavity of a coordination cage[443, 445], as well as the motion and dynamic behaviours of supramolecular systems (*e.g.* nanoparticles, tubules) in out-of-equilibrium conditions or under the effect of an external stimulus[446, 447]. Offering the opportunity to study molecular motion events occurring on long timescales at atomistic resolution and providing precious insights on the key steps or factors controlling them, these approaches are well suited also to study the dynamics of MIMs such as rotaxanes and molecular shuttles.
Here, we combine standard MD simulations and Well-Tempered Metadynamics (WT-MetaD)[285] techniques to investigate at the atomistic level the thermodynamic and kinetic behaviour of four known [2]rotaxanes[161, 448, 449, 426], covering different features and architectures in the framework of MIMs. With our *in silico* approach, we obtain a detailed insight into the thermodynamics of these systems, recognizing the enthalpic and entropic weight associated with the relevant conformations of MIMs during shuttling dynamics[292, 293], and outlining the mechanisms of motion, the kinetics of shuttling (found in perfect agreement with experiments). We also provide a flexible computational framework that can be used, in principle, to study a wide variety of mechanically interlocked molecules.

6.2 Results and Discussion

In this study, we have chosen to characterize, as representative examples, four types of [2]Rotaxanes in an effort to capture different aspects of MIMs common architectures and obtain general insight into the variegated dynamics of such systems.

Dynamics and shuttling rates measurement of $[R_4 - H_2]^{2+}$ [2]Rotaxane

As a first case study, we focus here on the H-shaped dicationic $[R_4 - H_2]^{2+}$ [2]Rotaxane (referred to hereinafter as **1**, Fig. 6.1) reported by Gholami et Al.[161], consisting of a rigid dumbbell with four phenyl rings between two benzimidazolium recognition sites, and a dibenzo[24]crown-8 ether (DB24C8) macrocycle, which instead is a relatively flexible element. We have chosen to study this system as it

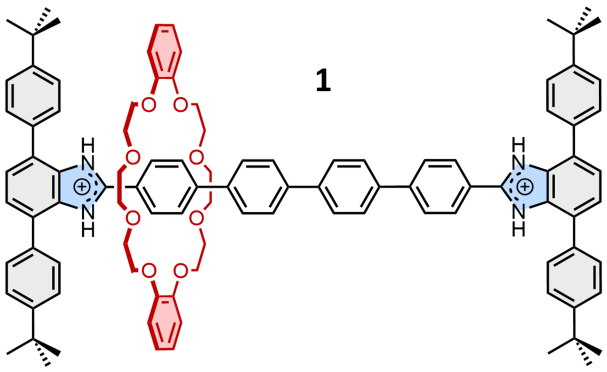

Fig. 6.1 Structural chemical formula of $[R_4 - H_2]^{2+}$ [2]Rotaxane. Adapted from Ref.[405].

is part of a influential study exploring the influence of axle length on the rate of shuttling in neutral and dicationic molecular shuttles. Moreover, the kinetic measurements reported in the reference paper[161] are well-suited for comparison with computational results.

Firstly, we built the all-atom (AA) model of **1** (displayed in Fig. 6.2a). The model system was equilibrated in explicit acetonitrile (ACN) (the solvent used in the experiments of Ref.[161]), adding PF_6^- counterions to neutralize the solution. This constituted the starting point of our investigation (see the Methods section or more details).

To reveal the translational dynamics and the mechanism of shuttling in **1**, we first performed WT-MetaD[285] simulations of this system. The extensive exploration enabled by WT-MetaD is essential to capture crucial features of MIMs thermodynamics at the atomistic level and to reconstruct the free-energy profile. Here, we reconstructed two-dimensional FES as a function of two chosen reaction variables (Fig. 6.2b), namely d, the location of the geometrical centre of DB24C8 along the rotaxane, and $d_{benz-benz}$, the distance between the centres of the two benzene rings of the macrocycle (see Fig. 6.2a).

The WT-MetaD calculations unequivocally reveal that the most stable conformations of the rotaxane correspond to states where the DB24C8 macrocycle is located close to the lateral recognition sites ($|d| \gtrsim 1.5$ nm), while the shuttling states are significantly less favoured (Fig. 6.2b). A deeper inspection of the structure of these minima shows that they are quite broad along the $d_{benz-benz}$ coordinate, signalling that the minimum is both characterized by states in which the macrocycle folds ($d_{benz-benz} \sim 0.8$ nm), either stacking its benzene rings with the phenyl rings of the stoppers (state A, in Fig. 6.2b) or with the first central phenyl ring of the axle (state C, in Fig. 6.2b), and

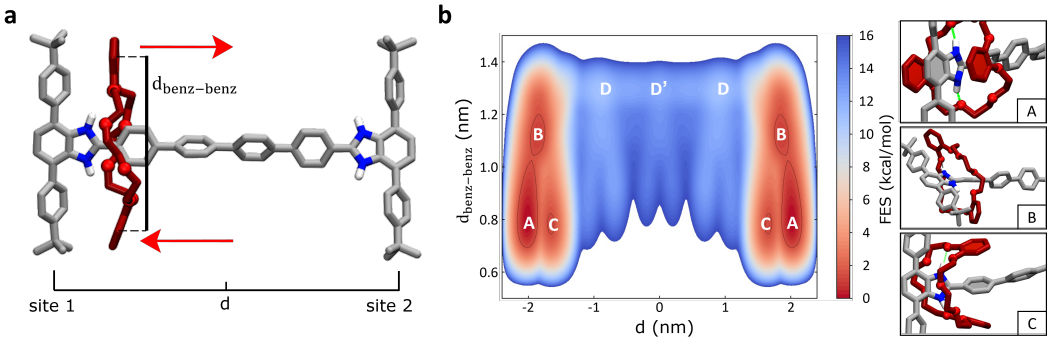

Fig. 6.2 (a) Atomistic molecular model of **1**. The collective variables (CVs) used to enhance dynamics in WT-MetaD simulations, namely $d_{benz-benz}$ and d (see text), are shown. Only the hydrogens of the recognition sites are displayed for clarity. (d) FES as a function of the two CVs biased during WT-MetaD simulations. Three representative snapshots of states A, B, and C are displayed on the lateral panels. Adapted from Ref.[405].

states in which the macrocycle is extended, but still bound to the recognition site (state B, in Fig. 6.2b). All these states are characterized by the presence of NH$\cdots$O hydrogen-bonds (HBs) between the macrocycle and the benzimidazolium groups at the recognition sites (Fig. 6.2b, right: insets). The shuttling process thus requires that DB24C8 breaks these HBs with benzimidazolium, reaching the "unbound state" (D, in Fig. 6.2b). As shown in Fig. 6.5a, the number of H-bonds drops from ~ 2 (at $|d| \gtrsim 1.5$) to 0 (for $0 \lesssim |d| \lesssim 1.5$ nm), while after breaking the interaction with the stoppers and leaving the bound states the DB24C8 ring diffuses along the axle. Interestingly, once in the unbound state (Fig. 6.2b: D and D'), the macrocycle does not maintain an open conformation and the ring shuttling is not resistance-free due to possible $\pi-\pi$ interactions between the aromatic rings of the rotaxane, and to residual solvophobic effects (see also Fig. 6.3a: plot of $d_{benz-benz}$ *vs.* d). The plot of the number of contacts between the DB24C8 macrocycle and phenyl rings of the axle of Fig. 6.3b indicates that the two components interact. During the shuttling, the DB24C8 macrocycle distorts assuming a clip-mode, promoting $\pi-\pi$ interactions between its benzene rings and the phenyl rings of the axle (in the central region of the axle, points where $d_{benz-benz}$ reduces, the number of DB24C8-axle $\pi-\pi$ contacts increases, and *vice versa*). This produces a step-wise diffusion of the macrocycle along the axle, which is consistent with DFT calculations[161].

WT-MetaD allows us to reconstruct the FES associated with the system, providing a reliable estimate of the free-energy differences between "bound" and "unbound" states. However, the ΔG profile obtained from the 1D projection of the FES (Fig. 6.5,

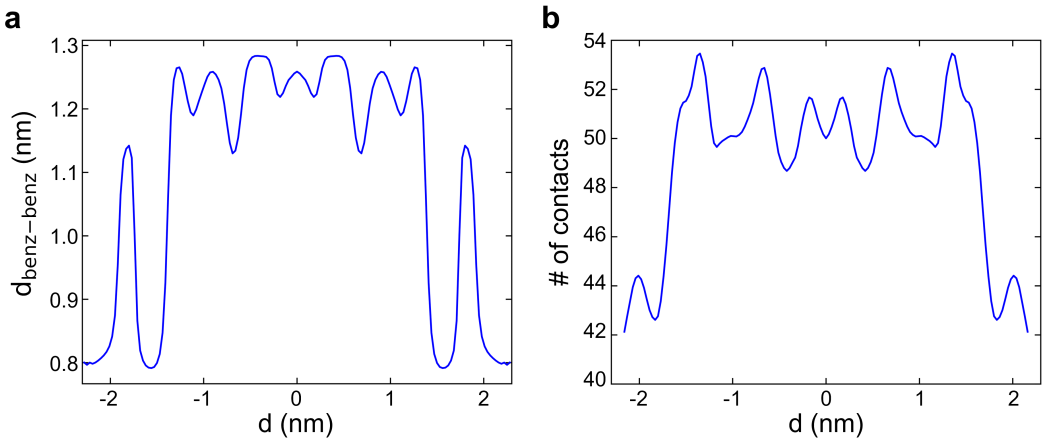

Fig. 6.3 (a) Mapping of $d_{benz-benz}$ along d. (b) Mapping of the number of contacts between DB24C8 and the phenyl rings of the axle along d. The mapping of the variables is obtained using the approach of Ref.[292]. Adapted from Ref.[405].

right panel, black curve) shows low barriers between intermediate states, suggesting that the set of variables is well suited to describe the translational motion of the macrocycle, but may not be optimal for describing the transition from the bound state. Therefore, we used a different strategy to estimate the transition barriers ($\Delta G^{\ddagger}$) and kinetic information on the shuttling dynamics, that is the "infrequent" WT-MetaD[289, 290]. In the "infrequent" variant, the Metadynamics strategy is adopted to accelerate a conformational transition between metastable states, but the bias is deposited infrequently, to avoid deposition over the transition state. By properly accounting for the associated acceleration of the dynamics, the unbiased rates of transition can be reconstructed (see Section 2.6.2). We here applied infrequent WT-MetaD to excite the transition to the "unbound" state (D, in Fig. 6.2b) starting from the "bound" state A. We observed that choosing A, B or C as a starting configuration provides very similar transition kinetics ($\sim$10s, $\sim$70s, $\sim$40s, respectively, see Appendix, Fig. A2a). The kinetics of transition among the "bound" states is indeed much faster than that of the passage to the "unbound" D state, which requires the rupture of the HBs with benzimidazolium, being the rate-limiting step toward shuttling from one station to the other. For this reason, from this point on we refer simply to the "bound" to "unbound" transition, as the distinction between A, B and C states is negligible for the shuttling dynamics.

As schematized in Fig. 6.4, the transition time, estimated via infrequent WT-MetaD, is found in the order of $\sim$ tens of seconds. The free-energy barrier associated with

this transition is 18.5 ± 0.4 kcal mol^{-1}. Experimental ^{1}H NMR investigations revealed that the shuttling dynamics in **1** (passage of the macrocycle from one binding site to the opposite one) is characterized by a free-energy barrier of approximately 19.8 kcal mol^{-1}[161], showing a significant agreement between the experimental and theoretical rates.

It is worth noting that infrequent WT-MetaD is employed to reconstruct the simple transition from "bound" to "unbound" states, whereas the experimental values refer to the shuttling between the two recognition sites (that is between the two global minima of Fig. 6.2b). In this sense, comparing the experimental and theoretical rates is meaningful if the "bound" to "unbound" transition is the rate-limiting step of the shuttling process. To test this hypothesis, we also performed unbiased MD simulations to gather statistics of the macrocycle "unbound" shuttling movement along the axle (from D to D', see Fig. 6.2b). The resulting transition rate is approximately nine orders of magnitude lower (in the nanoseconds range, see Appendix, Fig. A2b) than that associated with the bound-unbound transition, confirming that the departure from the recognition site is rate-determining for the shuttling kinetics. This observation agrees with the experimental results of the reference paper[161], where the reported kinetic rates are independent of the length of the axle (number of aromatic rings): with such kinetic differences, the shuttling dynamics of DB24C8 along the axis is in fact marginal. A complete schematic representation of the shuttling dynamics along the free-energy profile (as a function of *d*) is shown in Fig. 6.4, highlighting the times of transition between different states.

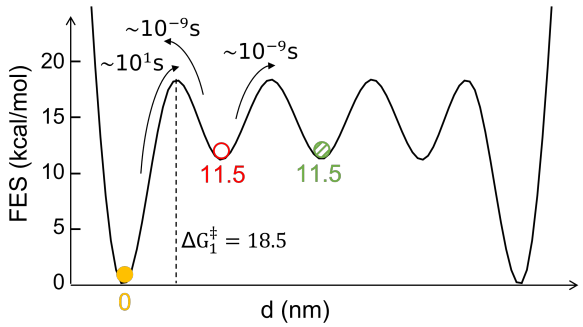

Fig. 6.4 Free-energy surface diagram as a function of *d* and relative transition times, combining the results obtained via WT-MetaD (ΔG of conformational states) and infrequent WT-MetaD (transition times). The free-energy value of each state (coloured circles) and the height of free-energy barriers are defined with respect to state A. Adapted from Ref.[405].

Free-energy decomposition and role of solvent on the shuttling of $[R_4 - H_2]^{2+}$ [2]Rotaxane

As recently shown via computational approaches[199, 206–210, 438, 439], the shuttling dynamics and mechanisms in such kinds of rotaxanes are significantly dictated by the influence of the environment (i.e. the solvent composition), we thus characterized and compared the free-energy profile associated to system **1** immersed in four different solvents, having different β parameter (the hydrogen-bond acceptor ability)[450]: chloroform ($CHCl_3$, $\beta = 0.10$), dichlorometane (DCM, $\beta = 0.10$), acetonitrile (ACN, $\beta = 0.40$) and dimethyl sulfoxide (DMSO, $\beta = 0.76$). While exploring the solvent impact, we enriched our analysis by separating the free-energy profiles in enthalpic and entropic contributions (using the approach described in Ref.[292] and recently applied in the study of the early stages of nucleation of metal organic frameworks[293]), to attain crucial information on the factors that regulate the equilibrium of the rotaxane.

Low β. Similarly to what we have seen in ACN solvent, the free-energy profile computed in $CHCl_3$ ($\beta = 0.10$, Fig. 6.5a, black line of the top panel) shows two minima at the recognition sites (at $|d| \gtrsim 1.5$ nm). The free-energy difference ΔG between the minima and the "unbound" state is estimated at around 14 kcal mol^{-1}. To capture the nature of this free-energy difference, we decompose the FES in enthalpic and entropic terms. Enthalpy (Fig. 6.5a, red line of the top panel) exhibits a global minimum in the configuration with the macrocycle clamped around the benzimidazolium ($d \sim \pm 2.1$ nm). As d increases, the system overcomes an enthalpy ramp (~ 10 kcal mol^{-1}) to reach the central region. This behaviour correlates with the evolution of the HBs established between the subcomponents of the rotaxane (Fig. 6.5a, bottom panel): the HBs between the binding site and macrocycle (that characterize the free-energy minima) are broken, indicating that the disruption of HBs contributes to the enthalpic cost as the macrocycle moves toward the centre of the axle. With the rupture of the HBs, the binding between the axle and macrocycle becomes less constrained, and the entropic contribution of these states becomes sensibly larger (minima of the blue line in Fig. 6.5a). Through the kinetic characterization, we also note that the access to the unbound state (D in Fig. 6.2b) from the global free-energy minimum is associated with a relatively high free-energy barrier ($\Delta G^{\ddagger} \simeq 21.2 \pm 0.4$ kcal mol^{-1}, estimated via infrequent MetaD, see also Appendix, Figs. A5-A6), which implies a slow shuttling rate (the transition time is found in the order of

minutes), on top of the thermodynamic penalty of the unbound state.
As additional tests, very similar results to those in $CHCl_3$ are obtained when studying the system in DCM, which has $\beta = 0.10$ (see Appendix, Figs. A5-A6).

High β. In strong contrast, the free-energy profile obtained in DMSO ($\beta = 0.76$) suggests a different scenario: as it can be observed from Fig. 6.5b, top panel, despite a similar shape in the global FES (see, *e.g.*, the position of the minima), the ΔG between unbound and bound states is significantly lower (9.5 kcal mol^{-1}) than in $CHCl_3$ solvent, as well as the free-energy barrier ($\Delta G^{\ddagger} \sim 16.5 \pm 0.3$ kcal mol^{-1}, computed via infrequent WT-MetaD). The decrease of ΔG and the faster kinetics between unbound and bound states are due to the high tendency of DMSO to accept HBs from the benzimidazolium site, thereby competing with the macrocycle, and favouring the departure of the latter from the recognition site. As shown in the bottom panel of Fig. 6.5b, the average number of HBs between the solvent molecules and the binding station increases by $\sim$2 units, once the macrocycle starts shuttling. The formation of HBs between the recognition sites and DMSO, not only compensates for the energy loss caused by the rupture of HBs between axle and macrocycle but also determines an overall favourable enthalpic configuration, as shown by the global enthalpy minimum that corresponds to the DB24C8 shuttling along the axle (Fig. 6.5b, top panel, red line).

Intermediate β. We can consider the ACN solvent ($\beta = 0.40$), discussed in the previous section, as an intermediate case between $CHCl_3$ and DMSO. In ACN, the free-energy profile as a function of d (Fig. 6.5c, top panel, black line) exhibits a $\Delta\Delta$G $\sim$12 kcal mol^{-1}, between unbound and bound states. Together with the free-energy barrier estimated via infrequent WT-MetaD ($\Delta G^{\ddagger} = 18.4 \pm 0.4$ kcal mol^{-1}), our simulations show that the ACN case has an intermediate behaviour between high and low β solvents, confirming that the dynamics of the system follows a clear trend dictated by the solvent propensity to form HBs (Fig. 6.5c, bottom panel). In ACN, the average number of HBs between solvent and benzimidazolium increases up to $\sim$2.9 units (vs 0 units in $CHCl_3$ and 3.5 in DMSO) as the macrocycle begins the shuttling. This compensates for the enthalpic loss due to the breaking of the interaction between the binding site and the wheel partially. As a result, differently from the DMSO case, the enthalpy here still favours the bound states (macrocycle located in the recognition sites).

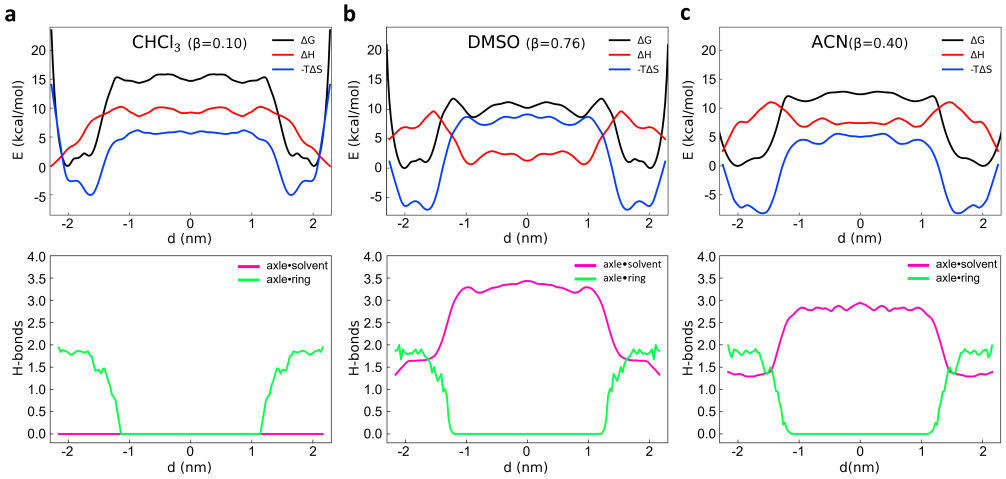

Fig. 6.5 Role of solvent in the shuttling mechanism of **1**. (a) top panel: free-energy (black), enthalpy (red) and entropy (blue) profiles of the Rotaxane **1** immersed in $CHCl_3$ solvent ($\beta = 0$) as a function of d; bottom panel: number of HBs between the axle and the macrocycle (green) and between the axle and the solvent molecules (purple line), as a function of d. (b) and (c) Same as (a), for DMSO ($\beta = 0.76$) and ACN solvent ($\beta = 0.40$). The results demonstrate how the thermodynamics may change as a function of the solvent. Adapted from Ref.[405].

Our simulations are very useful to obtain a molecular-level description of the shuttling mechanism and how the solvent nature may finely modulate the process. In particular, while confirming the *in silico* results obtained with a free-energy characterization of similar rotaxane systems[438], our approach allows us to quantitatively estimate shuttling rates in different conditions. This provides unprecedented details on the shuttling dynamics that enable the comparison with experimentally observed quantities[161], demonstrating a satisfactory agreement between modelling and experiments.

Dynamics and shuttling rates measurement of formamidinium [2]Rotaxane

The second test case analyzed is a formamidinium [2]Rotaxane, recently reported by Borodin et al.[448]. This rotaxane consists of an N,N'-disubstituted formamidinium ion and the flexible 24-crown-8 (24C8) ether (Fig. 6.6a), and it is part of a seminal work that describe the self-assembly of stimuli-responsive [2]rotaxanes, using the dynamic covalent reaction amidinium exchange[451]. This system (referred to hereinafter as **2**), allows us to probe our methods to a relevant case in which both the

axle and the macrocycle are flexible while focusing on relatively simple shuttling dynamics (the system is relatively small). We built an AA model for rotaxane **2** (Fig. 6.6b) immersed in explicit DCM, the typical solvent used in experimental analyses[448], neutralized with weakly coordinating PF_6^- anions. After an initial equilibration phase energy minimization and a short NPT equilibration, we studied the system through standard MD simulation. Enhanced sampling is unnecessary in this system, as the timescale of shuttling events is accessible within a practically attainable MD time. We study the translation of the macrocycle along the main axis of the axle molecule, indicated by the macrocycle coordinate d (see the schematic representation in Fig. 6.6c).

The free-energy associated with the system as a function of d, shown in Fig. 6.6d (black line) shows three stable states along this coordinate. The most stable minimum corresponds to the conformation with the macrocycle at the centre of the axle (d $\sim$0 nm). This "central" state is flanked by two lateral minima which are almost equivalent in free-energy (with a slight difference of 0.15 kcal mol^{-1}) in which the macrocycle approaches the lateral stoppers. The kinetics of transition between these states emerges as very fast, with a transition time of the order of hundreds of ps (Fig. A7b-c).

By inspecting the free-energy decomposition in enthalpic and entropic contributions (Fig. 6.6d, red and blue lines), we note that this free-energy landscape characterized by three, almost equivalent minima along d is the result of a subtle interplay among the two terms. We observe that, while both enthalpy and entropy favour lateral d values ($|d| \sim 0.4$ nm) there is a clear mismatch between the two contributions, as states, where the macrocycle is localized close to the stoppers, correspond to deep entropy minima ($|d| = 0.55$ nm), while the enthalpically favoured states are located around $|d| = 0.4$ nm. This results in less pronounced free-energy wells, having (slightly) lower stability than the central state ($|d| = 0$). In fact, in configurations where $|d| = 0$ the entropic and enthalpic terms mostly cooperate, with the result that local minima combine to form a global free-energy minimum. This shows how the subtly competing enthalpic and entropic contributions characterize the most likely positions of the macrocycle along the axle. We can relate the decomposition in Fig. 6.6d to the average number of HBs formed between the macrocycle and the formamidinium site at the centre of the axle (Fig. 6.6e). We note a clear correlation between this variable and d. The global minima of the entropic curve (Fig. 6.6d, blue line) are found at $|d| = 0.55$ nm, where the average HB number is relatively

low, enabling higher mobility of the macrocycle, while the repulsive walls between stoppers and macrocycle localize the longitudinal position of the latter [452], producing the tight entropic minima in Fig. 6.6d. The enthalpic contribution is instead modulated by both the formation of the HBs and the interaction with the solvent (see the Solvent Accessible Surface Area, SASA, as a function of d reported in Fig. 6.6e); The higher average number HBs (~ 1.7), combined with the relative reduction to solvent exposure (confirming that more compact geometry is more energetically accessible [205], determines the enthalpically favoured states.

