MOLECULAR DYNAMICS SIMULATIONS OF BIOPOLYMERS IN MATERIALS SCIENCE AND MEDICINE

By

JOHN W. WHITTAKER

A thesis submitted to the

Graduate School-Camden

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of Master of Science

Graduate Program in Computational & Integrative Biology

Written under the direction of

Dr. Luca Larini

And approved by

Dr. Luca Larini

Dr. David Salas-de la Cruz

Dr. Jinglin Fu

Camden, New Jersey

May 2018

THESIS ABSTRACT

Molecular Dynamics Simulations of Biopolymers in Materials Science and Medicine

by JOHN W. WHITTAKER

Thesis Director:

Dr. Luca Larini

Biopolymers are the molecular building blocks of the natural world and are a focal point of many areas of active research. Despite their universality, many of their properties are still poorly understood because these properties are associated with a level of detail that is far beyond experimental characterization. Molecular dynamics (MD) simulations offer the opportunity to study these ubiquitous macromolecules with femtosecond accuracy and angstrom resolution. In this work, we examine methodologies that are crucial to accurate, representative, and meaningful results to MD simulations of biopolymers. We conduct simulations using replica-exchange molecular dynamics, implicit and explicit solvent, and umbrella sampling in order to characterize these systems. We explore methods of validating empirical potential energy functions and use the results to characterize the effects of phosphorylation and its experimental counterpart pseudophosphorylation on the conformational ensemble of the tau-derived protein PHF43

in the early stages of aggregation in Alzheimer's disease. We then consider how MD methodologies can bolster theoretical models and be applied to the production of biomaterials from tunable biopolymer mixtures by modeling the dissolution and regeneration process of cellulose in ionic liquids.

Table of Contents

List of Figures

List of Tables

Notes on Publication

This work is an exposition of the following papers that have been published (or are currently in preparation/to be submitted):

1. **Impact of Phosphorylation and Pseudophosphorylation on the Early Stages of Aggregation of the Microtubule-Associated Protein Tau**. Dmitriy Prokopovich, John Whittaker, Micaiah Muthee, Azka Ahmed, and Luca Larini. *The Journal of Physical Chemistry B* **2017** 121 (9), 2095-2103
2. **Computational and Experimental Analysis of Short Peptide Motifs for Enzyme Inhibition**. Jinglin Fu, Luca Larini, Anthony Cooper, John Whittaker, Azka Ahmed, Junhao Dong, Minyoung Lee, Ting Zhang. PLoS One **2017** 12(8): e0182847
3. **A Theory for Designing Synthetic Compartmentalization of Enzymatic Systems**. Anthony Cooper, John Whittaker, Luca Larini. (To be submitted)
4. **A Hierarchical Model for the Processing of Polysaccharides/Protein-Based Films in Ionic Liquids.** Amnah Hadadi, John Whittaker, David Verrill, Luca Larini, David Salas-de la Cruz. (To be submitted)
5. **Evaluation of Force Field using Implicit Solvent in the Study of Intrinsically Disordered Proteins.** John Whittaker, Cory Trout, Gabriele Stankeviciute, Luca Larini. (In preparation)

1 Introduction

Effective modeling of complex molecular systems like biopolymers is crucial to the development, integration, and success of a broad range of fields including nanotechnology, medicine, and biomaterials. Understanding biological systems, though, at their microscopic level is an especially complicated task that involves contributions from a diverse fusion of physical and chemical theory that is actively being modified by the day. Advances in computational power over the past few decades have profoundly impacted the researcher's ability to both build more rigorous theoretical models and test and examine the results of these models. Today, molecular modeling techniques have become ubiquitous and their overall success hinges on the ability to translate the complex phenomena of nature into simple models and simple approaches to solve these models.

Classical molecular dynamics (MD) simulations are one particular theoretical tool that has been used extensively with great success.[1-3] MD simulation offers the unique ability to observe and probe molecular systems with atomic-mesoscopic resolution that exceeds the capabilities of modern experimental techniques. These simulations are primarily based on approximations that allow for the use of classical mechanics to describe the motion of fixed-charge nuclei while ignoring electronic degrees of freedom. Approximate methods make the exploration of the dynamics of a complex system computationally feasible as even with today's relatively massive and optimized computational resources, a purely quantum mechanical, statistically-relevant description

of a large system on a biological timescale is impossible at present.[4] A brief introduction to classical MD and approximations used in this work is given in **Chapter 2**.

The purpose of the rest of this work is to demonstrate how MD methods are invaluable to aiding our understanding of molecular medicine and the development and design of novel biomaterials. For many years computational methods have been an invaluable tool for rational drug design[5] in contexts such as elucidating electrostatic energies associated with ligand binding[6] and closely examining the conformational ensemble of medicinally-relevant macromolecules, like proteins.[7-8] **Chapters 3** describes a simple methodology for selecting an appropriate empirical MD force field in order to accurately simulate a protein system while **Chapter 4** uses the results from the previous chapter to perform an MD study investigating the protein tau and its role in Alzheimer's disease.

In the scope of design and development of biomaterials, MD has been used extensively to investigate biomaterial-biomolecule interfacial interaction[9] and the efficient production and harvesting of the abundant biopolymers that provide the building blocks for many biomaterials.[10] The latter of which is detailed in **Chapter 5** of this thesis.

In general, the methods that make research endeavors like those described above and later in this work possible are rooted in the theory of statistical mechanics. As Mark Tuckerman puts it, "…molecular simulation is the computational realization of statistical mechanics".[11] In this work we also delve into purely theoretical undertakings in an effort to provide simple, useful models that describe important properties of biomaterials. For example, the structural properties and synthesis of mixtures of biopolymers into novel,

naturally-derived biological materials are discussed in **Chapter 5** and **Appendix C,** while the implications on catalytic activity of enzymes encapsulated by simple DNA origami nanocages are detailed in **Appendix B**.

2 Background

In order to complement results in this thesis, we believe that an introduction to fundamental tools and methods of classical MD is necessary.

2.1 Classical Molecular Dynamics

Classical MD simulation of a molecular system is based on an iterative approach to numerical solutions of Newton's equations of motion:

$$m\ddot{\boldsymbol{r}}_i = \boldsymbol{f}_i \qquad \boldsymbol{f}_i = -\frac{\partial}{\partial \boldsymbol{r}_i} U(\boldsymbol{r}^N) \tag{2.1}$$

Where $U(\boldsymbol{r}^N)$ is a potential energy function whose argument $\boldsymbol{r}^N = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots \boldsymbol{r}_N)$ represents the complete set of $3N$ atomic coordinates. Integration of this system of differential equations yields a trajectory of the system as it evolves through time.

As mentioned in **Chapter 1**, classical molecular dynamics is built upon approximations of reality. Classical mechanics cannot perfectly describe the complexities associated with nature on the nanoscale, but clever methods have been developed over the past 50-60 years that show us these approximations are often good enough to tackle issues that arise in the microscopic world.

This does not mean that MD can be haphazardly applied to any microscopic system. Indeed, MD simulation should not be considered a “black box”. Great care must be taken in choosing the tools and methodologies to ensure accurate, useful results of the system under investigation.

In the most general sense, MD as a predictive tool suffers from two major limitations: inaccuracies in force fields and inadequate sampling of the system of interest’s properties.

2.2 Force Fields

Force fields are the MD realization of the potential energy function $U(\boldsymbol{r}^N)$ found in Equation 2.1. A commonly used additive functional form (that has been truncated slightly) of $U(\boldsymbol{r}^N)$ is given by[12]:

$$U = \sum_{bonds} \frac{1}{2} K_r (r - r_0)^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\varphi [1 - \cos(n\varphi + \delta)] + \sum_{non-bonded} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right) \quad (2.2)$$

This function is the sum of five contributions to the potential energy: bond stretching, bond angle opening and closing, rotation about dihedral angles, and non-bonded interactions encompassing van der Waals and electrostatic interactions.

A multitude of force fields have been developed and systematically improved since the 1960's.[13] These force fields, by and large, share a similar functional form and are empirically parameterized, meaning they derive parameters such as bond angles, partial charges, and atomic radii from experiments or *ab initio* quantum mechanical calculations.[14] The ultimate goal of an MD force field is to approximate the true quantum mechanical potential energy surface of a system of molecules but—as the fact that dozens of different force fields have been developed over the years suggests—there is no definite philosophy for determining and selecting force field parameters, and this goal can be accomplished in many different ways. Selecting a force field is heavily dependent upon the system under investigation. This component of choosing an effective potential is addressed in **Chapter 3** of this work.

2.3 Sampling Methods

Fully atomistic MD simulations—those that provide the highest level of detail—of large systems (i.e., large proteins solvated in many explicit solvent molecules) are immensely computationally expensive. Without using the enormously powerful brute-force approach of specially-designed supercomputers[15], "vanilla" MD methods are limited to nanosecond-microsecond time scales. Most interesting processes involving biopolymers, such as protein folding, occur on the order of many milliseconds, seconds, or even days. These processes are often characterized by multiple metastable, intermediate states separated from one another by large energetic barriers, making them exceedingly difficult to sample in full using MD.

To combat this issue, a variety of methods and approximations have been developed that aim to enhance the sampling capabilities of MD while retaining an appropriate level of accuracy with respect to a fully atomistic, continuous MD simulation.

2.3.1 Implicit Solvation

Atomistic MD simulations can generally be broken down into two broad categories: those that treat solvent *explicitly* and those that treat it *implicitly*.

Explicit solvation methods treat the solvent as discrete molecules whose interactions amongst themselves (solvent-solvent interactions) and with the solute (solvent-solute interactions) are computed explicitly. These methods are of course an area of major research and are rather efficient and accurate, but it still stands that the intense demands of computing the massive amount of interactions in a large explicitly-treated system is debilitating for many computational resources.

Implicit solvent methods, on the other hand, speed up MD simulations by replacing discrete solvent molecules with a dielectric continuum that provides a mean-field interaction with the solvent. Thus, implicit solvation drastically reduces the number of interactions computed with each step of an MD simulation. Furthermore, an additional speedup of the simulation occurs due to the low-viscosity nature of the implicit solvent environment, bolstering a system's ability to explore its conformational space in a timelier manner.

Our use of implicit solvent over explicit solvent in **Chapters 3 and 4** is motivated by the benefits described above. In those chapters, in order to successfully study large proteins which are particularly dynamic in nature, we must focus on utilizing a method that offers efficient and complete sampling of broad configurational spaces.

That said, despite implicit solvent models delivering a ~2-20 times[16] speedup in conformational sampling, for systems such as those studied in **Chapters 3 and 4**, it may be necessary to introduce a second method which enhances sampling further still. One method that works additively with implicit solvent is known as replica-exchange molecular dynamics.

2.3.2 Replica-Exchange Molecular Dynamics

The replica-exchange molecular dynamics (REMD) algorithm[17] is a widely used enhanced sampling method that works by applying the theory of parallel tempering[18] to molecular dynamics.

In simulations that involve protein folding especially, a room temperature system tends to become trapped in one of many local minimum energy states. In short, the REMD algorithm is based upon the fact that the energy barriers that separate the aforementioned energy wells on the potential energy surface of a biopolymer are much easier to overcome at higher temperatures than at lower temperatures. It consists of running multiple, non-interacting MD simulations (called replicas) in parallel, each at their own temperature (typically a sequence of increasing temperatures from ~298 K –

600 K). After a predetermined number of MD time steps, an attempt is made to exchange the coordinates of one replica with the coordinates of the adjacent (in temperature space) replica. The proposed exchange is accepted with probability given by a Metropolis Monte Carlo criterion given by equation (17) in Sugita et al.[17]

This criterion, based on the Boltzmann factors of both replicas considered, guarantees that exchanges occur only when a particular energy state of the system of interest is probable in both replicas (therefore, at both temperatures). In this way, replicas "wander" through temperature space and configurations from higher temperature simulations are made available to lower temperature simulations and vice versa. This results in a robust ensemble of conformations of your system that samples both low and high temperature configurations.

3 Validating Force Fields for Future Test Systems

A force field must perform a very sensitive balancing act in order to be applied to a wide swath of systems. In fact, the large majority of the many contemporary force fields differ by only a small number of torsional parameters. To accurately represent a system's conformational ensemble accurately, a force field must not only deliver on the ability to predict, sample, and maintain relevant secondary structure with respect to more structured proteins and peptides, but it must also straddle the line between a tendency for α or β bias.

Force fields are notorious for biases for or against certain secondary structure elements. The Cornell et al.[19] force field, ff94, was one of the first force fields to be widely adopted by researchers but was found to favor helical conformations.[20-23] Further refinements and improvements of this and other force fields over the years have yielded results that vary greatly depending on factors such as the nature of the peptide involved, the use of an implicit or explicit solvent model, and the sampling techniques utilized, among many others[21-27]. Thus, it is of utmost importance to carefully consider the force field that one will use to treat their system, else one may be left with results that in no way reflect reality.

In this chapter, we examine the ability of popular parameterizations of the AMBER force field family—ff96[28], ff99SB[24], and the more recent ff14SBonlysc[29] and ff14IDPSFF[30], in combination with the GB-OBC[31] (ff96, ff99SB) and GB-Neck2[32] (ff14SBonlysc, ff14IDPSFF) implicit solvent models—to accurately sample the conformational ensemble of the tau-derived peptide PHF43. The force field which best represents PHF43 will be used to examine it further in **Chapter 4**.

3.1 Methodology

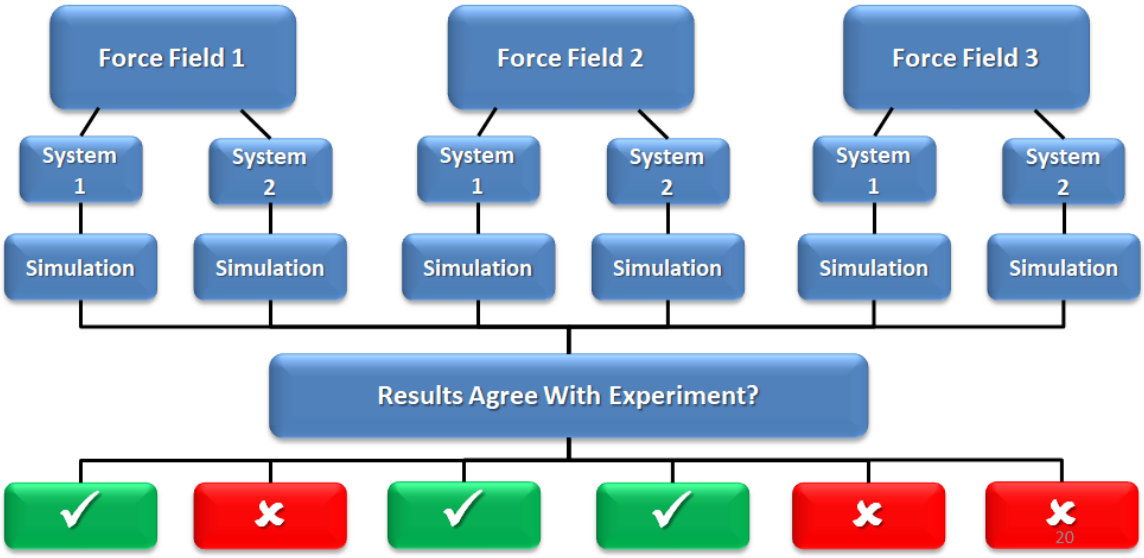

Figure 1. Schematic Representation of Simple Methodology for Choosing an Appropriate MD Force Field.

The basic methodology utilized in this chapter is given schematically by **Figure 1**. A series of force fields are applied to a series of systems and simulations are performed utilizing these parameters. Once the simulations have concluded, they are analyzed and

results are compared with relevant experimental data. From here we simply conclude whether or not the force field is worth using based on the agreement of relevant properties with experiment. Here, the series of force fields are again, AMBER ff96, ff99SB, ff14SBonlysc, and ff14IDPSFF and the system they are applied to is the tau-derived peptide PHF43.

A variety of methods of analysis are used to characterize the system and compare simulation results to experimental properties. Namely, simulations in this chapter and the following chapter are characterized by the peptide's radius of gyration, end-to-end distance, hydrogen bond content, and secondary structure content. These are described below.

Radius of Gyration and End-to-End Distance

First, in order to have a qualitative description of the entire set of conformations, we have decided to plot the radius of gyration of the molecules against its end-to-end distance. The radius of gyration provides a rough estimate of the size of the aggregate, while the end-to-end distance allows estimation of the entropy of each conformation. The radius of gyration can be thought of as the radius of a sphere that encloses the protein. However, for any given radius there can be multiple arrangements of the protein backbone. For this reason, we need a second parameter that characterizes this conformational ensemble. We chose the end-to-end distance, in agreement with the theory of ideal polymer chains[33]. Based on these considerations, we can expect that compact conformations will appear in the lower left corner of the plot, whereas the more extended conformations will populate the upper right corner. In addition, as only one conformation is available for a fully

extended chain, we expect that the entropy will decrease as the radius of gyration of the molecule increases. This means that the number of conformations and, as a consequence, the possible values of the end-to-end distance, decreases with increasing radius of gyration.

Cluster Analysis

To gain a better understanding of the conformations populated by the molecule, we performed a cluster analysis using the Daura algorithm[34] as implemented in the GROMACS package[35-38]. The clustering algorithm compares the conformations found in our simulation and groups them according to their root mean square deviation (RMSD). The idea is that structures that are very close to one another show a smaller RMSD. The RMSD of two molecules is defined as:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - B_i)^2} \qquad (3.1)$$

Where N is the number of atoms in one molecule, A_i and B_i are the i-th atom of molecule A and B, respectively. In our study, we consider structures with a RMSD between 3 and 5 Å to belong to the same cluster, depending on the needs imparted by differences in force fields. The algorithm also takes care of ranking the clusters according to their abundance, with the first cluster being the most abundant. A major reason to use cluster analysis is that while conformations may be very close in size (and so they appear in the

same area of the radius of gyration vs end-to-end distance plot), they may show a different secondary structure.

Hydrogen Bond Analysis

The number of hydrogen bonds is essential in order to estimate the amount of secondary structure in a protein. A hydrogen bond is considered present if the nitrogen N and the oxygen O in two different peptide bonds are at a distance smaller than 3.5 Å and the angle N - - O=C is bigger than 120°.

3.2 Simulation Details

All simulations were performed using the AMBER14[39] software package, using the ff96[28], ff99SB[24], ff14SBonlysc[29], and ff14IDPFF force fields. In association with the force fields utilized, we also examined two different implementations of the generalized Born (GB) implicit solvent model: GB-OBC[31] with both ff96 and ff99SB, and GB-Neck2[32] with ff14SBonlysc and ff14IDPSFF. GB solvent models were paired with the appropriate atomic radii: mbondi2 and mbondi3 for igb5 and igb8, respectively with the exception of simulations performed using ff96—these utilized mbondi, the default AMBER atomic radii for GB-OBC.[31] All peptides were generated, force fields selected, and atomic radii chosen using the tLEaP module included in the AMBER14[39] package.

In order to enhance conformational sampling of our system, replica exchange molecular dynamics (REMD) was employed for all production simulations. Each REMD simulation consisted of 16 replicas at temperatures spaced between 279 and 521 K. Prior to the production runs, each replica was simulated for 400 ps without exchange in order

to generate a random sample of initial conformations, eliminating the potential for bias of the system toward a particular starting conformation. All simulations utilizing ff14IDPFF began from the final configuration of the corresponding ff14SBonlysc simulation.

Production simulations were considered converged when radius of gyration (R_g) versus end-to-end (R_E) plots showed minimal change in adjacent temporal windows. The production run totaled at least 500 ns but some force fields required over 1 μs to be convincingly converged. In the case where simulation time < 1 μs, the final 150 ns of data collected at 300 K has been analyzed, whereas when simulation time > 1 μs, the final 300 ns collected at 300 K has been analyzed. Temperatures were kept constant using the Langevin thermostat with a collision frequency γ of 1 ps^{-1}. The equations of motion were integrated using the LeapFrog algorithm with a time step of 1 ps. Cutoff for both effective Born radii and van der Waals interactions was set to 2 nm. Hydrogen bond length constraints were introduced using the SHAKE[40] algorithm.

3.3 Results and Conclusions

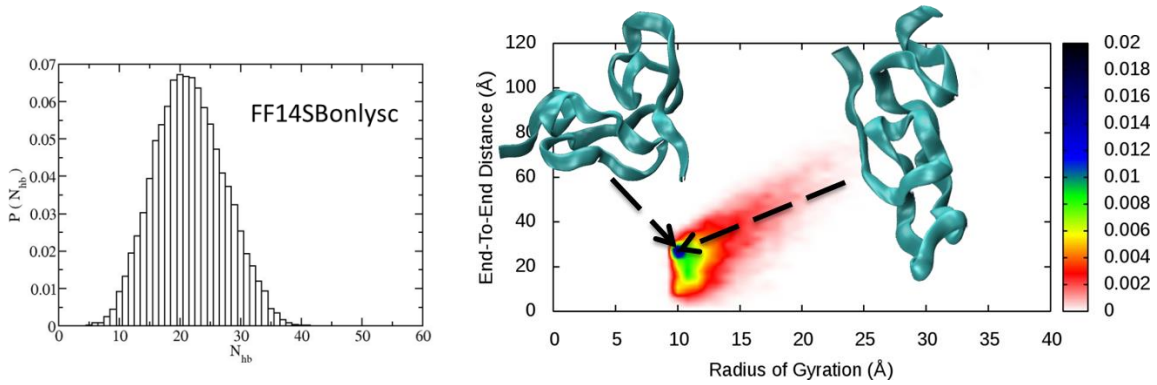

Figure 2. Radius of Gyration/End-to-end distance and Hydrogen Bond Distribution of PHF43/FF14SBonlysc. Hydrogen bond distribution (left). Normalized probability distribution of finding a chain with a specific end-to-end and radius of gyration value (right).

