DISSERTATION

DETERMINING DRIVING FORCES FOR SMALL MOLECULE AGGREGATION USING COMPUTATIONAL AND THEORETICAL METHODS

Submitted by

Jakob Edward Anderson

Department of Chemistry

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2022

Doctoral Committee:

Advisor: Anthony Rappé
Co-Advisor: Martin McCullagh

Alan Kennan
Eugene Chen
Patrick Shipman

ABSTRACT

DETERMINING DRIVING FORCES FOR SMALL MOLECULE AGGREGATION USING COMPUTATIONAL AND THEORETICAL METHODS

Molecular aggregation is largely dictated by noncovalent interactions and is a phenomenon found in a broad list of disciplines. Computational and theoretical methods, such as molecular dynamics simulations and Quantum Mechanical calculations, are well suited techniques to study the noncovalent association of various systems as they provide atomistic resolution and experimentally comparable results for the timescales on which association occurs. The studies found in this dissertation are introduced in the first chapter and are put in the context of using computational methods to study the noncovalent association and aggregation of small molecules. Chapters two, three, and four provide a foundation for the rational design of dipeptides for a given application. A wide range of potential applications for diphenylalanine (FF) have been proposed which would benefit from the development of design principles. Chapter two discusses the complexity of the noncovalent interactions at multiple stages in the FF self-assembly process. Specifically, we suggest the initial aggregation of FF is predominantly driven by electrostatics, and after a reorientation event, nanotube growth is suggested to be driven by solvent mediated forces. The results from this chapter use an array of generalized analyses enabling quantitative comparisons to future dipeptide studies. The impact of sidechain modification for either FF residue is studied in chapter three by considering valine-phenylalanine (VF) and phenylalanine-valine (FV). While the monomeric conformations are shown to sample the same states for these two dipeptides, the probabilities for state sampling as well as the water dynamics around the peptide bond are shown to differ. Chapter four connects chapters two and three by considering both the behavior of sequence dependence and dimerization of VF, FV, isoleucine-phenylalanine, and phenylalanine-isoleucine relative to that of FF. The modification of the C-terminus of FF to a smaller hydrophobic sidechain is hypothesized to

enable tighter packing from this study. Additionally, N-terminus FF modification is hypothesized to increase the solvent mediated forces during dimerization in agreement with the results from chapter three. While not a completed study, chapter four provides a foundation for the continued development of design principles for FF-derivatives. A novel approach to computing the free energy of association from Quantum Mechanical calculations is then described in chapter five. Due to the treatment of low energy frequencies as harmonic and a lack of temperature dependence, calculations of the entropy of associating molecules is inaccurate. The rigid-rotor-Gaussian-oscillator approximation proposed addresses these issues by treating low lying modes with anharmonic Gaussian potentials and wave functions as well as adding a temperature dependence to the partitioning between vibrational and translational/rotational modes. This approximation significantly reduces the error in computing the entropy of associating molecules resulting in a more accurate calculation of the total free energy. The results from these studies as well as future studies based on the work in this dissertation are then summarized in the final chapter.

ACKNOWLEDGEMENTS

I would like to thank my advisors Martin McCullagh and Anthony Rappé for their guidance during my time in graduate school. Martin your ambitious optimism and determination enabled me to persevere through disappointments and continue to strive towards achievement. Tony your patience, willingness to listen, and sage advice gave me the freedom to explore my scientific curiosities.

I would also like to thank my immediate family Kirk Anderson, Debi Olson, Dave Olson, and Ali Olson as well as my grandparents Bud and Barbara Toebbe. My family has always been a source of support and encouragement without which would have made graduate school much more difficult. I especially wanted to thank them for instilling in me the value of academia and motivating me to continue to read despite my early childhood rejections.

Although graduate school definitely had its difficulties, the friendships I've made enabled me to maintain some semblance of mental stability. Specifically, many thanks are given to Ryan Weber, Kelly Du Pont, Russell Davidson, Peter (Rex) Lake, Heidi Klem, Mortaza Derakshani, Matthew Lacroix, as well as Dan and Allie Mahar. My friends were always there to provide a laugh and an environment conducive to stimulating my imagination whether it was through DnD, wine nights, board games, or just hanging out. I count these friends a blessing and know that regardless of location or stage in life that I will be able to call them friends.