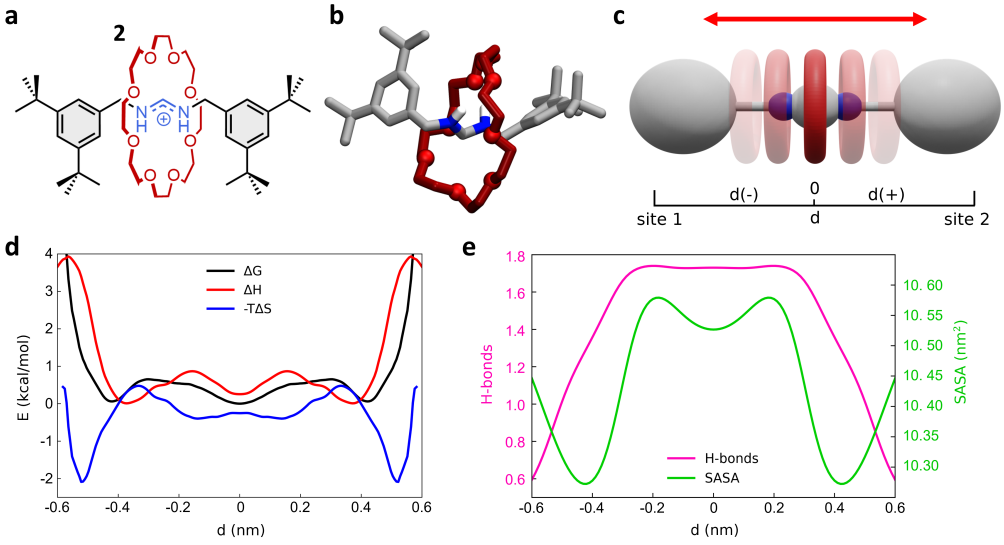

Fig. 6.6 Shuttling dynamics of the formamidinium [2]Rotaxane (**2**). (a) Structural, chemical formula of the formamidinium [2]Rotaxane, **2**. (b) AA molecular model of **2**. Only the hydrogens of the recognition sites are displayed for clarity. (c) Schematic representation of the mechanism of shuttling along the translational position d of the macrocycle. (d) Free-energy surface (black), enthalpy (red) and entropy (blue) profiles projected along d. (e) Number of HBs between the crown and the thread (purple line) and SASA of the rotaxane (green line) as a function of d. Adapted from Ref. [405].

The global enthalpic minima are found at $|d| = 0.4$ nm. A visual inspection of the trajectory showed that it is the ensemble of configurations where the flexible crown not only H-binds to the central formamidinium motif, but also appears to form CH$\cdots$O bonds with the lateral stoppers, stabilizing the states.

Overall, our analysis on system **2** provides detailed insight into the complexity of factors (solvent and conformational entropy, intermolecular interactions, ...) that can impact the shuttling motion of the macrocycle in this kind of flexible

system, exploiting the relative simplicity and fast dynamics of the formamidinium [2]Rotaxane. This level of characterization is an advantage of our *in silico* approach and is precluded to experiment [448].

Dynamics and shuttling rates measurement of [10]CPP-Fullerene [2]Rotaxane

The third test case chosen for our investigation is a [2]Rotaxane consisting of a ten para-substituted phenylene ring ([10]CPP]) that binds a central bis-adduct fullerene (C_{60}) with two bulky fullerenes (C_{60}) hexakis-adduct stoppers [449], separated by two flexible macrocyclic spacer groups (system **3**, Fig. 6.7a). In the last decade, π-conjugated molecules, and in particular CPP-fullerene complexes [453–456] has attracted enormous attention for its photophysical properties and potential applications. The rotaxane studied here is characterized by a stable "bound" state, in which the CPP binds the fullerene, and "unbound" states, in which the CPP lies on the spacer groups. While remarkable differences in the charge transport/recombination properties between those "bound" and "unbound" states have been shown in experiments, the relative dynamics of the system across these conformers remains elusive [449], providing a convenient testing ground for our methods.

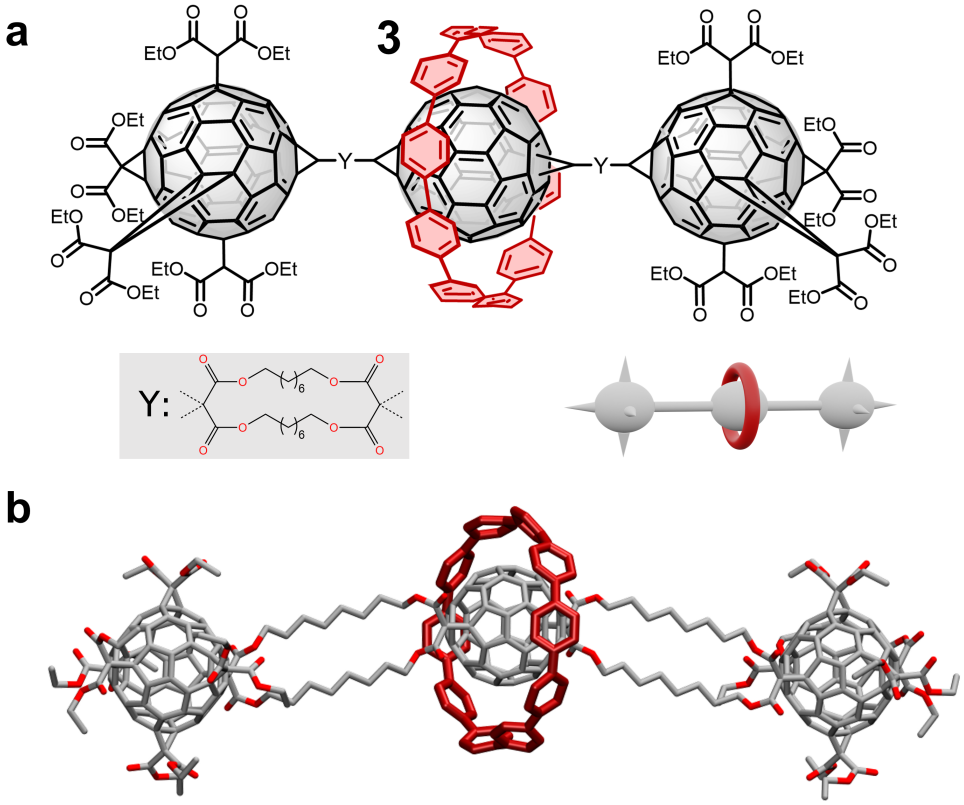

Fig. 6.7 (a) Structural chemical formula of [10]CPP-Fullerene [2]Rotaxane (**3**), Y: macrocyclic spacer and schematic representation of the rotaxane. (b) Atomistic molecular model of **3**, hydrogen atoms are omitted for clarity. Adapted from Ref. [405].

The FES in Fig. 6.8 (black line) as a function of d (see the schematic representation of the CV in the top panel) shows a very deep minimum at $d \sim 0$ nm, in which CPP surrounds the central fullerene ("bound" state). The time required to leave the central fullerene has been estimated via infrequent WT-MetaD, and it is found in the order of tens of seconds. It is informative to compare this transition time to the typical escape time of CPP from a [10]CPP-Pseudorotaxane system to the solution (Fig. A9). In such a system the C_{60} has only a single adduct so that the CPP can be released in solution directly from the bound state. This transition occurs in a typical time of $\sim$3–4 orders of magnitude faster than the time required for the macrocycle to leave the bound configuration in **3** [449]. The slower timing evaluated in **3** indicates the higher frustration CPP encounters in the system, as it can break the complex with the fullerene only by following the directions allowed by the substituents. This

reduces the allowed pathways that activate the shuttling along the flexible linkers, significantly slowing down the kinetics of the transition.

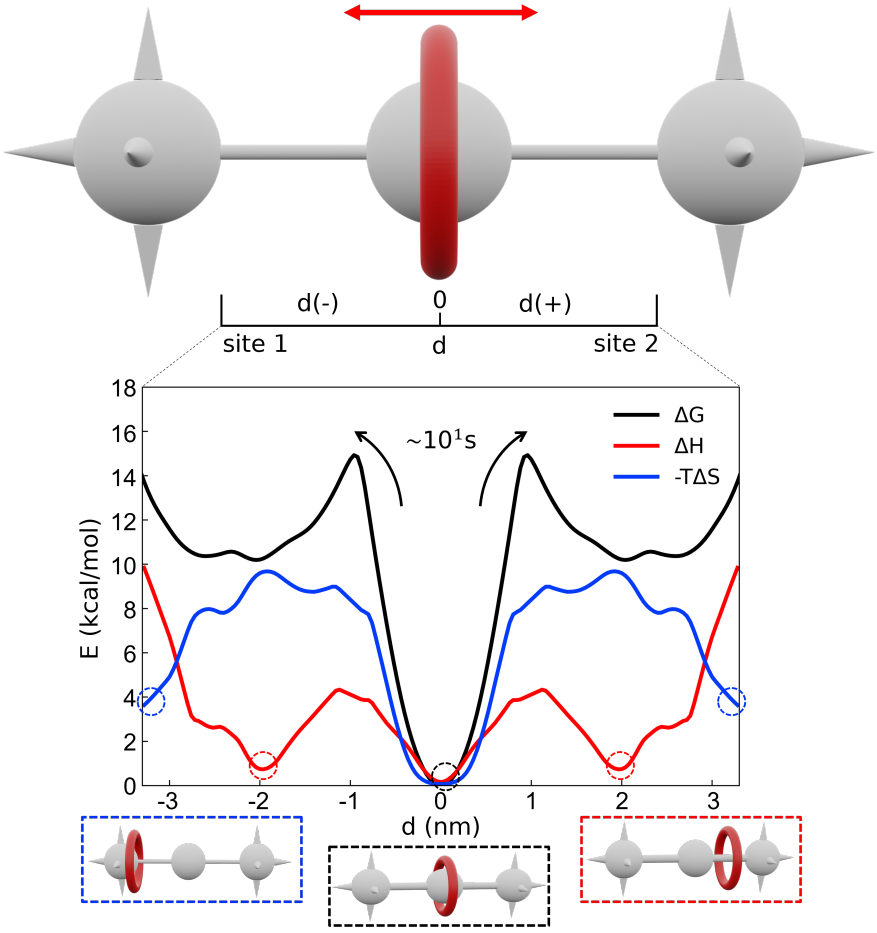

Fig. 6.8 Top: schematic representation of the shuttling mechanism and of the translational CV d used in WT-MetaD simulations. Bottom: free-energy profile (black), associated to the **3** system, with enthalpic (red) and entropic (blue) contributions. The boxes below show schematic representations of the relevant configurations, indicated by dashed circles on the free-energy profile. The arrows indicate the transition from the bound state (minimum) to the shuttling state, with the associated timescale. Adapted from Ref. [405].

Accordingly, in the infrequent WT-MetaD study of these transitions, we have applied the bias potential also on the angle between the CPP and the central fullerene (besides the translational CV d), triggering the exploration of different escape pathways until the allowed one was picked. Interestingly, the global minimum (bound state) of the FES is both enthalpically (π stacking interaction between [10]CPP and the central C_{60}) and entropically (rotation along the main axis of the dumbbell, "twisting" around the C_{60}-fullerene) favoured. During the translational motion, the rotaxane

must inevitably pay an entropic cost [457] (see Fig. 6.8, blue line) to have the macrocycle passing through the flexible linkers (already observed in similar [2]Rotaxanes with flexible linkers [425, 426]), as the possible conformations of the linkers are strongly reduced by the rigid ring encircling them; such entropic cost is partially compensated by the increasing interaction between CH_2 groups of the linkers and CPP, as also showed by ROESY NMR experiments [449], in an enthalpy-entropy compensated mechanism (red line) [458].

To obtain further information on the dynamics of [10]CPP in the bound state around the central fullerene, we also performed 1 μs unbiased MD simulation starting from this configuration, with the purpose of gathering insight into the motion of the CPP in this specific state. We thus decomposed the motion of [10]CPP around the fullerene in terms of pure rotation (ϕ, Fig. 6.9a, top panel) and tilting (ψ, bottom panel) components. The rotation can assume values between -π and π radians while the tilting ranges between $\pi/4$ and $3/4\pi$ radians. To better decompose and visualize this angular ring motion we collected the ϕ and ψ sines and cosines data, and operated a principal components (PC) analysis over these four variables. As expected, the first two PCs contain the majority of fluctuations (about 99%), because of the 2D nature of the motion. In Fig. 6.9b we show the projection of the sampled ring conformations onto the first two PCs, obtaining a FES associated with the angular distribution of the CPP when bound around the C_{60}. The free-energy shows 10 periodic minima arranged in a circle. The circle is associated with the rotation ϕ of CPP around the fullerene, while the width of the distribution around the circle is associated with the tilting motion. Each minimum is separated by a free-energy barrier of $\sim$2 kcal mol^{-1}. Interestingly, this result is in remarkable agreement with the energy barrier associated to the dynamic rotation of fullerene inside a carbon nanohoop, as reported by Matsuno et Al. [459]. The presence of ten minima indicates that, at the resolution of this simulation (Å, ns), the motion is step-wise, determined by the possible π-stacking of [10]CPP and fullerene benzene rings. We employed a clustering analysis (Probabilistic Analysis of Molecular Motifs [460]) to capture the transitions between these angular minima along the MD simulations, which characterize the kinetics of CPP rotation around the fullerene. We thus calculated a mean residence time in each minimum $\sim$0.1 ns, which signals a fast rotation of the macrocycle, characterized by a step-wise mechanism. These rotational degrees of freedom determine a crucial entropic contribution to the bound state, making it the complex favoured conformation. This combination of PC and clustering analysis

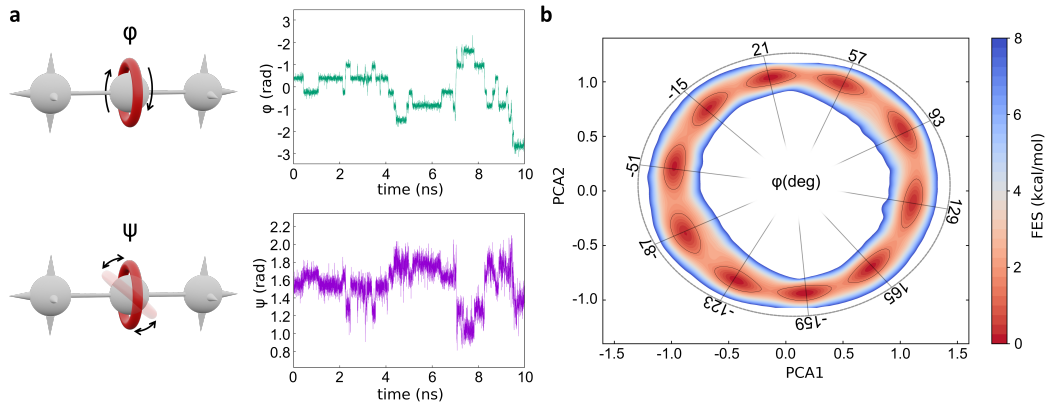

Fig. 6.9 (a) Left: the two angles of macrocycle rotation around the central fullerene. The rotaxane is treated as cylindrically symmetric, where ϕ is the rotation along the z-axis and ψ is the tilting angle (with axis in the in XY plane). Right: ϕ and ψ as a function of time, for a short portion of simulation. (b) Free-energy surface associated to the system in its bound state, as a function of CPP angular configurations ($\sin(\phi)$, $\cos(\phi)$ and $\sin(\psi)$, $\cos(\psi)$) projected along the first two PCs. The ϕ angle corresponding to each minimum is reported. Adapted from Ref. [405].

represents a further, useful strategy to characterize the dynamics of shuttling systems.

The proposed AA simulation study exhaustively characterized both translational and rotational degrees of freedom of system **3**, and revealed the dynamic nature of the interaction between CPP and fullerenes. We obtain a thermodynamic picture of the rotaxane that is in line with the experimental observations [449], providing atomistic details on the nature of the main conformers and of the fascinating kinetics of these fullerenes/CPP-based systems.

Dynamics and shuttling rates measurement of rigid bistable [2]Rotaxane

As the final example, we investigate a degenerate bistable [2]Rotaxane, consisting of a dumbbell containing two naphthalene (NP) recognition sites and the tetracationic macrocycle cyclobis(paraquat-p-phenylene) ($CBPQT^{4+}$), the so-called "blue-box" [426] (Fig. 6.10b). From now on, we will refer to this system as **4**. In this architecture, the rigid spacer between the two recognition sites determines simpler dynamics compared with more flexible systems, making this design a promising candidate for molecular switches. Nonetheless, a complete comprehension of the

on/off mechanism and kinetics is crucial for the rational utilization of rigid linkers in such systems. Also in this case, we developed an AA model of **4** (Fig. 6.10c, that was later immersed and equilibrated in explicit acetone solvent, to reproduce experimental conditions [426]. We then employed WT-MetaD simulation to explore the shuttling dynamics of the rotaxane, obtaining the free-energy landscape as a function of the translational position d of the blue-box macrocycle along the axle (Fig. 6.10d). The computed free-energy profile demonstrates the "all-or-nothing" mechanism of shuttling, in which the blue-box switches between two states that correspond to the binding of the cycle with the two symmetric recognition sites, separated by a high activation barrier [426]. Further insight into the movement of $CBPQT^{4+}$ along the dumbbell is provided by decomposing the free-energy in its enthalpic and entropic contributions. Fig. 6.10d shows clearly how the global minima of free-energy ($d \sim \pm$ 1.0 nm) are mostly entropically driven (also entropic minima correspond to $d \sim \pm$ 1.0 nm). On the other hand, four enthalpy minima have different locations along the axis of **4**. By inspecting the configurations corresponding to these entropic and enthalpic minima (Fig. 6.10d) we can elucidate the nature of these states. Indeed, the enthalpy minima correspond to configurations in which a ring pyridine in $CBPQT^{4+}$ is stacked on one of the NP rings. This creates an angle between the macrocycle plane and the NP, due to the spacing between pyridines in the blue-box (see snapshots in Fig. 6.10d). This determines two enthalpic minima per each NP site, either with the centre of mass heading outward (dot-dashed circles in Fig. 6.10d) or inward (dotted circles in Fig. 6.10d) to the centre of the axle. In contrast, the free-energy minima correspond to a more dynamic state, in which the bipyridine group of the blue-box fluctuates next to the NP without a well-defined stacking of the two aromatic rings. This "dynamic binding" originates the entropic stabilization of these states, which correspond to free-energy minima. The high entropic penalty when $CBPQT^{4+}$ surpasses the NP recognition states ($|d| > 1.5$) and is threaded by the lateral PEG chains agrees with the loss of conformational freedom observed with similar flexible chains (linkers/spacers) [425]. We then characterized the kinetics of the blue-box movements between the two main stable states, namely the dynamic binding with recognition sites, employing infrequent WT-MetaD (see scheme in Fig. 6.10e). For steric reasons, the NP site plane needs to be parallel to the $CBPQT^{4+}$ bipiridyne plane to allow the threading of the macrocycle and the binding with the recognition site. Provided the entropic nature of free-energy minima, we can impute the kinetic barrier of transition from one NP site to the other mostly to this

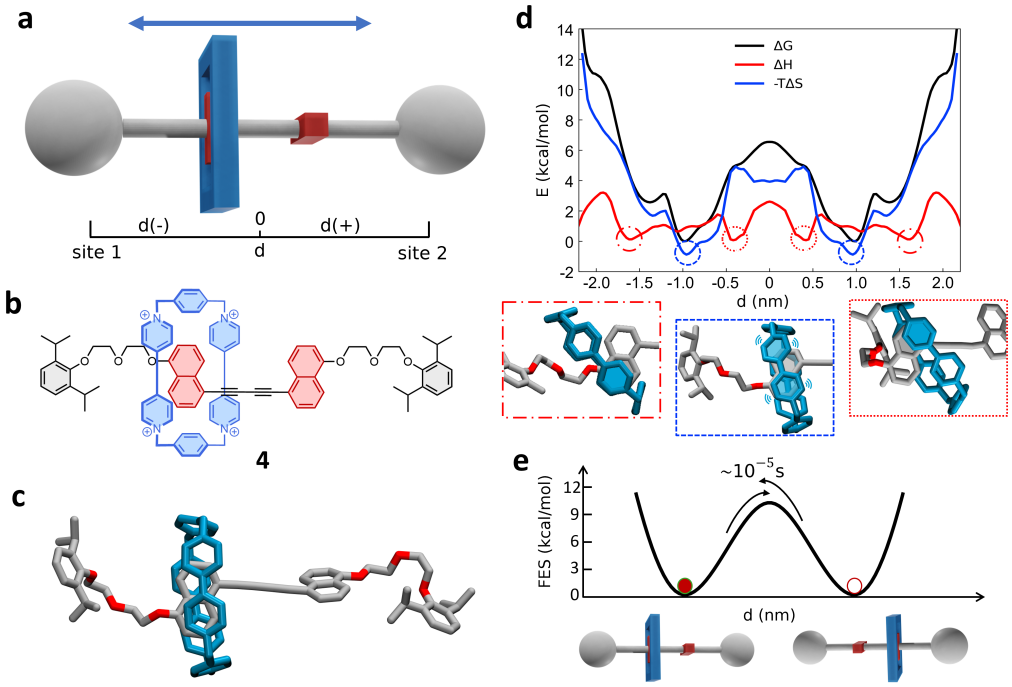

Fig. 6.10 Shuttling dynamics of rigid bistable [2]Rotaxane (**4**). (a) Schematic representation of the shuttling mechanism and the translational CV *d* used in WT-MetaD simulations. (b) Structural chemical formula and (c) atomistic molecular model of rigid bistable [2]Rotaxane (**4**). (d) Free-energy profile (black), associated with the **4** system, with enthalpic (red) and entropic (blue) contributions; representative snapshots of three relevant states (indicated in the plot by dashed, dotted and dot-dashed circles) are displayed in the three boxes below. (e) FES diagram showing the two free-energy minima in which the blue-box is located at the two recognition sites. The arrow indicates the transition of the blue-box between the two recognition sites, with the associated transition time. Adapted from Ref. [405].

constraint. We characterized the kinetics of the transition, obtaining a transition time of tenths of a second (see the Appendix, Fig. A11, corresponding to a free-energy barrier of $\sim$ 10.7 kcal mol^{-1}, in good agreement with the experimental result (9.7 kcal mol^{-1}).

Overall, our simulation approach allows characterizing this system with unprecedented detail, collecting data that complement the experimental results [426]. Interestingly, we demonstrated that the streamlined, bistable dynamics of this system result from subtle entropic effects, which need to be understood to rationally regulate the switching kinetics of the rotaxane.

6.3 Conclusions

In this last Chapter we have reported an extensive *in silico* investigation of the dynamics characterizing four notable examples of MIMs. By means of AA molecular modelling and enhanced sampling techniques based on standard and infrequent WT-MetaD approaches, we designed a protocol to access the dynamics of these rotaxanes with atomistic detail, in order to elucidate the main molecular factors guiding their shuttling motion. We have selected the four test cases presented herein in an effort to consider variable architectures and dynamics behaviours, that are common in the family of rotaxanes and of relevance for the various foreseen applications. The results of our simulations provide unprecedented insight into the free-energy characterization of this kind of system, allowing to determine the most favourable states that the interlocked molecules populate at the equilibrium. To reach a deeper comprehension into the factors that regulate the population of different molecular states, we have exploited the potential of the chosen computational techniques to uncover the enthalpy and entropy contributions to the free-energy associated to the system. This allowed us to discover how the most stable states of MIMs are often related to a dynamical regime, and how a delicate balance of intermolecular interactions, *e.g.* H-bonding, and conformational entropy determines the final performance of such molecular machines. We also pushed our analysis forward, capturing the kinetics of the transitions that connect the different states of such rotaxane molecules. The infrequent WT-MetaD protocol enabled the calculation of characteristic times associated to the most relevant transitions in each system. The results of our kinetic study are in considerable agreement with available experimental results, confirming the validity of our approach, which is thus extendable to a large variety of MIMs. This kinetic characterization is crucial in the context of MIMs, where a solid understanding of transition times is essential to exploit such molecular architectures for the foreseen engineering applications, such as molecular motors, nanowires and switches.

6.4 Methods

6.4.1 Simulation setup

All the model systems and ions were parametrized using the Generalized Amber Force Field (GAFF) [223], except for the β dihedral angle of **2**. The partial charges for the atomistic models have been calculated using the RESP [224] method, computed at Hartree-Fock (HF) level of theory, with 6-31G basis set, as implemented in Gaussian16 [300]. AmberTools21 [376] package was used for the parametrization of the systems. The solvents parameters were obtained from the *virtualchemistry.org* database [461, 462]. A proper amount of PF_6^- ions have been added to neutralize the total charge of the systems. All the systems simulated herein have been minimized via steepest-descent and equilibrated in NPT condition (constant N: number of particles, P: pressure, T: temperature). Production runs were conducted in the NPT ensemble for up to 2 μs. Temperature and pressure are kept constant (298 K and 1 bar for **1 2 3 4**) using the v-rescale thermostat[261] and the Berendsen[257], and Parrinello-Rahman[262] barostats (see A1 for details). Long-range electrostatic interactions have been treated using the Particle-Mesh-Ewald (PME)[347] and a cutoff of 1.0 nm. Van der Waals interactions were truncated at 1.0 nm. The LINCS algorithm[246] has been used to constrain all the bonds involving hydrogens. The leap frog integrator was used to propagate dynamics with a time step of 2 fs.
All the simulations were carried out using GROMACS[295] 2021.6, patched with PLUMED, version 2.7[296, 297].