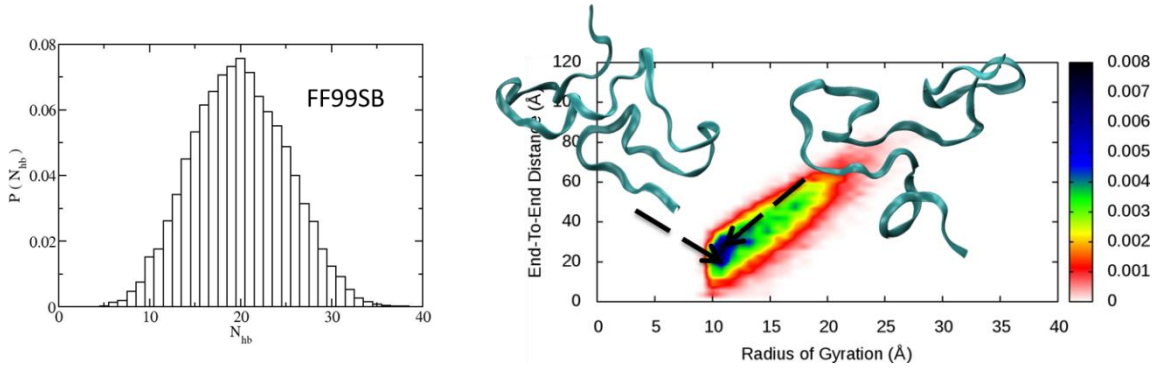

Figure 3. Radius of Gyration/End-to-end Distance and Hydrogen Bond Distribution of PHF43/FF99SB. Hydrogen bond distribution (left). Normalized probability distribution of finding a chain with a specific end-to-end and radius of gyration value (right)

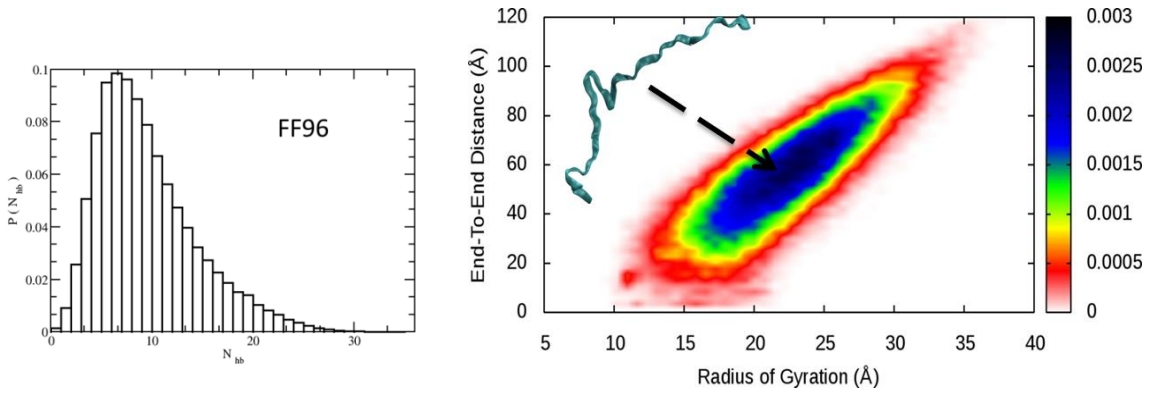

Figure 4. **Radius of Gyration/End-to-end Distance and Hydrogen Bond Distribution of PHF43/FF96**. Hydrogen bond distribution (left). Normalized probability distribution of finding a chain with a specific end-to-end and radius of gyration value (right)

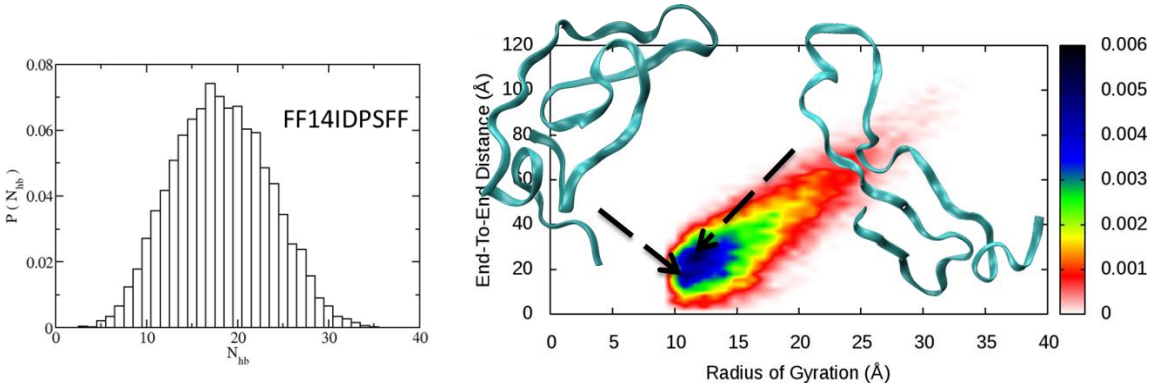

Figure 5. Radius of Gyration/End-to-end Distance and Hydrogen Bond Distribution of PHF43/FF14IDPSFF. Hydrogen bond distribution (left). Normalized probability distribution of finding a chain with a specific end-to-end and radius of gyration value (right)

Results from all force fields vary significantly in the case of PHF43. The most marked difference, which we will focus on, is between the performance of ff14SBonlysc and ff96 (**Figure 2** and **Figure 4**, respectively). Ff14SBonlysc presents top clusters which are notably collapsed and relatively rich in hydrogen bonds indicating some long lived globular structures whereas ff96 yields extended and compact structures with

comparatively few hydrogen bonds. A small, but long-lived, amount of clusters produced from ff96 simulations, though, do present β-structure as indicated by cluster analysis. Besides this small subset of conformations, very little secondary structure is found and the extent to which PHF43 explores its conformational ensemble has led us to decide that ff96 is the ideal candidate for simulation of this peptide. Experimental findings[41] indicate that extended, disordered conformations with small amounts of transient β-structure dominate the ensemble of PHF43. With these results from our simulations (of course, more can be said, but for the purposes of this thesis, we have seen enough), we can be confident that the results in the following chapter represent an accurate depiction of PHF43's conformational ensemble.

4 Impact of Phosphorylation and Pseudophosphorylation on the Early Stages of Aggregation of the Microtubule-Associated Protein Tau

This chapter utilizes the validation of AMBER ff96 as an effective force field for MD studies of the tau-derived peptide PHF43 found in the previous chapter. Here, we apply MD methodologies that have been discussed so far to investigate how the conformational ensemble of PHF43 is affected by phosphorylation and its experimental counterpart, pseudophosphorylation. Insights into the aggregation pathway of PHF43 in Alzheimer's disease are uncovered, as well.

The work presented in this chapter has been published.[42] Figures and some passages have been partially or fully adapted from the original publication.

Figure 6. **Sequence of the Microtubule Binding Region (MTBR) of Tau**. The four repeats M1-M4 are highlighted. The residues numbers are reported and based upon the longest isoform of tau. The fragment used in this study is shown at the bottom. The phosphorylated (and pseudophosphorylated) residues are underlined. (Reproduced from Ref.[42] with permission from the American Chemical Society, copyright © 2017, American Chemical Society)

4.1 Introduction

The microtubule associated protein (MAP) tau is found in the central nervous system, specifically in the axon of the neuron[43-45]. Tau is tasked with regulating the modifications of the cytoskeleton, namely the formation of and razing of microtubules, so neurons can form or prune connections to other neurons. Besides structural support, microtubules are also essential for the transport of molecules and organelles from their place of origin to their final destination in the axon. As a consequence, the tau protein is extremely important to the healthy function and development of the human brain. Thus, disruption of tau's normal function (whether through mutations or sporadic dysfunction) is extremely detrimental to the central nervous system.[43-49]

Tau is classified as an intrinsically disordered protein (IDP). IDPs are proteins that perform their function despite a lack of well-defined tertiary and secondary structure. IDPs are generally rich in charged residues that make these proteins extremely soluble and dynamic. They also typically contain a small amount of hydrophobic moieties, so hydrophobic collapse is normally prevented.[50-52] Therefore, one would expect that IDPs

such as tau would not be prone to aggregation under normal circumstances. Surprisingly, though, aggregated tau is involved in the pathogenesis of many diseases that lead to dementia and abnormal brain development. Most notably, aggregated tau is a hallmark of Alzheimer's disease, where it forms structures deemed neurofibrillary tangles. Tau aggregates present a "fuzzy coat" of disordered sequences that surrounds a region rich in β-sheet. The β-sheet rich region, which composes the microtubule binding region (MTBR) of the protein, shows a structure typical of amyloid fibers.[53-57] The MTBR consists of four pseudo-repeats (referred to as M1-M4 in **Figure 6**), which share similar amino acid sequences.

To perform its function, tau undergoes extensive post-translational modifications in the form of phosphorylation. Phosphorylation consists of adding a covalently bound phosphate group to the side chain of an amino acid.[58-61] The most common phosphorylation sites in the tau protein are the amino acids Serine, Threonine and Tyrosine. This corresponds to as many as 85 sites—about 20% of the total number of amino-acids—that are capable of being phosphorylated. However, only a few sites are phosphorylated at any given time. As the phosphorylation sites are turned over very quickly, it is extremely challenging to measure experimentally which sites are phosphorylated and when. It is estimated that about 18 sites are phosphorylated at a time in functional tau in a healthy cell, while this number of sites increases in cells affected by Alzheimer's disease (an effect that is called hyperphosphorylation).[45, 62-63]

In healthy cells, it is believed that multiple kinases,[64-66] are involved in the phosphorylation process. This suggests that the final phosphorylated state comes about as a consequence of multiple stages each encompassing further sub-events. Thus, at any

given time, multiple phosphorylated states of the same protein can exist at once. This is a formidable problem to deal with experimentally. In fact, to obtain a well-defined picture of the phosphorylated state of tau *in vitro*, it is required to know not only the kinases involved, but also in what order they perform their function. As mentioned above, in the case of Alzheimer's disease, tau is found in its hyperphosphorylated state in which the number of phosphorylated sites along the protein is found to be 3 to 4 times higher than normal.[67-69]

To overcome the issues associated with studying phosphorylated tau *in vitro* and in cell culture, experimentalists have turned to the method of pseudophosphorylation.[70-84] Pseudophosphorylation consists of mutating amino acids that are targets of phosphorylation into the amino acids Aspartate (Asp) or Glutamate (Glu). The rationale behind this technique is that tau interactions are thought to be mostly electrostatic in nature. As a consequence, replacing a negatively charged side chain (such as a phosphorylated one) with another negatively charged side chain (such as Asp or Glu) should have a relatively small impact on the properties of the protein.

In this study, we will use MD to characterize whether phosphorylation and pseudophosphorylation are actually equivalent at a molecular level. The strongest experimental support of equivalence between the two techniques comes from measurements of electrophoretic mobility, which includes information about both the size and conformation of monomers.[74, 80] These studies have shown that mobility is generally unaffected when experiments are performed using pseudophosphorylation or phosphorylation. For this reason, pseudophosphorylation is believed to be a reasonable approximation to phosphorylation that avoids the issue of being presented with a mixture

of tau phosphorylated at different sites. Of course, it must be noted that phosphorylation and pseudophosphorylation are not exactly equivalent. For instance, a phosphate group is large and bears two negatives charges, while Asp and Glu possess only one. In addition, the fact that experimental structures from both methods show a similar mobility does not necessarily conclude that in both cases the protein is adopting similar conformations. Since tau is an intrinsically disordered protein, different conformations may be adopted that have a similar size, but highly different internal structure. In this chapter we will show that this is the case.

Another reason for investigating their equivalence is to understand whether phosphorylated and pseudophosphorylated tau interact similarly during aggregation. Several studies have shown consistently that phosphorylated tau and pseudophosphorylated tau may lead to similar aggregation propensity or microtubule affinity.[77-80] However, again, phosphorylated tau found in pathological aggregates is a mixture of proteins that are phosphorylated at different locations and with a different number of phosphorylated sites. As a consequence, it is often impossible to assess if the pseudophosphorylated product is actually equivalent to a phosphorylated one.

4.2 Methods

4.2.1 Fragment Selection

In this chapter, we will focus on the fragment PHF43 of the tau protein. This fragment was shown to be able to aggregate and form amyloid fiber.[85] We will refer to the wild-type form of PHF43 as PHF43/WT. Computational and experimental studies have shown

that phosphorylation of Ser324 alone or phosphorylation of both Ser320 and Ser324 may lead to enhanced aggregation propensity.[86] However, these studies have been performed on fragments of tau much shorter than the ones considered in our study. In this chapter, we will consider both the phosphorylated (PH43/P) and pseudophosphorylated fragment, modifying both Ser320 and Ser324 at once. For the pseudophosphorylated fragment, we will compare the mutation of Ser320 and Ser324 into glutamic acid (PH43/PGlu) and aspartic acid (PHF43/PAsp) in respective simulations.

4.2.2 Simulation Details

Simulations were performed in the canonical NVT ensemble using the package AMBER14.[87-88] We used the AMBER ff96 force field in association with the GB-OBC[31]implict solvent model. This force field/implicit solvent combination was shown in the previous chapter to accurately reproduce the structure (or lack thereof) of PHF43 in solution. This model was also shown by Dill and coworkers[25] to correctly reproduce the secondary structure of other peptides. The choice of implicit solvent, as discussed in **Chapter 2**, **section 2.3.1**, is due to the large size of our system that would otherwise prevent a proper sampling of PHF43's ensemble of conformations. Phosphorylated amino acids used the force field developed by Homeyer et al.[89] The cutoff for both the van der Waals interactions and the implicit solvent model is 2 nm. A Langevin thermostat kept the temperature constant. The equations of motion were integrated using the leap-frog algorithm with a time step of 1 ps. Constraints for hydrogen atoms connected to heavy atoms were computed with the SHAKE algorithm.

To investigate the prion-like behavior of the peptide, we simulated a wild-type PHF43 dimer. The backbone of one of the chains was frozen in its original position, whereas the second chain was free to sample different conformations. To avoid the case wherein the chains separate and stop interacting, we added a restraint potential. This potential was zero as long as the distance between the centers of mass of the chains was below 4nm. Beyond this distance, a parabolic restraining potential was added. If the chains separated farther than 4.5 nm, a linear potential was implemented instead.

To accurately sample the conformational space of the peptide, replica-exchange molecular dynamics was employed. This method is described in **Chapter 2**, **section 2.3.2**. The temperatures ranged from 279K to 503K using 16 replicas. Only data collected at 300K has been analyzed. The simulations were at least 1.2μs long but in any case, only the final 300ns were analyzed. For a review of methods of analysis, please see **chapter 3, section 3.1**.

4.3 Results

4.3.1 Wild-type

We begin by analyzing the structure of the monomer of wild-type PHF43 (PHF43/WT) in solution. **Figure 7** (top left) shows that the monomer can easily populate multiple conformations ranging from compact (end-to-end distance about 10 Å) to extended structures (about 100 Å). As a comparison, a fully extended chain would be about 120 Å long. This is consistent with what we would expect of tau due to its classification as an intrinsically disordered protein. In order to understand the structure of PHF43/WT in detail, we computed the number of backbone hydrogen bonds. This analysis showed that

very little secondary structure is present, once again consistent with tau's intrinsically disordered nature.[43] However, this result is puzzling because it is in direct contrast with the β-rich conformations found in aggregated PHF43. As a result of this discrepancy, we decided to investigate how likely the peptide is to spontaneously adopt a β-rich structure.

We found that a small population of the conformational ensemble is indeed compact and folded (**Figure 7**, top left, highlighted region). A closer examination of this small population reveals that wild-type PHF43 forms a four-stranded β-sheet structure with a very high number of hydrogen bonds (not shown). This structure is additionally stabilized by two salt bridges, one between E43 and K22 and the other between E43 and K3.

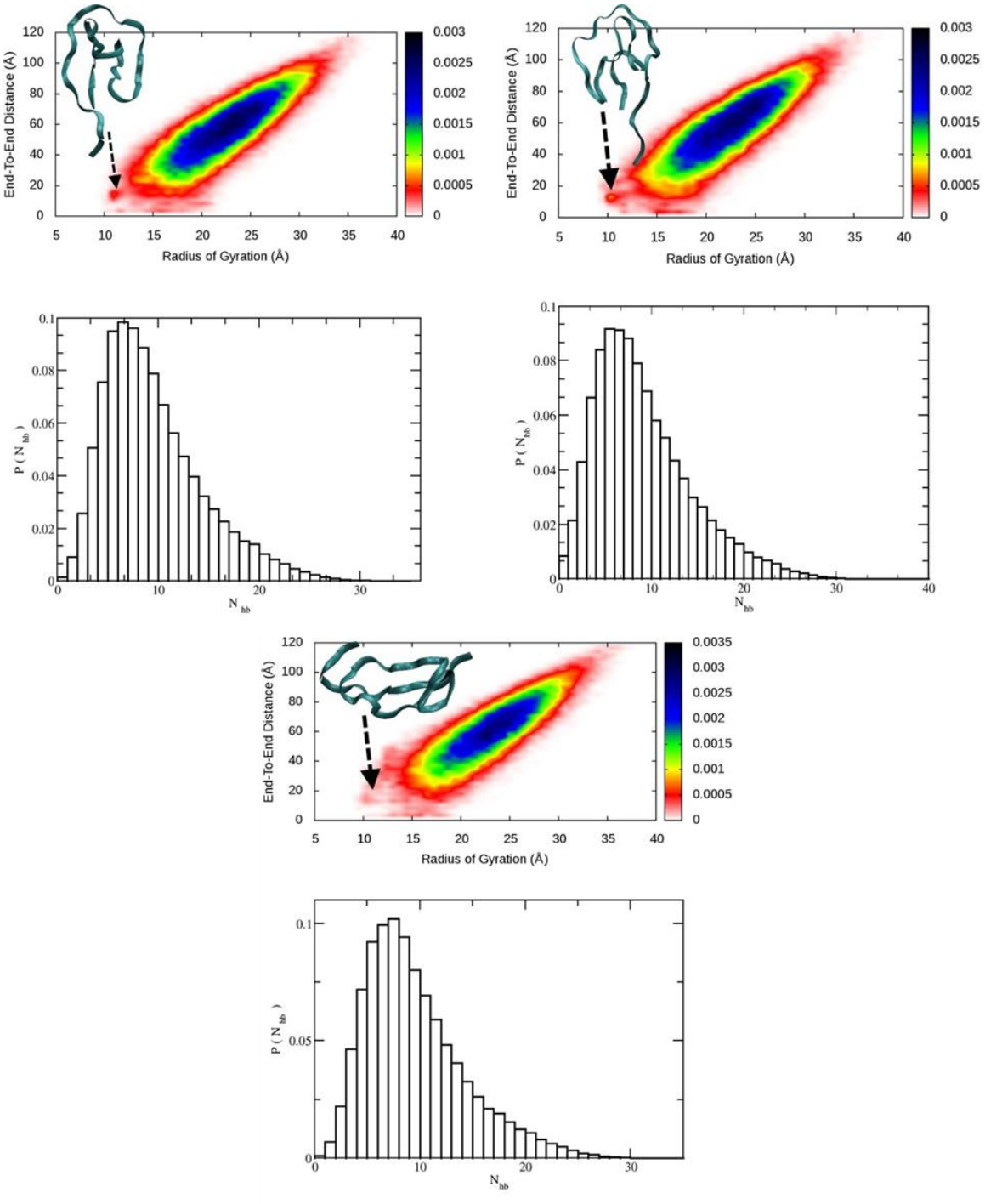

Figure 7. Radius of Gyration/End-to-end Distance and Hydrogen Bond Distributions of PHF43/WT, PGlu, and PAsp. Normalized probability distribution of finding a chain with a specific end-to-end and radius of gyration value (top) and hydrogen bond distribution (bottom). Clockwise from top left: PHF43/WT, PHF43/PGlu mutation, and PHF43/PAsp mutation. (Partially reproduced from Ref.[42] with permission from the American Chemical Society, copyright © 2017, American Chemical Society)

4.3.2 Pseudophosphorylation: Glutamate or Aspartate

Now, we turn to the case of pseudophosphorylated PHF43. Pseudophosphorylation is used extensively in the study of tau dynamics, so it is immensely important for us to understand how this mutation affects PHF43's structure in solution. A comparison of both psuedophosphorylation of PHF43 with glutamate (PHF43/PGlu) and psuedophosphorylation of PHF43 with aspartate (PHF43/PAsp) with PHF43/WT indicates that the overall ensemble of conformations is not affected by the double mutation. Again, a cluster of conformations from the PHF43/PGlu ensemble was identified with a radius of gyration of about 10Å (**Figure 7**, top right, highlighted region). These conformations were analyzed independently and were found to be comprised of 46.6% β-rich structures that look similar to those of the wild type in the same region. A similar, less defined population in the same area could be found in the PHF43/PAsp ensemble, but the amount of long-lived secondary structure present was found to be less than that in PHF43/WT and PHF43/PGlu.