Finally yet importantly, I would like to thank my beautiful and loving wife, Rae Anderson. While she could have also been included in the previous paragraph, Rae I especially appreciate you being the bestest of friends to me such that you deserve your own paragraph.

DEDICATION

I would like to dedicate this dissertation to the people that have taught me humility, the limits of knowledge, and the power of imagination.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Small Molecule Aggregation

The underlying forces of noncovalently interacting molecular species are a combination of dispersion and electrostatic forces between independent molecules. Processes which involve noncovalent interactions are ubiquitous in nature and include water uptake by plants, higher order structure of biomolecules, enzyme–substrate binding, the self-assembly of nanomaterials, and chemical synthesis. Molecular aggregation is the process for a number of independent molecules to form a supramolecular structure or aggregate through noncovalent interactions. Some examples of aggregates include viral capsids,[Cayman2019, 1] amyloid proteins,[3–5] peptide nanostructures,[6–8] and surfactants[9–11]. Aggregation is not only dependent on the molecular makeup of the solute but is also dependent on the choice of solvent,[12,13] temperature,[14,15] and concentration[16–18] of the system. There is not a general theory (e.g. classical nucleation theory)[19] which describes the mechanism for all aggregation in part due to the complexity of dependent variables and the range of degrees of noncovalent interactions. Therefore, to fully elucidate the aggregation behavior of a system of interest, focused study of the underlying noncovalent interactions giving rise to the aggregate is required. Additionally, by understanding the physiochemical interactions driving aggregation of specific molecules, one can predict the impact modification would have on the aggregated structure so as to be optimized for a given activity.

Computational and theoretical methods have been transformative in the study of noncovalent interactions as they enable atomistic resolution and direct measurement of molecular interactions. In molecular dynamics (MD) simulations, these advantages are extended to include time resolution as atom and molecule parameters found in input files are used to numerically solve Newton's equations of motion. Furthermore, MD enables the study of molecular conformational phase space as the parameters used are a combination of bonded (bond, angle, and dihedral) potentials and

non-bonded (van der Waals and Coulomb) potentials, thus modeling a non-stagnant visualization of natural systems. Assuming ergodic sampling, analysis of these simulations can provide experimental comparisons such as conformational or interaction lifetimes and relative association or conformational free energies. While the computational cost has been reduced significantly in recent years, most studies using MD simulations are limited to nm-length and μs-time scales for systems with fewer than five hundred thousand atoms. The parameters used for MD simulations are often determined using quantum mechanical (QM) calculations which give experimentally comparable free energies of individual molecules and accurate interaction energies for associating molecules.[20] With the advent and subsequent refinement of density functional theory (DFT), QM calculations are now routinely performed on systems on the order of hundreds of atoms. QM computations provide insight into small molecule association by providing energetically optimized molecular structures, experimentally comparable IR spectra, and often the parameters to perform MD simulations. Utilizing both MD simulations and QM calculations, the noncovalent interactions resulting in the aggregation of small molecules can be observed with atomic resolution and described with experimentally confirmable analyses. In doing so, the non-triviality of small molecule aggregation can be detailed using fundamental physics. Principles can then be formulated to rationally design optimized molecules for a given application.

1.2 Chapter Overview

The subsequent chapters of this dissertation contain the results from research I performed using a combination of MD simulations and QM calculations in the pursuit of understanding the underlying physics of noncovalent association and aggregation for small molecules. Chapters 2, 3, and 4 focus on understanding the driving forces for dipeptide aggregation and self-assembly. Specifically, work presented in Chapter 2 considers the driving forces for the self-assembly of diphenylalanine (FF) during dimerization, initial aggregation, and monomeric addition to a FF nanotube. The solvent structure and dynamics around valine-phenylalanine (VF) and phenylalanine-valine (FV) are investigated in Chapter 3 while the comparative dimerization propensities of FF, VF, FV,

isoleucine-phenylalanine (IF), and phenylalanine-isoleucine (FI) are compared in Chapter 4. A novel approach to computing the entropy for dimerization in a more rigorous manner that enables prediction of the total free energy of associating molecules computed using QM calculations is described in Chapter 5. The final chapter discusses overarching conclusions from the research presented in prior chapters as well as additional avenues for future research.