The choice of the CVs and all the technical details of the simulations, and data not included in the main text of this Thesis are discussed in the Appendix.

Chapter 7

Conclusions

This thesis collects different types of computational works that I have conducted along my PhD in the attempt to characterize and understand the dynamical behaviour of various supramolecular systems in various conditions and on different scales. Supramolecular polymers and superassemblies, as well as molecular machines, have been studied through the use of advanced molecular modelling techniques and all-atom and coarse-grained models. The multiscale approach allowed us to achieve a thorough comprehension of the underlying physical-chemical mechanisms spanning all the relevant time and space scales in a variety of systems of interest, unvealing the key factors controlling them.

Firstly, we studied the chain-capping and supramolecular polymerization of multi-component systems of porphyrin-based supramolecular polymers. This has been achieved by using Well-Tempered Metadynamics and coarse-grained models, validated through simulations at atomistic resolution. The predictions of these models helped to study the interaction between the monomers, quantifying the relative stability of the different aggregate species and probe the mechanism of supramolecular polymerization, revealing that multi-step processes (adsorb-slide-stack) are likely involved in the exchange phenomena.

Afterwards, climbing the hierarchical ladder, we investigated the interaction between isolated nanotubes in solution and the mechanism of nanotube intertwining. Our multiscale molecular modelling approach, combined with umbrella sampling simulations, revealed that the intertwining of the nanotubes is primarily driven by solvophobic effects. Through the use of coarse-grained models – carefully developed

to reproduce structural features of the atomistic one – we were able to simulate the large-scale folding of nanotubes and effectively reproduce different experimental conditions.

After gaining insights into the interaction between supramolecular monomers, we extended our investigations to their interactions with biological entities, with a particular focus on the potential of BTA-based supramolecular polymers as biomaterials. Our coarse-grained classical molecular dynamics and Metadynamics simulations revealed the interaction between BTA-OEG_4 and BSA and elucidated the mechanism of monomer exchange in this aggregate.

Lastly, in the main project of my PhD research, a flexible computational framework – based on Well-Tempered Metadynamics and its "infrequent" variant – has been succesfully applied to resolve the dynamics of a variety of mechanically interlocked molecules (*e.g.* rotaxanes, molecular shuttles). Enthalpy/entropy decomposition of the free-energy profiles allowed us to quantify the relative contributions to the motion at the molecular level. The effect of different solvents on the dynamics of molecular shuttles has also been evaluated.

Overall, the work reported in this Thesis has contributed to increasing our knowledge of the behaviour of supramolecular systems. By providing a detailed molecular-level understanding of the mechanisms governing self-assembly or the monomer exchange events in supramolecular fibres our simulations were instrumental in complementing and resolving unclear/ambiguous experimental findings and shedding light on the underlying principles of supramolecular polymers. The careful utilization of enhanced sampling methods allowed us to capture the slow dynamics (in the regime of all-atom simulations) observed in rotaxanes with sub-molecular resolution, providing unprecedented insights into their mechanism of motion. This work also represents, to the best of our knowledge, the first case of Metadynamics simulations in the field of rotaxanes and molecular shuttles, and – given the versatility of the approach – may be applied to a variety of MIMs.

However, despite its innovative aspects, this work is not without limitations. One is that the choice of the CVs are often based on chemical/physical intuitions and prior knowledge of the problem under investigation. This choice undoubtfully simplifies the description of the rare-event, leading to a smoothening of the potential energy surface. This might be one of the reason why some kinetic barriers are often underestimated using standard Well-Tempered Metadynamics, and we need to rely

on its *ad hoc* "infrequent" variant. For example, if the solvent drives or hinders some mechanisms (as in Chapter 4), a more complex coordinate that involves solvent degrees of freedom is needed[212]. One potential solution for overcoming this limitation and enabling new research directions is the application of machine-learning techniques to optimize the choice of our CVs[463–466]. For example, the group of Parrinello proposed the method called "deep-LDA", which combines neural networks and linear discriminant analysis (LDA) to – provided unbiased simulations from the two different basins – identify a linear combination of descriptors that distinguish between them. Lamim Ribeiro et al. combined Automatic Mutual Information Noise Omission (AMINO)[467] and Reweighted Autoencoded Variational Bayes for Enhanced sampling (RAVE) method[468] to learn optimal reaction coordinates for infrequent metadynamics simulations of drug dissociation[469]. Other approaches can be also useful to this end. One method is the "bias-exchange metadynamics" proposed by Piana and Laio[470]. This technique, based on the replica-exchange scheme[471], exchanges between replicas performed at the same temperature, but with the bias potential acting on different variables. With this approach, a large number of different variables can be simultaneously biased, improving the sampling of the multidimensional configurational space and overcoming barriers that are orthogonal to the reaction coordinates. The application of similar methodologies to derive a low-dimensional set of CVs capable of driving, *e.g.* the translational motion of MIMs without neglecting relevant slow modes can significantly extend the power of the computational approach illustrated in Chapter 4 for more challenging cases currently under study (see, *e.g.*, the system in Fig. 7.1c).

Another area where further improvement can be made is in the estimation of transition rates. In fact, the necessary conditions of ideal CVs and no bias deposition over the transition rates are not always met. The group of Pietrucci showed that Kramer's theory provide more accurate estimates of the transition rates between different states from biased simulations[472].

Further perspectives and related activities

From the research presented in this Thesis, several potential avenues emerge for further investigation. One opportunity is the development of a minimalistic coarse-grained model to study the self-assembly of porphyrin-based monomers. Porphyrin groups may coordinate different metal in their centres, which may impact their inter-

actions, aggregation pathways, and properties. By adjusting the interactions between the cores, it may be possible to model different metal centres or solvent conditions – as done in Chapter 2 – and capture the different aggregation pathways leading to J- or H- aggregates. Additionally, adjusting the strength of bonds or the length of arms could model changes in the structure of the monomers. Such a simplified model allows simulating a wide range of possible monomers in different conditions with limited computational efforts. This approach could significantly boost the rational design of these materials, leading to improvements in their performance in optoelectronic devices.

The development of customized minimalistic coarse-grained models is also currently being evaluated as a promising approach for investigating the later stages of MOF nucleation. While early stages of MOF nucleation are extensively studied at atomistic level[293, 473, 474], the later stages remain a challenge. To address this issue, we are currently collaborating with UCL on a project focused on developing minimalistic coarse-grained models of MLA, MLB, and MLC sub-units, which interact with each other via a simple LJ potential. These models will enable us to study the assembly of these MOFs on a larger scale and longer timescales, modelling various conditions and combinations of sub-units only by changing the strength of the interactions between the particles (Fig. 7.1a).

A relevant future work from this thesis would be also to develop coarse-grained models of porphyrin-based monomers compatible with the MARTINI 3 force field. In fact, while porphyrins are important both for proteins simulations and supramolecular chemistry, creating a coarse-grained model for this complex molecule is a non-trivial task. Using a bottom-up approach, the first big challenge involves selecting an appropriate level of theory to describe the metal-centre, due to the complex chemistry of these species. Then, one need to determine the force-field parameters to perform the atomistic simulations and generate the reference data. The final challenge is to map the atomistic structure, which is a relatively simpler task compared to the previous steps. Several mapping schemes have been proposed in the literature for porphyrins, and there are also automatic coarse-graining mapping algorithms that can be utilized[475–477] (Fig. 7.1b).
We plan to create a comprehensive framework for automating the coarse-graining process of metal porphyrins. Our proposed solution involves utilizing optimization algorithm (particle-swarm optimization is currently under evaluation) to iteratively adjust CG parameters, in order to match radial distribution functions (RDFs), order

parameters, dimerization free-energies, or other functions, from all-atom simulations. The ultimate goal of this research is to create a large database of CG-models of several metal-porphyrins, that will be shared with the community on the Martini Database server[478], thereby facilitating the preparation of porphyrin-based complex systems.

The computational approach presented in Chapter 4 can be extended to asymmetric interlocked architectures. Asymmetric rotaxanes have several advantages over symmetric rotaxanes, including greater control over the motion and increased functionality. This makes them a particularly relevant and promising target for experimentalist interested in developing more complex and functional molecular machines. Therefore, the atomistic-level details provided by our simulations can be fundamental for the engineering of molecular motors, switches, and molecular machines.

At last, we are currently collaborating with the prestigious group of Prof. Alberto Credi, a renowned expert in the field of molecular machines. This collaboration is aimed at elucidating the operation cycle of a chemically driven rotary motor based on a catenane. Through the use of the well-tested computational framework in Chapter 4, we will focus on determining the transition rates from one state to another, calculating the relative free-energies of the most relevant states and their thermodynamic balance (Fig. 7.1c).

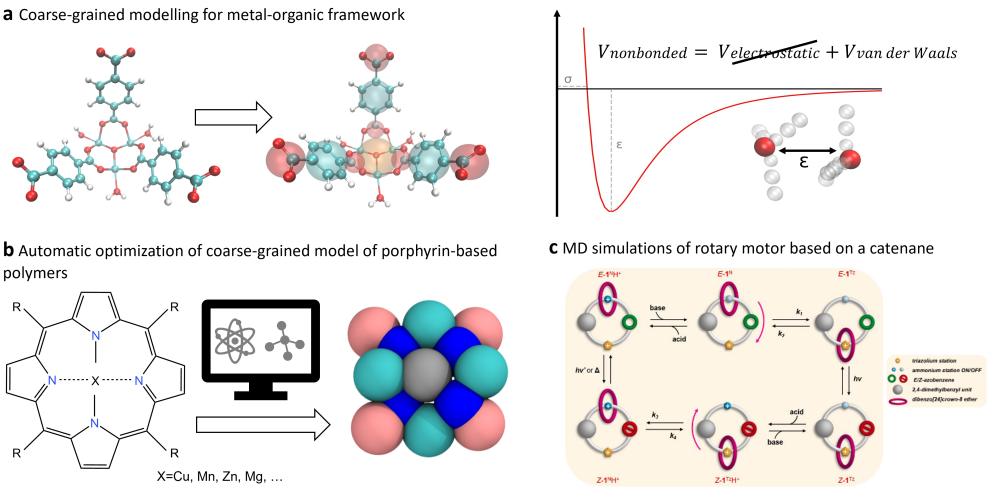

Fig. 7.1 Further perspective from this PhD Thesis: (a) Development of minimalistic coarse-grained models of MLA, MLB, and MLC sub-units, which interact with each other via a simple Lennard-Jones potential, to simulate the early stages of nucleation on larger scales and longer timescales. (b) Automatic parametrization of porphyrin-based monomers, using MARTINI 3 force field[242]. (c) Molecular simulations of a chemically driven rotary motor based on a catenane, a collaborative project with the group of Prof. A. Credi.

References

[1] Douglas Philp and J Fraser Stoddart. Self-assembly in natural and unnatural systems. *Angewandte Chemie International Edition in English*, 35(11):1154–1196, 1996.

[2] Philip Ball. *The self-made tapestry: pattern formation in nature*. Oxford University Press, Inc., 2001.

[3] George M Whitesides and Bartosz Grzybowski. Self-assembly at all scales. *Science*, 295(5564):2418–2421, 2002.

[4] Bartosz A Grzybowski, Christopher E Wilmer, Jiwon Kim, Kevin P Browne, and Kyle JM Bishop. Self-assembly: from crystals to cells. *Soft Matter*, 5(6):1110–1128, 2009.

[5] Ana C Mendes, Erkan T Baran, Rui L Reis, and Helena S Azevedo. Self-assembly in nature: using the principles of nature to create complex nanobiomaterials. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 5(6):582–612, 2013.

[6] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter, et al. Molecular biology of the cell. *Scandinavian Journal of Rheumatology*, 32(2):125–125, 2003.

[7] George M Whitesides, John P Mathias, and Christopher T Seto. Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. *Science*, 254(5036):1312–1319, 1991.

[8] George M Whitesides, Jennah K Kriebel, and Brian T Mayers. Self-assembly and nanostructured materials. In *Nanoscale assembly*, pages 217–239. Springer, 2005.

[9] Shuguang Zhang, Davide M Marini, Wonmuk Hwang, and Steve Santoso. Design of nanostructured biological materials through self-assembly of peptides and proteins. *Current opinion in chemical biology*, 6(6):865–871, 2002.

[10] Shuguang Zhang. Fabrication of novel biomaterials through molecular self-assembly. *Nature biotechnology*, 21(10):1171–1178, 2003.

[11] David B Amabilino, David K Smith, and Jonathan W Steed. Supramolecular materials. *Chemical society reviews*, 46(9):2404–2420, 2017.

[12] Samuel I Stupp, Vassou LeBonheur, K Walker, Li-Sheng Li, Kevin E Huggins, M Keser, and Amstutz Amstutz. Supramolecular materials: self-organized nanostructures. *Science*, 276(5311):384–389, 1997.

[13] Marek Grzelczak, Jan Vermant, Eric M Furst, and Luis M Liz-Marzán. Directed self-assembly of nanoparticles. *ACS nano*, 4(7):3591–3605, 2010.

[14] Abraham Ulman. Formation and structure of self-assembled monolayers. *Chemical reviews*, 96(4):1533–1554, 1996.

[15] Dennis E Discher and Adi Eisenberg. Polymer vesicles. *Science*, 297(5583):967–973, 2002.

[16] Darrin J Pochan, Zhiyun Chen, Honggang Cui, Kelly Hales, Kai Qi, and Karen L Wooley. Toroidal triblock copolymer assemblies. *Science*, 306(5693):94–97, 2004.

[17] Adam Blanazs, Steven P Armes, and Anthony J Ryan. Self-assembled block copolymer aggregates: from micelles to vesicles and their biological applications. *Macromolecular rapid communications*, 30(4-5):267–277, 2009.

[18] Hiroki Itabashi, Sougata Datta, Ryohei Tsukuda, Martin J Hollamby, and Shiki Yagai. Fine-tuning of the size of supramolecular nanotoroids suppresses the subsequent catenation of nano-[2] catenane. *Chemical Science*, 2023.

[19] Luc Brunsveld, Brigitte JB Folmer, Egbert Willem Meijer, and Rint P Sijbesma. Supramolecular polymers. *Chemical reviews*, 101(12):4071–4098, 2001.

[20] T Aida, EW Meijer, and SI Stupp. Functional supramolecular polymers. *Science*, 335(6070):813–817, 2012.

[21] Tom FA De Greef, Maarten MJ Smulders, Martin Wolffs, Albert PHJ Schenning, Rint P Sijbesma, and EW Meijer. Supramolecular polymerization. *Chemical Reviews*, 109(11):5687–5754, 2009.

[22] Takashi Kato and Jean MJ Frechet. A new approach to mesophase stabilization through hydrogen bonding molecular interactions in binary mixtures. *Journal of the American Chemical Society*, 111(22):8533–8534, 1989.

[23] Marius Wehner and Frank Würthner. Supramolecular polymerization through kinetic pathway control and living chain growth. *Nature Reviews Chemistry*, 4(1):38–53, 2020.

[24] Claudine Fouquey, Jean-Marie Lehn, and Anne-Marie Levelut. Molecular recognition directed self-assembly of supramolecular liquid crystalline polymers from complementary chiral components. *Advanced Materials*, 2(5):254–257, 1990.

[25] Ruijiao Dong, Yongfeng Zhou, Xiaohua Huang, Xinyuan Zhu, Yunfeng Lu, and Jian Shen. Functional supramolecular polymers for biomedical applications. *Advanced materials*, 27(3):498–526, 2015.

[26] Patricia YW Dankers and EW Meijer. Supramolecular biomaterials. a modular approach towards tissue engineering. *Bulletin of the Chemical Society of Japan*, 80(11):2047–2073, 2007.

[27] Rui Chen, Simona Neri, and Leonard J Prins. Enhanced catalytic activity under non-equilibrium conditions. *Nature Nanotechnology*, 15(10):868–874, 2020.

[28] Sundus Erbas-Cakmak, David A Leigh, Charlie T McTernan, and Alina L Nussbaumer. Artificial molecular machines. *Chemical reviews*, 115(18):10081–10206, 2015.

[29] Bernard R Brooks, Robert E Bruccoleri, Barry D Olafson, David J States, S a Swaminathan, and Martin Karplus. Charmm: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2):187–217, 1983.

[30] Xuzhou Yan, Feng Wang, Bo Zheng, and Feihe Huang. Stimuli-responsive supramolecular polymeric materials. *Chemical Society Reviews*, 41(18):6042–6065, 2012.

[31] Itamar Willner, Bernhard Basnar, and Bilha Willner. From molecular machines to microscale motility of objects: application as "smart materials", sensors, and nanodevices. *Advanced Functional Materials*, 17(5):702–717, 2007.

[32] Romain Barat, Thibaut Legigan, Isabelle Tranoy-Opalinski, Brigitte Renoux, Elodie Péraudeau, Jonathan Clarhaut, Pauline Poinot, Antony E Fernandes, Vincent Aucagne, David A Leigh, et al. A mechanically interlocked molecular system programmed for the delivery of an anticancer drug. *Chemical science*, 6(4):2608–2613, 2015.

[33] Emilie Moulin, Lara Faour, Christian C Carmona-Vargas, and Nicolas Giuseppone. From molecular machines to stimuli-responsive materials. *Advanced Materials*, 32(20):1906036, 2020.

[34] Vincenzo Balzani, Alberto Credi, and Margherita Venturi. Photochemical conversion of solar energy. *ChemSusChem: Chemistry & Sustainability Energy & Materials*, 1(1-2):26–58, 2008.

[35] *The cell: a molecular approach*, volume 4. 2007.

[36] Richard C Weisenberg. Microtubule formation in vitro in solutions containing low calcium concentrations. *Science*, 177(4054):1104–1105, 1972.

[37] Eva Nogales, Michael Whittaker, Ronald A Milligan, and Kenneth H Downing. High-resolution model of the microtubule. *Cell*, 96(1):79–88, 1999.

[38] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular biology of the cell*. Garland Science, 2002.

[39] Larissa KS von Krbek, Christoph A Schalley, and Pall Thordarson. Assessing cooperativity in supramolecular systems. *Chemical Society Reviews*, 46(9):2622–2637, 2017.

[40] Ivo AW Filot, Anja RA Palmans, Peter AJ Hilbers, Rutger A van Santen, Evgeny A Pidko, and Tom FA de Greef. Understanding cooperativity in hydrogen-bond-induced supramolecular polymerization: a density functional theory study. *The Journal of Physical Chemistry B*, 114(43):13667–13674, 2010.

[41] George A Jeffrey and Wolfram Saenger. *Hydrogen bonding in biological structures*. Springer Science & Business Media, 2012.

[42] Brandon H Toyama and Jonathan S Weissman. Amyloid structure: conformational diversity and consequences. *Annual review of biochemistry*, 80:557–585, 2011.

[43] Elisha Krieg, Maartje MC Bastings, Pol Besenius, and Boris Rybtchinski. Supramolecular polymers in aqueous media. *Chemical reviews*, 116(4):2414–2477, 2016.

[44] Tuomas PJ Knowles and Markus J Buehler. Nanomechanics of functional and pathological amyloid materials. *Nature nanotechnology*, 6(8):469–479, 2011.

[45] Jean-Marie Lehn. Supramolecular chemistry—scope and perspectives molecules, supermolecules, and molecular devices (nobel lecture). *Angewandte Chemie International Edition in English*, 27(1):89–112, 1988.

[46] Jean-Marie Lehn. Supramolecular polymer chemistry—scope and perspectives, 2002.

[47] Rint P Sijbesma, Felix H Beijer, Luc Brunsveld, Brigitte JB Folmer, JHK Ky Hirschberg, Ronald FM Lange, Jimmy KL Lowe, and EW Meijer. Reversible polymers formed from self-complementary monomers using quadruple hydrogen bonding. *Science*, 278(5343):1601–1604, 1997.

[48] Patricia YW Dankers, Thomas M Hermans, Travis W Baughman, Yuko Kamikawa, Roxanne E Kieltyka, Maartje MC Bastings, Henk M Janssen, Nico AJM Sommerdijk, Antje Larsen, Marja JA Van Luyn, et al. Hierarchical formation of supramolecular transient networks in water: a modular injectable delivery system. *Advanced materials*, 24(20):2703–2709, 2012.

[49] Maartje MC Bastings, Stefan Koudstaal, Roxanne E Kieltyka, Yoko Nakano, ACH Pape, Dries AM Feyen, Frebus J Van Slochteren, Pieter A Doevendans, Joost PG Sluijter, EW Meijer, et al. A fast ph-switchable and self-healing supramolecular hydrogel carrier for guided, local catheter injection in the infarcted myocardium. *Advanced healthcare materials*, 3(1):70–78, 2014.

[50] Philippe Cordier, François Tournilhac, Corinne Soulié-Ziakovic, and Ludwik Leibler. Self-healing and thermoreversible rubber from supramolecular assembly. *Nature*, 451(7181):977–980, 2008.

[51] Vesna Simic, Laurent Bouteiller, and Matthieu Jalabert. Highly cooperative formation of bis-urea based supramolecular polymers. *Journal of the American Chemical Society*, 125(43):13148–13154, 2003.

[52] Liulin Yang, Xinxin Tan, Zhiqiang Wang, and Xi Zhang. Supramolecular polymers: historical development, preparation, characterization, and functions. *Chemical reviews*, 115(15):7196–7239, 2015.

[53] Perry S Corbin and Steven C Zimmerman. Self-association without regard to prototropy. a heterocycle that forms extremely stable quadruply hydrogen-bonded dimers. *Journal of the American Chemical Society*, 120(37):9710–9711, 1998.

[54] Florian Herbst, Diana Döhler, Philipp Michael, and Wolfgang H Binder. Self-healing polymers via supramolecular forces. *Macromolecular rapid communications*, 34(3):203–220, 2013.

[55] Seda Cantekin, Tom FA de Greef, and Anja RA Palmans. Benzene-1, 3, 5-tricarboxamide: a versatile ordering moiety for supramolecular chemistry. *Chemical Society Reviews*, 41(18):6125–6137, 2012.

[56] AV Gorbunov, Tristan Putzeys, I Urbanavičiūtė, RAJ Janssen, Michael Wübbenhorst, RP Sijbesma, and Martijn Kemerink. True ferroelectric switching in thin films of trialkylbenzene-1, 3, 5-tricarboxamide (bta). *Physical Chemistry Chemical Physics*, 18(34):23663–23672, 2016.

[57] Indre Urbanaviciute, Xiao Meng, Tim D Cornelissen, Andrey V Gorbunov, Subham Bhattacharjee, Rint P Sijbesma, and Martijn Kemerink. Tuning the ferroelectric properties of trialkylbenzene-1, 3, 5-tricarboxamide (bta). *Advanced Electronic Materials*, 3(7):1600530, 2017.

[58] Xiao Meng, Andrey V Gorbunov, WS Christian Roelofs, Stefan CJ Meskers, René AJ Janssen, Martijn Kemerink, and Rint P Sijbesma. Ferroelectric switching and electrochemistry of pyrrole substituted trialkylbenzene-1, 3, 5-tricarboxamides. *Journal of Polymer Science Part B: Polymer Physics*, 55(8):673–683, 2017.

[59] Jianyun Wu, Takashi Takeda, Norihisa Hoshino, and Tomoyuki Akutagawa. Ferroelectric low-voltage on/off switching of chiral benzene-1, 3, 5-tricarboxamide derivative. *Journal of Materials Chemistry C*, 8(30):10283–10289, 2020.

[60] Pieter Fransen Martijn Arnoldus Johannes Veld Anja Rita Alberta Palmans Steinar Ouren, Henricus Marie Janssen.

[61] Neralagatta M Sangeetha and Uday Maitra. Supramolecular gels: Functions and uses. *Chemical Society Reviews*, 34(10):821–836, 2005.

[62] Andrew R Hirst, Beatriu Escuder, Juan F Miravet, and David K Smith. High-tech applications of self-assembling supramolecular nanostructured gel-phase materials: from regenerative medicine to electronic devices. *Angewandte Chemie International Edition*, 47(42):8002–8018, 2008.

[63] Pol Besenius, Joeri LM Heynens, Roel Straathof, Marko ML Nieuwenhuizen, Paul HH Bomans, Enzo Terreno, Silvio Aime, Gustav J Strijkers, Klaas Nicolay, and EW Meijer. Paramagnetic self-assembled nanoparticles as supramolecular mri contrast agents. *Contrast Media & Molecular Imaging*, 7(3):356–361, 2012.

[64] Maarten H Bakker, Cameron C Lee, EW Meijer, Patricia YW Dankers, and Lorenzo Albertazzi. Multicomponent supramolecular polymers as a modular platform for intracellular delivery. *ACS nano*, 10(2):1845–1852, 2016.