The hydrogen bond plots of both of the pseudophosphorylated peptides (**Figure 7**, top right (bottom panel) and bottom (bottom panel)) indicate that the most likely number of bonds is below 10. This is extremely similar to that observed in the analysis of PHF43/WT tau, once again supporting the ineffectiveness of the mutation in promoting secondary structure. Based on the above analysis, we believe that we can conclude pseudophosphorylation does not significantly enhance the presence of β-rich structures.

4.3.4 Actual Phosphorylation

The final set of simulations involving the PHF43 monomer includes the addition of a bona fide phosphate group to the serine side chains (PHF43/P). This modification causes a dramatically different conformational ensemble with respect to the wild type and pseudophosphorylated PHF43 monomers (**Figure 8**, top). The peptide collapses into a very compact structure and does not explore much of its conformational space outside this region. A phosphate group is larger in size and possesses a higher negative charge than the side chains of both aspartate and glutamate. Thus, each phosphate group is able to strongly interact via ionic bonds with multiple positively charged amino acids (in this case, lysines). As a consequence of these interactions, the backbone can adopt a very compact, stable conformation. These conformations are also stabilized by a large number of hydrogen bonds (**Figure 8**, bottom).

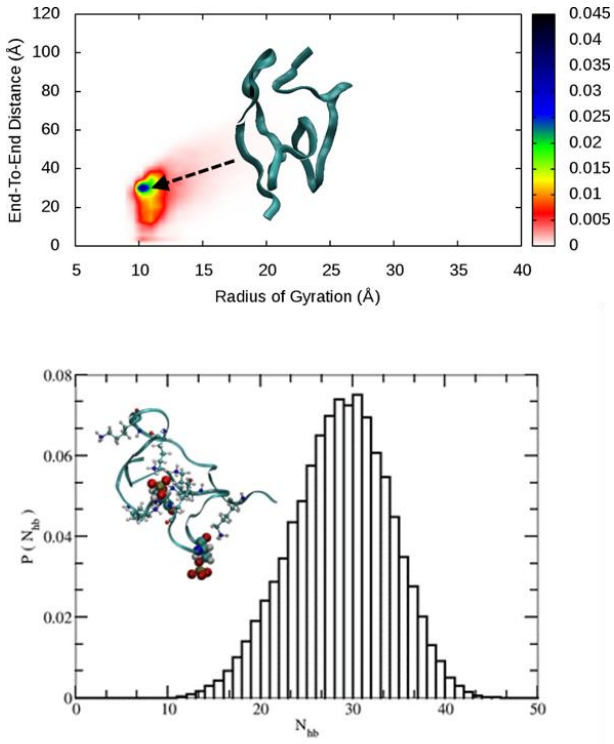

Figure 8. **Radius of Gyration/End-to-end Distance and Hydrogen Bond Distribution of PHF43/P**. Normalized probability distribution of finding a chain with a specific end-to-end and radius of gyration value (top). Hydrogen bond distribution (bottom). (Reproduced from Ref.[42] with permission from the American Chemical Society, copyright © 2017, American Chemical Society)

4.3.5 Dimerization

To study dimerization we selected a β-sheet conformation from the PHF43/WT ensemble. The structure selected is one of the most populated conformations that is rich in β-structure. It should be noted, though, that this conformation only represents about 1% of the total conformations and thus, it should be considered a rare intermediate state (although our simulations indicate that this state is long-lived). To reduce simulation time, the backbone of this compact conformation was frozen. Then, a second, fully

flexible PHF43/WT monomer was added. To prevent unwanted separation, the second was forced to lie within a sphere of radius 4 nm.

As shown in **Figure 9**, the flexible chain folds into a compact conformation as well. These conformations are characterized by a high number of hydrogen bonds. This indicates that the frozen chain actually acts as a template that influences the flexible chain in a way that makes it adopt a similar compact structure (**Figure 9**). However, the average distance between the two monomers suggests that these dimers are unstable and tend to dissociate (**Figure 10**, bottom). This is analogous to the behavior of a prion, suggesting that, while rare, these compact conformations can form and rapidly propagate similar to a prion infection.

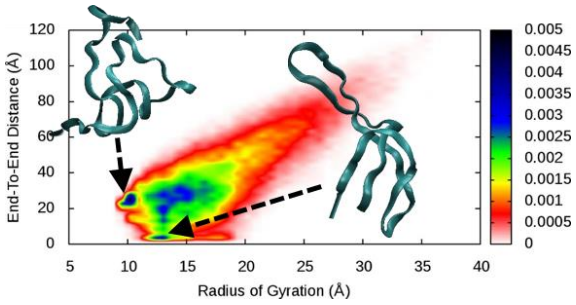

Figure 9. **Normalized Distribution of the End-to-end and Radius of Gyration for the Flexible Chain of a PHF43/WT Dimer**. (Partially reproduced from Ref.[42] with permission from the American Chemical Society, copyright © 2017, American Chemical Society)

4.4 Discussion

Experimentally, it is extremely difficult to characterize the structure of a single peptide chain, in particular when a well-defined native structure is not present. Our use of MD to analyze the conformational ensemble of PHF43 monomers suggests that

pseudophosphorylation and actual phosphorylation may lead to drastically different conformations in solution. However, experiments suggest that regardless, the aggregation products of these two states do not differ too much.[70-84] Consequently, we must attempt to explain how such a dramatic difference in monomer conformations can result in a similar end result.

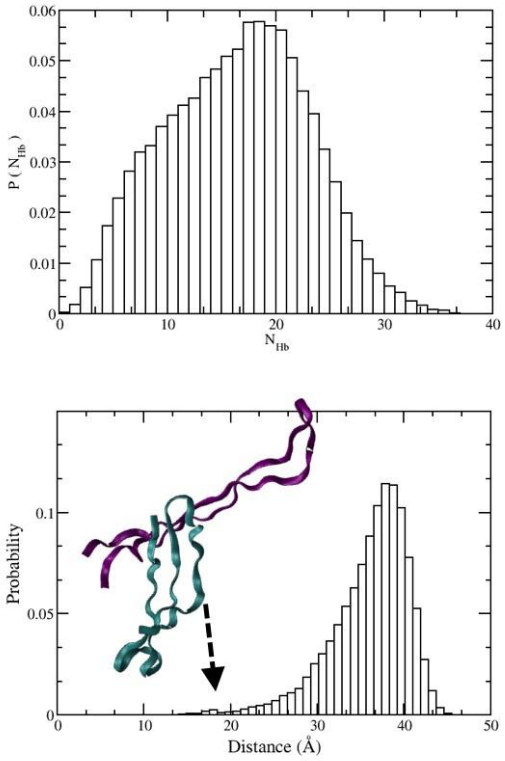

Figure 10. **Number of Hydrogen Bonds Presented in Flexible Chain and Center of Mass Distance for PHF43/WT Dimer**. (top) The number of hydrogen bonds for the incoming chain. (bottom) Distance between the centers of mass of the dimer. Most of the conformations are far apart. In purple, the frozen chain. (Partially reproduced from Ref.[42] with permission from the American Chemical Society, copyright © 2017, American Chemical Society**)**

Please see the published paper at ref.[42] for a full discussion. The following will primarily discuss the proposed aggregation pathway model.

From the results of our simulations, we can assume that once a PHF43 monomer adopts a compact, β-sheet conformation, it may coax a second monomer nearby to fold into a similar conformation. This is the idea of the "dock-lock" model,[90-96] where two chains first attach to each other ("dock" stage) and then a structural transition of the chains leads to a more stable aggregate ("lock" stage).

It appears that aggregation of wild-type PHF43 and pseudophosphorylated PHF43 can partially be described by this model. The monomers would exhibit prion-like behavior where very rare transitions into compact conformations trigger similar transitions in nearby peptides, which in turn propagate the structural transition to other chains. Interestingly, this prion-like behavior of tau has previously been reported in the literature.[97-99] As dimers, in our case, are unstable, probably due to electrostatic repulsion, we would expect that the solution would be populated by long living compact monomers that originate from the dissociation of these unstable dimers.

On the other hand, our simulations suggest that bona fide phosphorylation of PHF43 bypasses this "dock-lock" process entirely by immediately causing monomers in solution to adopt long-lived compact, β-rich structures. Speculatively, these monomers could also influence those around them, initiating a process similar to that described above.

Concluding, both pseudophosphorylation and phosphorylation have the ability to drive the formation of a large population of compact monomers as found experimentally, but the pathway followed to this point is quite different for both processes.

As a final remark, the model proposed above does not predict a direct pathway from pseudophosphorylation or phosphorylation that leads to the formation of amyloid fibers. Our model instead hypothesizes that there exists an intermediate state composed of compact PHF43 monomers. In particular, in the case of the wild-type or pseudophosphorylated chain, we would expect that these compact conformations would form around the original folded state. As the process propagates, a spherical cloud of compact monomers would form around this initial compact structure, as there is no preferred direction for the monomers to grow in. Once this region of crowded solution develops past a certain threshold, it would collapse into a stable aggregate. This is consistent with existing experimental data.[100] See **Figure 11** for a simple representation of the entire process.

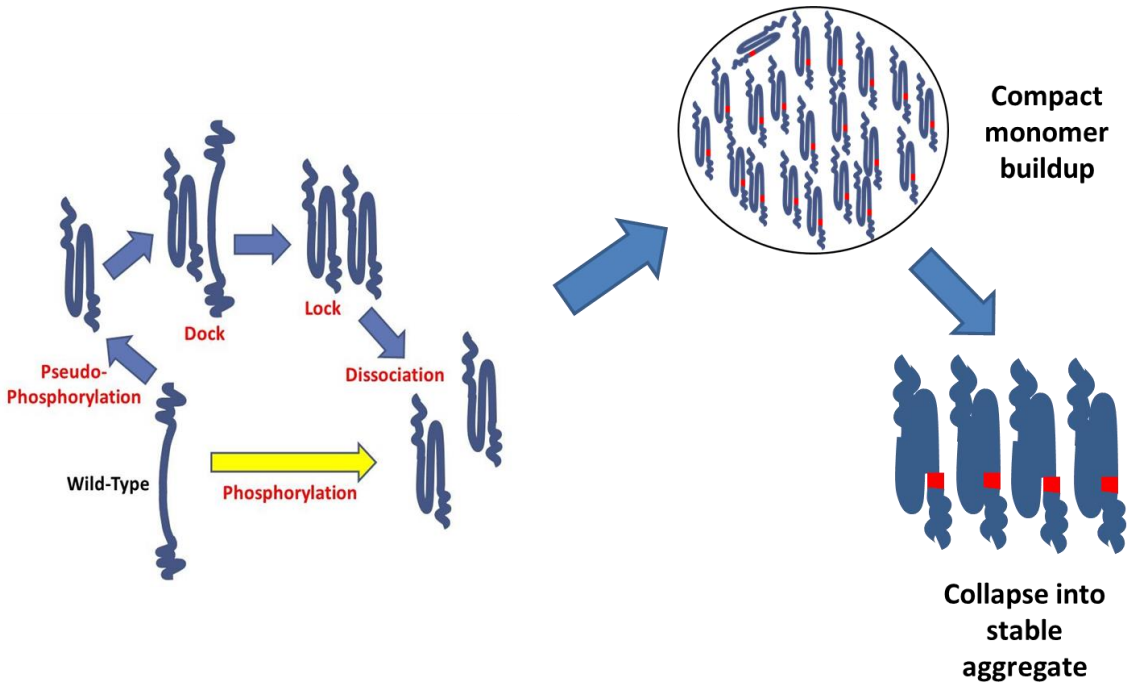

Figure 11. **Schematic of Aggregation Process via Pseudophosphorylation or Phosphorylation**. At left, the formation of compact monomers is shown via the dock-lock process and the more direct pathway of phosphorylation. Next, the compact monomers congregate around a central complex. Finally, once a critical threshold is reached, the buildup of monomers collapses into a stable aggregate. (Partially reproduced from Ref.[42] with permission from the American Chemical Society, copyright © 2017, American Chemical Society**)**

4.5 Conclusion

In this chapter, we have used MD to address the issue of phosphorylation and how it correlates with the aggregation propensity of the microtubule associated protein tau. Based on our results from simulation, we have shown that tau aggregation proceeds through multiple stages characterized by molecular rearrangements. In the wild-type and pseudophosphorylated states, these rearrangements are unlikely to take place, but may occur and drive formation of compact monomers through a prion-like propagation. Bona

fide phosphorylation on the other hand, bypasses intermediate rearrangement stages and produces compact monomers immediately. We have also shown that, while pseudophosphorylation and phosphorylation may not be exactly equivalent when considering a single chain in solution, when more chains are considered, the end product is essentially the same.

5 A Hierarchical Model for the Processing of Polysaccharides/Protein-Based Films in Ionic Liquids

In this chapter we examine the regeneration of biopolymer mixtures after dissolution in ionic liquid. Using a combination of MD simulation, theoretical modeling, and data from experiments, we propose a novel model that correlates a biomaterial's microscopic structure to its bulk properties. This study is to be submitted for publication and some figures and passages have been adapted from the original manuscript.

5.1 Introduction

Biomaterials are an expanding class of materials that combine the unique properties of biological systems with modern materials science. There are many advantages associated with an increased use of biomaterials, one of which is the fact that they are usually composed of biopolymers readily available in nature, offering a renewable source of material. For instance, cellulose[101] can be utilized as a byproduct of the wood industry and chitin[102] is one of the most abundant proteins in nature that composes the exoskeleton of many shellfish. In addition, biopolymers are extremely versatile as they can be mixed with other biopolymers to tune their properties without the use of harmful chemicals.

Examples of application varies from skin graft treatments for burn victims,[103] to the detection and eventual removal of pollutants from the atmosphere and water supplies.[104]

The properties of these biomaterials can easily be tuned by mixing together pure components or with other biopolymers.[105-106] For instance, a protein that shows good mixing properties with polysaccharides is silk.[106-109] In addition to these properties, silk has other advantages such as strong mechanical properties, oxygen and water vapor permeability, biocompatibility, and biodegradability. While in its native state, silk is quite brittle, its properties can be when modulated when mixed with other biopolymers.[110-112]

While these materials can, in principle, revolutionize modern materials science, there are still many open questions regarding their processability and the predicted properties of polymer mixtures. In this chapter, we will tackle this problem by proposing a coarse-grained model of the molecular interactions involved.

We use the MD methodology of umbrella sampling[113-115] to characterize the dissolution process of crystalline cellulose in ionic liquids: solvents composed of an organic cation and inorganic cation that have been shown to excellently dissolve cellulose[116]. In this study we use 1-ethyl-3-methylimidazolium ($EMIM^+$) as cation and Cl^- as anion. This pairing has been shown to dissolve cellulose quite well.[117] Our goal here is to find the most likely pathway by which cellulose dissolves.

Umbrella sampling is an especially useful technique for characterizing this process. This method involves generating a series of configurations along a reaction coordinate between two interacting macromolecules in the system. One of these macromolecules is used as a reference while the other is placed at increasing distance

away from the reference, held fixed at this distance by a harmonic potential. These configurations are deemed "sampling windows" and each window is its own independent simulation. This process produces an ensemble of independent simulations (thus, an ensemble of configurations) along the reaction coordinate (see **Figure 13** for a schematic of this process). From each of these sampling windows, values of the potential of mean force[118] (PMF) can be calculated. Then, using the weighted histogram analysis method[119] (WHAM), a PMF curve over the entire reaction coordinate can be generated.[120] We test two cases: one where layers of cellulose's crystalline structure become detached at a time, and another where a single chain detaches from the crystalline structure and leaves behind a smaller crystal. In both cases, the PMF curve generated will show the free energy needed to bring the macromolecule(s) from an associated state to a fully dissociated, ionic liquid-solvated state.

The results of these simulations are then used to bolster a theoretical model that describes properties of a general regenerated solution of a biopolymer mixture after solvent has been removed. The predicted properties are then compared with experiments performed by the Salas group at Rutgers-Camden who examined numerous properties of dissolved and regenerated polysaccharide and silk blends (samples summarized in **Table 1**). A schematic showing the crystalline form of cellulose and the process outlined above is given in **Figure 12**.

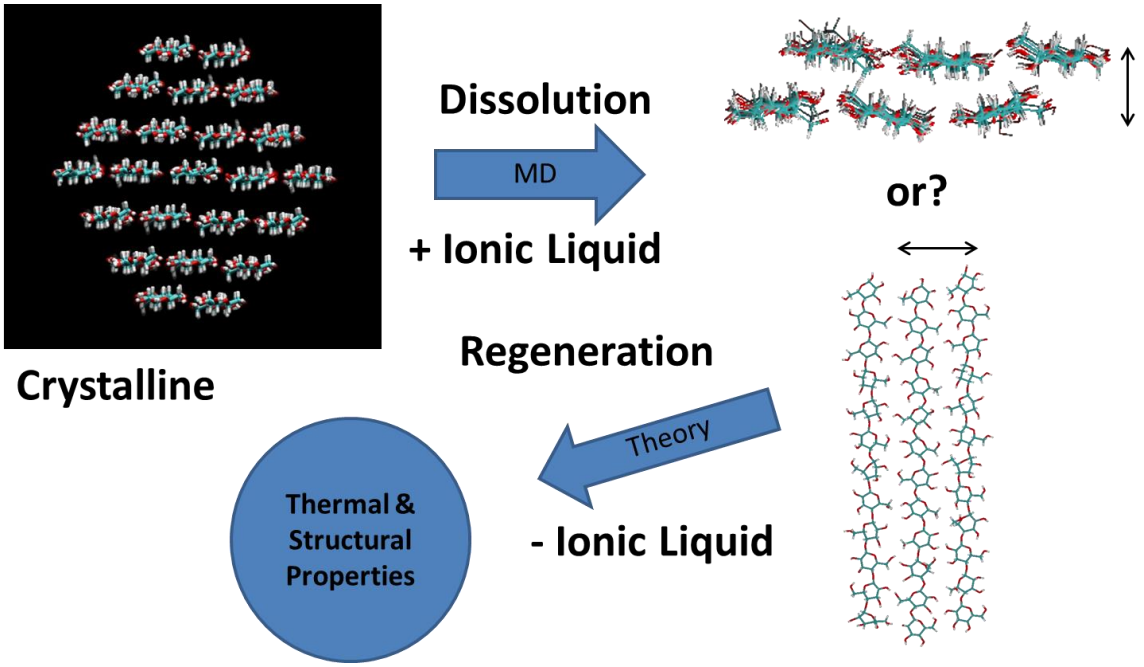

Figure 12. A Schematic Summary of the Process Used to Understand the Dissolution and Degeneration of Biopolymers. MD simulations are performed to characterize the dissolution pathway of cellulose in ionic liquid. We then use these results to develop a theory that describes the subsequent regeneration of the dissolved biopolymers once the ionic liquid is removed as well as some physical properties of the regenerated material.

Table 1. Summary of Samples Prepared and Analyzed Experimentally by the Salas Group for Comparison to Theoretical Predictions.

Polysaccharide	Silk	Abbreviations
Cellulose		
80%	20%	CES20
20%	80%	CES80
Chitin		
80%	20%	CHS20
20%	80%	CHS80
Chitosan		
80%	20%	CTS20
20%	80%	CTS80

5.2 Methods

5.2.1 Simulation details

Cellulose molecules were built using the GAG Builder by GLYCAM[121] and assembled into layers using VMD[122] while imidazolium-based chloride ionic liquid ([EMIM]Cl) was built and parameterized using Antechamber[123] and GAFF[124]. All simulations were conducted using GROMACS 5.1.4[35-38, 125-127] and parameters from the GLYCAM06[128] force field and GAFF were utilized throughout all simulations. Short-range non-bonded interactions were cut off at 1.2 nm, while long-range Coulombic interactions were calculated using the Particle Mesh Ewald method.[129-130] Long-range dispersion corrections were applied to energy and pressure. Periodic boundary conditions were applied in all directions for all simulations.

Starting structures of each system were first solvated with the cation and neutralized with chloride counterions, then subjected to steepest descents energy minimization. NPT and NVT ensemble equilibrations, in that order, were then carried out on each system with position restraints applied to each heavy atom making up the D-glucose units. NPT equilibration time to constant density varied depending on the system. Temperature was maintained using a Berendsen weakly coupled thermostat[131] set to a reference temperature of 363.15 K with a time constant of 0.1 ps^{-1}. Pressure was also maintained isotropically using a weakly coupled barostat at 1 bar with a time constant of 1 ps^{-1} and center-of-mass position restraint reference coordinate scaling. NVT equilibration simulations were then performed with velocities generated according to a Maxwell distribution at T=363.15 K and temperatures maintained using the same method as the NPT portion for 1 ns.

Following this final equilibration, production simulations began from these conformations. These simulations follow the same protocol as before, except that the Berendsen thermostat is replaced with the Nosé-Hoover thermostat.[132-133] Umbrella sampling[113-115] was performed on two separate cellulose systems: a two-layer configuration composed of 6 individual strands of cellulose (3 strands per layer) (**Figure 14**, inset) and a single layer of 3 cellulose strands (**Figure 15**, inset).