1.2.1 Design Principles for Dipeptide Self-Assembly

The self-assembly of FF has been suggested to have a wide variety of applications due to the stability and order of the macroscopically observed structures.[21–25] While the potential applications of FF nanostructures are numerous, having a concrete understanding of the noncovalent interactions driving FF dipeptide self-assembly is crucial in order to rationally optimize the activity of a FF-derivative for a given application. Previous experimental studies suggest that the hydrophobicity of the phenylalanine residues drives the self-assembly process of FF,[26] while other studies conclude electrostatic interactions drive FF self-assembly.[27] Computational and theoretical investigations have also resulted in conflicting hypotheses for which properties of FF promote the aggregation and ordering of FF.[28–32] Given the lack of consensus and the widespread potential applications, FF was investigated at varying points of the self-assembly process in Chapter 2. Using a combination of MD simulations and QM calculations, a variety of different analyses were performed in a generalized way. Although we demonstrate FF aggregation is initially dominated by electrostatic interactions, FF nanotube growth is also suggested to be dominated by solvent dispersion interactions. The methods and analyses from this study provide a foundation for characterizing the noncovalent interactions of associating FF derivatives relative to FF, from which a set of design principles can be developed.

Developing design principles for FF-derivatives requires not only determination of the noncovalent interactions that dictate the FF self-assembly process, but also explication of the affect for other tunable variables. Although solvent choice is an important variable to be conscious of,[12,13,16,33–36] chemical modification of FF significantly impacts the properties of, and propensity

for, self-assembled dipeptide nanostructures.[32] Residue substitution is a common tool used to investigate proteins as changing an amino acid in critical motifs can affect the activity of the protein. Due to the molecular size of a dipeptide, there is a dipeptide sequence dependence effect resulting from the amino acid stereocenters as well as the N- and C-termini distinction. While not surprising, the difference in noncovalent interactions giving rise to this sequence dependence is not well understood in the literature. As it has been shown that there is an aggregation propensity sequence dependence for dipeptides which contain one phenylalanine residue and either a valine, isoleucine, or tryptophan residue, these dipeptides are interesting initial targets for study.[26,32,37,38] In Chapter 3 it was determined that differences in monomeric conformational state sampling of VF and FV contributed to longer lived hydrogen bond interactions of waters to the carbonyl of the amide bond for VF relative to FV. This was determined by using both QM calculations and MD simulations combined with generalized analyses which enable direct comparisons to any dipeptide system. Our conclusions from Chapter 3 were supported by IR and 2D IR results from Amber Krummel and Christopher Kuhs. The full details in elucidating the noncovalent interaction differences between VF and FV and resulting hypotheses for FF modification are contained in Chapter 3.

A bottom-up approach to developing design principles for the self-assembly of a FF-derivative requires comparisons of the noncovalent interactions between systems where as few sources of change are modified as possible. While noncovalent interactions between the solvent and the dipeptide differ between monomers of VF and FV, description of noncovalent interactions during initial aggregation is still lacking. Additionally, the relative hydrophobicity and steric sidechain size is different between VF/FV and FF making it difficult to separate the origin of differences between the dipeptides. The initial aggregation, impact of sidechain steric size, and a trend in sidechain hydrophobicity can be considered by comparing the dimerization of VF, FV, IF, and FI relative to the dimerization of FF. In Chapter 4 it is shown that dimerization for each of the dipeptides considered is predominantly driven by electrostatic interactions. It is further hypothesized that FF modification of the C-terminal residue to another hydrophobic residue will increase the binding free energy relative to N-terminal substitution. Although not a completed study, Chapter 4

provides a foundation from which the noncovalent interactions of larger system sizes or the impact of solvent choice will be considered, enabling further development of design principles for FF-derivatives.