[65] Shinpei Hasegawa, Satoshi Horike, Ryotaro Matsuda, Shuhei Furukawa, Katsunori Mochizuki, Yoshinori Kinoshita, and Susumu Kitagawa. Three-dimensional porous coordination polymer functionalized with amide groups based on tridentate ligand: selective sorption and catalysis. *Journal of the American Chemical Society*, 129(9):2607–2614, 2007.

[66] Jiawei Gu, Yi Peng, Ting Zhou, Jiao Ma, Huan Pang, and Yusuke Yamauchi. Porphyrin-based framework materials for energy conversion. *Nano Research Energy*, 1(1):e9120009, 2022.

[67] Duc Duong La, Huu Hao Ngo, Dinh Duc Nguyen, Nam T Tran, Hoang Tung Vo, X Hoan Nguyen, Soon Woong Chang, Woo Jin Chung, and M Dac-Binh Nguyen. Advances and prospects of porphyrin-based nanomaterials via self-assembly for photocatalytic applications in environmental treatment. *Coordination Chemistry Reviews*, 463:214543, 2022.

[68] Zeyuan Zhang, Lingzhi Ma, Fang Fang, Yali Hou, Chenjie Lu, Chaoqun Mu, Yafei Zhang, Haifei Liu, Ke Gao, Ming Wang, et al. Porphyrin-based multicomponent metallacage: Host–guest complexation toward photooxidation-triggered reversible encapsulation and release. *JACS Au*, 2(6):1479–1487, 2022.

[69] Michael G Walter, Alexander B Rudine, and Carl C Wamser. Porphyrins and phthalocyanines in solar photovoltaic cells. *Journal of Porphyrins and Phthalocyanines*, 14(09):759–792, 2010.

[70] Lu-Lin Li and Eric Wei-Guang Diau. Porphyrin-sensitized solar cells. *Chemical society reviews*, 42(1):291–304, 2013.

[71] Maxence Urbani, Michael Gratzel, Mohammad Khaja Nazeeruddin, and Tomas Torres. Meso-substituted porphyrins for dye-sensitized solar cells. *Chemical Reviews*, 114(24):12330–12396, 2014.

[72] Asif Mahmood, Jian-Yong Hu, Bo Xiao, Ailing Tang, Xiaochen Wang, and Erjun Zhou. Recent progress in porphyrin-based materials for organic solar cells. *Journal of Materials Chemistry A*, 6(35):16769–16797, 2018.

[73] Elizabeth Huynh and Gang Zheng. Porphysome nanotechnology: a paradigm shift in lipid-based supramolecular structures. *Nano Today*, 9(2):212–222, 2014.

[74] Manivannan Ethirajan, Yihui Chen, Penny Joshi, and Ravindra K Pandey. The role of porphyrin chemistry in tumor imaging and photodynamic therapy. *Chemical Society Reviews*, 40(1):340–362, 2011.

[75] Dennis EJGJ Dolmans, Dai Fukumura, and Rakesh K Jain. Photodynamic therapy for cancer. *Nature reviews cancer*, 3(5):380–387, 2003.

[76] Aisling E O'Connor, William M Gallagher, and Annette T Byrne. Porphyrin and nonporphyrin photosensitizers in oncology: preclinical and clinical advances in photodynamic therapy. *Photochemistry and photobiology*, 85(5):1053–1074, 2009.

[77] Ethan D Sternberg, David Dolphin, and Christian Brückner. Porphyrin-based photosensitizers for use in photodynamic therapy. *Tetrahedron*, 54(17):4151–4202, 1998.

[78] Chanchan Yu, Lanju Xu, Yuanyuan Zhang, Peter S Timashev, Yuanyu Huang, and Xing-Jie Liang. Polymer-based nanomaterials for noninvasive cancer photothermal therapy. *ACS Applied Polymer Materials*, 2(10):4289–4305, 2020.

[79] Xuejiao Song, Qian Chen, and Zhuang Liu. Recent advances in the development of organic photothermal nano-agents. *Nano Research*, 8(2):340–354, 2015.

[80] Xiaolong Liang, Xiaoda Li, Lijia Jing, Xiuli Yue, and Zhifei Dai. Theranostic porphyrin dyad nanoparticles for magnetic resonance imaging guided photodynamic therapy. *Biomaterials*, 35(24):6379–6388, 2014.

[81] Lixin Zang, Huimin Zhao, Jianyu Hua, Feng Qin, Yangdong Zheng, Zhiguo Zhang, and Wenwu Cao. Water-soluble gadolinium porphyrin as a multifunctional theranostic agent: Phosphorescence-based oxygen sensing and photosensitivity. *Dyes and Pigments*, 142:465–471, 2017.

[82] Jiyun Shi, Tracy WB Liu, Juan Chen, David Green, David Jaffray, Brian C Wilson, Fan Wang, and Gang Zheng. Transforming a targeted porphyrin theranostic agent into a pet imaging probe for cancer. *Theranostics*, 1:363, 2011.

[83] Danlei Tao, Liangzhu Feng, Yu Chao, Chao Liang, Xuejiao Song, Hairong Wang, Kai Yang, and Zhuang Liu. Covalent organic polymers based on fluorinated porphyrin as oxygen nanoshuttles for tumor hypoxia relief and enhanced

photodynamic therapy. *Advanced Functional Materials*, 28(43):1804901, 2018.

[84] Dennis T Bong, Thomas D Clark, Juan R Granja, and M Reza Ghadiri. Self-assembling organic nanotubes. *Angewandte Chemie International Edition*, 40(6):988–1011, 2001.

[85] Masahito Sano, Ayumi Kamino, Junko Okamura, and Seiji Shinkai. Noncovalent self-assembly of carbon nanotubes for construction of "cages". *Nano Letters*, 2(5):531–533, 2002.

[86] Zhichang Liu, Guoliang Liu, Yilei Wu, Dennis Cao, Junling Sun, Severin T Schneebeli, Majed S Nassar, Chad A Mirkin, and J Fraser Stoddart. Assembly of supramolecular nanotubes from molecular triangles and 1, 2-dihalohydrocarbons. *Journal of the American Chemical Society*, 136(47):16651–16660, 2014.

[87] Mohamed Basel Bazbouz and George K Stylios. Novel mechanism for spinning continuous twisted composite nanofiber yarns. *European polymer journal*, 44(1):1–12, 2008.

[88] Jonathon Howard and RL Clark. Mechanics of motor proteins and the cytoskeleton. *Appl. Mech. Rev.*, 55(2):B39–B39, 2002.

[89] H Lee Sweeney and Anne Houdusse. Structural and functional insights into the myosin motor mechanism. *Annual review of biophysics*, 39:539–557, 2010.

[90] Kristen J Verhey, Neha Kaul, and Virupakshi Soppina. Kinesin assembly and movement in cells. *Annual review of biophysics*, 40:267–288, 2011.

[91] Anthony J Roberts, Takahide Kon, Peter J Knight, Kazuo Sutoh, and Stan A Burgess. Functions and mechanics of dynein motor proteins. *Nature reviews Molecular cell biology*, 14(11):713–726, 2013.

[92] Wolfgang Junge and Nathan Nelson. Atp synthase. *Annual review of biochemistry*, 84:631–657, 2015.

[93] Paul D Boyer. The atp synthase—a splendid molecular machine. *Annual review of biochemistry*, 66(1):717–749, 1997.

[94] Daniela Stock, Clyde Gibbons, Ignacio Arechaga, Andrew GW Leslie, and John E Walker. The rotary mechanism of atp synthase. *Current opinion in structural biology*, 10(6):672–679, 2000.

[95] Vincenzo Balzani, Alberto Credi, Françisco M Raymo, and J Fraser Stoddart. Artificial molecular machines. *Angewandte Chemie International Edition*, 39(19):3348–3391, 2000.

[96] Nina PM Huck, Wolter F Jager, Ben De Lange, and Ben L Feringa. Dynamic control and amplification of molecular chirality by circular polarized light. *Science*, 273(5282):1686–1688, 1996.

[97] Ben L Feringa, Richard A van Delden, Nagatoshi Koumura, and Edzard M Geertsema. Chiroptical molecular switches. *Chemical Reviews*, 100(5):1789–1816, 2000.

[98] Ben L Feringa. In control of motion: from molecular switches to molecular motors. *Accounts of chemical research*, 34(6):504–513, 2001.

[99] Manuel Natali and Silvia Giordani. Molecular switches as photocontrollable "smart" receptors. *Chemical Society Reviews*, 41(10):4010–4029, 2012.

[100] Devens Gust, Thomas A Moore, and Ana L Moore. Molecular switches controlled by light. *Chemical Communications*, (11):1169–1178, 2006.

[101] Salma Kassem, Thomas van Leeuwen, Anouk S Lubbe, Miriam R Wilson, Ben L Feringa, and David A Leigh. Artificial molecular motors. *Chemical Society Reviews*, 46(9):2592–2621, 2017.

[102] Nagatoshi Koumura, Robert WJ Zijlstra, Richard A van Delden, Nobuyuki Harada, and Ben L Feringa. Light-driven monodirectional molecular rotor. *Nature*, 401(6749):152–155, 1999.

[103] J Fraser Stoddart. Mechanically interlocked molecules (mims)—molecular shuttles, switches, and machines (nobel lecture). *Angewandte Chemie International Edition*, 56(37):11094–11125, 2017.

[104] Jean-Pierre Sauvage and Christiane Dietrich-Buchecker. *Molecular catenanes, rotaxanes and knots: a journey through the world of molecular topology*. John Wiley & Sons, 2008.

[105] He-Ye Zhou, Ying Han, and Chuan-Feng Chen. ph-controlled motions in mechanically interlocked molecules. *Materials Chemistry Frontiers*, 4(1):12–28, 2020.

[106] Carson J Bruns and J Fraser Stoddart. Rotaxane-based molecular muscles. *Accounts of chemical research*, 47(7):2186–2199, 2014.

[107] M Consuelo Jiménez, Christiane Dietrich-Buchecker, and Jean-Pierre Sauvage. Towards synthetic molecular muscles: Contraction and stretching of a linear rotaxane dimer. *Angewandte Chemie International Edition*, 39(18):3284–3287, 2000.

[108] Maria Consuelo Jimenez-Molero, Christiane Dietrich-Buchecker, and Jean-Pierre Sauvage. Chemically induced contraction and stretching of a linear rotaxane dimer. *Chemistry–A European Journal*, 8(6):1456–1466, 2002.

[109] Jovica D Badjic, Vincenzo Balzani, Alberto Credi, Serena Silvi, and J Fraser Stoddart. A molecular elevator. *Science*, 303(5665):1845–1849, 2004.

[110] Chun-Ju Chuang, Wan-Sheung Li, Chien-Chen Lai, Yi-Hung Liu, Shie-Ming Peng, Ito Chao, and Sheng-Hsien Chiu. A molecular cage-based [2] rotaxane that behaves as a molecular muscle. *Organic Letters*, 11(2):385–388, 2009.

[111] Shouichi Tsukagoshi, Atsuhisa Miyawaki, Yoshinori Takashima, Hiroyasu Yamaguchi, and Akira Harada. Contraction of supramolecular double-threaded dimer formed by α-cyclodextrin with a long alkyl chain. *Organic Letters*, 9(6):1053–1055, 2007.

[112] Ryan E Dawson, Stephen F Lincoln, and Christopher J Easton. The foundation of a light driven molecular muscle based on stilbene and α-cyclodextrin. *Chemical communications*, (34):3980–3982, 2008.

[113] Guangyan Du, Emilie Moulin, Nicolas Jouault, Eric Buhler, and Nicolas Giuseppone. Muscle-like supramolecular polymers: integrated motion from thousands of molecular machines. *Angewandte Chemie*, 124(50):12672–12676, 2012.

[114] Yuya Tachibana, Nobuhiro Kihara, and Toshikazu Takata. Asymmetric benzoin condensation catalyzed by chiral rotaxanes tethering a thiazolium salt moiety via the cooperation of the component: can rotaxane be an effective reaction field? *Journal of the American Chemical Society*, 126(11):3438–3439, 2004.

[115] David A Leigh, Vanesa Marcos, and Miriam R Wilson. Rotaxane catalysts, 2014.

[116] Jose Berna, Mateo Alajarin, and Raul-Angel Orenes. Azodicarboxamides as template binding motifs for the building of hydrogen-bonded molecular shuttles. *Journal of the American Chemical Society*, 132(31):10741–10747, 2010.

[117] Gaku Hattori, Tetsuharu Hori, Yoshihiro Miyake, and Yoshiaki Nishibayashi. Design and preparation of a chiral ligand based on a pseudorotaxane skeleton: Application to rhodium-catalyzed enantioselective hydrogenation of enamides. *Journal of the American Chemical Society*, 129(43):12930–12931, 2007.

[118] Yong Li, Yu Feng, Yan-Mei He, Fei Chen, Jie Pan, and Qing-Hua Fan. Supramolecular chiral phosphorous ligands based on a [2] pseudorotaxane complex for asymmetric hydrogenation. *Tetrahedron Letters*, 49(18):2878–2881, 2008.

[119] Victor Blanco, Armando Carlone, Kevin D Hänni, David A Leigh, and Bartosz Lewandowski. A rotaxane-based switchable organocatalyst. *Angewandte Chemie*, 124(21):5256–5259, 2012.

[120] Victor Blanco, David A Leigh, Urszula Lewandowska, Bartosz Lewandowski, and Vanesa Marcos. Exploring the activation modes of a rotaxane-based switchable organocatalyst. *Journal of the American Chemical Society*, 136(44):15775–15780, 2014.

[121] Alberto Martinez-Cuezva, Adrian Saura-Sanmartin, Tomas Nicolas-Garcia, Cristian Navarro, Raul-Angel Orenes, Mateo Alajarin, and Jose Berna. Photoswitchable interlocked thiodiglycolamide as a cocatalyst of a chalcogeno-baylis–hillman reaction. *Chemical science*, 8(5):3775–3780, 2017.

[122] Victor Blanco, David A Leigh, and Vanesa Marcos. Artificial switchable catalysts. *Chemical Society Reviews*, 44(15):5341–5370, 2015.

[123] Massimo Baroncini, Serena Silvi, and Alberto Credi. Photo-and redox-driven artificial molecular motors. *Chemical reviews*, 120(1):200–268, 2019.

[124] Adrien Bessaguet, Quentin Blancart-Remaury, Pauline Poinot, Isabelle Opalinski, and Sébastien Papot. Stimuli-responsive catenane-based catalysts. *Angewandte Chemie International Edition*, 2022.

[125] David A Leigh, Jenny KY Wong, François Dehez, and Francesco Zerbetto. Unidirectional rotation in a mechanically interlocked molecular rotor. *Nature*, 424(6945):174–179, 2003.

[126] José V Hernández, Euan R Kay, and David A Leigh. A reversible synthetic rotary molecular motor. *Science*, 306(5701):1532–1537, 2004.

[127] Chun-Hua Lu, Alessandro Cecconello, Johann Elbaz, Alberto Credi, and Itamar Willner. A three-station dna catenane rotary motor with controlled directionality. *Nano letters*, 13(5):2303–2308, 2013.

[128] Anthony R Pease, Jan O Jeppesen, J Fraser Stoddart, Yi Luo, C Patrick Collier, and James R Heath. Switching devices based on interlocked molecules. *Accounts of chemical research*, 34(6):433–444, 2001.

[129] Ali Coskun, Jason M Spruell, Gokhan Barin, William R Dichtel, Amar H Flood, Youssry Y Botros, and J Fraser Stoddart. High hopes: can molecular electronics realise its potential? *Chemical society reviews*, 41(14):4827–4859, 2012.

[130] Charles P Collier, Gunter Mattersteig, Eric W Wong, Yi Luo, Kristen Beverly, José Sampaio, Francisco M Raymo, J Fraser Stoddart, and James R Heath. A [2] catenane-based solid state electronically reconfigurable switch. *Science*, 289(5482):1172–1175, 2000.

[131] Yu Liu, Guo-Song Chen, Yong Chen, Ning Zhang, Jing Chen, and Yan-Li Zhao. Bundle-shaped cyclodextrin- tb nano-supramolecular assembly mediated by c60: intramolecular energy transfer. *Nano letters*, 6(10):2196–2200, 2006.

[132] Zhe Zhang, Yiming Li, Bo Song, Yuan Zhang, Xin Jiang, Ming Wang, Ryan Tumbleson, Changlin Liu, Pingshan Wang, Xin-Qi Hao, et al. Intra-and intermolecular self-assembly of a 20-nm-wide supramolecular hexagonal grid. *Nature chemistry*, 12(5):468–474, 2020.

[133] Yiming Li, Gui-Fei Huo, Bingqing Liu, Bo Song, Yuan Zhang, Xiaomin Qian, Heng Wang, Guang-Qiang Yin, Alexander Filosa, Wenfang Sun, et al. Giant concentric metallosupramolecule with aggregation-induced phosphorescent emission. *Journal of the American Chemical Society*, 142(34):14638–14648, 2020.

[134] Mark Burnworth, Liming Tang, Justin R Kumpfer, Andrew J Duncan, Frederick L Beyer, Gina L Fiore, Stuart J Rowan, and Christoph Weder. Optically healable supramolecular polymers. *Nature*, 472(7343):334–337, 2011.

[135] Janne Ruokolainen, R Makinen, Mika Torkkeli, T Makela, Ritva Serimaa, G ten Brinke, and Olli Ikkala. Switching supramolecular polymeric materials with multiple length scales. *Science*, 280(5363):557–560, 1998.

[136] Freek JM Hoeben, Pascal Jonkheijm, EW Meijer, and Albertus PHJ Schenning. About supramolecular assemblies of π-conjugated systems. *Chemical reviews*, 105(4):1491–1546, 2005.

[137] Andreas Peil, Pengfei Zhan, and Na Liu. Dna origami catenanes templated by gold nanoparticles. *Small*, 16(6):1905987, 2020.

[138] Jonathan List, Elisabeth Falgenhauer, Enzo Kopperger, Günther Pardatscher, and Friedrich C Simmel. Long-range movement of large mechanically interlocked dna nanostructures. *Nature communications*, 7(1):1–7, 2016.

[139] Lorenzo Albertazzi, Daan van der Zwaag, Christianus MA Leenders, Robert Fitzner, Remco W van der Hofstad, and EW Meijer. Probing exchange pathways in one-dimensional aggregates with super-resolution microscopy. *Science*, 344(6183):491–495, 2014.

[140] David M Tiede, Ruitian Zhang, Lin X Chen, Lianhe Yu, and Jonathan S Lindsey. Structural characterization of modular supramolecular architectures in solution. *Journal of the American Chemical Society*, 126(43):14054–14062, 2004.

[141] Shiki Yagai, Yuichi Kitamoto, Sougata Datta, and Bimalendu Adhikari. Supramolecular polymers capable of controlling their topology. *Accounts of chemical research*, 52(5):1325–1335, 2019.

[142] Giacomo Mariani, Jean-Rémy Colard-Itté, Emilie Moulin, Nicolas Giuseppone, and Eric Buhler. Structural properties of contractile gels based on light-driven molecular motors: a small-angle neutron and x-ray study. *Soft Matter*, 16(16):4008–4023, 2020.

[143] Niklas Geue, Richard EP Winpenny, and Perdita E Barran. Structural characterisation methods for supramolecular chemistry that go beyond crystallography. *Chemical Society Reviews*, 2022.

[144] Martin J Hollamby, Keisuke Aratsu, Brian R Pauw, Sarah E Rogers, Andrew J Smith, Mitsuaki Yamauchi, Xu Lin, and Shiki Yagai. Simultaneous saxs and sans analysis for the detection of toroidal supramolecular polymers composed of noncovalent supermacrocycles in solution. *Angewandte Chemie*, 128(34):10044–10047, 2016.

[145] Antonio Fernandez, Jesus Ferrando-Soria, Eufemio Moreno Pineda, Floriana Tuna, Inigo J Vitorica-Yrezabal, Christiane Knappke, Jakub Ujma, Christopher A Muryn, Grigore A Timco, Perdita E Barran, et al. Making hybrid [n]-rotaxanes as supramolecular arrays of molecular electron spin qubits. *Nature communications*, 7(1):1–6, 2016.

[146] Antoine Goujon, Thomas Lang, Giacomo Mariani, Emilie Moulin, Gad Fuks, Jesus Raya, Eric Buhler, and Nicolas Giuseppone. Bistable [c 2] daisy chain rotaxanes as reversible muscle-like actuators in mechanically active gels. *Journal of the American Chemical Society*, 139(42):14825–14828, 2017.

[147] Michael Krieg, Gotthold Fläschner, David Alsteens, Benjamin M Gaub, Wouter H Roos, Gijs JL Wuite, Hermann E Gaub, Christoph Gerber, Yves F Dufrêne, and Daniel J Müller. Atomic force microscopy-based mechanobiology. *Nature Reviews Physics*, 1(1):41–57, 2019.

[148] Michael A Beuwer, MF Knopper, Lorenzo Albertazzi, Daan van der Zwaag, Wouter G Ellenbroek, EW Meijer, Menno WJ Prins, and Peter Zijlstra. Mechanical properties of single supramolecular polymers from correlative afm and fluorescence microscopy. *Polymer Chemistry*, 7(47):7260–7268, 2016.

[149] A Kis, S Kasas, B Babić, AJ Kulik, W Benoit, GAD Briggs, C Schönenberger, S Catsicas, and L Forro. Nanomechanics of microtubules. *Physical review letters*, 89(24):248101, 2002.

[150] Sung Ho Jung, Davide Bochicchio, Giovanni M Pavan, Masayuki Takeuchi, and Kazunori Sugiyasu. A block supramolecular polymer and its kinetically enhanced stability. *Journal of the American Chemical Society*, 140(33):10570–10577, 2018.

[151] Tomoya Fukui, Takayuki Uchihashi, Norihiko Sasaki, Hiroki Watanabe, Masayuki Takeuchi, and Kazunori Sugiyasu. Direct observation and manipulation of supramolecular polymerization by high-speed atomic force microscopy. *Angewandte Chemie*, 130(47):15691–15696, 2018.

[152] Rob van Der Weegen, Abraham JP Teunissen, and EW Meijer. Directing the self-assembly behaviour of porphyrin-based supramolecular systems. *Chemistry–A European Journal*, 23(15):3773–3783, 2017.

[153] Elisabeth Weyandt, Ivo AW Filot, Ghislaine Vantomme, and EW Meijer. Consequences of amide connectivity in the supramolecular polymerization of porphyrins: spectroscopic observations rationalized by theoretical modelling. *Chemistry–A European Journal*, 27(37):9700–9707, 2021.

[154] Yongbin Han, Kejun Cheng, Karen A Simon, Yanmei Lan, Preeti Sejwal, and Yan-Yeung Luk. A biocompatible surfactant with folded hydrophilic head group: Enhancing the stability of self-inclusion complexes of ferrocenyl in a β-cyclodextrin unit by bond rigidity. *Journal of the American Chemical Society*, 128(42):13913–13920, 2006.

[155] Liangliang Zhu, Dong Zhang, Dahui Qu, Qiaochun Wang, Xiang Ma, and He Tian. Dual-controllable stepwise supramolecular interconversions. *Chemical communications*, 46(15):2587–2589, 2010.

[156] Yu Liu, Yan-Li Zhao, Heng-Yi Zhang, Zhi Fan, Guo-Dong Wen, and Fei Ding. Spectrophotometric study of inclusion complexation of aliphatic alcohols by β-cyclodextrins with azobenzene tether. *The Journal of Physical Chemistry B*, 108(26):8836–8843, 2004.

[157] Mitsuaki Yamauchi, Bimalendu Adhikari, Deepak D Prabhu, Xu Lin, Takashi Karatsu, Tomonori Ohba, Nobutaka Shimizu, Hideaki Takagi, Rie Haruki, Shin-ichi Adachi, et al. Supramolecular polymerization of supermacrocycles: effect of molecular conformations on kinetics and morphology. *Chemistry–A European Journal*, 23(22):5270–5280, 2017.

[158] Aurelia Pastor and Eloísa Martínez-Viviente. Nmr spectroscopy in coordination supramolecular chemistry: A unique and powerful methodology. *Coordination Chemistry Reviews*, 252(21-22):2314–2345, 2008.

[159] Rina Rosenzweig and Lewis E Kay. Bringing dynamic molecular machines into focus by methyl-trosy nmr. *Annual review of biochemistry*, 83:291–315, 2014.

[160] Pier Lucio Anelli, Neil Spencer, and J Fraser Stoddart. A molecular shuttle. *Journal of the American Chemical Society*, 113(13):5131–5133, 1991.

[161] Ghazale Gholami, Kelong Zhu, Giorgio Baggi, Eduardo Schott, Ximena Zarate, and Stephen J Loeb. Influence of axle length on the rate and mechanism of shuttling in rigid h-shaped [2] rotaxanes. *Chemical science*, 8(11):7718–7723, 2017.

[162] Hans Wolfgang Spiess. 50th anniversary perspective: The importance of nmr spectroscopy to macromolecular science. *Macromolecules*, 50(5):1761–1777, 2017.