The pulling protocol was different for each system. In the case of the double layer, the goal was to separate the two layers from each other as if one was pulling two sheets of paper stacked on top of one another apart while also making an effort to keep the pieces of paper as flat as possible. To do this, we restrained the lower layer as

described for equilibration but removed restraints from the top layer. From here, the free top layer was held fixed (0.0 nm ps^{-1} pull rate, spring constant of 10^5 kJ mol^{-1} nm^{-2}), subject to an umbrella potential, at varying distances *d* (**Figure 14**, inset) away from the bottom layer using a combination of three-cylinder geometry groups with radii equaling about $1/6^{th}$ of the length of the long axis of the layer. In order to further ensure the unrestrained layer remained flat with respect to the fixed reference layer, a series of fixed-distance pull groups (spring constant of 10^6 kJ mol^{-1} nm^{-2}) across the width of the layer were implemented. This was only to keep the layer as "taut" as possible; any information from these pulls was not used for analysis. Simulations were initially performed in increments of 0.25 Å in the y direction from the near-equilibrium center of mass (COM) separation between bottom and top layer of approx. 0.5 nm to 1.35 nm. Inspection of the resulting data revealed that in certain adjacent sampling windows, the respective COM separations along the y-coordinate did not overlap adequately and thus, halfway between these windows, an extra simulation window was added to fill this gap (e.g., if position data from 0.50 nm and 0.525 nm COM separation did not overlap, a third window fixed at 0.5125 nm was performed). Eventually, sampling windows were also added at an increment of 0.20 Å from 0.44 nm to 0.48 nm COM distance in order to emphasize the energy minimum of ~0.50 nm. This spacing culminated in 46 windows—each 200 ns—resulting in ~9.2 μs of total simulation time.

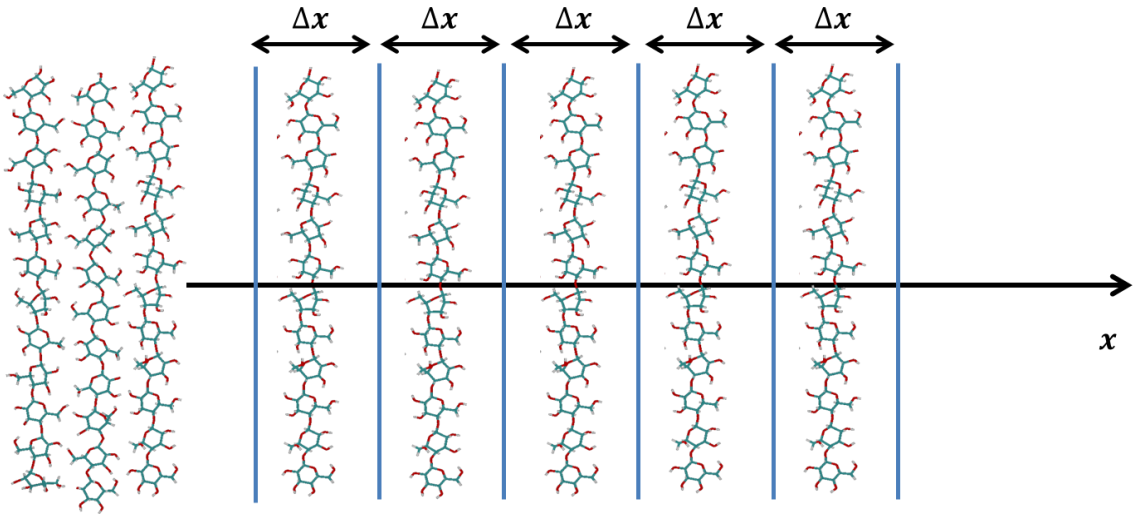

Figure 13. Schematic Representation of Umbrella Sampling in the Case of Removing a Single Chain of Cellulose from a Layer in the Presence of Ionic Liquid. Adjacent windows are separate simulations that do not interact. The reaction coordinate is given as the x direction towards the right.

In the case of the single layer, the goal was to "pluck" a chain from the edge of the layer and thus, a more routine protocol was implemented. Of the three chains that compose the layer, two were frozen (one edge chain and the middle, see **Figure 15**, inset) using the same position restraints mentioned previously. Five reference COM groups were assigned to oxygen atoms in closest proximity to five COM groups on the free, pulled chain which were located on glycosidic oxygens along the chain's axis from end to end. Along the z coordinate connecting these paired groups, the edge chain was held fixed (0.0 nm ps^{-1} pull rate, 10^6 kJ mol^{-1} nm^{-2} spring constant) at COM separation distances ranging from 0.40 nm to 2.00 nm with a sampling window spacing of 0.125 Å. Eventually, spacing of 0.2 Å was employed from 1.5-2.0 Å as at this point, these windows were

sufficient for sampling. This spacing culminated in 114 windows, again each totaling 200 ns, resulting in ~22.8 μs of total simulation time.

Analysis was performed using the weighted histogram analysis method[119] (WHAM) as implemented in GROMACS 5.1.4.

5.3 Results and Discussion

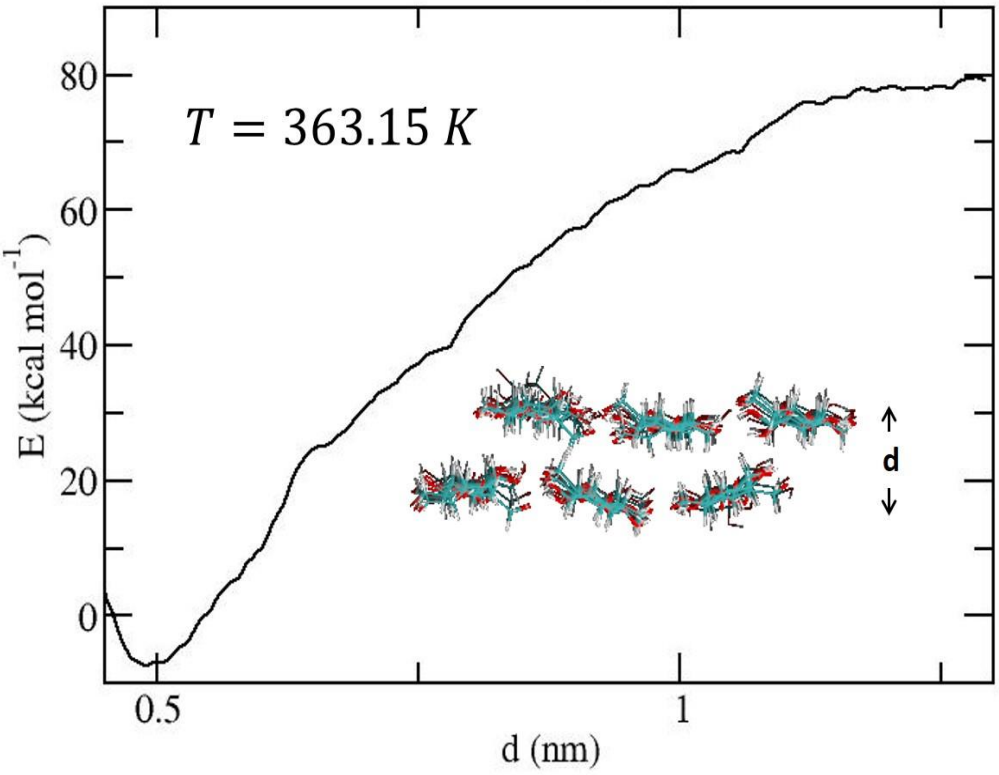

Figure 14. **Free Energy to Separate Two Layers of Cellulose in Ionic Liquid**.

5.3.1 Dissolution and Regeneration

As discussed in the introduction, ionic liquids are known to dissolve biopolymers, mostly by breaking the hydrogen bonds between chains. However, under normal experimental conditions, it is unlikely that the dissolution reaches completion, as the inter-layer dispersion forces would resist the dissolution. In addition, polymers are expected to adopt a random coil conformation in good solvent with a consequent increase in volume. In

order to investigate this process, we built a cellulose bilayer and we computed the free energy difference between fully associated and dissociated layers. As **Figure 14** shows, the energy that is required to separate entire layers is prohibitively large. On the other hand, separating one single chain from a layer by disrupting intermolecular hydrogen bonds requires a much more reasonable energy (**Figure 15**). In both cases, the mobility of the layer or chain increases once detached. As a consequence, the chain or layer can adopt conformations that are forbidden within a rigid crystalline structure. Ideally, this increased freedom can be exploited to select conformations that better match the surface of other materials, such as silk, acting like a sort of glue between polysaccharide and protein surfaces.

Based on the result just discussed, we propose the following model. Starting from a macrofibril, the biomolecules at the edge linked by hydrogen bonds are the first to detach. In this model single chains are slowly released into the solution, until the macrofibril breaks down into layers. In principle, given enough time, all the crystals will be broken down and only single fibers will be left in solution. However, in practice, for a finite time, the solution will contain a mixture of small crystals and single chains. When the solvent is removed, the small crystals will act as seeds that trigger crystallization of the free chains. This process is represented schematically by **Figure 16**.

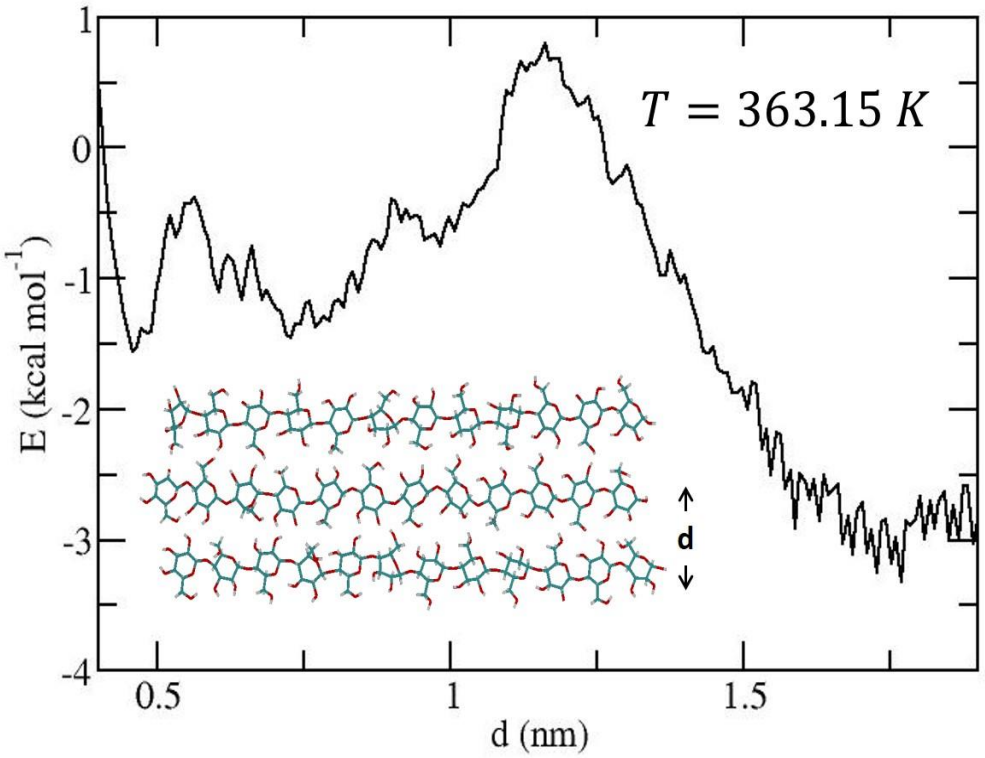

Figure 15. Free Energy to Separate One Chain of Cellulose from an Existing Layer in Ionic Liquid.

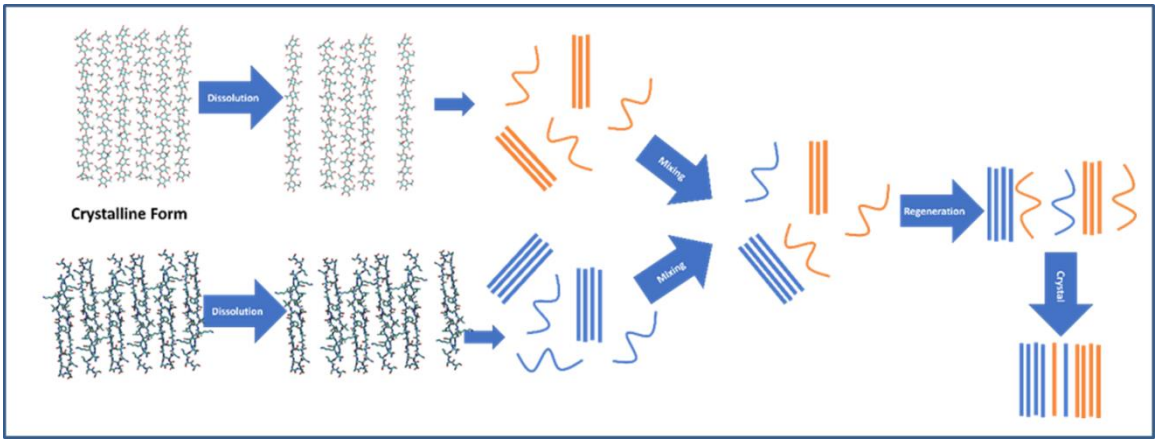

Figure 16. Schematic Representation of the Dissolution, Mixture, and Regeneration Process.

For brevity's sake, the theoretical model focusing on the structural arrangement of mixtures of biopolymers and its derivation are presented in **Appendix C**. Given here are the major results from that portion of the study.

5.3.2 Stacking Model

If we assume that the system is a binary mixture of species A and B, this predicts that the cohesion energy ϵ_A for an A-A interface and ϵ_B for an B-B interface is much larger that the energy ϵ_{AB} for the interface A-B. Under these conditions, let's assume that the system is composed of N_A layers A, N_B layers B and that N_{AB} interface A-B are present. As each interface is composed of one-layer A and one-layer B, then the layers that can form contiguous A and B sections of the crystals are $N'_A = N_A - N_{AB}$ and $N'_B = N_B - N_{AB}$. Based on these considerations, we aim to find the configuration of layers that minimizes the free energy *F* of the system.

The result of this calculation (with some generalizations, see **Appendix C)** yields the number of interfaces N_{AB}:

$$N_{AB}^2 \approx 2N_A N_B \exp\left(-\frac{2\varepsilon - 2a\varepsilon}{k_B T}\right) \tag{5.1}$$

5.3.2 Model Predictions and Comparison with Experiment

Based on these values we can make the following predictions:

1. Even if a = 0, it is predicted that a few interfaces are formed. This is consistent with the formation of impurities even if the two materials are unlikely to blend.
2. As *a* increases, more and more interfaces are formed. As these interfaces have a lower cohesion energy than the A-A and B-B surfaces, it is expected that they are less stable as the temperature rises. This fact will affect the thermal stability of the crystal. For instance, it is expected that these interfaces will start decomposing at a lower temperature than the pure substance. The consequence of this fragmentation is that the crystal will break into smaller chunks. Each chunk is also easier to vaporize as the layers at the edges are the first to be dislodged.
3. The formula is symmetric for concentration. That means that 80% A and 20% B has the same number of interfaces as 80% B and 20% A. As a consequence, they will exhibit similar decomposition temperature.

These trends derived from the theory given correctly predict trends observed experimentally by the Salas group.

For example, as point number 3 above states, the model predicts that decomposition temperatures should be very similar for 80%-20% and 20%-80% compositions. To validate this prediction experimentally, we turned to thermogravimetric analysis (TGA) performed by the Salas group, the results of which are given in **Table 2**.

Table 2. Temperatures of Highest Decomposition Rate for Pure and Blended Biopolymer Samples.

Sample	Temperature (℃)
Pure	
Cellulose	363
Chitin	287, 396
Chitosan	288
Silk	322
20% silk	
CES20	255, 268
CHS20	266, 386
CTS20	217, 254
80% silk	
CES80	258
CHS80	276
CTS80	224, 256

Chitin reveals a two-stage decomposition process, with a first maximum at 287 ℃ and a second maximum at 396℃ in agreement with existing studies.[134] Pure cellulose and chitosan, on the other hand, show a one stage degradation process, with maxima at 363℃ and 288℃, respectively. Concerning the mixtures, we notice that the 20% silk films have two decomposition temperatures, whereas the 80% silk/chitosan film (CTS80) is the only one that shows two decomposition temperatures. In general, the decomposition temperatures of the blended films occur at lower temperatures than the corresponding pure material. Contrary to this observation, though, we note that an increase of the silk

content from 20% to 80% leads to a slight increase in the temperature of thermal degradation (below 10℃). This is in excellent agreement with our theoretical prediction that the change in decomposition temperature should be minimal.

Based on our theoretical model for the dissolution process we would predict that higher decomposition temperatures should be related to a higher crystallinity as well. Indeed, we expect that some small crystalline regions will survive the dissolution and that these regions will trigger the aggregation during regeneration. The size of these seeds would depend mostly on the strength of the dispersion forces between stacked layers, which can be indirectly estimated by the decomposition temperature. Ideally, a higher decomposition temperature is related to stronger dispersion forces, which in turn are linked to larger crystallites. Thus, we expected that material with a higher decomposition temperature will show a higher crystallinity upon regeneration. In order to compare these predictions with experimental data, the Salas group characterized the crystallinity (defined as the amount of β-sheet from the IR spectrum[135]) of the regenerated films. The results of these calculations can be found in **Table 3**.

Table 3. The Percentage of Crystallinity of β-sheet Blended Films at 20 and 80% Silk and the Crystallinity Differences.

Polysaccharide	Crystallinity of β-sheet in 20% silk film	Crystallinity of β-sheet in 80% silk film	The difference in crystallinity
Cellulose	28.4%	37.6%	9.2%
Chitin	22.4%	42.2%	19.8%
Chitosan	18.1%	31.3%	13.2%

This trend is respected except for blends with 80% silk. The reason is that our model assumes that each strand is a single molecule. However, proteins can fold multiple times, so that adjacent strands will generally belong to the same molecule. Thus, breaking the hydrogen bonds from one strand, will not necessarily remove an entire molecule from the crystal. As a consequence, each crystal is expected to form a "fuzzy" coat around the core of the crystal that will further slow disassembly of the crystal. Thus, samples high in silk content will generally show higher crystallinity than other samples with the same decomposition temperature.

5.4 Conclusions

We have proposed a novel model that describes the local structure of biopolymer blends composed of polysaccharides and proteins like silk. This model states that upon dissolution, layer-like materials (a common feature of crystalline biopolymers), form a

mixture of free polymers and small crystalline seeds. When the solvent is removed, these seeds would act as nuclei that grow by adding both free polymer chains and whole seeds. Our model correctly predicts trends in the material's properties including its crystallinity and decomposition temperature.

6 Conclusion

The research discussed in this thesis emphasizes the multidisciplinary path that the CCIB has offered me during my time as a Master's student at Rutgers—Camden. The work presented in **Chapter 4** began as a stand-alone project that would examine the effect of pseudophosphorylation and phosphorylation on the conformational ensemble of the protein tau in the early pathogenesis of Alzheimer's disease.

It became apparent to us early on, though, that due to tau's highly dynamic, intrinsically disordered nature and the large variety of classical force fields available today, this study would benefit from validation that shows that these force fields can accurately represent the experimental behavior of tau. This notion led to the results presented in **Chapter 3** of this work. This verification is important because as interest in the simulation community shifts towards MD simulations of disordered biopolymers, we must closely scrutinize results from today's empirical force fields, many of which were developed and validated with much more structured macromolecules in mind.

We have also addressed, in **Chapter 5**, the complementary nature of MD simulation, pen-and-paper theory, and experiment. We showed that results from simulation of the dissolution of biopolymers in ionic liquids can reinforce predictions made by theory rooted in statistical mechanics that extends to the full dissolution and regeneration process of biopolymer mixtures. This theory predicts the trends that are observed experimentally by the Salas group.

6.1 Future Work

A natural extension of the work presented in **Chapter 3** is to examine the ability of classical force fields to predict the conformational ensemble of more intrinsically disordered peptide systems. Our simulations show that many force fields cannot reproduce the unstructured ensemble of the tau-derived peptide PHF43. As the simulation community grows more and more wary of the skewed results presented by empirical force fields in the simulation of disordered peptides,[27, 30, 136] a comprehensive evaluation of their ability to predict disordered structures is necessary. Results from a concerted effort to tackle this issue can be used to identify undesirable trends in popular force fields and guide the development of more accurate parameterizations that are reliable for disordered systems such as that presented in **Chapter 4**.

There are numerous directions that the work presented in **Chapter 5** could take, as this issue of biopolymer mixtures and the model shown are relatively recent developments. Our MD simulations unfortunately characterize only a small percentage of the wide range of possible inter-biopolymer and biopolymer-solvent interactions that occur during the dissolution and regeneration phase of the mixing process. An extension of this work on a similar scale to that presented here would be to simulate free chains of cellulose (or another polysaccharide) and silk in ionic liquid simultaneously to examine the structures that do or do not develop. From here, ionic liquid molecules could be removed and replaced with explicit water molecules systematically in order to investigate what occurs on the molecular level as the IL is washed away. Of course, these

simulations would be extremely computationally demanding, and methods such as course-graining may be necessary, especially as we endeavor to characterize long-range structural characteristics of these mixtures with MD.

Appendix A

Presented here is work that has been published[137] previously. Figures are adapted from the original publication. All experimental work was performed by the Fu group at Rutgers—Camden.