1.2.2 A More Rigorous Approach to Computing Association Free Energies from QM Calculations

The binding free energy of associating molecules can be described as the energetic mean of the sum of noncovalent interactions acting on the system. The free energy of association is a useful quantity for drug design, MD simulation force field development, and for predicting mechanistic pathways. Current QM calculations accurately compute a systems potential energy, however, the lowest potential energy state does not imply the lowest free energy state and is thus not necessarily reflective of a system *in situ*. Although QM calculations have become an important tool for scientific study, an important and often overlooked assumption of QM calculations is that the computed thermodynamic properties are only valid for systems with non-interacting particles. The limitation to ideal gas systems is generally agreed to originate in the determination of the entropy. This has led to a variety of different approaches to computing the entropy,[39–54] most popular of which utilizing the rigid-rotor-harmonic-oscillator approximation.[55–59] Nevertheless, when the six intermolecular degrees of freedom of each molecule are converted to six degrees of freedom of the complex during association, they are often low energy vibrational modes. Due to the harmonic approximation being notoriously inaccurate for low lying frequencies, frequencies less than 300 cm^{-1} are often not considered in determining the entropy, resulting in an underestimate of the systems entropy. In Chapter 5, a novel approximation which uses a rigid-rotor-Gaussian-oscillator is developed and utilized on a set of associating molecules. This approximation is non-harmonic and thus all frequencies are considered, resulting in association free energies which are in better agreement to experimental values. This chapter has broad applications in many fields and enables further study of the potential noncovalent interactions within a system.

Chapter 2

The Initial Aggregation and Ordering Mechanism of Diphenylalanine From Microsecond All-Atom Molecular Dynamics Simulations.[1]

2.1 Overview

Self-assembled diphenylalanine (FF) nanostructures have recently been demonstrated to be interesting materials for antibacterial and anticancer applications. These applications, among others, seek to take advantage of the high-order and resulting appealing physical properties of FF nanostructures while modifying the peptide in some way to achieve specific functionality. In order to rationally design modifications to the dipeptide that allow for this behavior, the driving forces of FF self-assembly must be understood. Molecular simulations have been utilized to assess these properties but have yielded conflicting conclusions due to inconsistencies in models chosen as well as a lack of quantitative analyses on the specific driving forces. Here we present an all-atom explicit solvent molecular dynamics based study on different length-scales of FF aggregation. We utilize a free-energy decomposition analysis as well as a dimer cluster analysis to identify the initial aggregation driving force to be FF intermolecular electrostatics while solvent mediated forces drive crystal growth. This data is consistent with the hypothesis that all hydrophobic dipeptides will have a similar initial aggregation mechanism until a critical aggregate size is reached at which point crystallization occurs and subsequent crystal growth is dominated by solvent mediated forces.

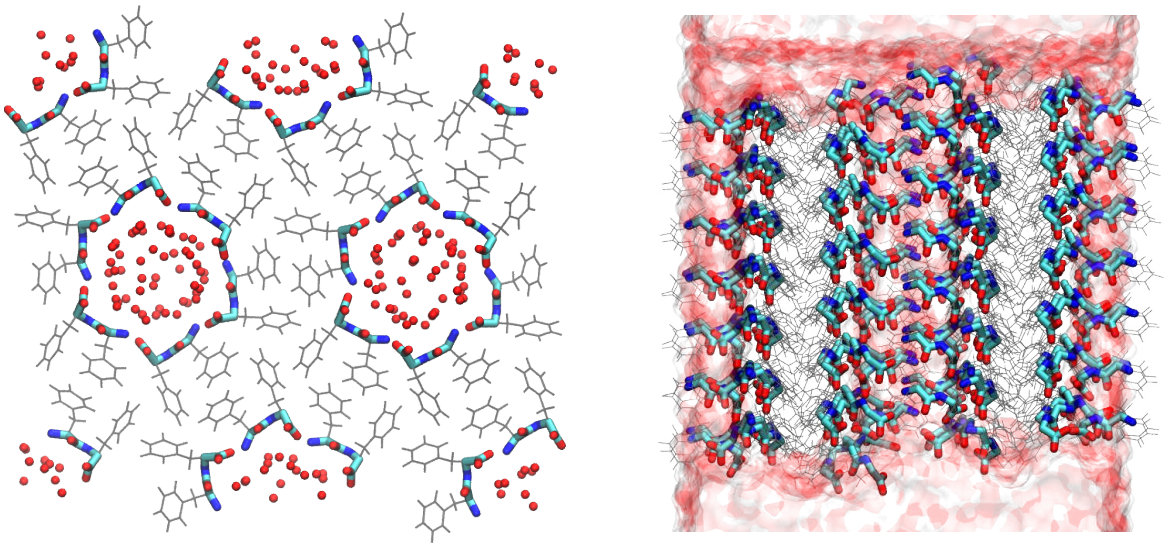
[1] Jakob Anderson[a], Peter T. Lake[a], Martin McCullagh[a]; [a]Department of Chemistry, Colorado State University, Fort Collins, CO, USA; Reproduced with permission from Anderson, J., & McCullagh, M. (2018). Initial Aggregation and Ordering Mechanism of Diphenylalanine from Microsecond All-Atom Molecular Dynamics Simulations. Journal of Physical Chemistry B, 122(51), 12331–12341.https://doi.org/10.1021/acs.jpcb.8b10335. Copyright 2018 American Chemical Society