[163] Kelong Zhu, Giorgio Baggi, and Stephen J Loeb. Ring-through-ring molecular shuttling in a saturated [3] rotaxane. *Nature chemistry*, 10(6):625–630, 2018.

[164] Aritra Sarkar, Tejmani Behera, Ranjan Sasmal, Riccardo Capelli, Charly Empereur-Mot, Jaladhar Mahato, Sarit S Agasti, Giovanni M Pavan, Arindam Chowdhury, and Subi J George. Cooperative supramolecular block copolymerization for the synthesis of functional axial organic heterostructures. *Journal of the American Chemical Society*, 142(26):11528–11539, 2020.

[165] Amy Y Shih, Anton Arkhipov, Peter L Freddolino, Stephen G Sligar, and Klaus Schulten. Assembly of lipids and proteins into lipoprotein particles, 2007.

[166] Benedict EK Snodin, Flavio Romano, Lorenzo Rovigatti, Thomas E Ouldridge, Ard A Louis, and Jonathan PK Doye. Direct simulation of the self-assembly of a small dna origami. *ACS nano*, 10(2):1724–1737, 2016.

[167] Clément Arnarez, Jaakko J Uusitalo, Marcelo F Masman, Helgi I Ingólfsson, Djurre H De Jong, Manuel N Melo, Xavier Periole, Alex H De Vries, and Siewert J Marrink. Dry martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *Journal of chemical theory and computation*, 11(1):260–275, 2015.

[168] One-Sun Lee, Vince Cho, and George C Schatz. Modeling the self-assembly of peptide amphiphiles into fibers using coarse-grained molecular dynamics. *Nano letters*, 12(9):4907–4913, 2012.

[169] Goundla Srinivas, Dennis E Discher, and Michael L Klein. Self-assembly and properties of diblock copolymers by coarse-grain molecular dynamics. *Nature materials*, 3(9):638–644, 2004.

[170] Michael L Klein and Wataru Shinoda. Large-scale molecular dynamics simulations of self-assembling systems. *Science*, 321(5890):798–800, 2008.

[171] Daniel J Cerasale, Dominic C Ward, and Timothy L Easun. Mofs in the time domain. *Nature Reviews Chemistry*, 6(1):9–30, 2022.

[172] Peter S Kim and Robert L Baldwin. Intermediates in the folding reactions of small proteins. *Annual review of biochemistry*, 59(1):631–660, 1990.

[173] Matteo Cioni, Daniela Polino, Daniele Rapetti, Luca Pesce, Massimo Delle Piane, and Giovanni M Pavan. Innate dynamics and identity crisis of a metal surface unveiled by machine learning of atomic environments. *The Journal of Chemical Physics*.

[174] Jens Kehlet Nørskov, Thomas Bligaard, Jan Rossmeisl, and Claus Hviid Christensen. Towards the computational design of solid catalysts. *Nature chemistry*, 1(1):37–46, 2009.

[175] Fabrizio Marinelli, Fabio Pietrucci, Alessandro Laio, and Stefano Piana. A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS computational biology*, 5(8):e1000452, 2009.

[176] Pilar Cossio, Fabrizio Marinelli, Alessandro Laio, and Fabio Pietrucci. Optimizing the performance of bias-exchange metadynamics: folding a 48-residue lysm domain using a coarse-grained model. *The Journal of Physical Chemistry B*, 114(9):3259–3265, 2010.

[177] Metadynamics studies of crystal nucleation. *IUCrJ*, 2(2):256–266, 2015.

[178] Matteo Salvalaglio, Thomas Vetter, Federico Giberti, Marco Mazzotti, and Michele Parrinello. Uncovering molecular details of urea crystal growth in the presence of additives. *Journal of the American Chemical Society*, 134(41):17221–17233, 2012.

[179] Rodrigo Q Albuquerque, Andreas Timme, Roman Kress, Jürgen Senker, and Hans-Werner Schmidt. Theoretical investigation of macrodipoles in supramolecular columnar stackings. *Chemistry–A European Journal*, 19(5):1647–1657, 2013.

[180] Chidambar Kulkarni, Sandeep Kumar Reddy, Subi J George, and Sundaram Balasubramanian. Cooperativity in the stacking of benzene-1, 3, 5-tricarboxamide: the role of dispersion. *Chemical Physics Letters*, 515(4-6):226–230, 2011.

[181] Davide Bochicchio and Giovanni M Pavan. Molecular modelling of supramolecular polymers. *Advances in Physics: X*, 3(1):1436408, 2018.

[182] Divya B Korlepara and S Balasubramanian. Molecular modelling of supramolecular one dimensional polymers. *RSC advances*, 8(40):22659–22669, 2018.

[183] Ion Danila, Francois Riobe, Flavia Piron, Josep Puigmartí-Luis, John D Wallis, Mathieu Linares, Hans Ågren, David Beljonne, David B Amabilino, and Narcis Avarvari. Hierarchical chiral expression from the nano-to mesoscale in synthetic supramolecular helical fibers of a nonamphiphilic c 3-symmetrical π-functional molecule. *Journal of the American Chemical Society*, 133(21):8344–8353, 2011.

[184] Patrick Brocorens, Mathieu Linares, Carine Guyard-Duhayon, Regis Guillot, Bruno Andrioletti, Dominique Suhr, Benjamin Isare, Roberto Lazzaroni, and Laurent Bouteiller. Conformational plasticity of hydrogen bonded bis-urea supramolecular polymers. *The Journal of Physical Chemistry B*, 117(17):5379–5386, 2013.

[185] Bruno Giordano Alvarenga, Kalil Bernardino, André Farias de Moura, and Edvaldo Sabadini. Two different pathways for assembling bis-urea in benzene and toluene. *Journal of Molecular Modeling*, 24:1–10, 2018.

[186] Matthew B Baker, Lorenzo Albertazzi, Ilja K Voets, Christianus MA Leenders, Anja RA Palmans, Giovanni M Pavan, and EW Meijer. Consequences of chirality on the dynamics of a water-soluble supramolecular polymer. *Nature communications*, 6(1):6234, 2015.

[187] Matteo Garzoni, Matthew B Baker, Christianus MA Leenders, Ilja K Voets, Lorenzo Albertazzi, Anja RA Palmans, EW Meijer, and Giovanni M Pavan. Effect of h-bonding on order amplification in the growth of a supramolecular polymer in water. *Journal of the American Chemical Society*, 138(42):13985–13995, 2016.

[188] Carlos-Andres Palma, Paolo Samorì, and Marco Cecchini. Atomistic simulations of 2d bicomponent self-assembly: from molecular recognition to self-healing. *Journal of the American Chemical Society*, 132(50):17880–17885, 2010.

[189] One-Sun Lee, Samuel I Stupp, and George C Schatz. Atomistic molecular dynamics simulations of peptide amphiphile self-assembly into cylindrical nanofibers. *Journal of the American Chemical Society*, 133(10):3677–3683, 2011.

[190] Christopher Maffeo, Jejoong Yoo, and Aleksei Aksimentiev. De novo reconstruction of dna origami structures through atomistic molecular dynamics simulation. *Nucleic acids research*, 44(7):3013–3019, 2016.

[191] Fatima Chami and Mark R Wilson. Molecular order in a chromonic liquid crystal: a molecular simulation study of the anionic azo dye sunset yellow. *Journal of the American Chemical Society*, 132(22):7794–7802, 2010.

[192] Karteek K Bejagam, Giacomo Fiorin, Michael L Klein, and Sundaram Balasubramanian. Supramolecular polymerization of benzene-1, 3, 5-tricarboxamide: a molecular dynamics simulation study. *The Journal of Physical Chemistry B*, 118(19):5218–5228, 2014.

[193] Jerome Henin, Giacomo Fiorin, Christophe Chipot, and Michael L Klein. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *Journal of chemical theory and computation*, 6(1):35–47, 2010.

[194] Hadi H Arefi and Takeshi Yamamoto. Communication: Self-assembly of a model supramolecular polymer studied by replica exchange with solute tempering. *The Journal of Chemical Physics*, 147(21):211102, 2017.

[195] Karteek K Bejagam and Sundaram Balasubramanian. Supramolecular polymerization: a coarse grained molecular dynamics study. *The Journal of Physical Chemistry B*, 119(17):5738–5746, 2015.

[196] Davide Bochicchio and Giovanni M Pavan. From cooperative self-assembly to water-soluble supramolecular polymers using coarse-grained simulations. *ACS nano*, 11(1):1000–1011, 2017.

[197] Davide Bochicchio, Matteo Salvalaglio, and Giovanni M Pavan. Into the dynamics of a supramolecular polymer at submolecular resolution. *Nature Communications*, 8(1):147, 2017.

[198] Martina Crippa, Claudio Perego, Anna L de Marco, and Giovanni M Pavan. Molecular communications in complex systems of dynamic supramolecular polymers. *Nature Communications*, 13(1):2162, 2022.

[199] Haohao Fu, Xueguang Shao, Christophe Chipot, and Wensheng Cai. The lubricating role of water in the shuttling of rotaxanes. *Chemical science*, 8(7):5087–5094, 2017.

[200] Andranik Kazaryan, Jos CM Kistemaker, Lars V Schafer, Wesley R Browne, Ben L Feringa, and Michael Filatov. Understanding the dynamics behind the photoisomerization of a light-driven fluorene molecular rotary motor. *The Journal of Physical Chemistry A*, 114(15):5058–5067, 2010.

[201] Arjen Cnossen, Jos CM Kistemaker, Tatsuo Kojima, and Ben L Feringa. Structural dynamics of overcrowded alkene-based molecular motors during thermal isomerization. *The Journal of Organic Chemistry*, 79(3):927–935, 2014.

[202] Changfeng Fang, Baswanth Oruganti, and Bo Durbeej. Computational study of the working mechanism and rate acceleration of overcrowded alkene-based light-driven rotary molecular motors. *RSC Advances*, 4(20):10240–10251, 2014.

[203] Seung Soon Jang, Yun Hee Jang, Yong-Hoon Kim, William A Goddard, Jang Wook Choi, James R Heath, Bo W Laursen, Amar H Flood, J Fraser Stoddart, Kasper Nørgaard, et al. Molecular dynamics simulation of amphiphilic bistable [2] rotaxane langmuir monolayers at the air/water interface. *Journal of the American Chemical Society*, 127(42):14804–14816, 2005.

[204] Sandra M Mendoza, Caroline M Whelan, Jukka-Pekka Jalkanen, Francesco Zerbetto, Francesco G Gatti, Euan R Kay, David A Leigh, Monika Lubomska, and Petra Rudolf. Experimental and theoretical study of the adsorption of fumaramide [2] rotaxane on au (111) and ag (111) surfaces. *The Journal of chemical physics*, 123(24):244708, 2005.

[205] Costantino Zazza, Giordano Mancini, Giuseppe Brancato, Nico Sanna, and Vincenzo Barone. Neutral molecular shuttle in acetonitrile dilute solution investigated by molecular dynamics and density functional theory. *Computational and Theoretical Chemistry*, 985:53–61, 2012.

[206] Peng Liu, Wensheng Cai, Christophe Chipot, and Xueguang Shao. Thermodynamic insights into the dynamic switching of a cyclodextrin in a bistable molecular shuttle. *The Journal of Physical Chemistry Letters*, 1(12):1776–1780, 2010.

[207] Ying Liu, Christophe Chipot, Xueguang Shao, and Wensheng Cai. How does the solvent modulate shuttling in a pillararene/imidazolium [2] rotaxane? insights from free energy calculations. *The Journal of Physical Chemistry C*, 120(11):6287–6293, 2016.

[208] Shuangshuang Wang, Xueguang Shao, and Wensheng Cai. Solvent and structure effects on the shuttling in pillar [5] arene/triazole rotaxanes. *The Journal of Physical Chemistry C*, 121(45):25547–25553, 2017.

[209] Shuangli Du, Haohao Fu, Xueguang Shao, Christophe Chipot, and Wensheng Cai. Water-controlled switching in rotaxanes. *The Journal of Physical Chemistry C*, 122(16):9229–9234, 2018.

[210] Haohao Fu, Christophe Chipot, Wensheng Cai, and Xueguang Shao. Repurposing existing molecular machines through accurate regulation of cooperative motions. *The Journal of Physical Chemistry Letters*, 12(1):613–619, 2020.

[211] A.R. Leach. *Molecular Modelling: Principles and Applications*. Prentice Hall, 2001.

[212] Mark Tuckerman. *Statistical mechanics: theory and molecular simulation*. Oxford university press, 2010.

[213] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[214] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.

[215] Leslie Greengard and Vladimir Rokhlin. A fast algorithm for particle simulations. *Journal of computational physics*, 73(2):325–348, 1987.

[216] John A Barker and Robert O Watts. Monte carlo studies of the dielectric properties of water-like models. *Molecular Physics*, 26(3):789–792, 1973.

[217] Hendrik Antoon Lorentz. Ueber die anwendung des satzes vom virial in der kinetischen theorie der gase. *Annalen der physik*, 248(1):127–136, 1881.

[218] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Computational science. Elsevier Science, 2001.

[219] Michael JS Dewar, Eve G Zoebisch, Eamonn F Healy, and James JP Stewart. Development and use of quantum mechanical molecular models. 76. am1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, 107(13):3902–3909, 1985.

[220] Chris Oostenbrink, Alessandra Villa, Alan E Mark, and Wilfred F Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *Journal of computational chemistry*, 25(13):1656–1676, 2004.

[221] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[222] William L Jorgensen and Julian Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[223] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

[224] Christopher I Bayly, Piotr Cieplak, Wendy Cornell, and Peter A Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.

[225] Araz Jakalian, Bruce L Bush, David B Jack, and Christopher I Bayly. Fast, efficient generation of high-quality atomic charges. am1-bcc model: I. method. *Journal of computational chemistry*, 21(2):132–146, 2000.

[226] Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-grained protein models and their applications. *Chemical reviews*, 116(14):7898–7936, 2016.

[227] W Schommers. Pair potentials in disordered many-particle systems: A study for liquid gallium. *Physical Review A*, 28(6):3599, 1983.

[228] Alexander P Lyubartsev and Aatto Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Physical Review E*, 52(4):3730, 1995.

[229] Furio Ercolessi and James B Adams. Interatomic potentials from first-principles calculations: the force-matching method. *EPL (Europhysics Letters)*, 26(8):583, 1994.

[230] Aviel Chaimovich and M Scott Shell. Coarse-graining errors and numerical optimization using a relative entropy framework. *The Journal of chemical physics*, 134(9):094112, 2011.

[231] E Brini and NFA Van der Vegt. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *The Journal of Chemical Physics*, 137(15):154113, 2012.

[232] Victor Ruhle, Christoph Junghans, Alexander Lukyanov, Kurt Kremer, and Denis Andrienko. Versatile object-oriented toolkit for coarse-graining applications. *Journal of chemical theory and computation*, 5(12):3211–3223, 2009.

[233] Nicholas JH Dunn, Kathryn M Lebold, Michael R DeLyser, Joseph F Rudzinski, and WG Noid. Bocs: Bottom-up open-source coarse-graining software. *The Journal of Physical Chemistry B*, 122(13):3363–3377, 2017.

[234] Alexander Mirzoev, Lars Nordenskiöld, and Alexander Lyubartsev. Magic v. 3: An integrated software package for systematic structure-based coarse-graining. *Computer Physics Communications*, 237:263–273, 2019.

[235] James A Graham, Jonathan W Essex, and Syma Khalid. Pycgtool: automated generation of coarse-grained molecular dynamics models from atomistic trajectories. *Journal of chemical information and modeling*, 57(4):650–656, 2017.

[236] Charly Empereur-Mot, Luca Pesce, Giovanni Doni, Davide Bochicchio, Riccardo Capelli, Claudio Perego, and Giovanni M Pavan. Swarm-cg: automatic parametrization of bonded terms in martini-based coarse-grained models of simple to complex molecules via fuzzy self-tuning particle swarm optimization. *ACS omega*, 5(50):32823–32843, 2020.

[237] Charly Empereur-Mot, Riccardo Capelli, Mattia Perrone, Cristina Caruso, Giovanni Doni, and Giovanni M Pavan. Automatic multi-objective optimization of coarse-grained lipid force fields using swarmcg. *The Journal of Chemical Physics*, 156(2):024801, 2022.

[238] Charly Empereur-mot, Kasper B Pedersen, Cristina Caruso, Mattia Perrone, Martina Crippa, Paulo CT Souza, Siewert J Marrink, and Giovanni M Pavan. On the automatic optimization of lipid models in the martini force field using swarmcg. 2023.

[239] Marco S Nobile, Paolo Cazzaniga, Daniela Besozzi, Riccardo Colombo, Giancarlo Mauri, and Gabriella Pasi. Fuzzy self-tuning pso: A settings-free algorithm for global optimization. *Swarm and evolutionary computation*, 39:70–85, 2018.

[240] Siewert J Marrink, Alex H De Vries, and Alan E Mark. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B*, 108(2):750–760, 2004.

[241] Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. The martini force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B*, 111(27):7812–7824, 2007.

[242] Paulo CT Souza, Riccardo Alessandri, Jonathan Barnoud, Sebastian Thallmair, Ignacio Faustino, Fabian Grünewald, Ilias Patmanidis, Haleh Abdizadeh, Bart MH Bruininks, Tsjerk A Wassenaar, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature methods*, 18(4):382–388, 2021.

[243] Xavier Periole, Marco Cavalli, Siewert-Jan Marrink, and Marco A Ceruso. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *Journal of chemical theory and computation*, 5(9):2531–2543, 2009.

[244] Adolfo B Poma, Marek Cieplak, and Panagiotis E Theodorakis. Combining the martini and structure-based coarse-grained approaches for the molecular dynamics studies of conformational transitions in proteins. *Journal of Chemical Theory and Computation*, 13(3):1366–1374, 2017.

[245] Xavier Periole and Siewert-Jan Marrink. *The Martini Coarse-Grained Force Field*, pages 533–565. Humana Press, Totowa, NJ, 2013.

[246] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of computational chemistry*, 18(12):1463–1472, 1997.

[247] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341, 1977.

[248] Loup Verlet. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical review*, 159(1):98, 1967.

[249] William C Swope, Hans C Andersen, Peter H Berens, and Kent 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*, 76(1):637–649, 1982.

[250] Roger Williams Hockney, SP Goel, and JW Eastwood. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14(2):148–158, 1974.

[251] Ryogo Kubo, Morikazu Toda, and Natsuki Hashitsume. *Statistical physics II: nonequilibrium statistical mechanics*, volume 31. Springer Science & Business Media, 2012.

[252] Robert Zwanzig. *Nonequilibrium statistical mechanics*. Oxford university press, 2001.

[253] WF Van Gunsteren and HJC Berendsen. Algorithms for brownian dynamics. *Molecular Physics*, 45(3):637–647, 1982.

[254] Wilfred F Van Gunsteren and Herman JC Berendsen. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, 1(3):173–185, 1988.

[255] MP Allen. Brownian dynamics simulation of a chemical reaction in solution. *Molecular Physics*, 40(5):1073–1087, 1980.

[256] N Goga, AJ Rzepiela, AH De Vries, SJ Marrink, and HJC Berendsen. Efficient algorithms for langevin and dpd dynamics. *Journal of chemical theory and computation*, 8(10):3637–3649, 2012.

[257] Herman JC Berendsen, JPM van Postma, Wilfred F Van Gunsteren, ARHJ DiNola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.

[258] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52(2):255–268, 1984.

[259] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*, 31(3):1695, 1985.

[260] Hans C Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics*, 72(4):2384–2393, 1980.

[261] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[262] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

[263] Shuichi Nosé and ML Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50(5):1055–1076, 1983.

[264] Peter G Bolhuis, David Chandler, Christoph Dellago, and Phillip L Geissler. Transition path sampling: Throwing ropes. *Annu. Rev. Phys. Chem*, 53:291–318, 2002.

[265] Christophe Chipot and Andrew Pohorille. Free energy calculations : theory and applications in chemistry and biology. 2007.

[266] Anna S Kamenik, Stephanie M Linker, and Sereina Riniker. Enhanced sampling without borders: on global biasing functions and how to reweight them. *Physical Chemistry Chemical Physics*, 24(3):1225–1236, 2022.

[267] Jérôme Hénin, Tony Lelièvre, Michael R Shirts, Omar Valsson, and Lucie Delemotte. Enhanced sampling methods for molecular dynamics simulations. *arXiv preprint arXiv:2202.04164*, 2022.

[268] Omar Valsson, Pratyush Tiwary, and Michele Parrinello. Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annual review of physical chemistry*, 67:159–184, 2016.

[269] Eric Darve and Andrew Pohorille. Calculating free energies using average force. *The Journal of chemical physics*, 115(20):9169–9183, 2001.

[270] Sanghyun Park and Klaus Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *The Journal of chemical physics*, 120(13):5946–5961, 2004.

[271] EA Carter, Giovanni Ciccotti, James T Hynes, and Raymond Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chemical Physics Letters*, 156(5):472–477, 1989.

[272] Michiel Sprik and Giovanni Ciccotti. Free energy from constrained molecular dynamics. *The Journal of chemical physics*, 109(18):7737–7744, 1998.

[273] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

[274] PA Bash, UC Singh, R Langridge, and PA Kollman. Free energy calculations by computer simulation. *Science*, 236(4801):564–568, 1987.

[275] Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[276] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[277] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):826–843, 2011.

[278] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[279] Tomasz Róg, Mykhailo Girych, and Alex Bunker. Mechanistic understanding from molecular dynamics in pharmaceutical research 2: lipid membrane in drug design. *Pharmaceuticals*, 14(10):1062, 2021.

[280] Jan Domanski, George Hedger, Robert B Best, Phillip J Stansfeld, and Mark SP Sansom. Convergence and sampling in determining free energy landscapes for membrane protein association. *The journal of physical chemistry B*, 121(15):3364–3375, 2017.

[281] Bradley Efron. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer, 1992.

[282] Henrik Flyvbjerg and Henrik Gordon Petersen. Error estimates on averages of correlated data. *The Journal of Chemical Physics*, 91(1):461–466, 1989.

[283] Jochen S Hub, Bert L De Groot, and David van der Spoel. g_wham a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of chemical theory and computation*, 6(12):3713–3720, 2010.

[284] Fangqiang Zhu and Gerhard Hummer. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *Journal of computational chemistry*, 33(4):453–465, 2012.

[285] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical review letters*, 100(2):020603, 2008.

[286] Guido Tiana. Estimation of microscopic averages from metadynamics. *The European Physical Journal B*, 63(2):235–238, 2008.

[287] Massimiliano Bonomi, Alessandro Barducci, and Michele Parrinello. Reconstructing the equilibrium boltzmann distribution from well-tempered metadynamics. *Journal of computational chemistry*, 30(11):1615–1621, 2009.

[288] Pratyush Tiwary and Michele Parrinello. A time-independent free energy estimator for metadynamics. *The Journal of Physical Chemistry B*, 119(3):736–742, 2015.

[289] Pratyush Tiwary and Michele Parrinello. From metadynamics to dynamics. *Physical review letters*, 111(23):230602, 2013.

[290] Matteo Salvalaglio, Pratyush Tiwary, and Michele Parrinello. Assessing the reliability of the dynamics reconstructed from metadynamics. *Journal of chemical theory and computation*, 10(4):1420–1425, 2014.

[291] Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

[292] Ilaria Gimondi, Gareth A Tribello, and Matteo Salvalaglio. Building maps in collective variable space. *The Journal of chemical physics*, 149(10):104104, 2018.

[293] Loukas Kollias, David C Cantu, Vassiliki-Alexandra Glezakou, Roger Rousseau, and Matteo Salvalaglio. On the role of enthalpic and entropic contributions to the conformational free energy landscape of mil-101 (cr) secondary building units. *Advanced Theory and Simulations*, 3(12):2000092, 2020.

[294] Herman JC Berendsen, David van der Spoel, and Rudi van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer physics communications*, 91(1-3):43–56, 1995.

[295] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[296] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Computer physics communications*, 185(2):604–613, 2014.

[297] Promoting transparency and reproducibility in enhanced molecular simulations. *Nature methods*, 16(8):670–673, 2019.

[298] Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, 4(1):1–17, 2012.

[299] David A Case, Thomas E Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz Jr, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The amber biomolecular simulation programs. *Journal of computational chemistry*, 26(16):1668–1688, 2005.

[300] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian~16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

[301] Araz Jakalian, David B Jack, and Christopher I Bayly. Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation. *Journal of computational chemistry*, 23(16):1623–1641, 2002.

[302] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[303] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.

[304] Elisabeth Weyandt, Luigi Leanza, Riccardo Capelli, Giovanni M Pavan, Ghislaine Vantomme, and EW Meijer. Controlling the length of porphyrin supramolecular polymers via coupled equilibria and dilution-induced supramolecular polymerization. *Nature Communications*, 13, 2022.