Computational and Experimental Analysis of Short Peptide Motifs for Enzyme Inhibition

Introduction

Enzymes are crucial to cellular function, being responsible for the production of molecules and the derivation of energy that is vital to metabolism and reproduction of living things.[138-139] Regulation of enzyme activity is central to therapeutics and drug discovery.[140-141] Experimental methods to identify ligands that can bind to proteins and modulate their functions[142-147] are generally labor-intensive and costly, requiring either multiple selection cycles over large chemical libraries (10^6 or more) or the structural information of proteins. Recent developments of microarray technology have allowed the screening of small molecules, peptides, proteins and nucleic acids for identifying ligands that can bind to a protein target.[148-150] Peptides represent a promising class of potential enzyme modulators[151] due to the large chemical diversity[152], and well-established methods for library synthesis.[153] Peptides and their derivatives are found to play critical roles in modulating enzymes and mediating cellular uptakes, which are increasingly

valuable in therapeutics.[154] In this work, we and the Fu group presented a methodology that combines molecular dynamic (MD) simulations and point-variant screening to identify short peptide motifs for inhibiting enzymes.

Results

The functional motifs were predicted based on the simulated conformations of lead peptides. Two 20-mer lead peptides (PEP-1: RVFKRKRWLHVSRYYFGSC; PEP-2: PASMFSYFKKQGYYYKLGSC) were previously selected from a microarray of 10,000 peptides for inhibiting β-Gal with IC_{50} values ~ 1.6 μM and 13 μM[150]. Both of the two peptides were stronger inhibitors than phenylethyl β-D-thiogalactoside (PETG, IC_{50} ~ 35 μM), a known competitive inhibitor of β-Gal[155].

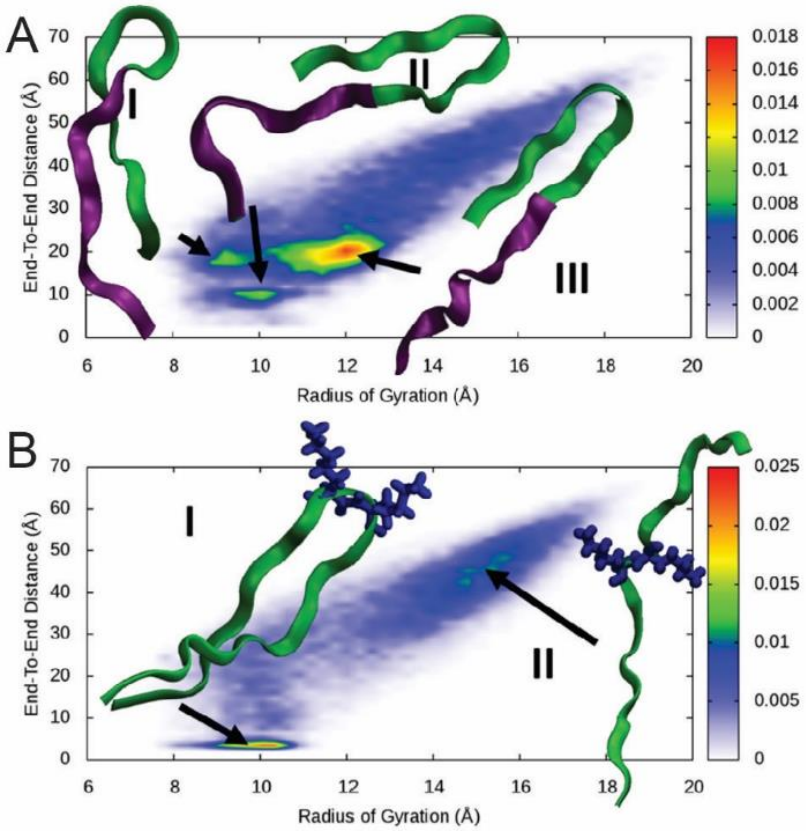

Figure 17. Normalized Distribution of the End-to-end and Radius of Gyration Plot for PEP-1 (A) and PEP-2 (B). PEP-1 populates three major clusters characterized by β-hairpin conformations. The N-terminus (RVFKRYKRW, labelled in purple) is usually fully exposed to the solvent. PEP-2 conformations are dominated by the repulsion of the two central lysines. The most stable conformations are characterized by the two lysines pointing in opposite direction (side chains reported in blue), so to minimize their repulsion. (Reproduced from Ref.[137] under the terms of the Creative Commons Attribution License, copyright © 2017 Fu et al)

To analyze the critical motifs for inhibiting enzymes, MD was used to simulate the conformations of two selected peptides in aqueous solution. As shown in **Figure 17A**, PEP-1 showed three most populated conformations, all with highly positively charged N-terminus fragments (labelled with purple) that were fully exposed to the solvent. The C-terminus of PEP-1 (labelled with green) adopted a β-hairpin conformation that was rich

in hydrophobic moieties, which induced its compact structure. In contrast, the abundance of positively charged residues in N-terminus fragments created a repulsive force that extended the linear structure. In these populated conformations, an interesting exposable N-terminal motif (RVFKRYKRW) was observed, which were suggested to be responsible for inhibiting β-Gal. As shown in **Figure 17B**, PEP-2 was mainly populated by two conformations: conformation I tended to adopt a β-hairpin while conformation II tended to adopt an extended structure. The transition between two conformations happened via a two-lysine portion (KK) in the middle of the sequence. The two lysine residues pointed to the opposite direction that constrained the adopted conformations of PEP-2. It was expected that the removal of these two lysines would affect the adopted conformations of PEP-2 for inhibiting β-Gal.

Since β-Gal was a negatively charged enzyme (pI ~ 4.7) at neutral pH condition, we suggested that positively charged residues of peptides (e.g. lysine, K; or arginine, R) would be more critical for interacting with the enzyme via charge interaction. Based on the sequences of two peptides, N-terminal motif of PEP-1 (RVFKRYKRW) contained most of the positively charged residues, thus, it was predicted to be critical for inhibiting β-Gal. PEP-2 was divided into two motifs by KK residues: N-terminal PASMFSYF and C-terminal KKQGYYYKL. C-terminal motif of PEP-2 was suggested to be more critical for inhibiting enzyme due to the rich K residues.

To validate the above modelling results, the Fu group performed a set of truncation and alanine-scan analysis. Their methods and results can be found in the original paper[137]. In summary, as was predicted by our MD simulations, inhibition of β-

Gal was found not to be dependent on the full sequence of the two peptides, but was instead largely dependent on much shorter motifs. For example, truncation analysis revealed that a 12-mer "RVFKRYKRWGSC" (with "GSC" as a linker) showed similar inhibition of β-Gal with respect to the 20-mer PEP-1's ability to inhibit the enzyme. This truncated sequence was exactly the same as the N-terminus motif that was identified in our simulations. Furthermore, this 12-mer was further reduced to a 10-mer motif "FKRYKRWGSC" as the N-terminus "RV" residues were shown to not significantly affect the inhibition of β-Gal. Point-variant screening results showed further that the inhibitory ability of the shortened motifs could be optimized more so by replacement of single residues with others that were hypothesized to be able to play a role in inhibition (charged residues, those with bulky side chains, etc…). This resulted in a fully optimized 7-mer inhibitory sequence. Again, for a more detailed description of experimental results and methods, please see ref.[137].

Discussion

In summary, we and the Fu group described a method that combined MD simulations and experimental validations to rapidly identify short peptide motifs for inhibiting enzymes. Based on the simulated conformations of lead peptides, short peptide motifs were predicted for inhibiting enzymes with low μM IC_{50} values. The point-variant screening performed by the Fu group showed that a small, 7-mer peptide motif can be optimized with these methods described here. Compared with other MD simulations of ligand-

protein interactions, our presented approach does not require the detailed structural information of the enzyme target and the ligand-protein complexes beforehand, thereby this method reduces the effort associated with selecting these structures as well as resources and time needed. The identified short peptide motifs can be easily synthesized in large scale with high purity. This approach appears to be applicable to both natural and artificial peptides. With more discovered short motifs, they can be exploited for modulating biocatalysis, and developing new medicine.

Methods

Experimental methods and materials utilized by the Fu group can be found in the original paper.[137]

Molecular Dynamics: The simulations were performed using the AMBER 14 package[87] with the AMBER96 force field[88] in an implicit solvent model. As an implicit solvent model we chose the IGB = 5 set of parameters associated with gbsa = 1.[156] Simulations were performed in pure water. A cutoff of 20Å was used for the implicit solvent model, the dispersion forces and the electrostatics. The temperature was kept constant with a Langevin bath with a characteristic time of 1 ps^{-1}. In order to increase the sampling of the conformations we have applied Replica Exchange Molecular Dynamics (REMD).[157] Exchanges are attempted every 3 ps and we use 8 temperatures (282, 300, 325, 353, 383, 415, 448, 484 K). The equations of motion were integrated using a time step of 1 fs. The

bonds between hydrogen atoms and heavy atoms were kept constant using the SHAKE algorithm.[158] Simulations were 600 ns long and we analyzed only the final 300 ns.

Appendix B

This appendix is a reproduction of a manuscript that is to be submitted.

A Theory for Designing Synthetic Compartmentalization of Enzymatic Systems

Background

The biochemical processes that govern the function of a cell typically take place within organelle—essential, compartment-like subunits that perform specific cellular duties. Organelles are just one method by which cells are able to compartmentalize chemical reactions.[159-161] They provide a way for certain molecules to be chemically and spatially isolated from the cell proper, allowing for a more favorable environment in which to undergo whatever process they are intended. We often find that organelles in eukaryotic cells are encapsulated by a membrane, which acts as a barrier and a filter of sorts.

In this case, these organelles typically allow for the entrance of some molecules and the exclusion of others. Generally, proteins embedded in the membrane are selective with respect to these molecules and either actively or passively facilitate transit across the membrane (for example, ion channels). Organelles delimited by a membrane are probably the most advanced form of intracellular compartmentalization due to their extreme selectivity and their ability to alter their behavior according to external cues, repair the membrane itself, and replicate.

Cells, however, exhibit simpler systems of compartmentalization than membrane-bound organelles. Cells form many types of proteinaceous and membraneless organelles such as ribosomes, so-called “nuage granules”, and perhaps the most widely-studied, nucleoli.[162-163] These “phase-separated” compartments are found throughout the cell where they remain isolated from their immediate environment as if they were oil droplets in water. The absence of a proper membrane makes these organelles highly dynamic and extremely sensitive to environmental changes, lending them the ability to rapidly appear and disappear contingent on minute cues from their surroundings.[162-165]

Biochemical reactions typically produce yields that cannot be matched by today's artificial catalysts. As a consequence, many groups have made an attempt to mimic nature's behavior.[159-161] In these studies it has been shown that an enzyme enclosed within a scaffold, even in one different from its natural cellular environment, increases the enzyme's activity. Thus, in the current study, we ask whether encapsulation alone may provide any advantage from both a technological and evolutionary point of view.

Collecting enzymes and other proteins into clusters of interacting units yields several advantages for a cell. The most obvious of which is that compartmentalization limits the presence of contaminants or toxic species and avoids unwanted cross-interaction among pathways. Additionally, these structures tend to exhibit enhanced activity. Thus, it would be desirable to mimic and exploit the same type of mechanism by building artificial “compartments” that perform a requested chemical reaction at maximum yield. In this study, we will focus on a particular class of such structures consisting of an enzyme encapsulated within a DNA scaffold.

From a technological point of view, another approach to the problem consists in designing novel enzymes with enhanced capabilities. This technique has shown some success[166-168] and is definitely promising. However, designing novel enzymes requires knowing the primary structure, the genes involved, and which region corresponds to the active sites. In addition, it requires extensive trial and error, as some mutations may lead to misfolding and even aggregation. On the other hand, encapsulated enzymes are far easier to produce, as they can be extracted from any source and it is relatively straightforward to build a cage with the desired properties using DNA origami techniques. Of course, the optimal situations would be a combination of these two approaches. Furthermore, it should be noted that enzymes have evolved to work in an enclosed environment, so understanding how confinement works would have direct application to understanding evolution as well as improving drug discovery and delivery.

From a practical point of view, the DNA origami methodology and other techniques for building encapsulated enzymes, lack a sound theoretical framework to predict the observed trends. While, in principle, we must have accurate numerical simulations in order to have quantitative agreement with experiments, here we will prefer a more practical approach (e.g. refs.[169-175]). This methodology relates observed trends to quantities that can easily be controlled during experiments. The approach is completely general and relies only on geometric properties of the nanostructure. Due to the generality of the approach, our theory proves that compartmentalization has the intrinsic advantage of enhancing the yield of a reaction, independent of the tightly regulated cellular environment at large. It is appealing to speculate that from an evolutionary point of view, single molecules that have been encapsulated within any rigid scaffold may have found a

more favorable environment for self-replication. And only later on, these structures have been collected into more complex structures such as cells.

System

To keep the theory concrete, we will consider a practical application of this framework. We will focus on the system composed of an enzyme encapsulated within a DNA cage. This nanocage is built using DNA origami techniques[176], which allow high-level control of the shape, size, and porosity of the cage. A schematic of the structure is shown in **Figure 18** and the geometric variables used in the theory are defined there as well. Details about the construction and validation of the structure obtained is provided in greater detail in the paper by Zhao et al.[177] Currently, no theory exists that explains the experimental behavior observed therein.

Model

We begin our modeling by noticing that, due to thermal motion, the cage will rotate about its center of mass, tracing a sphere of radius $R_{ex} = 54\ nm/2 = 27\ nm$ (**Figure 18A**). To enter the cage, the substrate must strike one of the porous sides. These sides have a square cross-section with edge 27 nm. However, only a fraction of the edge length $L_{in} = 20\ nm$ is actually connected to the internal cavity. Thus, the probability that the substrate can actually enter the cage is:

$$P_a = \frac{2L_{in}^2}{4\pi R_{ex}^2} \qquad (B.1)$$

The flux of substrate through an arbitrary surface is the same in the case of both the free and encapsulated enzyme. As a consequence, the probability $P^{(e)}$ that the substrate reaches the encapsulated enzyme is proportional to the flux of substrate through the surface of the sphere of radius R_{ex} times the probability P_a that it strikes one of the porous sides. For the free enzyme, the probability $P^{(f)}$ of substrate approaching the enzyme surface is proportional to the flux of ions that reaches the surface itself, assumed to be a sphere with radius R_E equal to its hydrodynamic radius. Hence:

$$\frac{P^e}{P^f} \propto \frac{P_a R_{ex}^2}{R_E^2} \qquad (B.2)$$

A further aspect to keep in mind, though, is the possibility that a particle may collide with the correct surface but still bounce back into solution due to the presence of the DNA helices. However, this notion will probably not be the case here. Once a molecule of substrate is consumed by the free enzyme, other molecules can rapidly fill this void. However, the box is separated from the solvent by a thin membrane. This membrane delays the exchange of substrate and product between the cavity and the solution. As a result, every time a substrate is consumed and a molecule of product is released, the product will linger within the cavity. At the same time, as the cavity is relatively small, the substrate concentration drops considerably with respect to the free enzyme. As a consequence of an imbalance of concentration of product and substrate between the interior and the exterior of the box, an osmotic pressure is created that will push the

product out of the box while attracting substrate into the box. Thus, substrate entering the box does not do so merely according to diffusion, but according to the convective motion that is generated, as well. This is equivalent to assume that a particle is always forced inside the cavity by the osmotic pressure.[169]

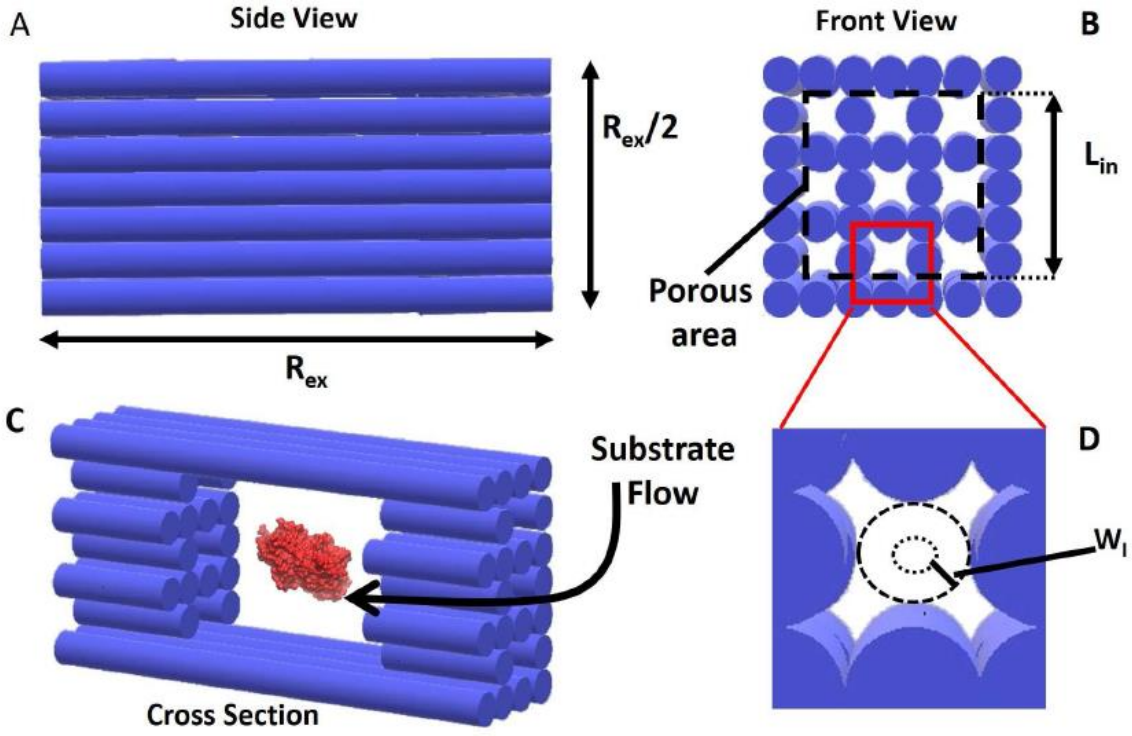

Figure 18. DNA Nanocage with Parameters. The figure shows a schematic picture of the nanocage (not to scale). The scaffold is built using DNA origami techniques. Each cylinder in the picture represents a DNA double helix with a radius of about 1 nm. A) The cage is open only on two sides (left and right in the picture). B) Only a small portion of these two sides is actually connected to the internal cavity. This area has a square cross section of edge $\boldsymbol{L_{in}} = 20$ nm. The porosity of each side is measured in helix per area $\boldsymbol{\rho_{DNA}}$. C) An enzyme with hydrodynamic radius $\boldsymbol{R_E}$ is located in the center. The internal cavity is a cube of edge 20 nm. D) Due to water strongly bound to the DNA surface, the effective area of each pore that can be actually traversed is smaller by $\boldsymbol{W_l}$. Details about the fabrication can be found in Ref.[177].

A second aspect we must add to our model is the fact that, once inside the cavity, the substrate may strike the enzyme with a probability $P^{(0)}$. If the substrate misses the enzyme during the first passage, it may be reflected by the opposite surface of the cage

with probability P_r. This ricochet effect could take place multiple times, leading to a converging geometric series:

$$P_{hit} = P^{(0)} \sum_{n=0}^{\infty} \left[\left(1 - P^{(0)}\right) P_r\right]^n = P^{(0)} \frac{1}{(1 - (1 - P^0) P_r} \quad \text{(B.3)}$$

Where P_{hit} is the likelihood of striking the enzyme after multiple ricochets. Next, we must estimate $P^{(0)}$ and P_r. If we assume that the enzyme is a hard sphere of radius R_E and the substrate has a radius R_s, then:

$$P^{(0)} = \frac{\pi (R_E + R_s)^2}{L_{in}^2} \quad \text{(B.4)}$$

More aspects to keep into account, though, when considering $P^{(0)}$ are the internal volume interactions. The cage is built using DNA chains, which are highly charged polymers. As a consequence, its electric field will affect the behavior of other molecules that surround it. What we are interested in is the ability of the substrate to displace water molecules during its motion. It has been shown that the exact number of water molecules bound to DNA depends on their location along its surface[178]; the minor groove affects two water layers (about 0.6 nm), while the effect of the major groove is limited to one water shell (about 0.3 nm). The extent to which water is bound to the DNA is also affected by the ionic strength of the solution (discussed further in a later section). As a consequence, it is unlikely that the substrate would approach the wall as the water next to the DNA would be very difficult to displace. Thus, we would expect that the substrate may approach the DNA somewhere around the second water layer (0.3-0.6 nm) on average. For now, we

will be assuming a minimum distance from the wall of about $W_l = 0.43\,nm$ (see Supplementary Information for details):

$$P^{(0)} = \frac{\pi(R_E + R_s)^2}{(L_{in} - 2W_l)^2} \qquad (B.5)$$

Now, we evaluate P_r which accounts for internal reflection with the cage walls. There are two contributions. The first relates to the substrate striking one of the helices, but in addition, we must also consider the case when the substrate misses the helix but collides with the layer of water bonded to the DNA. We start by considering the latter contribution and we will reason as follows. The openings in the porous sides of the cage are of the size of a DNA chain (**Figure 18B**). Thus, we can think of these channels as empty sockets for DNA chains. The effect of the water layer is then equivalent to shrinking the opening, which is assumed circular (**Figure 18B**, smaller dotted circle). Therefore, an empty socket, without bound water molecules, has a cross-sectional area of $A_h = \pi\ nm^2$ (the radius of a DNA helix is 1 nm). Accounting for the water layer, this becomes $A_{hl} = (1 - W_l)^2 \pi\ nm^2$. Hence:

$$P_r = \rho_{dna} A_h + \frac{(1 - \rho_{dna} A_h)(A_h - A_{hl})}{A_h} \qquad (B.6)$$

Thus, we can evaluate the expression:

$$\frac{P^{(e)}}{P^{(f)}} \propto P_a P_{hit} \frac{R_{ex}^2}{R_E^2} \qquad (B.7)$$

We will now estimate the coefficient of proportionality. The analysis so far has focused on factors involving how the substrate will approach the enzyme. However, catalytic enhancement depends on the activation barrier for binding to the enzyme, as well. Transition state theory states that the ratio between the two rates will be

$$P_{TST} = \frac{\exp(F^{(e)}/RT)}{\exp(F^{(f)}/RT)} = \exp\left(\frac{F^{(f)} - F^{(e)}}{RT}\right) \quad (B.8)$$

where R is the gas constant, T the temperature, and $F^{(f)}$ and $F^{(e)}$ are the free energy of activation of the free and encapsulated enzyme, respectively. It should be pointed out that this contribution pertains to the enzyme itself and not the geometry of the box. It is well-established that enzymes[179-181] as well as other biomolecules[182-183] in confined environments will behave differently. While this ratio is very difficult to measure both experimentally and computationally, Zhao et al.[177] were able to measure a value of about 2.5 for the enzyme G6pDH with glucose-6-phosphate as substrate. Of course, this value is not expected to be the same for all enzymes. To estimate a value for the other enzymes involved we use the fixed-charge-density framework.[184-185] This framework connects the size of the enzyme with its activation energy. The leading term of the activation energy can be written as:

$$F = \lambda_1 + \frac{\lambda_2}{R_E} \quad (B.9)$$

where λ_1 is a parameter that depends on the dielectric properties of the active site of the enzyme, whereas λ_2 depends on the properties of the surface of the active site and its surrounding environment (local geometry as well as dielectric environment). It is

reasonable to assume that the value of λ_1 does not change if the enzyme is encapsulated, so that it will disappear when the difference is computed in Eq. B.8. Hence:

$$P_{TST} = \exp\left(\frac{F^{(f)} - F^{(e)}}{RT}\right) = \exp\left(\frac{\lambda}{R_E RT}\right) \quad \text{(B.10)}$$

where $\lambda = \lambda_2^{(f)} - \lambda_2^{(e)}$. The value of λ can be fixed knowing that $P_{TST} = 2.5$ when R_E is the radius of G6pDH.