We demonstrate that this proposed mechanism is testable by infrared spectroscopy focusing on the blueshift of the Amide I peak as well as the ordering of the carboxylate peak.

2.2 Introduction

The self-assembly of diphenylalanine (FF) has garnered significant interest due to the versatility and stability of the assembled material. Originally investigated for its role as a key structural motif in amyloid-β (Phe^{19}-Phe^{20}),[60] FF has since been proposed as a useful material for a diverse set of applications[21] including being used as an antibacterial agent[61], a drug delivery vehicle,[62] and a piezoelectric material.[63,64] FF microtubes are one of the strongest biomaterials presently known with a Young's modulus of 19 GPa[21–23] and have been shown to persist through changes in solvent environment and heating.[24] The adoption of this material to a broad spectrum of applications is predicated on the idea of chemically modifying FF while maintaining the assembly behavior. It is of interest, therefore, to investigate the molecular features of FF that dictate self-assembly.

FF has been shown to form macroscopic structures such as nanopores and nanowires and is described as an exemplar of small molecule self-assembly.[21,24,28,29,62,65–67] The self-assembly has been described as hierarchical[21,67] in that the molecules assemble into sheets of nanopores and these sheets assemble into microtubes.[21] The FF crystal structure has six dipeptides in a helical structure around a water pore with the hydrophobic phenyl groups contained in the pore wall, as seen in the representative structure in Figure 2.1a. A variety of other hydrophobic dipeptides have also been crystallized with water pores, including dileucine, leucine-phenylalanine and phenylalanine-leucine.[33] These four crystal structures demonstrate similar behavior with the dipeptide backbones creating a water pore of increasing size from dileucine to FF. The similarity in these crystal structures, coupled with evidence that a variety of dipeptides aggregate in water,[68] suggests similarities in the assembly mechanism of dipeptides. While comparisons can be made, it should be stressed that FF is unique in the degree of hierarchical order and therefore likely contains a unique self-assembly mechanism.



(a) **(b)**

Figure 2.1: The FF nanotube crystal simulated seen from the top-down (a) with water molecules within 5.2 Å of the peptides represented as red circles and from a side view (b) with water represented as a red surface. The system is simulated with periodic boundaries such that the crystal lattice is infinite in the x and y dimensions with water layers separating periodic images in the z dimension.

Previous experimental investigations into the molecular level driving forces of FF assembly generally include the study of chemical modifications to FF. Non-natural modifications of the peptide coupled with solvent changes have indicated the importance of electrostatic interactions in the assembled structures of FF derivative substances.[27] Natural modifications of the FF dipeptide, namely the mutation of either residue to isoleucine or valine, demonstrate how the aromatic sidechains of FF lead to more ordered structures than these hydrophobic, but non-aromatic counterparts.[26] Additionally, bioinformatic studies have also demonstrated the potential importance of aromatic side-chains on the assembly of small peptides.[69–71] Fourier-transformed infrared spectroscopy has been applied to the self-assembled FF nanostructure[37,72] in aqueous solutions as well as under different solvent environments.[13] Time resolved measurements of the assembly mechanism remain a challenge due to the lack of solubility of the FF molecule. Further atomistic level mechanistic insight into the assembly mechanism of FF would benefit the bottom-up hypotheses development from these types of studies.