[305] Johannes AAW Elemans, Richard Van Hameren, Roeland JM Nolte, and Alan E Rowan. Molecular materials by self-assembly of porphyrins, phthalocyanines, and perylenes. *Advanced Materials*, 18(10):1251–1266, 2006.

[306] Harry L Anderson. Building molecular wires from the colours of life: conjugated porphyrin oligomers. *Chemical Communications*, (23):2323–2330, 1999.

[307] Peter A Korevaar, Tom FA de Greef, and EW Meijer. Pathway complexity in π-conjugated materials. *Chemistry of Materials*, 26(1):576–586, 2014.

[308] Soichiro Ogi, Kazunori Sugiyasu, Swarup Manna, Sadaki Samitsu, and Masayuki Takeuchi. Living supramolecular polymerization realized through a biomimetic approach. *Nature chemistry*, 6(3):188–195, 2014.

[309] Tomoya Fukui, Shinnosuke Kawai, Satoko Fujinuma, Yoshitaka Matsushita, Takeshi Yasuda, Tsuneaki Sakurai, Shu Seki, Masayuki Takeuchi, and Kazunori Sugiyasu. Control over differentiation of a metastable supramolecular assembly in one and two dimensions. *Nature chemistry*, 9(5):493–499, 2017.

[310] Pol Besenius. Controlling supramolecular polymerization through multicomponent self-assembly. *Journal of Polymer Science Part A: Polymer Chemistry*, 55(1):34–78, 2017.

[311] Mizuki Endo, Tomoya Fukui, Sung Ho Jung, Shiki Yagai, Masayuki Takeuchi, and Kazunori Sugiyasu. Photoregulated living supramolecular polymerization established by combining energy landscapes of photoisomerization and nucleation–elongation processes. *Journal of the American Chemical Society*, 138(43):14347–14353, 2016.

[312] Bala NS Thota, Xianwen Lou, Davide Bochicchio, Tim FE Paffen, René PM Lafleur, Joost LJ van Dongen, Svenja Ehrmann, Rainer Haag, Giovanni M Pavan, Anja RA Palmans, et al. Supramolecular copolymerization as a strategy to control the stability of self-assembled nanofibers. *Angewandte Chemie International Edition*, 57(23):6843–6847, 2018.

[313] Jiheong Kang, Daigo Miyajima, Tadashi Mori, Yoshihisa Inoue, Yoshimitsu Itoh, and Takuzo Aida. A rational strategy for the realization of chain-growth supramolecular polymerization. *Science*, 347(6222):646–651, 2015.

[314] Joe B Gilroy, Torben Gädt, George R Whittell, Laurent Chabanne, John M Mitchels, Robert M Richardson, Mitchell A Winnik, and Ian Manners. Monodisperse cylindrical micelles by crystallization-driven living self-assembly. *Nature chemistry*, 2(7):566–570, 2010.

[315] Beatrice Adelizzi, Antonio Aloi, Albert J Markvoort, Huub MM Ten Eikelder, Ilja K Voets, Anja RA Palmans, and EW Meijer. Supramolecular block copolymers under thermodynamic control. *Journal of the American Chemical Society*, 140(23):7168–7175, 2018.

[316] Aritra Sarkar, Ranjan Sasmal, Charly Empereur-Mot, Davide Bochicchio, Srinath VK Kompella, Kamna Sharma, Shikha Dhiman, Balasubramanian Sundaram, Sarit S Agasti, Giovanni M Pavan, et al. Self-sorted, random, and block supramolecular copolymers via sequence controlled, multicomponent self-assembly. *Journal of the American Chemical Society*, 142(16):7606–7617, 2020.

[317] Srinu Kotha, Mathijs FJ Mabesoone, Dasari Srideep, Rahul Sahu, Sandeep K Reddy, and Kotagiri Venkata Rao. Supramolecular depolymerization in the mixture of two poor solvents: mechanistic insights and modulation of supramolecular polymerization of ionic π-systems. *Angewandte Chemie*, 133(10):5519–5526, 2021.

[318] Goutam Ghosh and Suhrit Ghosh. Solvent dependent pathway complexity and seeded supramolecular polymerization. *Chemical Communications*, 54(45):5720–5723, 2018.

[319] Mathijs FJ Mabesoone, Albert J Markvoort, Motonori Banno, Tomoko Yamaguchi, Floris Helmich, Yuki Naito, Eiji Yashima, Anja RA Palmans, and EW Meijer. Competing interactions in hierarchical porphyrin self-assembly introduce robustness in pathway complexity. *Journal of the American Chemical Society*, 140(25):7810–7819, 2018.

[320] Peter A Korevaar, Charley Schaefer, Tom FA de Greef, and EW Meijer. Controlling chemical self-assembly by solvent-dependent dynamics. *Journal of the American Chemical Society*, 134(32):13482–13491, 2012.

[321] Kotagiri Venkata Rao, Daigo Miyajima, Atsuko Nihonyanagi, and Takuzo Aida. Thermally bisignate supramolecular polymerization. *Nature chemistry*, 9(11):1133–1139, 2017.

[322] Kotagiri Venkata Rao, Mathijs FJ Mabesoone, Daigo Miyajima, Atsuko Nihonyanagi, EW Meijer, and Takuzo Aida. Distinct pathways in "thermally bisignate supramolecular polymerization": Spectroscopic and computational studies. *Journal of the American Chemical Society*, 142(1):598–605, 2019.

[323] Floris Helmich, Cameron C Lee, Marko ML Nieuwenhuizen, Jeroen C Gielen, Peter CM Christianen, Antje Larsen, George Fytas, Philippe ELG Leclère, Albertus PHJ Schenning, and EW Meijer. Dilution-induced self-assembly of porphyrin aggregates: a consequence of coupled equilibria. *Angewandte Chemie International Edition*, 49(23):3939–3942, 2010.

[324] Yuuya Nagata, Ryohei Takeda, and Michinori Suginome. Asymmetric catalysis in chiral solvents: chirality transfer with amplification of homochirality through a helical macromolecular scaffold. *ACS Central Science*, 5(7):1235–1240, 2019.

[325] Matthieu Raynal, François Portier, Piet WNM Van Leeuwen, and Laurent Bouteiller. Tunable asymmetric catalysis through ligand stacking in chiral rigid rods. *Journal of the American Chemical Society*, 135(47):17687–17690, 2013.

[326] Alaric Desmarchelier, Xavier Caumes, Matthieu Raynal, Anton Vidal-Ferran, Piet WNM van Leeuwen, and Laurent Bouteiller. Correlation between the selectivity and the structure of an asymmetric catalyst built on a chirally amplified supramolecular helical scaffold. *Journal of the American Chemical Society*, 138(14):4908–4916, 2016.

[327] Miriam de Torres, Richard van Hameren, Roeland JM Nolte, Alan E Rowan, and Johannes AAW Elemans. Photocatalytic oxidation of stilbene by self-assembled stacks of manganese porphyrins. *Chemical Communications*, 49(92):10787–10789, 2013.

[328] Shaji Varghese, Bram Spierenburg, Jeroen PJ Bruekers, Anne Swartjes, Paul B White, Johannes AAW Elemans, and Roeland JM Nolte. Effect of chirality on the binding of viologen guests in porphyrin macrocycles. *European Journal of Organic Chemistry*, 2019(21):3525–3533, 2019.

[329] Pieter J Gilissen, Annemiek D Slootbeek, Jiangkun Ouyang, Nicolas Vanthuyne, Rob Bakker, Johannes AAW Elemans, and Roeland JM Nolte. Enantioselective synthesis of chiral porphyrin macrocyclic hosts and kinetic enantiorecognition of viologen guests. *Chemical Science*, 12(5):1661–1667, 2021.

[330] Krishnachary Salikolimi, Vakayil K Praveen, Achalkumar Ammathnadu Sudhakar, Kuniyo Yamada, Noriko Nishizawa Horimoto, and Yasuhiro Ishida. Helical supramolecular polymers with rationally designed binding sites for chiral guest recognition. *Nature Communications*, 11(1):2311, 2020.

[331] Aritra Sarkar, Shikha Dhiman, Aditya Chalishazar, and Subi J George. Visualization of stereoselective supramolecular polymers by chirality-controlled energy transfer. *Angewandte Chemie International Edition*, 56(44):13767–13771, 2017.

[332] Bhawani Narayan, Karteek K Bejagam, Sundaram Balasubramanian, and Subi J George. Autoresolution of segregated and mixed p-n stacks by stereoselective supramolecular polymerization in solution. *Angewandte Chemie International Edition*, 54(44):13053–13057, 2015.

[333] Hanna Jedrzejewska and Agnieszka Szumna. Making a right or left choice: chiral self-sorting as a tool for the formation of discrete complex structures. *Chemical reviews*, 117(6):4863–4899, 2017.

[334] Markus Hecht, Pawaret Leowanawat, Tabea Gerlach, Vladimir Stepanenko, Matthias Stolte, Matthias Lehmann, and Frank Würthner. Self-sorting supramolecular polymerization: helical and lamellar aggregates of tetrabay-acyloxy perylene bisimide. *Angewandte Chemie International Edition*, 59(39):17084–17090, 2020.

[335] Marius Wehner, Merle Insa Silja Röhr, Vladimir Stepanenko, and Frank Würthner. Control of self-assembly pathways toward conglomerate and racemic supramolecular polymers. *Nature Communications*, 11(1):5460, 2020.

[336] Will R Henderson, Guancen Liu, Khalil A Abboud, and Ronald K Castellano. Tuning supramolecular polymer assembly through stereoelectronic interactions. *Journal of the American Chemical Society*, 143(32):12688–12698, 2021.

[337] Mark M Green, Bruce A Garetz, Beth Munoz, HePing Chang, Steven Hoke, and R Graham Cooks. Majority rules in the copolymerization of mirror image isomers. *Journal of the American Chemical Society*, 117(14):4181–4182, 1995.

[338] Maarten MJ Smulders, Patrick JM Stals, Tristan Mes, Tim FE Paffen, Albertus PHJ Schenning, Anja RA Palmans, and EW Meijer. Probing the limits of the majority-rules principle in a dynamic supramolecular polymer. *Journal of the American Chemical Society*, 132(2):620–626, 2010.

[339] Michihisa Ueda, Tsubasa Aoki, Takayoshi Akiyama, Takayuki Nakamuro, Keitaro Yamashita, Haruaki Yanagisawa, Osamu Nureki, Masahide Kikkawa, Eiichi Nakamura, Takuzo Aida, et al. Alternating heterochiral supramolecular copolymerization. *Journal of the American Chemical Society*, 143(13):5121–5126, 2021.

[340] Jatish Kumar, Hiroyuki Tsumatori, Junpei Yuasa, Tsuyoshi Kawai, and Takuya Nakashima. Self-discriminating termination of chiral supramolecular polymerization: Tuning the length of nanofibers. *Angewandte Chemie*, 127(20):6041–6045, 2015.

[341] Souvik Sarkar, Aritra Sarkar, and Subi J George. Stereoselective seed-induced living supramolecular polymerization. *Angewandte Chemie International Edition*, 59(45):19841–19845, 2020.

[342] Souvik Sarkar, Aritra Sarkar, Arka Som, Sarit S Agasti, and Subi J George. Stereoselective primary and secondary nucleation events in multicomponent seeded supramolecular polymerization. *Journal of the American Chemical Society*, 143(30):11777–11787, 2021.

[343] Floris Helmich, Maarten MJ Smulders, Cameron C Lee, Albertus PHJ Schenning, and EW Meijer. Effect of stereogenic centers on the self-sorting, depolymerization, and atropisomerization kinetics of porphyrin-based aggregates. *Journal of the American Chemical Society*, 133(31):12238–12246, 2011.

[344] Floris Helmich, Cameron C Lee, Albertus PHJ Schenning, and EW Meijer. Chiral memory via chiral amplification and selective depolymerization of porphyrin aggregates. *Journal of the American Chemical Society*, 132(47):16753–16755, 2010.

[345] Piero Gasparotto, Davide Bochicchio, Michele Ceriotti, and Giovanni M Pavan. Identifying and tracking defects in dynamic supramolecular polymers. *The Journal of Physical Chemistry B*, 124(3):589–599, 2019.

[346] Andrea Torchi, Davide Bochicchio, and Giovanni M Pavan. How the dynamics of a supramolecular polymer determines its dynamic adaptivity and stimuli-responsiveness: structure–dynamics–property relationships from coarse-grained simulations. *The Journal of Physical Chemistry B*, 122(14):4169–4178, 2018.

[347] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[348] Kalipada Koner, Shayan Karak, Sharath Kandambeth, Suvendu Karak, Neethu Thomas, Luigi Leanza, Claudio Perego, Luca Pesce, Riccardo Capelli, Monika Moun, et al. Porous covalent organic nanotubes and their assembly in loops and toroids. *Nature Chemistry*, 14(5):507–514, 2022.

[349] Irving Langmuir. The arrangement of electrons in atoms and molecules. *Journal of the American Chemical Society*, 41(6):868–934, 1919.

[350] RH Wentorf, Robert C DeVries, and FP Bundy. Sintered superhard materials. *Science*, 208(4446):873–880, 1980.

[351] Tomokazu Tozawa, James TA Jones, Shashikala I Swamy, Shan Jiang, Dave J Adams, Stephen Shakespeare, Rob Clowes, Darren Bradshaw, Tom Hasell, Samantha Y Chong, et al. Porous organic cages. *Nature materials*, 8(12):973–978, 2009.

[352] Fernando J Uribe-Romo, Joseph R Hunt, Hiroyasu Furukawa, Cornelius Klock, Michael O'Keeffe, and Omar M Yaghi. A crystalline imine-linked 3-d porous covalent organic framework. *Journal of the American Chemical Society*, 131(13):4570–4571, 2009.

[353] Adrien P Cote, Annabelle I Benin, Nathan W Ockwig, Michael O'Keeffe, Adam J Matzger, and Omar M Yaghi. Porous, crystalline, covalent organic frameworks. *science*, 310(5751):1166–1170, 2005.

[354] Chenguang Qiu, Zhiyong Zhang, Mengmeng Xiao, Yingjun Yang, Donglai Zhong, and Lian-Mao Peng. Scaling carbon nanotube complementary transistors to 5-nm gate lengths. *Science*, 355(6322):271–276, 2017.

[355] Kangkang Guo, Baojuan Xi, Ruchao Wei, Haibo Li, Jinkui Feng, and Shenglin Xiong. Hierarchical microcables constructed by cop@ c carbon framework intertwined with carbon nanotubes for efficient lithium storage. *Advanced energy materials*, 10(12):1902913, 2020.

[356] Yong Zhou, Masaki Kogiso, Masumi Asakawa, Sijun Dong, Ryoiti Kiyama, and Toshimi Shimizu. Antimicrobial nanotubes consisting of ag-embedded peptidic lipid-bilayer membranes as delivery vehicles. *Advanced Materials*, 21(17):1742–1745, 2009.

[357] Sumio Iijima. Helical microtubules of graphitic carbon. *nature*, 354(6348):56–58, 1991.

[358] Sumio Iijima and Toshinari Ichihashi. Single-shell carbon nanotubes of 1-nm diameter. *nature*, 363(6430):603–605, 1993.

[359] Donald S Bethune, Ch H Kiang, MS De Vries, G Gorman, R_ Savoy, J Vazquez, and R Beyers. Cobalt-catalysed growth of carbon nanotubes with single-atomic-layer walls. *Nature*, 363(6430):605–607, 1993.

[360] Toshiaki Kato and Rikizo Hatakeyama. Direct growth of short single-walled carbon nanotubes with narrow-chirality distribution by time-programmed plasma chemical vapor deposition. *ACS nano*, 4(12):7395–7400, 2010.

[361] Avetik R Harutyunyan, Gugang Chen, Tereza M Paronyan, Elena M Pigos, Oleg A Kuznetsov, Kapila Hewaparakrama, Seung Min Kim, Dmitri Zakharov, Eric A Stach, and Gamini U Sumanasekera. Preferential growth of single-walled carbon nanotubes with metallic conductivity. *Science*, 326(5949):116–120, 2009.

[362] Feng Yang, Xiao Wang, Daqi Zhang, Juan Yang, Da Luo, Ziwei Xu, Jiake Wei, Jian-Qiang Wang, Zhi Xu, Fei Peng, et al. Chirality-specific growth of single-walled carbon nanotubes on solid alloy catalysts. *Nature*, 510(7506):522–524, 2014.

[363] Naohiro Kameta, Hiroyuki Minamikawa, Mitsutoshi Masuda, Go Mizuno, and Toshimi Shimizu. Controllable biomolecule release from self-assembled organic nanotubes with asymmetric surfaces: ph and temperature dependence. *Soft Matter*, 4(8):1681–1687, 2008.

[364] Naohiro Kameta, Hiroyuki Minamikawa, and Mitsutoshi Masuda. Supramolecular organic nanotubes: how to utilize the inner nanospace and the outer space. *Soft Matter*, 7(10):4539–4561, 2011.

[365] Christopher R Murdock and David M Jenkins. Isostructural synthesis of porous metal–organic nanotubes. *Journal of the American Chemical Society*, 136(31):10983–10988, 2014.

[366] Nuria Rodríguez-Vázquez, M Amorín, and JR Granja. Recent advances in controlling the internal and external properties of self-assembling cyclic peptide nanotubes and dimers. *Organic & Biomolecular Chemistry*, 15(21):4490–4505, 2017.

[367] Abbie Trewin and Andrew I Cooper. Porous organic polymers: distinction from disorder? *Angewandte Chemie International Edition*, 49(9):1533–1535, 2010.

[368] Christian Reece, David J Willock, and Abbie Trewin. Modelling analysis of the structure and porosity of covalent triazine-based frameworks. *Physical Chemistry Chemical Physics*, 17(2):817–823, 2015.

[369] Jia-Xing Jiang, Abbie Trewin, Fabing Su, Colin D Wood, Hongjun Niu, James TA Jones, Yaroslav Z Khimyak, and Andrew I Cooper. Microporous poly (tri (4-ethynylphenyl) amine) networks: synthesis, properties, and atomistic simulation. *Macromolecules*, 42(7):2658–2666, 2009.

[370] Jens MH Thomas and Abbie Trewin. Amorphous paf-1: Guiding the rational design of ultraporous materials. *The Journal of Physical Chemistry C*, 118(34):19712–19722, 2014.

[371] Pierre Fayon and Abbie Trewin. Formation mechanism of ultra porous framework materials. *Physical Chemistry Chemical Physics*, 18(25):16840–16847, 2016.

[372] Andrew Tarzia, James EM Lewis, and Kim E Jelfs. High-throughput computational evaluation of low symmetry pd2l4 cages to aid in system design. *Angewandte Chemie International Edition*, 60(38):20879–20887, 2021.

[373] Andrew Tarzia, Masahide Takahashi, Paolo Falcaro, Aaron W Thornton, Christian J Doonan, and David M Huang. High-throughput screening of metal–organic frameworks for macroscale heteroepitaxial alignment. *ACS applied materials & interfaces*, 10(47):40938–40950, 2018.

[374] Andrew Tarzia, Aaron W Thornton, Christian J Doonan, and David M Huang. Molecular insight into assembly mechanisms of porous aromatic frameworks. *The Journal of Physical Chemistry C*, 121(30):16381–16392, 2017.

[375] Emma H Wolpert and Kim E Jelfs. Coarse-grained modelling to predict the packing of porous organic cages. *Chemical Science*, 13(45):13588–13599, 2022.

[376] Junmei Wang, Wei Wang, Peter A Kollman, and David A Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling*, 25(2):247–260, 2006.

[377] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[378] Silvia Varela-Aramburu, Giulia Morgese, Lu Su, Sandra MC Schoenmakers, Mattia Perrone, Luigi Leanza, Claudio Perego, Giovanni M Pavan, Anja RA Palmans, and EW Meijer. Exploring the potential of benzene-1, 3, 5-tricarboxamide supramolecular polymers as biomaterials. *Biomacromolecules*, 21(10):4105–4115, 2020.

[379] Olga JGM Goor, Simone IS Hendrikse, Patricia YW Dankers, and EW Meijer. From supramolecular polymers to multi-component biomaterials. *Chemical Society Reviews*, 46(21):6621–6637, 2017.

[380] Matthew J Webber, Eric A Appel, EW Meijer, and Robert Langer. Supramolecular biomaterials. *Nature materials*, 15(1):13–26, 2016.

[381] Christian Frantz, Kathleen M Stewart, and Valerie M Weaver. The extracellular matrix at a glance. *Journal of cell science*, 123(24):4195–4200, 2010.

[382] Yiliu Liu, Hui Yang, Zhiqiang Wang, and Xi Zhang. Cucurbit [8] uril-based supramolecular polymers. *Chemistry–An Asian Journal*, 8(8):1626–1632, 2013.

[383] Jianxiang Zhang and Peter X Ma. Host–guest interactions mediated nanoassemblies using cyclodextrin-containing hydrophilic polymers and their biomedical applications. *Nano today*, 5(4):337–350, 2010.

[384] Thomas Choisnet, David Canevet, Marc Sallé, Erwan Nicol, Frédérick Niepceron, Jacques Jestin, and Olivier Colombani. Robust supramolecular nanocylinders of naphthalene diimide in water. *Chemical Communications*, 55(64):9519–9522, 2019.

[385] Daniel Görl, Xin Zhang, Vladimir Stepanenko, and Frank Würthner. Supramolecular block copolymers by kinetically controlled co-self-assembly of planar and core-twisted perylene bisimides. *Nature communications*, 6(1):7009, 2015.

[386] Kanya Rajangam, Heather A Behanna, Michael J Hui, Xiaoqiang Han, James F Hulvat, Jon W Lomasney, and Samuel I Stupp. Heparin binding nanostructures to promote growth of blood vessels. *Nano letters*, 6(9):2086–2090, 2006.

[387] Jeffrey D Hartgerink, Elia Beniash, and Samuel I Stupp. Self-assembly and mineralization of peptide-amphiphile nanofibers. *science*, 294(5547):1684–1688, 2001.

[388] Jeffrey D Hartgerink, Elia Beniash, and Samuel I Stupp. Peptide-amphiphile nanofibers: a versatile scaffold for the preparation of self-assembling materials. *Proceedings of the National Academy of Sciences*, 99(8):5133–5138, 2002.

[389] Samuel I Stupp, Tristan D Clemons, James K Carrow, Hiroaki Sai, and Liam C Palmer. Supramolecular and hybrid bonding polymers. *Israel Journal of Chemistry*, 60(1-2):124–131, 2020.

[390] Simone IS Hendrikse, Sjors PW Wijnands, René PM Lafleur, Maarten J Pouderoijen, Henk M Janssen, Patricia YW Dankers, and EW Meijer. Controlling and tuning the dynamic nature of supramolecular polymers in aqueous solutions. *Chemical Communications*, 53(14):2279–2282, 2017.

[391] Maarten H Bakker, Roxanne E Kieltyka, Lorenzo Albertazzi, and Patricia YW Dankers. Modular supramolecular ureidopyrimidinone polymer carriers for intracellular delivery. *RSC advances*, 6(112):110600–110603, 2016.

[392] Simone IS Hendrikse, Sergio Spaans, EW Meijer, and Patricia YW Dankers. Supramolecular platform stabilizing growth factors. *Biomacromolecules*, 19(7):2610–2617, 2018.

[393] Marion K Müller and Luc Brunsveld. A supramolecular polymer as a self-assembling polyvalent scaffold. *Angewandte Chemie International Edition*, 48(16):2921–2924, 2009.

[394] David Straßburger, Natascha Stergiou, Moritz Urschbach, Hajime Yurugi, Daniel Spitzer, Dieter Schollmeyer, Edgar Schmitt, and Pol Besenius. Mannose-decorated multicomponent supramolecular polymers trigger effective uptake into antigen-presenting cells. *ChemBioChem*, 19(9):912–916, 2018.

[395] Xianwen Lou, René PM Lafleur, Christianus MA Leenders, Sandra MC Schoenmakers, Nicholas M Matsumoto, Matthew B Baker, Joost LJ Van Dongen, Anja RA Palmans, and EW Meijer. Dynamic diversity of synthetic supramolecular polymers in water as revealed by hydrogen/deuterium exchange. *Nature communications*, 8(1):15420, 2017.

[396] Christianus MA Leenders, Gijs Jansen, Martijn MM Frissen, René PM Lafleur, Ilja K Voets, Anja RA Palmans, and EW Meijer. Monosaccharides as versatile units for water-soluble supramolecular polymers. *Chemistry–A European Journal*, 22(13):4608–4615, 2016.

[397] Matthew B Baker, Ronald PJ Gosens, Lorenzo Albertazzi, Nicholas M Matsumoto, Anja RA Palmans, and EW Meijer. Exposing differences in monomer exchange rates of multicomponent supramolecular polymers in water. *ChemBioChem*, 17(3):207–213, 2016.

[398] Simone IS Hendrikse, Lu Su, Tim P Hogervorst, Rene PM Lafleur, Xianwen Lou, Gijsbert A van der Marel, Jeroen DC Codee, and EW Meijer. Elucidating the ordering in self-assembled glycocalyx mimicking supramolecular copolymers in water. *Journal of the American Chemical Society*, 141(35):13877–13886, 2019.