Experimentally, the catalytic activity of an enzyme is measured by the turnover number K_{cat} that represents the number of molecules of substrate converted by an enzyme into product per unit of time. The ratio of the turnover numbers of the free enzyme $K_{cat}^{(f)}$ and the encapsulated enzyme is $K_{cat}^{(e)}$:

$$\alpha = \frac{K_{cat}^{(e)}}{K_{cat}^{(f)}} = P_{TST} P_a P_{hit} \frac{R_{ex}^2}{R_E^2} \quad \text{(B.11)}$$

Results

Enzyme activity

Our model does not require any fitting parameters. However, the hydrodynamic radii are not available for all substrates that may be involved.} Here, we have assumed that glucose and glucose-6-phosphate have the same size. For NAD+ we used the radius of gyration in solution[186] and converted into the hydrodynamic radius assuming a ratio of 0.775[187]. NAD+ and NADH are assumed to have similar size. All values are reported in **Table 4**. With respect to the original paper by Zhao et al.[177], we have omitted the enzyme β-Galactosidase (whose size is similar to the inner cavity) and the substrate hydrogen peroxide (this can replace water molecules bound to the DNA cage). In **Table 4** we have compared the ratio α reported in the literature[177] to the theoretical prediction. The agreement is generally good, with the exception of G6pDH with Glucose-6-phosphate. In this case, the negatively charged phosphate group attached to the glucose is likely repelled by the charged DNA cage. In addition, we have also compared the concentration at which the chemical reaction reaches half speed K_m, another commonly computed quantity. As α is concentration independent (the flux J cancels out when the ratio in Eq. B.11 is computed), we would expect this value not to change. This prediction is not respected by the larger enzymes. This is due to the orientation of the active site of the enzyme within the cage. For small enzymes, it is reasonable to assume that the active site is generally exposed to the stream of substrate in the same way as the free enzyme during reorientation.[170] However, in larger enzymes—as they cannot be modeled as easily as

perfect spheres—some orientations may position the active site either too close to the cage wall, effectively deactivating the enzyme, or may expose it directly to the porous area, so that more substrate reaches the active sites.

Table 4. List of the Hydrodynamic Radii (values with * are estimates) and Experimental and Theoretical Values for α and K_m

Enzyme	Substrate	R_E (nm)	R_s(nm)	α (exp.)	α (th.)	K_m^f (μM)	K_m^e (μM)
HRP[188]	ABTS[189]	3	0.67	9.5 ± 0.5	10.3	2600 ± 400	2500 ± 200
MDH[190]	NADH	3.62	0.65*	9.0 ± 1.3	6.9	180 ± 50	270 ± 50
G6pDH[191]	Glucose-6-Phosphate	4.3	0.4*	3.5 ± 0.2	4.6	220 ± 20	310 ± 30
G6pDH	NAD+	4.3	0.65*	4.8 ± 0.2	4.9	510 ± 50	590 ± 40
LDH[192]	NADH	4.3	0.65*	4.1 ± 0.3	4.9	7.2 ± 1.3	17 ± 1.5
GOx[193]	Glucose[194]	4.3	0.4	5.4 ± 0.4	4.6	6200 ± 900	3000 ± 600

Helix Density

Another factor that has been evaluated experimentally is the density of packing for the helices on the porous sides of the cage. The only case available experimentally is the enzyme G6pDH with either glucose-6-phosphate or NAD+ as substrate. For a density of 0.12 helices/nm (typical of a honeycomb structure), $\alpha = 4.6$ for glucose-6-phosphate and 4.9 for NAD+. If the density reaches 0.16 helices/nm, these values become 5.09 and 5.35, respectively. Thus, we predict an increase of 1.1 times, slightly smaller than the experimental 1.3.[177]

NaCl Concentration

NaCl is expected to screen the electrostatic effects of DNA, so W_l should decrease as the salt concentration increases. Assuming a dependance similar to the Debye length[195], then $W_l \propto \sqrt{I}$, where I is the ionic strength. Thus, if we define the value of W_l' at $I = I'$ and W_l at $I = I' + I_{Na}$, then

$$\frac{W_l}{W_l'} = \frac{1}{\sqrt{1 + \frac{I_{Na}}{I'}}} \qquad (B.12)$$

with I_{Na} being the ionic strength associated with a concentration c_{Na} (Mol/Liter) of the added salt. For NaCl: $I_{Na} = 0.5c_{Na}$. A typical PBS buffer used in experiments contains about 0.1 M of Na. Thus,

$$\frac{W_l}{W_l'} = \frac{1}{\sqrt{1 + \frac{c_{Na}}{0.1}}} \qquad (B.13)$$

As some of the data points are measured at very high concentrations of NaCl, it is more appropriate to write the equations in terms of the activity a_{Na}:

$$\frac{W_l}{W_l'} = \frac{1}{\sqrt{1 + \frac{a_{Na}}{0.1}}} \qquad (B.14)$$

Figure 19 shows a comparison between theory and experiment.[177] The experiments are reported (squares) as well as two sets of theoretical predictions (circles and triangles). The circles represent a direct application of Eq. B.14. These values are reasonably close to the experiments for low concentration, but they depart at higher concentration. This happens because so far we have been assuming that all of the enzymes stay active at high salt concentration. However, that is never the case.[196-198] The second curve (triangle) is the prediction if the enzyme inhibition is taken into account (data from Ref.[197]). In this case we used the data for the free enzyme in solution. The fact that this prediction is a bit below the experimentally measured yield most likely results from the cage offering some protection from salt inhibition. This is probably a consequence of the negative charge of the DNA chains, which repel chloride ions. Indeed, it has been shown that both NaCl and KCl inhibit the enzyme in the same way, suggesting that the anion is responsible.[199]

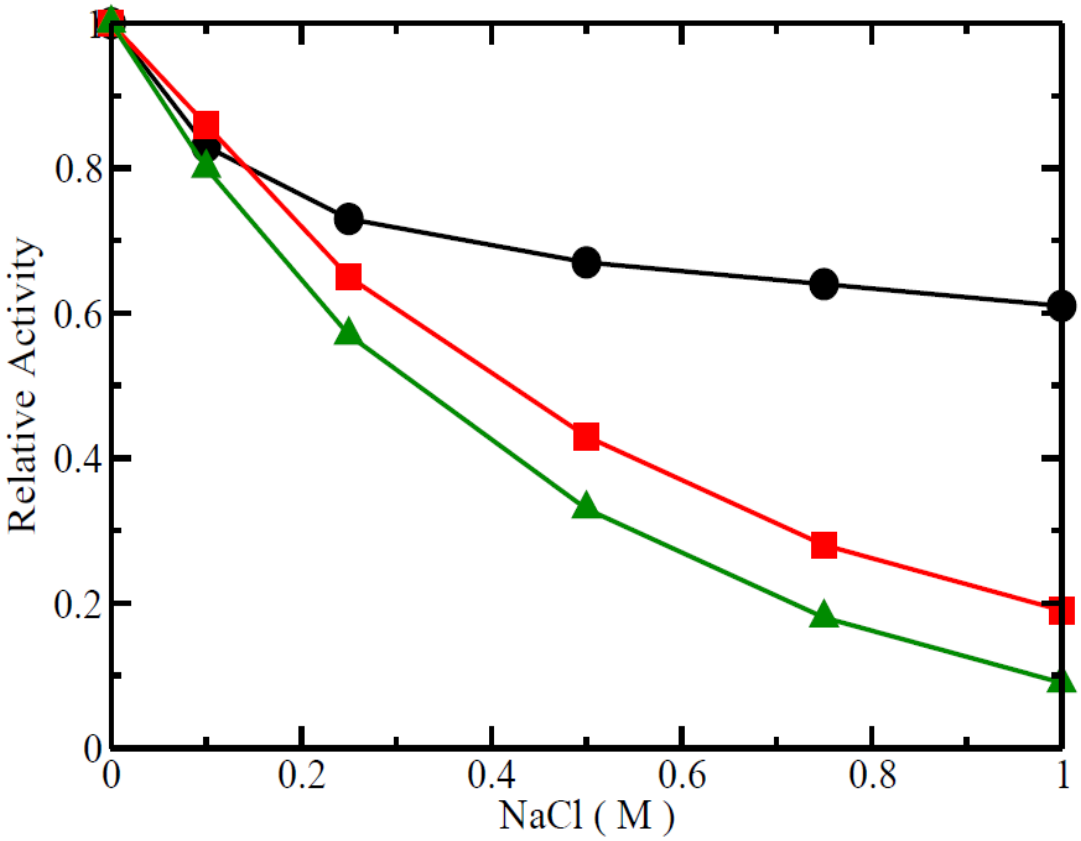

Figure 19. Effect of the Concentration of NaCl on the Activity of G6pDH and Glucose. The figure compares the experimental data (square), with the theoretical prediction in case enzymes are not (circles) or are inhibited (triangles) by the salt.

Implications for Evolution

As a final part of our discussion, it is worthwhile to postulate a connection between our and other's studies and the origin of life. The classic experiment by Miller and Urey[200] showed that, under the proper conditions, a mixture of inorganic compounds may be converted into organic molecules. The subsequent step for life would be the development of self-replicating entities that eventually evolve to the point of the complex cells we see today. It is generally accepted that RNA would have been the best candidate to enact this

evolutionary maneuver because, in addition to its ability to self-replicate, it can also perform regulatory tasks upon itself.[201-202] Our theory shows that if a molecule of RNA were to be arbitrarily encapsulated by a membrane, it would enhance its ability to secure nucleic acids as well as replicate. At this stage, as there is no mechanism to guarantee a perfect copy of the RNA strand, this process would not be free from error, therefore mutations would be very common. As encapsulated RNA replicates faster, a large pool of mutants would soon become available. Eventually, mutants that interact strongly with the membrane will appear. Following this evolutionary factor, the ability of replicating not just the RNA, but the entire RNA and membrane system, would emerge. In fact, as the RNA continues to replicate within a membrane, it would eventually reach a critical internal concentration where the electrostatic repulsion between RNA chains will lead to the rupture of the existing membrane and the formation of two smaller membranes with residual RNA molecules remaining in each. RNA molecules with higher binding affinity for the membrane molecules will rapidly be encapsulated once again and start replicating faster than free RNA. An analytical model that is consistent with this hypothesis was proposed by Amir[203] that successfully modeled the replication of bacteria. The model relies on a buildup of one molecule until a critical concentration is reached before replication starts. As the model is very general in its assumptions, it applies equally well to the RNA/membrane system discussed here. In a study by Golding and Cox[204] on bacteria, RNA showed subdiffusive dynamics. The authors argue that RNA's movement is retarded by crowding of the cellular environment, an effect that would be present in our RNA/membrane system, as well. Molecules affected by subdiffusive motion will linger for longer within the cellular milieu, so the escape rate will be slower. Thus, RNA

remains within the membrane longer, leading to more replication. Taken together, this representation describes a bottom-up evolutionary concept wherein the formation of primitive replicating organelle-like structures is possible based on unsophisticated physical considerations alone, without considering the complex machinery associated with modern cells. A possible flaw that one may notice in this discussion is that the system will evolve from a disordered (fragments of RNA and membrane) into an ordered state, with an apparent net loss of entropy. In reality, however, this is not the case. A recent paper by Kachman et al.[205] has shown that if a particular structure harvests energy more efficiently than other configurations composed of the same building blocks, it will be thermodynamically stabilized due to a concurrent increase in the heat dissipated, which in turn causes the total entropy to grow. In our case, as encapsulated enzymes exhibit a higher K_{cat}, they perform chemical work more effciently than free enzymes, and so would be stabilized according to the same principle

Conclusion

Concluding, we have proposed a model that has its roots in statistical mechanics and can correctly reproduce the trends of encapsulated enzymes. The model does not rely on any fitting parameters and all the quantities associated with it can be measured experimentally. This model has a direct application to the study of catalysis, as methods to produce structures considered herein are available today, while also giving way to implications for the formation of self-replicating organisms during the early stages of life.

Appendix C

This appendix provides a detailed explanation and derivation of the theoretical model that culminates with equation (5.1) in **Chapter 5**. As the paragraph that precedes the introduction to **Chapter 5** states, this section is reproduced from a manuscript that is to be submitted for publication.

Supplemental Theoretical Model for Chapter 5

We will start by assuming that the crystalline form of the biopolymer is composed of a set of layers stacked on each other (See **Figure 20**). It should be noted that the term "layer" is used in a very broad sense and that we do not make any assumption about the nature of the layer itself. For instance, on the one hand, each section in the figure can be considered as a β-sheet of each biopolymer, with the interface composed of hydrogen bonds between molecules of different types. In this case, each strand in the picture would represent the backbone of one biopolymer. On the other hand, the model works equally well if we assume that each strand is actually an entire β-sheet structure, so each layer is stacked and held together by non-bonded interactions instead of hydrogen bonds. At the same time, we do not make any assumptions about the size of each strand. For instance, each strand may represent a single polymer, a section of a polymer, an entire β-sheet or even a short fibril (even if the latter is probably unlikely in many cases). Thanks to this generality the model can be employed to study mixtures of molecules that belong to different chemical groups, such as proteins and sugars. As a consequence, the model is

useful to determine whether the material under examination is composed of alternating disposition of "blocks" with uniform properties or not. While this may not necessarily represent every possible material, we will show that it is consistent with the polymer blend studied in this work. We will test this assumption in more detail in the simulations discussed later on.

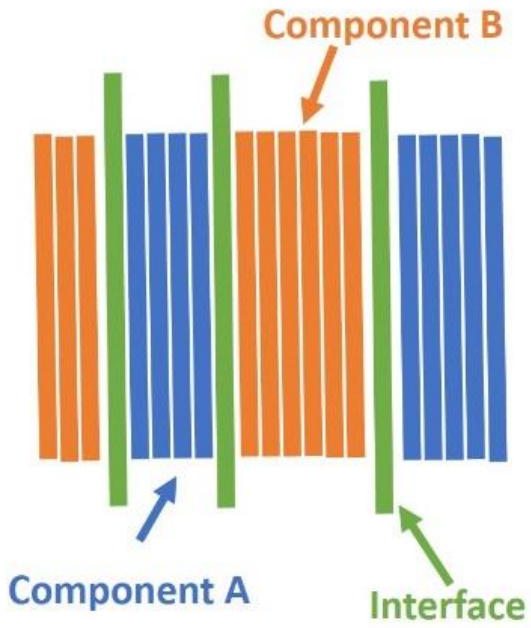

Figure 20. **Schematic of an Idealized Local Crystalline Structure**. We assume that locally the material is composed of crystals composed of chunks of component A and component B separated by surfaces. The model describes the entropy of this system as the number of possible way the interfaces can be arranged.

Stacking model

We expect that the cohesion energy will depend on the ability of each layer to properly adhere to the next layer. It is reasonable to expect that naturally occurring materials have

been selected to maximize this property. If we assume that the system is a binary mixture of species A and B, this predicts that the cohesion energy ϵ_A for an A-A interface and ϵ_B for an B-B interface is much larger that the energy ϵ_{AB} for the interface A-B. Under these conditions, let's assume that the system is composed of N_A layers A, N_B layers B and that N_{AB} interface A-B are present. As each interface is composed of one-layer A and one-layer B, then the layers that can form contiguous A and B sections of the crystals are $N'_A = N_A - N_{AB}$ and $N'_B = N_B - N_{AB}$. Based on these considerations, we aim to find the configuration of layers that minimizes the free energy *F* of the system. The free energy *F* is defined as:

$$F = E - TS \qquad (C.1)$$

Where *E* is the internal energy, *T* the temperature and *S* the entropy. The energy of the system is written as:

$$E = \left(N'_A + \frac{N_{AB}}{2}\right)\epsilon_A + \left(N'_B + \frac{N_{AB}}{2}\right)\epsilon_B + N_{AB}\epsilon_{AB} \qquad (C.2)$$

$$E = \left(N_A - \frac{N_{AB}}{2}\right)\epsilon_A + \left(N_A - \frac{N_{AB}}{2}\right)\epsilon_B + N_{AB}\epsilon_{AB} \qquad (C.3)$$

To compute the entropy, we will use the same strategy adopted for computing the specific heat of a group of harmonic oscillators (namely the Einstein model for the specific heat of a crystal). In this approach, we think of these interfaces as the delimiter of each energy level. The number of "quanta of energy" in each level is now replaced by the number of layers. In addition, each species can only occupy half of the available levels. With this in mind, we can write the entropy as:

$$S_x/k_B = \left(N_x' + \frac{N_{AB}}{2}\right)\ln\left(N_x' + \frac{N_{AB}}{2}\right) - N_x'\ln N_x' - \frac{N_{AB}}{2}\ln\frac{N_{AB}}{2} \quad (C.4)$$

$$S_x/k_B = \left(N_x - \frac{N_{AB}}{2}\right)\ln\left(N_x - \frac{N_{AB}}{2}\right) - (N_x - N_{AB})\ln(N_x - N_{AB}) - \frac{N_{AB}}{2}\ln\frac{N_{AB}}{2} \quad (C.5)$$

Where x stands for either A or B and k_B is the Boltzmann constant. Finally

$$F = \left(N_A - \frac{N_{AB}}{2}\right)\epsilon_A + \left(N_B - \frac{N_{AB}}{2}\right)\epsilon_B + N_{AB}\epsilon_{AB} - k_BT[\,S_A + S_B] \quad (C.6)$$

To find the equilibrium structure, we need the minimum of the free energy:

$$\frac{\partial F}{\partial N_{AB}} = -\frac{1}{2}(\epsilon_A + \epsilon_B - 2\epsilon_{AB}) - k_BT\left[-\frac{1}{2}\ln\left(N_A - \frac{N_{AB}}{2}\right) - \frac{1}{2}\ln\left(N_B - \frac{N_{AB}}{2}\right) + \ln(N_A - N_{AB}) + \ln(N_B - N_{AB}) - \ln\frac{N_{AB}}{2}\right] \quad (C.7)$$

Assuming that N_{AB} is smaller than N_A and N_B:

$$\frac{\partial F}{\partial N_{AB}} \approx -\frac{1}{2}(\epsilon_A + \epsilon_B - 2\epsilon_{AB}) - k_BT\left[-\frac{1}{2}\ln(N_A) - \frac{1}{2}\ln(N_B) + \ln(N_A) + \ln(N_B) - \ln\frac{N_{AB}}{2}\right] \quad (C.8)$$

$$\frac{\partial F}{\partial N_{AB}} \approx -\frac{1}{2}(\epsilon_A + \epsilon_B - 2\epsilon_{AB}) - k_BT\left[\frac{1}{2}\ln(N_A) + \frac{1}{2}\ln(N_B) - \ln\frac{N_{AB}}{2}\right] \quad (C.9)$$

$$\frac{\partial F}{\partial N_{AB}} \approx -\frac{1}{2}(\epsilon_A + \epsilon_B - 2\epsilon_{AB}) - \frac{1}{2}k_BT\left[\ln\left(2\frac{N_AN_B}{N_{AB}^2}\right)\right] \quad (C.10)$$

The minimum is found for

$$\frac{\partial F}{\partial N_{AB}} \approx -\frac{1}{2}(\epsilon_A + \epsilon_B - 2\epsilon_{AB}) - \frac{1}{2}k_BT\left[\ln\left(2\frac{N_AN_B}{N_{AB}^2}\right)\right] = 0 \quad (C.11)$$

$$\frac{N_AN_B}{N_{AB}^2} = \frac{1}{2}\exp\left(-\frac{\epsilon_A + \epsilon_B - 2\epsilon_{AB}}{k_BT}\right) \quad (C.12)$$

To assess the validity of the model, we can make some reasonable guesses about the values of the parameters needed. We know that species A and B form crystals. Thus, their energies ϵ_A and ϵ_B must be negative. In addition, we can guess that ϵ_{AB} is negative (otherwise the surface will not form) and smaller than the other energies because an A-B layer does not come together as favorably as an A-A or B-B layer. Thus:

$$\epsilon_A = \epsilon_B = -\varepsilon \tag{C.13}$$

$$\epsilon_{AB} = -a\varepsilon \tag{C.14}$$

The reasoning here is that the only factor that matters is that ϵ_A and ϵ_B are bigger than ϵ_{AB}--their exact values are uninfluential. The factor a expresses ϵ_{AB} as a fraction of ε. Hence:

$$\frac{N_A N_B}{N_{AB}^2} \approx \frac{1}{2}\exp\left(\frac{2\varepsilon - 2a\varepsilon}{k_B T}\right) \tag{C.15}$$

$$N_{AB}^2 \approx 2N_A N_B \exp\left(-\frac{2\varepsilon - 2a\varepsilon}{k_B T}\right) \tag{C.16}$$

Based on these values we can make the following predictions:

1. Even if a=0, it is predicted that a few interfaces are formed. This is consistent with the formation of impurities even if the two materials are unlikely to blend.
2. As a increases, more and more interfaces are formed. As these interfaces have a lower cohesion energy than the A-A and B-B surfaces, it is expected that they are less stable as the temperature rises. This fact will affect the thermal stability of the crystal. For instance, it is expected that these interfaces will start decomposing at a lower temperature than the pure substance. The consequence of this

fragmentation is that the crystal will break into smaller chunks. Each chunk is also easier to vaporize as the layers at the edges are the first to be dislodged.