[399] Sjors PW Wijnands, Wouter Engelen, René PM Lafleur, EW Meijer, and Maarten Merkx. Controlling protein activity by dynamic recruitment on a supramolecular polymer platform. *Nature Communications*, 9(1):65, 2018.

[400] Christianus MA Leenders, Tristan Mes, Matthew B Baker, Marcel ME Koenigs, Pol Besenius, Anja RA Palmans, and EW Meijer. From supramolecular polymers to hydrogel materials. *Materials Horizons*, 1(1):116–120, 2014.

[401] Anna Bujacz. Structures of bovine, equine and leporine serum albumin. *Acta Crystallographica Section D: Biological Crystallography*, 68(10):1278–1289, 2012.

[402] Djurre H de Jong, Gurpreet Singh, WF Drew Bennett, Clement Arnarez, Tsjerk A Wassenaar, Lars V Schafer, Xavier Periole, D Peter Tieleman, and Siewert J Marrink. Improved parameters for the martini coarse-grained protein force field. *Journal of chemical theory and computation*, 9(1):687–697, 2013.

[403] Luca Monticelli, Senthil K Kandasamy, Xavier Periole, Ronald G Larson, D Peter Tieleman, and Siewert-Jan Marrink. The martini coarse-grained force field: extension to proteins. *Journal of chemical theory and computation*, 4(5):819–834, 2008.

[404] Monique M Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical review letters*, 77(9):1905, 1996.

[405] Luigi Leanza, Claudio Perego, Luca Pesce, Max von Delius, Matteo Salvalaglio, and Giovanni M Maria Pavan. Into the dynamics of rotaxanes at atomistic resolution. *Chemical Science*, 2023.

[406] Tania A Baker and Stephen P Bell. Polymerases and the replisome: machines within machines. *Cell*, 92(3):295–305, 1998.

[407] James A Spudich. How molecular motors work. *Nature*, 372(6506):515–518, 1994.

[408] Josh E Baker, Ingrid Brust-Mascher, Sampath Ramachandran, Leslie EW LaConte, and David D Thomas. A large and distinct rotation of the myosin light chain domain occurs upon muscle contraction. *Proceedings of the National Academy of Sciences*, 95(6):2944–2949, 1998.

[409] James R Sellers. Myosins: a diverse superfamily. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1496(1):3–22, 2000.

[410] AF Huxley. Mechanics and models of the myosin motor. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1396):433–440, 2000.

[411] Ryohei Yasuda, Hiroyuki Noji, Kazuhiko Kinosita, and Masasuke Yoshida. F1-atpase is a highly efficient molecular motor that rotates with discrete 120 steps. *Cell*, 93(7):1117–1124, 1998.

[412] George Oster and Hongyun Wang. Atp synthase: two motors, two fuels. *Structure*, 7(4):R67–R72, 1999.

[413] C Mavroidis, A Dubey, and ML Yarmush. Molecular machines. *Annu. Rev. Biomed. Eng.*, 6:363–395, 2004.

[414] Roberto Ballardini, Vincenzo Balzani, Alberto Credi, Maria Teresa Gandolfi, and Margherita Venturi. Artificial molecular-level machines: which energy to make them work? *Accounts of Chemical Research*, 34(6):445–455, 2001.

[415] J Fraser Stoddart. The chemistry of the mechanical bond. *Chemical Society Reviews*, 38(6):1802–1820, 2009.

[416] Lei Fang, Mark A Olson, Diego Benítez, Ekaterina Tkatchouk, William A Goddard III, and J Fraser Stoddart. Mechanically bonded macromolecules. *Chemical Society Reviews*, 39(1):17–29, 2010.

[417] J Fraser Stoddart. Putting mechanically interlocked molecules (mims) to work in tomorrow's world, 2014.

[418] Guzmán Gil-Ramírez, David A Leigh, and Alexander J Stephens. Catenanes: fifty years of molecular links. *Angewandte Chemie International Edition*, 54(21):6110–6150, 2015.

[419] Min Xue, Yong Yang, Xiaodong Chi, Xuzhou Yan, and Feihe Huang. Development of pseudorotaxanes and rotaxanes: from synthesis to stimuli-responsive motions to applications. *Chemical Reviews*, 115(15):7398–7501, 2015.

[420] Jonathan E Green, Jang Wook Choi, Akram Boukai, Yuri Bunimovich, Ezekiel Johnston-Halperin, Erica DeIonno, Yi Luo, Bonnie A Sheriff, Ke Xu, Young Shik Shin, et al. A 160-kilobit molecular electronic memory patterned at 1011 bits per square centimetre. *Nature*, 445(7126):414–417, 2007.

[421] Victor Blanco, David A Leigh, Vanesa Marcos, Jose A Morales-Serna, and Alina L Nussbaumer. A switchable [2] rotaxane asymmetric organocatalyst that utilizes an acyclic chiral secondary amine. *Journal of the American Chemical Society*, 136(13):4905–4908, 2014.

[422] Jack Beswick, Victor Blanco, Guillaume De Bo, David A Leigh, Urszula Lewandowska, Bartosz Lewandowski, and Kenji Mishiro. Selecting reactions and reactants using a switchable rotaxane organocatalyst with two different active sites. *Chemical Science*, 6(1):140–143, 2015.

[423] Yusuf Cakmak, Sundus Erbas-Cakmak, and David A Leigh. Asymmetric catalysis with a mechanically point-chiral rotaxane. *Journal of the American Chemical Society*, 138(6):1749–1751, 2016.

[424] Philip G Young, Keiji Hirose, and Yoshito Tobe. Axle length does not affect switching dynamics in degenerate molecular shuttles with rigid spacers. *Journal of the American Chemical Society*, 136(22):7899–7906, 2014.

[425] Seogshin Kang, Scott A Vignon, Hsian-Rong Tseng, and J Fraser Stoddart. Molecular shuttles based on tetrathiafulvalene units and 1, 5-dioxynaphthalene ring systems. *Chemistry–A European Journal*, 10(10):2555–2564, 2004.

[426] Sune Nygaard, Ken C-F Leung, Ivan Aprahamian, Taichi Ikeda, Sourav Saha, Bo W Laursen, Soo-Young Kim, Stinne W Hansen, Paul C Stein, Amar H Flood, et al. Functionally rigid bistable [2] rotaxanes. *Journal of the American Chemical Society*, 129(4):960–970, 2007.

[427] Albert C Fahrenbach, Carson J Bruns, Dennis Cao, and J Fraser Stoddart. Ground-state thermodynamics of bistable redox-active donor–acceptor mechanically interlocked molecules. *Accounts of chemical research*, 45(9):1581–1592, 2012.

[428] Albert C Fahrenbach, Carson J Bruns, Hao Li, Ali Trabolsi, Ali Coskun, and J Fraser Stoddart. Ground-state kinetics of bistable redox-active donor–acceptor mechanically interlocked molecules. *Accounts of chemical research*, 47(2):482–493, 2014.

[429] X Grabuleda and C Jaime. Molecular shuttles. a computational study (mm and md) on the translational isomerism in some [2] rotaxanes. *The Journal of Organic Chemistry*, 63(26):9635–9643, 1998.

[430] Xavi Grabuleda, Petko Ivanov, and Carlos Jaime. Computational studies on pseudorotaxanes by molecular dynamics and free energy perturbation simulations. *The Journal of Organic Chemistry*, 68(4):1539–1547, 2003.

[431] Xavi Grabuleda, Petko Ivanov, and Carlos Jaime. Shuttling process in [2] rotaxanes. modeling by molecular dynamics and free energy perturbation simulations. *The Journal of Physical Chemistry B*, 107(31):7582–7588, 2003.

[432] Naresh K Jena and N Arul Murugan. Solvent-dependent conformational states of a [2] rotaxane-based molecular machine: A molecular dynamics perspective. *The Journal of Physical Chemistry C*, 117(47):25059–25068, 2013.

[433] Seung Soon Jang, Yun Hee Jang, Yong-Hoon Kim, William A Goddard, Amar H Flood, Bo W Laursen, Hsian-Rong Tseng, J Fraser Stoddart, Jan O Jeppesen, Jang Wook Choi, et al. Structures and properties of self-assembled monolayers of bistable [2] rotaxanes on au (111) surfaces from molecular dynamics simulations validated with experiment. *Journal of the American Chemical Society*, 127(5):1563–1575, 2005.

[434] Giordano Mancini, Costantino Zazza, Massimiliano Aschi, and Nico Sanna. Conformational analysis and uv/vis spectroscopic properties of a rotaxane-based molecular machine in acetonitrile dilute solution: when simulations meet experiments. *Physical Chemistry Chemical Physics*, 13(6):2342–2349, 2011.

[435] Elena Kolodzeiski and Saeed Amirjalayer. On-the-fly training of atomistic potentials for flexible and mechanically interlocked molecules. *Journal of Chemical Theory and Computation*, 17(11):7010–7020, 2021.

[436] Elena Kolodzeiski and Saeed Amirjalayer. Dynamic network of intermolecular interactions in metal-organic frameworks functionalized by molecular machines. *Science Advances*, 8(26):eabn4426, 2022.

[437] Tim Hempel, Simon Olsson, and Frank Noé. Markov field models: scaling molecular kinetics approaches to large molecular machines. *Current Opinion in Structural Biology*, 77:102458, 2022.

[438] Zhen Wu, Shuangshuang Wang, Zilin Zhang, Yanjun Zhang, Yanzhen Yin, Haixin Shi, and Shufei Jiao. Solvent effects on the motion of a crown ether/amino rotaxane. *RSC Advances*, 12(47):30495–30500, 2022.

[439] Peng Liu, Christophe Chipot, Xueguang Shao, and Wensheng Cai. Solvent-controlled shuttling in a molecular switch. *The Journal of Physical Chemistry C*, 116(7):4471–4476, 2012.

[440] Pratyush Tiwary, Vittorio Limongelli, Matteo Salvalaglio, and Michele Parrinello. Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proceedings of the National Academy of Sciences*, 112(5):E386–E391, 2015.

[441] Matteo Salvalaglio, Pratyush Tiwary, Giovanni Maria Maggioni, Marco Mazzotti, and Michele Parrinello. Overcoming time scale and finite size limitations to compute nucleation rates from small scale well tempered metadynamics simulations. *The Journal of chemical physics*, 145(21):211925, 2016.

[442] Mickael Lelimousin, Vittorio Limongelli, and Mark SP Sansom. Conformational changes in the epidermal growth factor receptor: Role of the transmembrane domain investigated by coarse-grained metadynamics free energy calculations. *Journal of the American Chemical Society*, 138(33):10611–10622, 2016.

[443] Massimo Delle Piane, Luca Pesce, Matteo Cioni, and Giovanni M Pavan. Reconstructing reactivity in dynamic host–guest systems at atomistic resolution: amide hydrolysis under confinement in the cavity of a coordination cage. *Chemical Science*, 13(37):11232–11245, 2022.

[444] Anna L de Marco, Davide Bochicchio, Andrea Gardin, Giovanni Doni, and Giovanni M Pavan. Controlling exchange pathways in dynamic supramolecular polymers by controlling defects. *ACS nano*, 15(9):14229–14241, 2021.

[445] Luca Pesce, Claudio Perego, Angela B Grommet, Rafal Klajn, and Giovanni M Pavan. Molecular factors controlling the isomerization of azobenzenes in the cavity of a flexible coordination cage. *J. Am. Chem. Soc.*, 142(21):9792–9802, 2020.

[446] Davide Bochicchio, Supaporn Kwangmettatam, Tibor Kudernac, and Giovanni M Pavan. How defects control the out-of-equilibrium dissipative evolution of a supramolecular tubule. *ACS nano*, 13(4):4322–4334, 2019.

[447] Chiara Lionello, Andrea Gardin, Annalisa Cardellini, Davide Bochicchio, Manisha Shivrayan, Ann Fernandez, S Thayumanavan, and Giovanni M Pavan. Toward chemotactic supramolecular nanoparticles: from autonomous surface motion following specific chemical gradients to multivalency-controlled disassembly. *ACS nano*, 15(10):16149–16161, 2021.

[448] Oleg Borodin, Yevhenii Shchukin, Craig C Robertson, Stefan Richter, and Max von Delius. Self-assembly of stimuli-responsive [2] rotaxanes by amidinium exchange. *Journal of the American Chemical Society*, 143(40):16448–16457, 2021.

[449] Youzhi Xu, Ramandeep Kaur, Bingzhe Wang, Martin B Minameyer, Sebastian Gsanger, Bernd Meyer, Thomas Drewello, Dirk M Guldi, and Max von Delius. Concave–convex π–π template approach enables the synthesis of [10] cycloparaphenylene–fullerene [2] rotaxanes. *Journal of the American Chemical Society*, 140(41):13413–13420, 2018.

[450] Mortimer J Kamlet and RW Taft. The solvatochromic comparison method. i. the. beta.-scale of solvent hydrogen-bond acceptor (hba) basicities. *Journal of the American chemical Society*, 98(2):377–383, 1976.

[451] Roger W Alder, Michael E Blake, Simone Bufali, Craig P Butts, A Guy Orpen, Jan Schütz, and Stuart J Williams. Preparation of tetraalkylformamidinium salts and related species as precursors to stable carbenes. *Journal of the Chemical Society, Perkin Transactions 1*, (14):1586–1593, 2001.

[452] Ronald Beckett and J Calvin Giddings. Entropic contribution to the retention of nonspherical particles in field-flow fractionation. *Journal of colloid and interface science*, 186(1):53–59, 1997.

[453] Takahiro Iwamoto, Yoshiki Watanabe, Tatsuya Sadahiro, Takeharu Haino, and Shigeru Yamago. Size-selective encapsulation of c 60 by [10] cycloparaphenylene: Formation of the shortest fullerene-peapod. *Angew. Chem.*, 36(123):8492–8494, 2011.

[454] Takahiro Iwamoto, Yoshiki Watanabe, Hikaru Takaya, Takeharu Haino, Nobuhiro Yasuda, and Shigeru Yamago. Size-and orientation-selective encapsulation of c70 by cycloparaphenylenes. *Chem. Eur. J.*, 19(42):14061–14068, 2013.

[455] Takahiro Iwamoto, Zdenek Slanina, Naomi Mizorogi, Jingdong Guo, Takeshi Akasaka, Shigeru Nagase, Hikaru Takaya, Nobuhiro Yasuda, Tatsuhisa Kato, and Shigeru Yamago. Partial charge transfer in the shortest possible metallofullerene peapod, la@c82⊂[11] cycloparaphenylene. *Chem. Eur. J.*, 20(44):14403–14409, 2014.

[456] Youzhi Xu, Bingzhe Wang, Ramandeep Kaur, Martin B Minameyer, Michael Bothe, Thomas Drewello, Dirk M Guldi, and Max von Delius. A supramolecular [10] cpp junction enables efficient electron transfer in modular porphyrin–[10] cpp fullerene complexes. *Angewandte Chemie*, 130(36):11723–11727, 2018.

[457] Mark R Smyda and Stephen C Harvey. The entropic cost of polymer confinement. *The Journal of Physical Chemistry B*, 116(35):10928–10934, 2012.

[458] Lei Liu and Qing-Xiang Guo. Isokinetic relationship, isoequilibrium relationship, and enthalpy- entropy compensation. *Chemical Reviews*, 101(3):673–696, 2001.

[459] Taisuke Matsuno, Yusuke Nakai, Sota Sato, Yutaka Maniwa, and Hiroyuki Isobe. Ratchet-free solid-state inertial rotation of a guest ball in a tight tubular host. *Nature Communications*, 9(1):1907, 2018.

[460] Piero Gasparotto, Robert Horst Meißner, and Michele Ceriotti. Recognizing local and global structural motifs at the atomic scale. *Journal of chemical theory and computation*, 14(2):486–498, 2018.

[461] Carl Caleman, Paul J Van Maaren, Minyan Hong, Jochen S Hub, Luciano T Costa, and David 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*, 8(1):61–74, 2012.

[462] David van der Spoel, Paul J van Maaren, and Carl Caleman. Gromacs molecule & liquid database. *Bioinformatics*, 28(5):752–753, 2012.

[463] Ao Ma and Aaron R Dinner. Automatic method for identifying reaction coordinates in complex systems. *The Journal of Physical Chemistry B*, 109(14):6769–6779, 2005.

[464] Christoph Wehmeyer and Frank Noé. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of chemical physics*, 148(24):241703, 2018.

[465] Luigi Bonati, Valerio Rizzi, and Michele Parrinello. Data-driven collective variables for enhanced sampling. *The journal of physical chemistry letters*, 11(8):2998–3004, 2020.

[466] Jutta Rogal, Elia Schneider, and Mark E Tuckerman. Neural-network-based path collective variables for enhanced sampling of phase transformations. *Physical Review Letters*, 123(24):245701, 2019.

[467] Pavan Ravindra, Zachary Smith, and Pratyush Tiwary. Automatic mutual information noise omission (amino): generating order parameters for molecular systems. *Molecular Systems Design & Engineering*, 5(1):339–348, 2020.

[468] Yihang Wang, João Marcelo Lamim Ribeiro, and Pratyush Tiwary. Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nature communications*, 10(1):3573, 2019.

[469] João Marcelo Lamim Ribeiro, Davide Provasi, and Marta Filizola. A combination of machine learning and infrequent metadynamics to efficiently predict kinetic rates, transition states, and molecular determinants of drug dissociation from g protein-coupled receptors. *The Journal of Chemical Physics*, 153(12):124105, 2020.

[470] Stefano Piana and Alessandro Laio. A bias-exchange approach to protein folding. *The journal of physical chemistry B*, 111(17):4553–4559, 2007.

[471] Koji Hukushima and Koji Nemoto. Exchange monte carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65(6):1604–1608, 1996.

[472] Karen Palacio-Rodriguez, Hadrien Vroylandt, Lukas S Stelzl, Fabio Pietrucci, Gerhard Hummer, and Pilar Cossio. Transition rates and efficiency of collective variables from time-dependent biased simulations. *The Journal of Physical Chemistry Letters*, 13(32):7490–7496, 2022.

[473] Loukas Kollias, David C Cantu, Marcus A Tubbs, Roger Rousseau, Vassiliki-Alexandra Glezakou, and Matteo Salvalaglio. Molecular level understanding of the free energy landscape in early stages of metal–organic framework nucleation. *Journal of the American Chemical Society*, 141(14):6073–6081, 2019.

[474] Loukas Kollias, Roger Rousseau, Vassiliki-Alexandra Glezakou, and Matteo Salvalaglio. Understanding metal–organic framework nucleation from a solution with evolving graphs. *Journal of the American Chemical Society*, 144(25):11099–11109, 2022.

[475] Thomas D Potter, Elin L Barrett, and Mark A Miller. Automated coarse-grained mapping algorithm for the martini force field and benchmarks for membrane–water partitioning. *Journal of Chemical Theory and Computation*, 17(9):5777–5791, 2021.

[476] Tristan Bereau and Kurt Kremer. Automated parametrization of the coarse-grained martini force field for small organic molecules. *Journal of chemical theory and computation*, 11(6):2783–2791, 2015.

[477] Dimitris Nasikas, Eleonora Ricci, George Giannakopoulos, Vangelis Karkaletsis, Doros N Theodorou, and Niki Vergadou. Investigation of machine learning-based coarse-grained mapping schemes for organic molecules. In *Proceedings of the 12th Hellenic Conference on Artificial Intelligence*, pages 1–8, 2022.

[478] Cécile Hilpert, Louis Beranger, Paulo CT Souza, Petteri A Vainikka, Vincent Nieto, Siewert J Marrink, Luca Monticelli, and Guillaume Launay. Facilitating cg simulations with mad: The martini database server. *Journal of chemical information and modeling*, 63(3):702–710, 2023.

Appendix A

This appendix contains additional analysis and information that have been omitted from Chapter 6 for the sake of conciseness. Parts of these results are part of the published article:
L. Leanza, C. Perego, L. Pesce, M. von Delius, M. Salvalaglio and G. M. Pavan. "Into the dynamics of rotaxanes at atomistic resolution" Chem. Sci. 2023, DOI:10.1039/D3SC01593A

Shuttling and kinetics of $[R_4 - H_2]^{2+}$ [2]Rotaxane

To study the translational dynamics of system **1**, we employed WT-MetaD simulations, biasing: 1) the difference d of the distance between the centre of geometry of the oxygen atoms of the ring and the centre of geometry of one benzimidazolium (d1) and between the centre of geometry of the oxygen atoms of the ring and the second benzimidazolium (d2); 2) the distance $d_{benz-benz}$ between the two benzene rings of the dibenzo[24]crown-8 ether macrocycle (see the schematic representation of CVs in Fig. A1).

We initially considered the position of the center of geometry of all the heavy atoms of the macrocycle as a collective variable (CV). However, the visual inspection of the trajectories showed that changes in the CV often were the result of fluctuations of the benzene rings of the macrocycle, that used to form aromatic stackings with the phenyl rings along the dumbbell or with the benzimidazolium site. These changes were not indicative of the actual translation of the macrocycle that we aimed to study, but rather of faster conformational motions. Therefore, we decided to consider the center of geometry of the oxygen atoms of the macrocycle only, and include a second CV describing the benzene rings opening-closing.

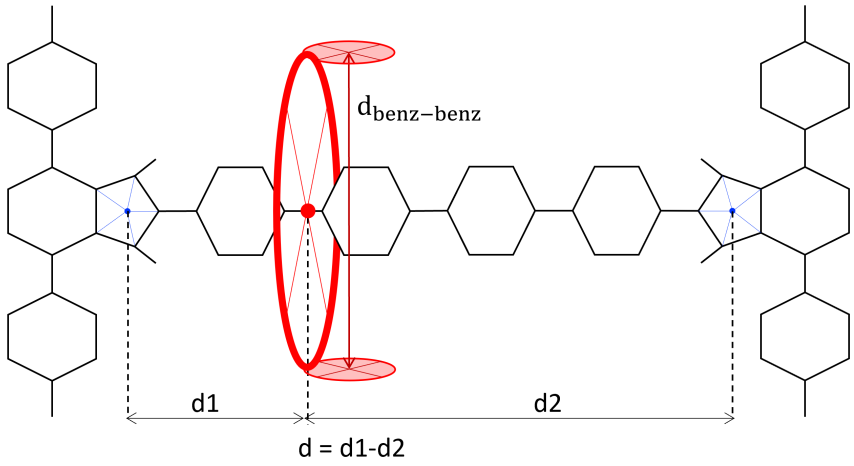

Fig. A1 Schematic representation of WT-MetaD simulation setup for system **1**. The two selected CVs d and $d_{benz-benz}$ are displayed. Adapted from Ref.[405].

The bias has been updated every 500 time steps (1 ps of simulation time) using Gaussians of initial height 1 kJ/mol, the width of 0.1 nm and 0.05 nm for CV1 and CV2, respectively, and a bias factor of 20 in DMSO, 30 in ACN, and 40 in DCM and $CHCl_3$. For each system, we ran a minimum of 1.5 μs, allowing us to record multiple shuttling events across stations.

For the calculation of the shuttling rates in **1** from bound to unbound we ran $\sim$25 infrequent WT-MetaD simulations, starting from the bounded state and biasing the distance between the centre of geometry of the ring and the centre of geometry of the closest benzimidazolium ($d1$, CV1) and the distance ($d_{benz-benz}$, CV2) between the two benzene rings of the dibenzo[24]crown-8 ether macrocycle. Each run ends when CV1 falls in basin D (see 2D-FES in Fig. 6.2b). In **ACN**, the bias has been added every 20 ps using Gaussians of initial height 1 kJ/mol, width of 0.1 nm and 0.05 nm for CV1 and CV2, respectively, and a bias factor of 20.

The times for transitions D $\rightarrow$ A and D $\rightarrow$ D' in ACN are collected via multiple classical MD runs. Every 500 simulation steps, PLUMED checks whether CV1 is in basin A or D' and, eventually, stops the simulation.

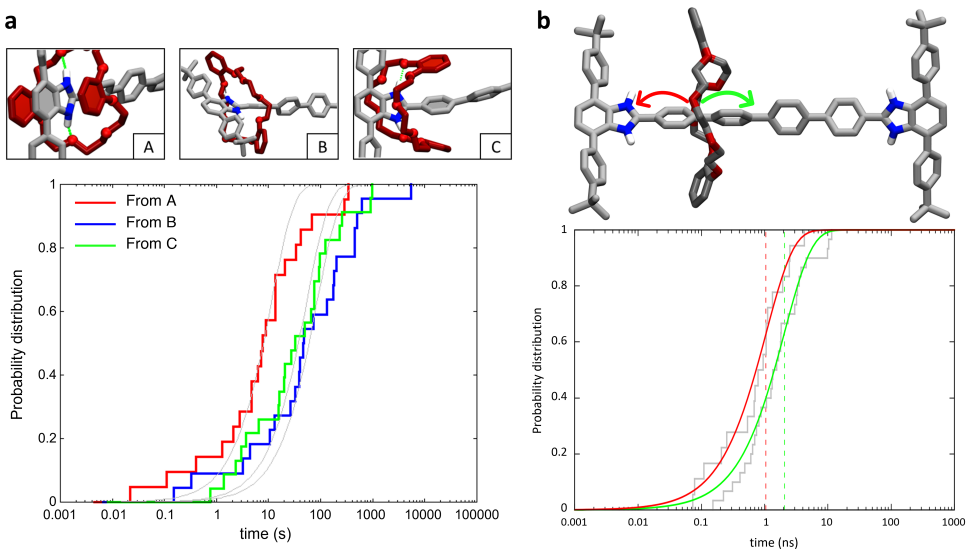

Fig. A2 (a) The transition times for the macrocycle shuttling are estimated via multiple infrequent WT-MetaD simulations starting from different starting points: state A (red steps), from state B (blue steps), and state C (green steps). The fitting of ECDF via the Poisson function (grey curves) leads to very similar results in the three cases ($\sim$10s, $\sim$70s, $\sim$40s, respectively), suggesting that the disruption of the h-bonds between macrocycle and binding site is the rate-limiting step, and it is barely affected from the aromatic stacking between rings. The starting points (reported in Fig. 6.2b) are reported at the top of the panel for clearer interpretation. (b) Top panel: atomistic molecular structure of **1**, state D. The red and green arrows indicate the two possible directions of motion of the macrocycle. Bottom: Transition times estimated via multiple MD simulations (grey steps) and Poissonian probability distributions (red and green curves) for the D$\rightarrow$A and D$\rightarrow$D' motion of the macrocycle along the dumbbell. The vertical dashed red and green lines represent the characteristic timescale τ associated with the events and are estimated by fitting the ECDF with the Poissonian distributions. Adapted from Ref.[405].

In **$CHCl_3$** and **DCM**, the bias has been added every 50 ps, using Gaussians of initial height 2 kJ/mol, a width of 0.1 nm and 0.05 nm for CV1 and CV2, respectively, and a bias factor of 45. In **DMSO**, the bias has been added every 80 ps, using Gaussians of initial height 2.5 kJ/mol, width of 0.1 nm and 0.05 nm for CV1 and CV2, respectively, and a bias factor of 25.

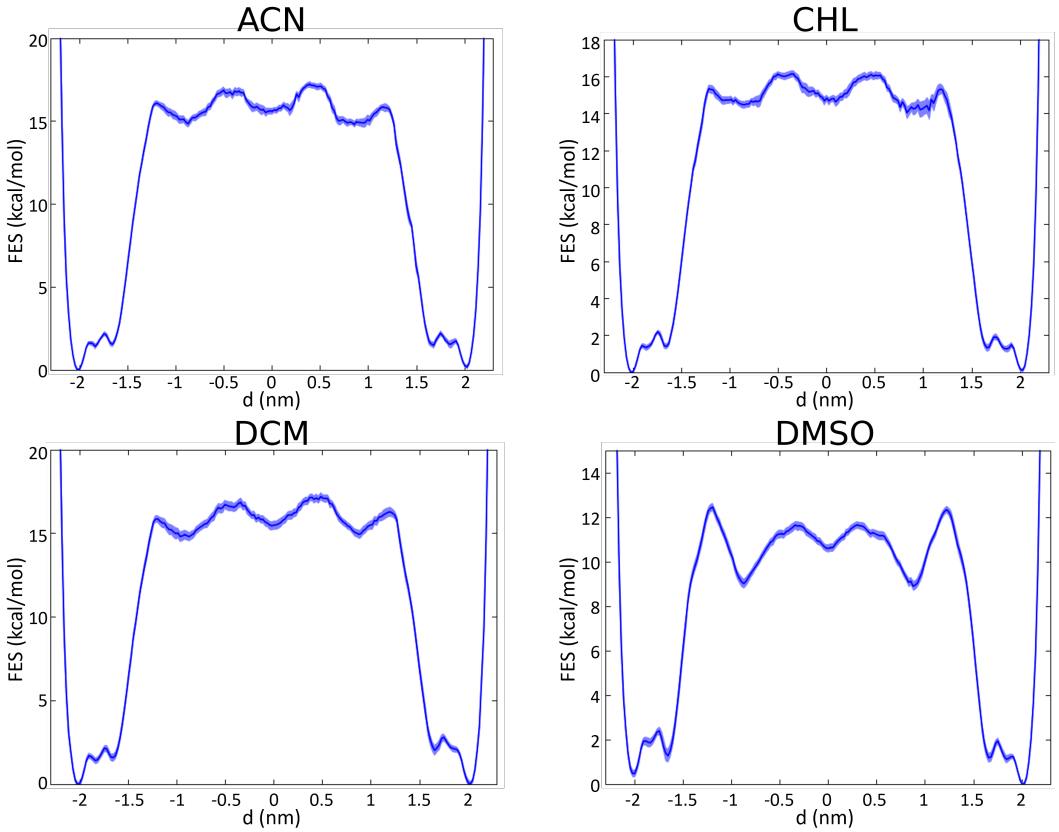

Fig. A3 FESs from WT-MetaD simulations in the four solvents. The associated error bars are calculated using block averages[282].

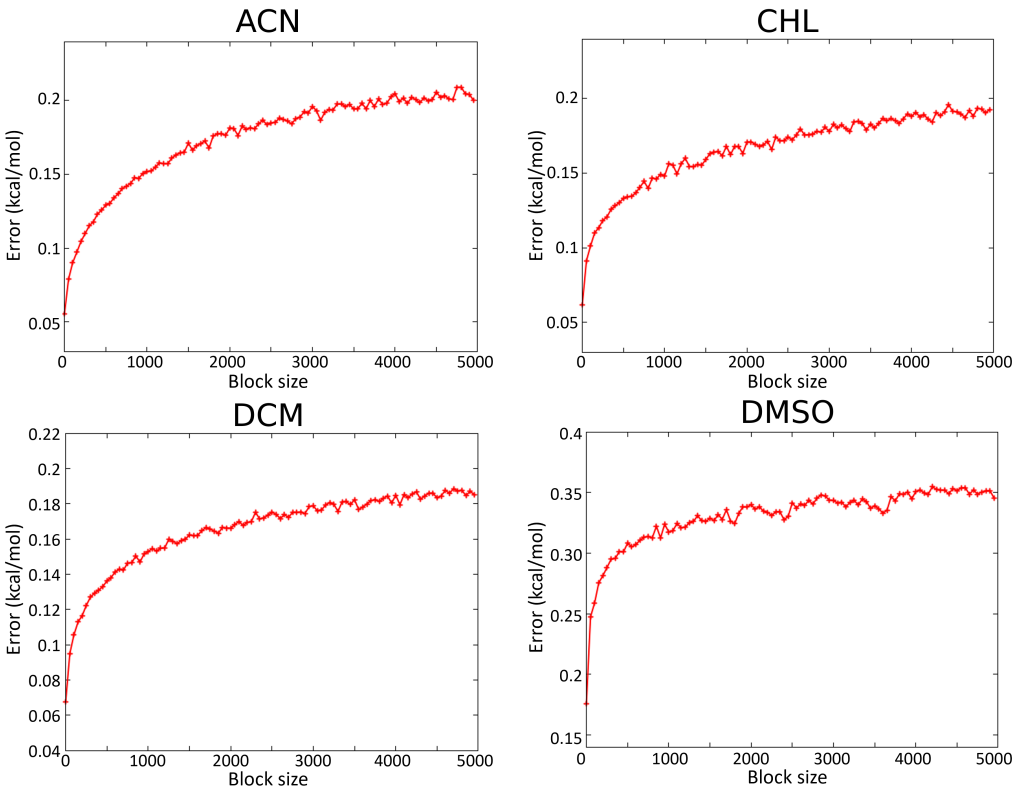

Fig. A4 Block analysis of the WT-MetaD simulations.

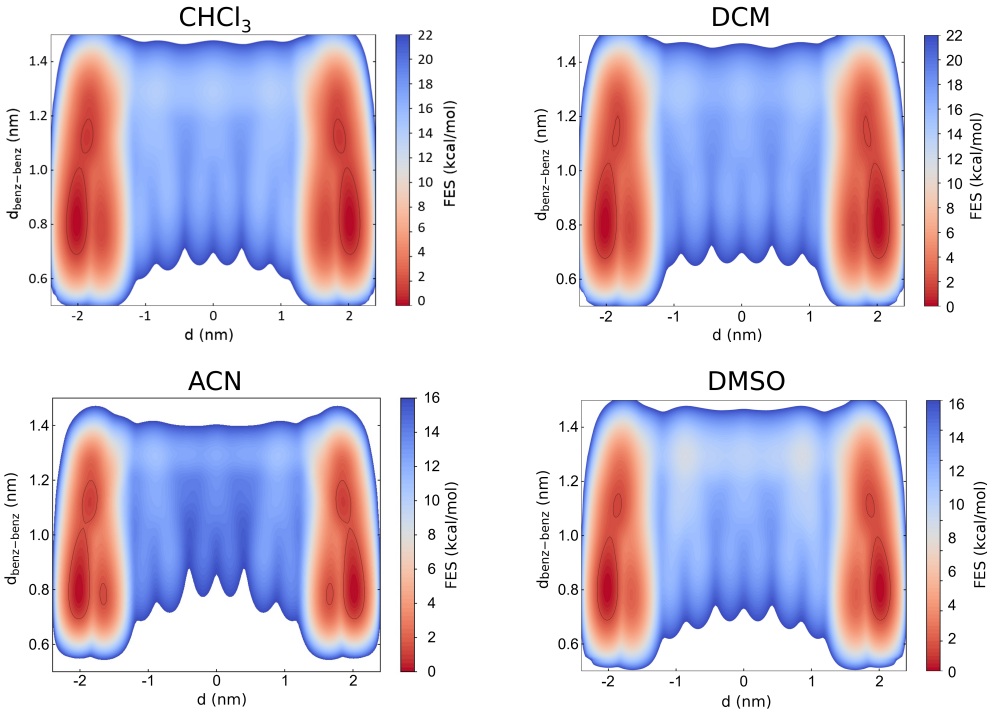

Fig. A5 FESs as a function of the two CVs biased during WT-MetaD simulations in all the four solvents investigated in this study. The free-energy profiles are symmetrized with respect to $d = 0$, exploiting the symmetry of the system. Adapted from Ref.[405].

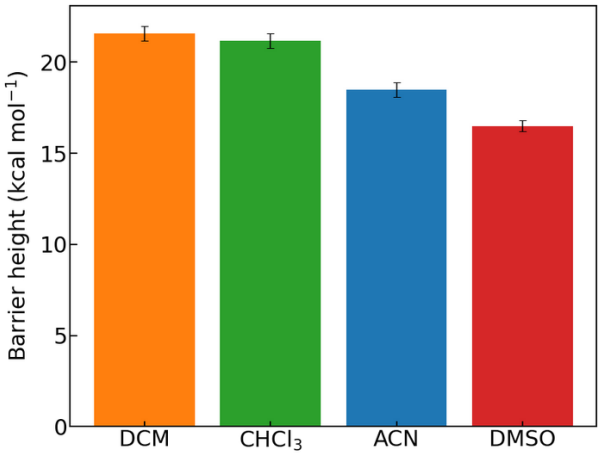

Fig. A6 Barrier heights of unbinding from the recognition site in various solvents, estimated from infrequent Metadynamics. Adapted from Ref.[405].

Shuttling and kinetics of formamidinium [2]Rotaxane

In formamidinium [2]Rotaxane[448], we run 1μs of classical MD simulation, in NPT conditions. The 2D-FES has been computed along the difference between the distance between the centre of geometry of the macrocycle and the centre of geometry of one stopper and the distance between the centre of geometry of the macrocycle and the second stopper.

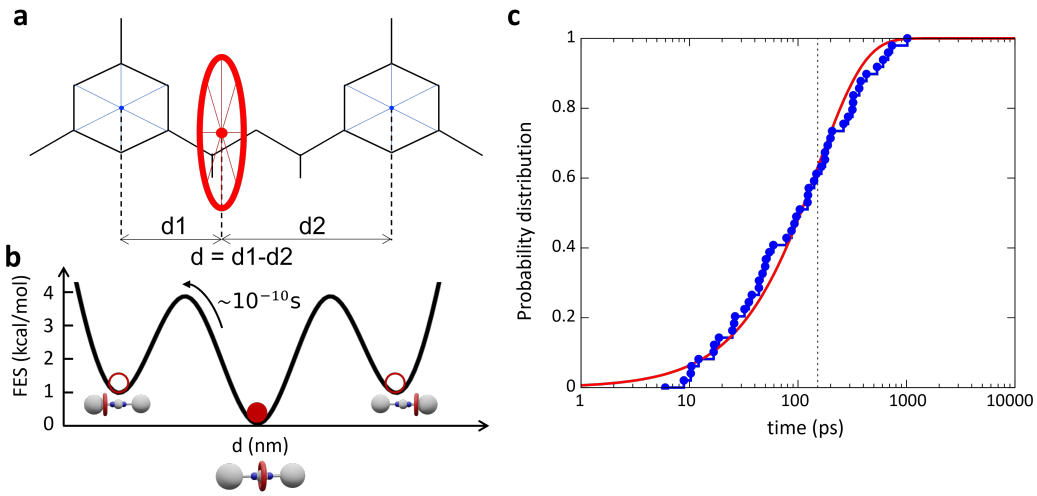

Fig. A7 (a) Schematic representation of **2**. The CV d used for reweighting the free-energy surfaces is displayed. (b) Free-energy surface diagram as a function of d and relative transition times. The free-energy value of each state (coloured circles) and the height of free-energy barriers are defined with respect to the central minimum. (c) Transition times obtained from multiple single-transition MD replicas. Adapted from Ref.[405].

Shuttling and kinetics of [10]CPP Fullerene Pseudorotaxane and [10]CPP Fullerene [2]Rotaxane

The rates for the decomplexation of [10]CPP-Fullerene Pseudorotaxane[449] have been calculated running 20 infrequent WT-MetaD using the distance between the centre of geometry of [10]CPP and the centre of geometry of the fullerene as CV, depositing Gaussians every 40 ps, 0.5 kJ/mol of initial height, 0.1 nm width and a bias factor of 10.

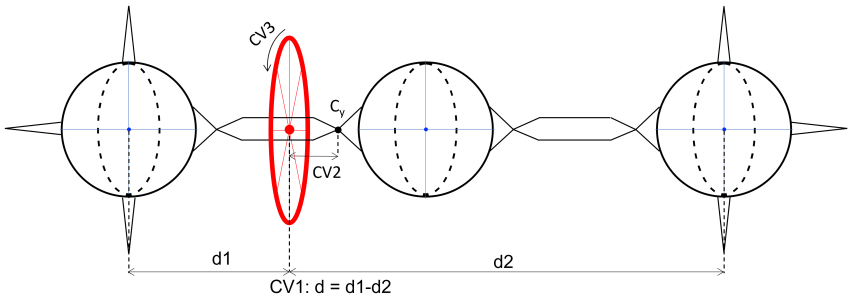

Fig. A8 Schematic representation of WT-MetaD simulation setup for system **3**. The selected CVs (CV1, CV2, CV3) are displayed. Adapted from Ref.[405].

We used here a WT-MetaD bias acting on the difference *d* between the distance between the centre of geometry of the [10]CPP and the centre of geometry of one stopper (d1) and the distance between the centre of geometry of the [10]CPP and the second stopper (d2); Gaussians has been added every ps, with an initial height of 1 kJ/mol, gaussian width of 0.3 nm, a bias factor of 25. To prevent the *trans*-3 fullerene bisadduct from closing under the effect of the Metadynamics bias, we imposed an external wall potential that acted on the centre of geometry of the two fullerene hexakis-adduct stoppers when their displacement from reference is greater than 1 nm, with a force constant equal to 200 (kJ/mol)/nm^2.

For a quantitative estimation of the characteristic timescale associated with a transition from “bound“ to “unbound“ states in [10]CPP-Fullerene [2]Rotaxane, three CVs were used: 1) The distance between the centre of geometry of [10]CPP and the centre of geometry of the central fullerene (CV1); 2) The distance between the centre of geometry of [10]CPP and the first carbon atom of the substituent C_y (CV2); 3) The relative rotation between [10]CPP and the central fullerene (CV3). The bias has been added every 40 ps using Gaussians of initial height 0.5 kJ/mol, width of 0.1 nm for CV1 and CV2 and 0.3 radians for CV3, and a bias factor of 10.

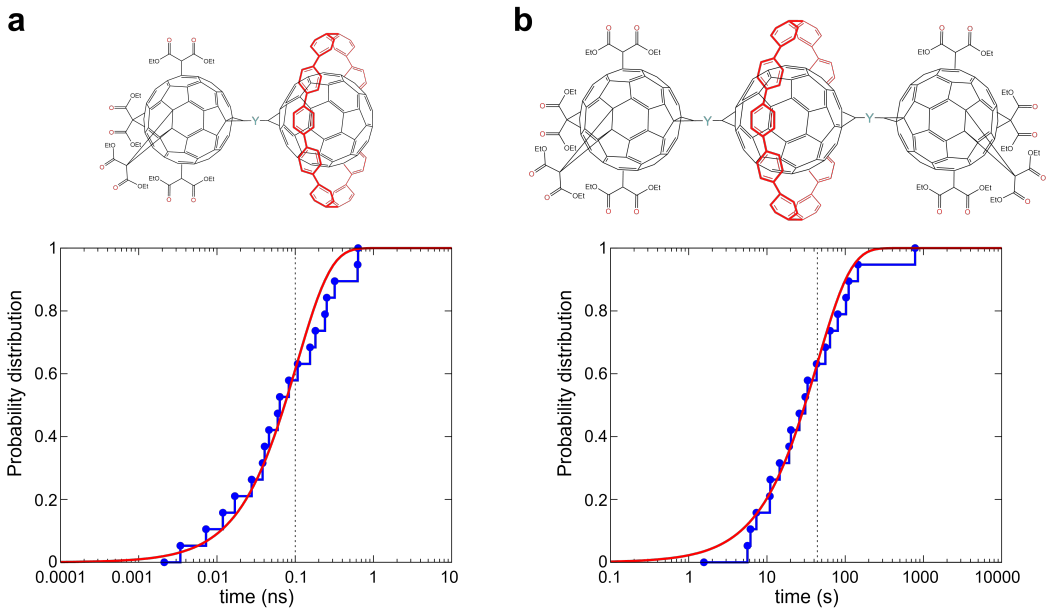

Fig. A9 (a) Transition times for [10]CPP Fullerene Pseudorotaxane decomplexation (jump in solution) estimated via multiple infrequent WT-MetaD. (b) Estimated times for [10]CPP to leave the central Fullerene ("bound" state) and jump on one lateral stopper ("unbound" state), collected via multiple infrequent WT-MetaD). Adapted from Ref.[405].

Calculation of the rotation in [10]CPP-Fullerene [2]Rotaxane

The analysis of the dynamics of the [10]CPP over the central Fullerene was done by decomposing its rotation (ϕ – rotation axis of the [10]CPP parallel to cylindrical symmetry axis of the molecule itself) from the tilting (ψ – tilting axis orthogonal to the cylindrical symmetry axis of [10]CPP) (see Fig. 4d in main text). Performing principal component analysis (PCA) over $\cos(\phi)$, $\sin(\phi)$, $\cos(\psi)$ and $\sin(\psi)$ sampled during unbiased MD, the first two principal components, PC1 and PC2 (the new coordinates obtained from the analysis), transform the displacement of the measured values into ten blobs over a circle (see Fig. 4e in main text the $-K_BT\log(\rho)$ map, where ρ is the density map of the configurations). The position along the circle is directly related to ϕ as shown in Fig. 4e, while the radial position correlates with ψ. By performing cluster analysis by the means of Probabilistic Analysis of Molecular Motifs (PAMM)[460], we could identify ten clusters (states), and by analysing the dynamics time evolution of the labelling of each data-point, we could estimate the ~ 0.1ns residence time in each cluster (configurations). These clusters identify periodic configurations of the macrocycle around the fullerene, which demonstrate

that the rotation of the latter is fast in these conditions (occurring in the nanoseconds scale). From this analysis it is clear that the most relevant movement of [10]CPP is the rotation, while the tilting mainly fluctuates about its equilibrium position.

Shuttling and kinetics of rigid bistable [2]Rotaxane

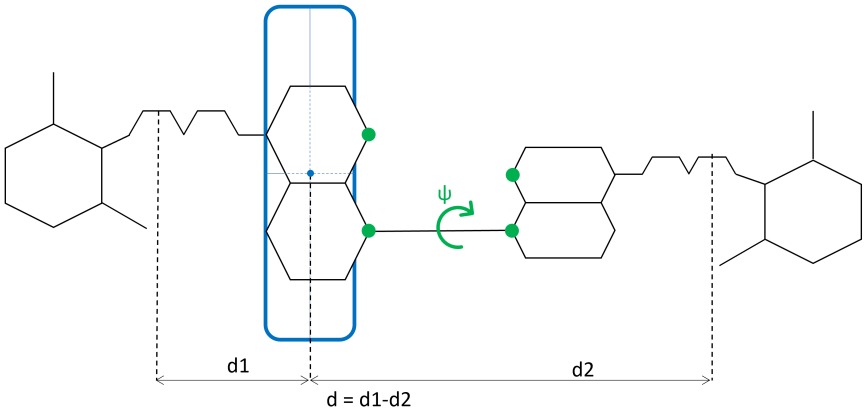

Fig. A10 Schematic representation of WT-MetaD simulation setup for system **4**. The two selected CVs d and ψ are displayed. Adapted from Ref.[405].

The WT-MetaD simulation of system **4**[426] has been run using the difference between the distance between the centre of geometry of $CBPQT^{4+}$ and the oxygen atom of the last PEG of one lateral chain and the distance between the centre of geometry of the ring and the oxygen atom of the last PEG of the other lateral chain (CV1) and the torsional angle between the two naphthalenes (CV2) as CVs. The bias has been added every 2 ps, using Gaussians with an initial height of 1 kJ/mol, a width of 0.2 nm for CV1 and 0.2 rad for CV2, and a bias factor of 20.

For estimating the kinetic rates, forty-five infrequent WT-MetaD have been run, using the distance between the centre of geometry of $CBPQT^{4+}$ and the centre of geometry of the closest naphthalene (CV1) and the torsional angle between the two naphthalenes (CV2) as collective variables. The bias has been added every 20 ps, using Gaussians with an initial height of 0.239 kJ/mol, a width of 0.1 nm for CV1 and 0.2 rad for CV2, and a bias factor of 20. The simulations ended when $CBPQT^{4+}$ jumped from station 1 to station 2.

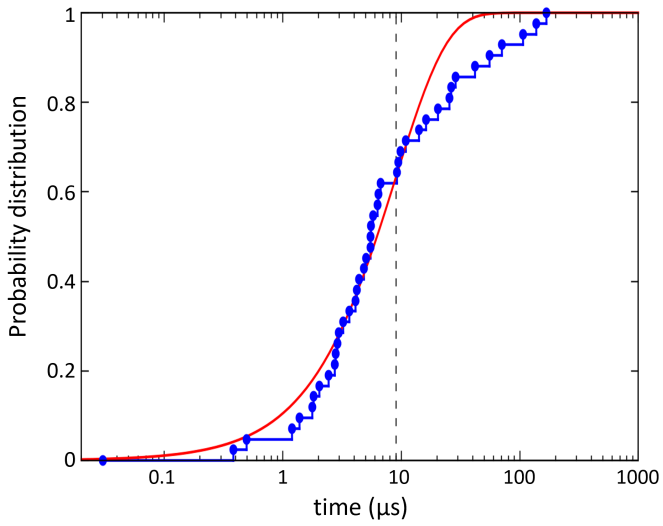

Fig. A11 Transition times estimated via multiple infrequent Metadynamics simulations (blue steps) and Poissonian probability distribution (red curve) for the transition of $CBPQT^{4+}$ from one naphthalene to the other. The vertical dashed line represents the characteristic timescale τ associated with the event and is estimated by fitting the ECDF with the Poissonian distribution. Adapted from Ref.[405].

Table A1 Details of Molecular Systems simulated in Chapter 4

System	**Solvent**	**Number of atoms**	**Size of the box** (Å)	**Barostat**	**Simulation time** (μs)
1	ACN	26162	8.2x7.4x6.7	Parr-Rahm	2
1	$CHCl_3$	26621	10.6x9.1x7.6	Parr-Rahm	2
1	DCM	22676	9.5x8.1x6.8	Berendsen	2
1	DMSO	21732	7.5x6.4x5.3	Parr-Rahm	1.6
2	DCM	5495	5.6x4.8x4.5	Berendsen	1
3	$CHCl_3$	21849	9.7x7.8x7.8	Berendsen	2
4	ACE	11724	8.9x6.9x6.9	Parr-Rahm	2