3. The formula is symmetric for concentration. That means that 80% A and 20% B has the same number of interfaces as 80% B and 20% A. As a consequence, they will exhibit similar decomposition temperature.

References

1. Childers, M. C.; Daggett, V., Insights from molecular dynamics simulations for computational protein design. *Molecular Systems Design & Engineering* **2017,** *2* (1), 9.

2. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biology* **2011,** *9*, 71.

3. Klein, M. L.; Shinoda, W., Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems. *Science* **2008,** *321* (5890), 798.

4. Tozzini, V., Multiscale Modeling of Proteins. *Accounts of Chemical Research* **2010,** *43* (2), 220.

5. Zhao, H.; Caflisch, A., Molecular dynamics in drug design. *European Journal of Medicinal Chemistry* **2015,** *91*, 4.

6. Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W., Modeling electrostatic effects in proteins. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006,** *1764* (11), 1647.

7. Salsbury, F. R., Molecular Dynamics Simulations of Protein Dynamics and their relevance to drug discovery. *Current opinion in pharmacology* **2010,** *10* (6), 738.

8. Salsbury, F. R.; Clodfelter, J. E.; Gentry, M. B.; Hollis, T.; Scarpinato, K. D., The molecular mechanism of DNA damage recognition by MutS homologs and its consequences for cell death response. *Nucleic Acids Research* **2006,** *34* (8), 2173.

9. Qun, W.; Meng-hao, W.; Ke-feng, W.; Yaling, L.; Hong-ping, Z.; Xiong, L.; Xing-dong, Z., Computer simulation of biomolecule–biomaterial interactions at surfaces and interfaces. *Biomedical Materials* **2015,** *10* (3), 032001.

10. Li, Y.; Liu, X.; Zhang, S.; Yao, Y.; Yao, X.; Xu, J.; Lu, X., Dissolving process of a cellulose bunch in ionic liquids: a molecular dynamics study. *Physical Chemistry Chemical Physics* **2015,** *17* (27), 17894.

11. Tuckerman, M. E.; Oxford University, P., *Statistical mechanics : theory and molecular simulation*. Oxford University Press: Oxford, 2017.

12. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995,** *91* (1), 1.

13. Levitt, M., The birth of computational structural biology. *Nature Structural Biology* **2001,** *8*, 392.

14. Monticelli, L.; Tieleman, D. P., Force Fields for Classical Molecular Dynamics. In *Biomolecular Simulations: Methods and Protocols*, Monticelli, L.; Salonen, E., Eds. Humana Press: Totowa, NJ, 2013; pp 197.

15. Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Istv; #225; Kolossv, n.; #225; ry; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C., Anton, a special-purpose machine for molecular dynamics simulation. *Communication ACM* **2008,** *51* (7), 91.

16. Anandakrishnan, R.; Drozdetski, A.; Walker, Ross C.; Onufriev, Alexey V., Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophysical Journal* **2015,** *108* (5), 1153.

17. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999,** *314* (1–2), 141.

18. Earl, D. J.; Deem, M. W., Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics* **2005,** *7* (23), 3910.

19. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995,** *117* (19), 5179.

20. Wickstrom, L.; Okur, A.; Simmerling, C., Evaluating the Performance of the ff99SB Force Field Based on NMR Scalar Coupling Data. *Biophysical Journal* **2009,** *97* (3), 853.

21. Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C., Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. *Journal of Computational Chemistry* **2003,** *24* (1), 21.

22. Zhou, R., Free energy landscape of protein folding in water: Explicit vs. implicit solvent. *Proteins: Structure, Function, and Bioinformatics* **2003,** *53* (2), 148.

23. Yoda, T.; Sugita, Y.; Okamoto, Y., Comparisons of force fields for proteins by generalized-ensemble simulations. *Chemical Physics Letters* **2004,** *386* (4–6), 460.

24. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006,** *65* (3), 712.

25. Shell, M. S.; Ritterson, R.; Dill, K. A., A Test on Peptide Stability of AMBER Force Fields with Implicit Solvation. *The Journal of Physical Chemistry B* **2008,** *112* (22), 6878.

26. García, A. E.; Sanbonmatsu, K. Y., α-Helical stabilization by side chain shielding of backbone hydrogen bonds. *Proceedings of the National Academy of Sciences* **2002,** *99* (5), 2782.

27. Maffucci, I.; Contini, A., An Updated Test of AMBER Force Fields and Implicit Solvent Models in Predicting the Secondary Structure of Helical, β-Hairpin, and Intrinsically Disordered Peptides. *Journal of Chemical Theory and Computation* **2016,** *12* (2), 714.

28. Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A., The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J., Eds. Springer Netherlands: Dordrecht, 1997; pp 83.

29. Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C., Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *Journal of the American Chemical Society* **2014,** *136* (40), 13959.

30. Song, D.; Luo, R.; Chen, H.-F., The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling* **2017,** *57* (5), 1166.

31. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **2004,** *55* (2), 383.

32. Nguyen, H.; Roe, D. R.; Simmerling, C., Improved Generalized Born Solvent Model Parameters for Protein Simulations. *Journal of Chemical Theory and Computation* **2013,** *9* (4), 2020.

33. Doi, M.; Edwards, S. F., *The Theory of Polymer Dynamics*. Cambridge University Press: Cambridge, 1988.

34. Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E., Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International* **1999,** *38* (1/2), 236.

35. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995,** *91*, 43.

36. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulations. *Journal of Chemical Theory and Computation* **2008,** *4*, 435.

37. Lindahl, E.; Hess, B.; van der Spoel, D., GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *Journal of Molecular Modeling* **2001,** *7*, 306.

38. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, Flexible and Free. *Journal of Computational Chemistry* **2005,** *26*, 1701.

39. D.A. Case, V. B., J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E.; Duke, H. G., A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko,; T.S. Lee, S. L., T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui,; R. Salomon-Ferrer, G. S., C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X.; Kollman, W. a. P. A. *AMBER 14*, University of California, San Francisco: 2014.

40. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23* (3), 327.

41. Mukrasch, M. D.; Bibow, S.; Korukottu, J.; Jeganathan, S.; Biernat, J.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M., Structural Polymorphism of 441-Residue Tau at Single Residue Resolution. *PLoS Biology* **2009,** *7* (2), e1000034.

42. Prokopovich, D. V.; Whittaker, J. W.; Muthee, M. M.; Ahmed, A.; Larini, L., Impact of Phosphorylation and Pseudophosphorylation on the Early Stages of Aggregation of the Microtubule-Associated Protein Tau. *The Journal of Physical Chemistry B* **2017,** *121* (9), 2095.

43. Kolarova, M.; García-Sierra, F.; Bartos, A.; Ricny, J.; Ripova, D., Structure and Pathology of Tau Protein in Alzheimer Disease. *International Journal of Alzheimers Disease* **2012,** *2012*, 1.

44. Avila, J.; Lucas, J. J.; Pérez, M.; Hernández, F., Role of Tau Protein in Both Physiological and Pathological Conditions *Physiological Review* **2004,** *84* (2), 361.

45. Lee, V. M.-Y.; Goedert, M.; Trojanowski, J. Q., Neurodegenerative Tauopathies. *Annual Review of Neuroscience* **2001,** *24* (1), 1121.

46. Kanaan, N. M.; Cox, K.; Alvarez, V. E.; Stein, T. D.; Poncil, S.; McKee, A. C., Characterization of Early Pathological Tau Conformations and Phosphorylation in Chronic Traumatic Encephalopathy. *Journal of Neuropathology and Experimental Neurology* **2016,** *75* (1), 19.

47. Medina, M.; Avila, J., Further understanding of tau phosphorylation: implications for therapy. *Expert Review of Neurotherapeutics* **2015,** *15* (1), 115.

48. Ando, K.; Maruko-Otake, A.; Ohtake, Y.; Hayashishita, M.; Sekiya, M.; Iijima, K. M., Stabilization of Microtubule-Unbound Tau via Tau Phosphorylation at Ser262/356 by Par-1/MARK Contributes to Augmentation of AD-Related Phosphorylation and Aβ42-Induced Tau Toxicity. *PLOS Genetics* **2016,** *12* (3), e1005917.

49. Meyer, V.; Dinkel, P. D.; Luo, Y.; Yu, X.; Wei, G.; Zheng, J.; Eaton, G. R.; Ma, B.; Nussinov, R.; Eaton, S. S.; Margittai, M., Single Mutations in Tau Modulate the Populations of Fibril Conformers through Seed Selection. *Angewandte Chemie International Edition* **2014,** *53* (6), 1590.

50. Uversky, V. N., A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Science* **2013,** *22* (6), 693.

51. Dias, C. L.; Karttunen, M.; Chan, H. S., Hydrophobic interactions in the formation of secondary structures in small peptides. *Physical Review E* **2011,** *84* (4), 041931.

52. Narayanan, C.; Dias, C. L., Hydrophobic interactions and hydrogen bonds in β-sheet formation. *The Journal of Chemical Physics* **2013,** *139* (11), 115103.

53. Jakes, R.; Novak, M.; Davison, M.; Wischik, C. M., Identification of 3- and 4-repeat tau isoforms within the PHF in Alzheimer's disease. *The EMBO Journal* **1991,** *10* (10), 2725.

54. Wischik, C. M.; Novak, M.; Edwards, P. C.; Klug, A.; Tichelaar, W.; Crowther, R. A., Structural characterization of the core of the paired helical filament of Alzheimer disease. *Proceedings of the National Academy of Sciences* **1988,** *85* (13), 4884.

55. Conde, C.; Cáceres, A., Microtubule Assembly, Organization and Dynamics in Axons and Dendrites. *Nature Review Neuroscience* **2009,** *10* (5), 319.

56. Luo, Y.; Ma, B.; Nussinov, R.; Wei, G., Structural Insight into Tau Protein's Paradox of Intrinsically Disordered Behavior, Self-Acetylation Activity, and Aggregation. *The Journal of Physical Chemistry Letters* **2014,** *5* (17), 3026.

57. Qi, R.; Luo, Y.; Wei, G.; Nussinov, R.; Ma, B., Aβ "Stretching-and-Packing" Cross-Seeding Mechanism Can Trigger Tau Protein Aggregation. *The Journal of Physical Chemistry Letters* **2015,** *6* (16), 3276.

58. Stoothoff, W. H.; Johnson, G. V. W., Tau phosphorylation: physiological and pathological consequences. *BBA - Molecular Basis of Disease* **2005,** *1739* (2–3), 280.

59. Noble, W.; Hanger, D. P.; Miller, C. C.; Lovestone, S., The importance of tau phosphorylation for neurodegenerative diseases. *Frontiers in Neurology* **2013,** *4*, 1.

60. Šimić, G.; Babić Leko, M.; Wray, S.; Harrington, C.; Delalle, I.; Jovanov-Milošević, N.; Bažadona, D.; Buée, L.; de Silva, R.; Di Giovanni, G.; Wischik, C.; Hof, P., Tau Protein Hyperphosphorylation and Aggregation in Alzheimer's Disease and Other Tauopathies, and Possible Neuroprotective Strategies. *Biomolecules* **2016,** *6* (1), 1.

61. Schwalbe, M.; Kadavath, H.; Biernat, J.; Ozenne, V.; Blackledge, M.; Mandelkow, E.; Zweckstetter, M., Structural Impact of Tau Phosphorylation at Threonine 231. *Structure* **2015,** *23* (8), 1448.

62. Hanger, D. P.; Anderton, B. H.; Noble, W., Tau Phosphorylation: The Therapeutic Challenge for Neurodegenerative Disease. *Trends in Molecular Medicine* **2009,** *15* (3), 112.

63. Buée, L.; Bussière, T.; Buée-Scherrer, V.; Delacourte, A.; Hof, P. R., Tau Protein Isoforms, Phosphorylation and Role in Neurodegenerative Disorders. *Brain Research Review* **2000,** *33* (1), 95.

64. Drewes, G.; Lichtenberg-Kraag, B.; Döring, F.; Mandelkow, E. M.; Biernat, J.; Goris, J.; Dorée, M.; Mandelkow, E., Mitogen activated protein (MAP) kinase transforms tau protein into an Alzheimer-like state. *The EMBO Journal* **1992,** *11* (6), 2131.

65. Hanger, D. P.; Hughes, K.; Woodgett, J. R.; Brion, J.-P.; Anderton, B. H., Glycogen synthase kinase-3 induces Alzheimer's disease-like phosphorylation of tau: Generation of paired helical filament epitopes and neuronal localisation of the kinase. *Neuroscience Letters* **1992,** *147* (1), 58.

66. Baumann, K.; Mandelkow, E.-M.; Biernat, J.; Piwnica-Worms, H.; Mandelkow, E., Abnormal Alzheimer-like phosphorylation of tau-protein by cyclin-dependent kinases cdk2 and cdk5. *FEBS Letters* **1993,** *336* (3), 417.

67. Köpke, E.; Tung, Y. C.; Shaikh, S.; Alonso, A. C.; Iqbal, K.; Grundke-Iqbal, I., Microtubule-Associated Protein Tau. Abnormal Phosphorylation of a Non-Paired Helical Filament Pool in Alzheimer Disease. *Journal of Biological Chemistry* **1993,** *268* (32), 24374.

68. Kenessey, A.; Yen, S.-H. C., The Extent of Phosphorylation of Fetal Tau is Comparable to that of PHF-Tau from Alzheimer Paired Helical Filaments. *Brain Res.* **1993,** *629* (1), 40.

69. Wang, J.-Z.; Grundke-Iqbal, I.; Iqbal, K., Kinases and phosphatases and tau sites involved in Alzheimer neurofibrillary degeneration. *European Journal of Neuroscience* **2007,** *25* (1), 59.

70. Eidenmüller, J.; Fath, T.; Hellwig, A.; Reed, J.; Sontag, E.; Brandt, R., Structural and Functional Implications of Tau Hyperphosphorylation: Information from Phosphorylation-Mimicking Mutated Tau Proteins. *Biochemistry* **2000,** *39* (43), 13166.

71. Jeganathan, S.; Hascher, A.; Chinnathambi, S.; Biernat, J.; Mandelkow, E. M.; Mandelkow, E., Proline-directed pseudo-phosphorylation at AT8 and PHF1 epitopes induces a compaction of the paperclip folding of Tau and generates a pathological (MC-1) conformation. *Journal of Biological Chemistry* **2008,** *283*.

72. Haase, C.; Stieler, J. T.; Arendt, T.; Holzer, M., Pseudophosphorylation of tau protein alters its ability for self-aggregation. *Journal of Neurochemistry* **2004,** *88* (6), 1509.

73. Fath, T.; Eidenmüller, J.; Brandt, R., Tau-Mediated Cytotoxicity in a Pseudohyperphosphorylation Model of Alzheimer's Disease. *Journal of Neuroscience* **2002,** *22* (22).

74. Léger, J.; Kempf, M.; Lee, G.; Brandt, R., Conversion of Serine to Aspartate Imitates Phosphorylation-induced Changes in the Structure and Function of Microtubule-associated Protein Tau. *Journal of Biological Chemistry* **1997,** *272* (13), 8441.

75. Necula, M.; Kuret, J., Site-specific pseudophosphorylation modulates the rate of tau filament dissociation. *FEBS Letters* **2005,** *579* (6), 1453.

76. Huang, W.; Erikson, R. L., Constitutive activation of Mek1 by mutation of serine phosphorylation sites. *Proceedings of the National Academy of Science U. S. A.* **1994,** *91* (19), 8960.

77. EidenmÜLler, J.; Fath, T.; Maas, T.; Pool, M.; Sontag, E.; Brandt, R., Phosphorylation-mimicking glutamate clusters in the proline-rich region are sufficient to simulate the functional deficiencies of hyperphosphorylated tau protein. *Biochemical Journal* **2001,** *357* (3), 759.

78. Necula, M.; Kuret, J., Pseudophosphorylation and Glycation of Tau Protein Enhance but Do Not Trigger Fibrillization in Vitro. *Journal of Biological Chemistry* **2004,** *279* (48), 49694.

79. Sun, Q.; Gamblin, T. C., Pseudohyperphosphorylation Causing AD-like Changes in Tau Has Significant Effects on Its Polymerization. *Biochemistry* **2009,** *48* (25), 6002.

80. Rankin, C. A.; Sun, Q.; Gamblin, T. C., Pseudo-phosphorylation of tau at Ser202 and Thr205 affects tau filament formation. *Molecular Brain Research* **2005,** *138* (1), 84.

81. Fischer, D.; Mukrasch, M. D.; Biernat, J.; Bibow, S.; Blackledge, M.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M., Conformational Changes Specific for Pseudophosphorylation at Serine 262 Selectively Impair Binding of Tau to Microtubules. *Biochemistry* **2009,** *48* (42), 10047.

82. Ding, H.; Matthews, T. A.; Johnson, G. V. W., Site-specific Phosphorylation and Caspase Cleavage Differentially Impact Tau-Microtubule Interactions and Tau Aggregation. *Journal of Biological Chemistry* **2006,** *281* (28), 19107.

83. Combs, B.; Voss, K.; Gamblin, T., Pseudohyperphosphorylation has differential effects on polymerization and function of tau isoforms. *Biochemistry* **2011,** *50* (44), 9446.

84. Bibow, S.; Ozenne, V.; Biernat, J.; Blackledge, M.; Mandelkow, E.; Zweckstetter, M., Structural Impact of Proline-Directed Pseudophosphorylation at AT8, AT100, and PHF1 Epitopes on 441-Residue Tau. *Journal of the American Chemical Society* **2011,** *133* (40), 15842.

85. von Bergen, M.; Friedhoff, P.; Biernat, J.; Heberle, J.; Mandelkow, E.-M.; Mandelkow, E., Assembly of τ protein into Alzheimer paired helical filaments depends on a local sequence motif (306VQIVYK311) forming β structure. *Proceedings of the National Academy of Sciences U.S.A.* **2000,** *97* (10), 5129.

86. Mendieta, J.; Fuertes, M. A.; Kunjishapatham, R.; Santa-María, I.; Moreno, F. J.; Alonso, C.; Gago, F.; Muñoz, V.; Avila, J.; Hernández, F., Phosphorylation modulates the alpha-helical structure and polymerization of a peptide from the third tau microtubule-binding repeat. *Biochim. Biophys. Acta - Gen. Subj.* **2005,** *1721* (1-3), 16.

87. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerruti, T. E.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Javonowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, J. C.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A., *AMBER 14*. University of California: San Francisco, 2014.

88. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179.

89. Homeyer, N.; Horn, A. C.; Lanig, H.; Sticht, H., AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *Journal of Molecular Modeling* **2006,** *12* (3), 281.

90. Serio, T. R.; Cashikar, A. G.; Kowal, A. S.; Sawicki, G. J.; Moslehi, J. J.; Serpell, L.; Arnsdorf, M. F.; Lindquist, S. L., Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant. *Science* **2000,** *289* (5483), 1317.

91. Cheon, M.; Chang, I.; Mohanty, S.; Luheshi, L. M.; Dobson, C. M.; Vendruscolo, M.; Favrin, G., Structural Reorganisation and Potential Toxicity of Oligomeric Species Formed during the Assembly of Amyloid Fibrils. *PLoS Computational Biology* **2007,** *3* (9), e173.

92. Esler, W. P.; Stimson, E. R.; Jennings, J. M.; Vinters, H. V.; Ghilardi, J. R.; Lee, J. P.; Mantyh, P. W.; Maggio, J. E., Alzheimer's Disease Amyloid Propagation by a Template-Dependent Dock-Lock Mechanism. *Biochemistry* **2000,** *39* (21), 6288.

93. O'Brien, E. P.; Okamoto, Y.; Straub, J. E.; Brooks, B. R.; Thirumalai, D., Thermodynamic Perspective on the Dock−Lock Growth Mechanism of Amyloid Fibrils. *The Journal of Physical Chemistry B* **2009,** *113* (43), 14421.

94. Massi, F.; Straub, J. E., Energy landscape theory for Alzheimer's amyloid β-peptide fibril elongation. *Proteins: Structure, Function, and Bioinformatics* **2001,** *42* (2), 217.

95. Straub, J. E.; Thirumalai, D., Toward a Molecular Theory of Early and Late Events in Monomer to Amyloid Fibril Formation. *Annual Review of Physical Chemistry* **2011,** *62* (1), 437.

96. Larini, L.; Shea, J.-E., Role of β-Hairpin Formation in Aggregation: The Self-Assembly of the Amyloid-β(25–35) Peptide. *Biophysical Journal* **2012,** *103* (3), 576.

97. Vergara, C.; Ordóñez-Gutiérrez, L.; Wandosell, F.; Ferrer, I.; del Río, J. A.; Gavín, R., Role of PrPC Expression in Tau Protein Levels and Phosphorylation in Alzheimer's Disease Evolution. *Molecular Neurobiology* **2015,** *51* (3), 1206.

98. Lasagna-Reeves, C. A.; Castillo-Carranza, D. L.; Sengupta, U.; Guerrero-Munoz, M. J.; Kiritoshi, T.; Neugebauer, V.; Jackson, G. R.; Kayed, R., Alzheimer brain-derived tau oligomers propagate pathology from endogenous tau. *Scientific Reports* **2012,** *2*, 700.

99. Holmes, B. B.; Diamond, M. I., Prion-like Properties of Tau Protein: The Importance of Extracellular Tau as a Therapeutic Target. *Journal of Biological Chemistry* **2014,** *289* (29), 19855.

100. Larini, L.; Gessel, M. M.; LaPointe, N. E.; Do, T. D.; Bowers, M. T.; Feinstein, S. C.; Shea, J.-E., Initiation of assembly of tau(273-284) and its ΔK280 mutant: an experimental and computational study. *Physical Chemistry Chemical Physics* **2013,** *15* (23), 8916.

101. Abdul Khalil, H. P. S.; Davoudpour, Y.; Islam, M. N.; Mustapha, A.; Sudesh, K.; Dungani, R.; Jawaid, M., Production and modification of nanofibrillated cellulose using various mechanical processes: A review. *Carbohydrate Polymers* **2014,** *99*, 649.

102. Hamed, I.; Özogul, F.; Regenstein, J. M., Industrial applications of crustacean by-products (chitin, chitosan, and chitooligosaccharides): A review. *Trends in Food Science & Technology* **2016,** *48*, 40.

103. Ha, T. L. B.; Quan, T. M.; Vu, D.; Si, D.; Andrades, J., *Naturally derived biomaterials: Preparation and application*. InTech: 2013; p 247.

104. Ling, S.; Jin, K.; Kaplan, D. L.; Buehler, M. J., Ultrathin Free-Standing Bombyx mori Silk Nanofibril Membranes. *Nano Letters* **2016,** *16* (6), 3795.

105. Lewis, A.; Waters, J. C.; Stanton, J.; Hess, J.; Salas-de la Cruz, D., Macromolecular Interactions Control Structural and Thermal Properties of Regenerated Tri-Component Blended Films. *International Journal of Molecular Sciences* **2016,** *17* (12), 1989.

106. Silva, S. S.; Santos, T. C.; Cerqueira, M. T.; Marques, A. P.; Reys, L. L.; Silva, T. H.; Caridade, S. G.; Mano, J. F.; Reis, R. L., The use of ionic liquids in the processing of chitosan/silk hydrogels for biomedical applications. *Green Chemistry* **2012,** *14* (5), 1463.

107. Jin, J.; Hassanzadeh, P.; Perotto, G.; Sun, W.; Brenckle, M. A.; Kaplan, D.; Omenetto, F. G.; Rolandi, M., A Biomimetic Composite from Solution Self-Assembly of Chitin Nanofibers in a Silk Fibroin Matrix. *Advanced Materials* **2013,** *25* (32), 4482.

108. Sharma, C.; Dinda, A. K.; Potdar, P. D.; Mishra, N. C., Fabrication of quaternary composite scaffold from silk fibroin, chitosan, gelatin, and alginate for skin regeneration. *Journal of Applied Polymer Science* **2015,** *132* (44), 1.

109. Fernandes, L. L.; Resende, C. X.; Tavares, D. S.; Soares, G. A.; Castro, L. O.; Granjeiro, J. M., Cytocompatibility of chitosan and collagen-chitosan scaffolds for tissue engineering. *Polimeros* **2011,** *21* (1), 1.

110. Tian, D.; Li, T.; Zhang, R.; Wu, Q.; Chen, T.; Sun, P.; Ramamoorthy, A., Conformations and Intermolecular Interactions in Cellulose/Silk Fibroin Blend Films: A Solid-State NMR Perspective. *The Journal of Physical Chemistry B* **2017,** *121* (25), 6108.

111. Jin, J.; Hassanzadeh, P.; Perotto, G.; Sun, W.; Brenckle, M. A.; Kaplan, D.; Omenetto, F. G.; Rolandi, M., A Biomimetic Composite from Solution Self-Assembly of Chitin Nanofibers in a Silk Fibroin Matrix. *Advanced Materials* **2013,** *25* (32), 4482.

112. Bhardwaj, N.; Kundu, S. C., Silk fibroin protein and chitosan polyelectrolyte complex porous scaffolds for tissue engineering applications. *Carbohydrate Polymers* **2011,** *85* (2), 325.

113. Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977,** *23* (2), 187.

114. Patey, G. N.; Valleau, J. P., The free energy of spheres with dipoles: Monte Carlo with multistage sampling. *Chemical Physics Letters* **1973,** *21* (2), 297.

115. Torrie, G. M.; Valleau, J. P., Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chemical Physics Letters* **1974,** *28* (4), 578.

116. Mohd, N.; Draman, S. F. S.; Salleh, M. S. N.; Yusof, N. B., Dissolution of cellulose in ionic liquid: A review. *AIP Conference Proceedings* **2017,** *1809* (1), 020035.

117. Mäki-Arvela, P.; Anugwom, I.; Virtanen, P.; Sjöholm, R.; Mikkola, J. P., Dissolution of lignocellulosic materials and its constituents using ionic liquids—A review. *Industrial Crops and Products* **2010,** *32* (3), 175.

118. Kirkwood, J. G., Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics* **1935,** *3* (5), 300.

119. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992,** *13* (8), 1011.

120. Lemkul, J. A.; Bevan, D. R., Assessing the Stability of Alzheimer's Amyloid Protofibrils Using Molecular Dynamics. *The Journal of Physical Chemistry B* **2010,** *114* (4), 1652.

121. Woods Group GLYCAM Web. http://glycam.org.

122. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996,** *14* (1), 33.

123. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **2006,** *25* (2), 247.

124. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004,** *25* (9), 1157.

125. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1–2*, 19.

126. Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E., Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In *Solving Software Challenges for Exascale: International Conference on Exascale Applications and Software, EASC 2014, Stockholm, Sweden, April 2-3, 2014, Revised Selected Papers*, Markidis, S.; Laure, E., Eds. Springer International Publishing: Cham, 2015; pp 3.

127. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29* (7), 845.

128. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry* **2008,** *29* (4), 622.

129. Darden, T.; York, D.; L. Pedersen, L., Particle mesh Ewald: An N • log(N) method for Ewald sums in large systems. *Journal of Chemical Physics* **1993,** *98*, 10089.

130. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics* **1995,** *103* (19), 8577.

131. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984,** *81* (8), 3684.

132. Hoover, W. G., Canonical Dynamics: Equilibrium Phase-Space Distributions. *Physical Review A* **1985,** *31*, 1695.

133. Nosé, S., A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Molecular Physics* **1984,** *52*, 255.

134. Köll, P.; Borchers, G.; Metzger, J., Thermal degradation of chitin and cellulose. *Journal of Analytical and Applied Pyrolysis* **1991,** *19*, 119.

135. Hu, X.; Kaplan, D.; Cebe, P., Determining Beta-Sheet Crystallinity in Fibrous Proteins by Thermal Analysis and Infrared Spectroscopy. *Macromolecules* **2006,** *39* (18), 6161.

136. Henriques, J.; Cragnell, C.; Skepö, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *Journal of Chemical Theory and Computation* **2015,** *11* (7), 3420.

137. Fu, J.; Larini, L.; Cooper, A. J.; Whittaker, J. W.; Ahmed, A.; Dong, J.; Lee, M.; Zhang, T., Computational and experimental analysis of short peptide motifs for enzyme inhibition. *PLOS ONE* **2017,** *12* (8), e0182847.

138. Jørgensen, K.; Rasmussen, A. V.; Morant, M.; Nielsen, A. H.; Bjarnholt, N.; Zagrobelny, M.; Bak, S.; Møller, B. L., Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Current Opinion in Plant Biology* **2005,** *8* (3), 280.

139. Voet, D.; Voet, J. G., Biochemistry, 4-th Edition. *NewYork: John Wiley& SonsInc* **2011**, 492.

140. Drews, J., Drug Discovery: A Historical Perspective. *Science* **2000,** *287* (5460), 1960.

141. Gordon, G. H.; Wu, C.-W., Regulation of Enzyme Activity. *Science* **1971,** *172* (3989), 1205.

142. Roberts, R. W.; Szostak, J. W., RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences* **1997,** *94* (23), 12297.

143. Smith, G., Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **1985,** *228* (4705), 1315.

144. Liu, T.; Qian, Z.; Xiao, Q.; Pei, D., High-Throughput Screening of One-Bead-One-Compound Libraries: Identification of Cyclic Peptidyl Inhibitors against Calcineurin/NFAT Interaction. *ACS Combinatorial Science* **2011,** *13* (5), 537.

145. Guntas, G.; Mansell, T. J.; Kim, J. R.; Ostermeier, M., Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (32), 11224.

146. Blind, M.; Blank, M., Aptamer Selection Technology and Recent Advances. *Molecular Therapy - Nucleic Acids* **2015,** *4*, e223.

147. Murray, C. W.; Rees, D. C., The rise of fragment-based drug discovery. *Nature Chemistry* **2009,** *1* (3), 187.

148. Vegas, Arturo J.; Fuller, J. H.; Koehler, Angela N., Small-molecule microarrays as tools in ligand discovery. *Chemical Society Review* **2008,** *37*, 1385.

149. Schreiber, Stuart L., The small-molecule approach to biology. *Chemical Engineering News* **2003,** *81*, 51.

150. Fu, J.; Cai, K.; Johnston, S. A.; Woodbury, N. W., Exploring Peptide Space for Enzyme Modulators. *Journal of the American Chemical Society* **2010,** *132* (18), 6419.

151. Troitskaya, L. A.; Kodadek, T., Peptides as modulators of enzymes and regulatory proteins. *Methods* **2004,** *32* (4), 406.

152. Devlin, J.; Panganiban, L.; Devlin, P., Random peptide libraries: a source of specific protein binding molecules. *Science* **1990,** *249* (4967), 404.

153. Fodor, S.; Read, J.; Pirrung, M.; Stryer, L.; Lu, A.; Solas, D., Light-directed, spatially addressable parallel chemical synthesis. *Science* **1991,** *251* (4995), 767.

154. Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D., The Future of Peptide-based Drugs. *Chemical Biology & Drug Design* **2013,** *81* (1), 136.

155. Xu, H.; Ewing, A. G., A rapid enzyme assay for β-galactosidase using optically gated sample introduction on a microfabricated chip. *Analytical and Bioanalytical Chemistry* **2004,** *378* (7), 1710.

156. Onufriev, A.; Bashford, D.; Case, D. A., Modification of the generalized Born model suitable for macromolecules. *Journal of Physical Chemistry B* **2000,** *104* (15), 3712.

157. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999,** *314*, 141.

158. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23* (3), 327.

159. Sontag, E. M.; Vonk, W. I. M.; Frydman, J., Sorting out the trash: the spatial nature of eukaryotic protein quality control. *Current opinion in cell biology* **2014,** *26*, 139.

160. Hurtley, S., Location, Location, Location. *Science* **2009,** *326* (5957), 1205.

161. Chen, A. H.; Silver, P. A., Designing biological compartmentalization. *Trends in Cell Biology* **2012,** *22* (12), 662.

162. Shorter, J., Phasing in and out. *Nature Chemistry* **2016,** *8*, 528.

163. Brangwynne, Clifford P.; Tompa, P.; Pappu, Rohit V., Polymer physics of intracellular phase transitions. *Nature Physics* **2015,** *11*, 899.

164. Brangwynne, C. P., Phase transitions and size scaling of membrane-less organelles. *The Journal of Cell Biology* **2013,** *203* (6), 875.

165. Guo, L.; Shorter, J., It's Raining Liquids: RNA Tunes Viscoelasticity and Dynamics of Membraneless Organelles. *Molecular Cell* **2015,** *60* (2), 189.

166. Zhang, Y. H. P., Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnology Advances* **2011,** *29* (6), 715.

167. Dunn, M. F., Allosteric Regulation of Substrate Channeling and Catalysis in the Tryptophan Synthase Bienzyme Complex. *Archives of biochemistry and biophysics* **2012,** *519* (2), 154.

168. Xu, Y.; Abeles, R. H., Inhibition of tryptophan synthase by (1-fluorovinyl)glycine. *Biochemistry* **1993,** *32* (3), 806.

169. Pierse, C. A.; Dudko, O. K., Distinguishing Signatures of Multipathway Conformational Transitions. *Physical Review Letters* **2017,** *118* (8), 088101.

170. Shoup, D.; Szabo, A., Role of diffusion in ligand binding to macromolecules and cell-bound receptors. *Biophysical Journal* **1982,** *40* (1), 33.

171. Dudko, O. K.; Graham, T. G. W.; Best, R. B., Locating the Barrier for Folding of Single Molecules under an External Force. *Physical Review Letters* **2011,** *107* (20), 208301.

172. Chin, S. L.; Lu, Q.; Dane, E. L.; Dominguez, L.; McKnight, C. J.; Straub, J. E.; Grinstaff, M. W., Combined Molecular Dynamics Simulations and Experimental Studies of the Structure and Dynamics of Poly-Amido-Saccharides. *Journal of the American Chemical Society* **2016,** *138* (20), 6532.

173. Małolepsza, E.; Straub, J. E., Empirical Maps For The Calculation of Amide I Vibrational Spectra of Proteins From Classical Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2014,** *118* (28), 7848.

174. Su, Z.; Mahmoudinobar, F.; Dias, C. L., Effects of Trimethylamine-N-oxide on the Conformation of Peptides and its Implications for Proteins. *Physical Review Letters* **2017,** *119* (10), 108102.

175. Levine, Z. A.; Larini, L.; LaPointe, N. E.; Feinstein, S. C.; Shea, J.-E., Regulation and aggregation of intrinsically disordered peptides. *Proceedings of the National Academy of Sciences* **2015,** *112* (9), 2758.

176. Douglas, S. M.; Dietz, H.; Liedl, T.; Högberg, B.; Graf, F.; Shih, W. M., Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **2009,** *459*, 414.

177. Zhao, Z.; Fu, J.; Dhakal, S.; Johnson-Buck, A.; Liu, M.; Zhang, T.; Woodbury, N. W.; Liu, Y.; Walter, N. G.; Yan, H., Nanocaged enzymes with enhanced catalytic activity and increased stability against protease digestion. *Nature Communications* **2016,** *7*, 10619.

178. Pal, S. K.; Zewail, A. H., Dynamics of Water in Biological Recognition. *Chemical Reviews* **2004,** *104* (4), 2099.

179. Kim, J.; Grate, J. W.; Wang, P., Nanostructures for enzyme stabilization. *Chemical Engineering Science* **2006,** *61* (3), 1017.

180. Tagliazucchi, M.; Szleifer, I., How Does Confinement Change Ligand–Receptor Binding Equilibrium? Protein Binding in Nanopores and Nanochannels. *Journal of the American Chemical Society* **2015,** *137* (39), 12539.

181. Küchler, A.; Yoshimoto, M.; Luginbühl, S.; Mavelli, F.; Walde, P., Enzymatic reactions in confined environments. *Nature Nanotechnology* **2016,** *11*, 409.

182. Martinez, A. V.; Małolepsza, E.; Rivera, E.; Lu, Q.; Straub, J. E., Exploring the role of hydration and confinement in the aggregation of amyloidogenic peptides Aβ(16−22) and Sup35(7−13) in AOT reverse micelles. *The Journal of Chemical Physics* **2014,** *141* (22), 22D530.

183. Simone, N.; Emmanouil, G.; Nicholas, B. T., Glass transition of polymers in bulk, confined geometries, and near interfaces. *Reports on Progress in Physics* **2017,** *80* (3), 036602.

184. Krishtalik, L. I.; Topolev, V. V., Effects of medium polarization and pre-existing field on activation energy of enzymatic charge-transfer reactions. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2000,** *1459* (1), 88.

185. Kharkats, Y. I.; Krishtalik, L. I., Medium reorganization energy and enzymatic reaction activation energy. *Journal of Theoretical Biology* **1985,** *112* (2), 221.

186. Smith, P. E., Molecular Dynamics Simulations of NAD+ in Solution. *Journal of the American Chemical Society* **1999,** *121* (37), 8637.

187. Burchard, W., Solution Properties of Branched Macromolecules. In *Branched Polymers II*, Roovers, J., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 1999; pp 113.

188. Fu, B.; Curry, F.-R. E.; Adamson, R. H.; Weinbaum, S., A model for interpreting the tracer labeling of interendothelial clefts. *Annals of Biomedical Engineering* **1997,** *25* (2), 375.

189. Di Fusco, M.; Favero, G.; Mazzei, F., Polyazetidine-Coated Microelectrodes: Electrochemical and Diffusion Characterization of Different Redox Substrates. *The Journal of Physical Chemistry B* **2011,** *115* (5), 972.

190. Kagawa, T.; Bruno, P. L., NADP-malate dehydrogenase from leaves of Zea mays: Purification and physical, chemical, and kinetic properties. *Archives of biochemistry and biophysics* **1988,** *260* (2), 674.

191. Rothe, G. M., *Electrophoresis*. 1988; Vol. 9.

192. Andersson, M. M.; Breccia, J. D.; Hatti-Kaul, R., Stabilizing effect of chemical additives against oxidation of lactate dehydrogenase. *Biotechnology and Applied Biochemistry* **2000,** *32* (3), 145.

193. Shenoy, R.; Bowman, C. N., Kinetics of interfacial radical polymerization initiated by a glucose-oxidase mediated redox system. *Biomaterials* **2012,** *33* (29), 6909.

194. Sabek, O. M.; Ferrati, S.; Fraga, D. W.; Sih, J.; Zabre, E. V.; Fine, D. H.; Ferrari, M.; Gaber, A. O.; Grattoni, A., Characterization of a nanogland for the autotransplantation of human pancreatic islets. *Lab on a Chip* **2013,** *13* (18), 3675.

195. Dessinges, M. N.; Maier, B.; Zhang, Y.; Peliti, M.; Bensimon, D.; Croquette, V., Stretching Single Stranded DNA, a Model Polyelectrolyte. *Physical Review Letters* **2002,** *89* (24), 248102.

196. Munns, R.; Tester, M., Mechanisms of Salinity Tolerance. *Annual Review of Plant Biology* **2008,** *59* (1), 651.

197. Truesdell, A. H., Activity Coefficients of Aqueous Sodium Chloride from 15° to 50°C Measured with a Glass Electrode. *Science* **1968,** *161* (3844), 884.

198. Zhao, H., Effect of ions and other compatible solutes on enzyme activity, and its implication for biocatalysis using ionic liquids. *Journal of Molecular Catalysis B: Enzymatic* **2005,** *37* (1), 16.

199. Greenway, H.; Osmond, C. B., Salt Responses of Enzymes from Species Differing in Salt Tolerance. *Plant Physiology* **1972,** *49* (2), 256.

200. Miller, S. L., A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **1953,** *117* (3046), 528.

201. Cech, T. R., Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Bioscience Reports* **1990,** *10* (3), 239.

202. Higgs, P. G.; Lehman, N., The RNA World: molecular cooperation at the origins of life. *Nature Reviews Genetics* **2014,** *16*, 7.

203. Amir, A., Cell Size Regulation in Bacteria. *Physical Review Letters* **2014,** *112* (20), 208102.

204. Golding, I.; Cox, E. C., Physical Nature of Bacterial Cytoplasm. *Physical Review Letters* **2006,** *96* (9), 098102.

205. Kachman, T.; Owen, J. A.; England, J. L., Self-Organized Resonance during Search of a Diverse Chemical Space. *Physical Review Letters* **2017,** *119* (3), 038001.