Computational and theoretical methods are ideally suited to investigate the molecular level mechanism of initial FF aggregation. Electronic structure methods employed to study the crystal structure have highlighted the need for dispersion to accurately reproduce the Young's modulus predicted by experiment.[30] Aggregation behavior has been studied by both coarse-grained (CG) and all-atom (aa) molecular dynamics (MD) simulations. Previous aaMD simulations of FF have been limited to hundreds of nanoseconds and fewer than twenty FF molecules. Studies on these time and length scales are insufficient to observe nanotube formation, but have demonstrated the importance of electrostatic[31,72] and, for one study, hydrophobic[29] interactions on the initial aggregation behavior. CG models, using the MARTINI force field, are able to observe initial aggregation and ordering of FF.[28,32] Guo *et al.* highlight the importance of T-shaped $\pi - \pi$ stacking as well as backbone head–tail (electrostatic) interactions for the ordered structure. The CG models, however, also result in the formation of lamellar like phases as opposed to the tubular structures observed in the crystal. Thus, due to inconsistencies in the models chosen, simulation procedures, and analyses performed, previous simulation results provide conflicting conclusions on the driving forces of FF self-assembly.

In order to provide a coherent picture of FF self-assembly, we employ extensive all-atom explicit solvent molecular dynamics simulations. Aggregation is probed from the length-scale of dimerization up to that of crystal growth. In the next two sections we discuss the variety of systems and techniques employed in our simulation study. This is followed by a description of the driving forces behind dimerization, initial aggregation and finally crystal growth. Our results suggest that initial aggregation and ordering is driven by electrostatics, while crystal growth is driven by both electrostatics and solvent mediated forces.

2.3 Methods

2.3.1 Force Field

Previous aaMD studies of FF have used the ff99SB,[29,73] the GROMOS53a6,[31,74,75] and the CHARMM22[72,76] force fields. We tested the ability of these and five other commonly used force

fields to reproduce dimerization energies computed using electronic structure calculations. Four FF dimers, shown in Figure A1 of the Supporting Information, are chosen that generalize possible types of FF dimers and consist of backbone–backbone, head–tail, sidechain–sidechain, and both backbone–backbone and sidechain–sidechain interactions. Each dimer is first geometry optimized using the ωB97xD density functional theory method and a 6-311+G* basis set with the Gaussian09 package.[77,78] This functional is chosen due to the inclusion of long-range dispersion corrections which has been shown to be important for FF.[30] The dimer energy is then obtained accounting for basis set superposition error which, to the authors knowledge, can only be done *in vacuo*. These values are then compared to the interaction energies produced from the set of chosen force fields. The ff14ipq force field is used for all simulations as it has the lowest mean unsigned percent error (7.28%) relative to the quantum mechanical calculations.[79] The dimer interaction energies and the error relative to quantum mechanical calculations can be found in Table A1 of Supporting Information.

The FF nanotube has been experimentally shown to be incredibly stable. Therefore, a simulation of the FF nanotube with the chosen force field should maintain the organized structure. We measured the RMSD of FF nanotube selections after simulating in both the ff14ipq and ff14SB force fields for 1 μs. The FF nanotube has lower RMSD's of trimer and hexamer structures (Figure S2 and Figure S3) with the ff14ipq as opposed to the ff14SB force field, further supporting our choice in using the ff14ipq force field. Moreover, the FF nanotube is similar to a globular protein in that FF orders with hydrophobic and hydrophilic layers. The IpolQ force fields more closely reproduce the NMR chemical shifts relative to other Amber force fields of large globular proteins further supporting our choice in force field.[80]

2.3.2 System Setup

Models of individual FF molecules in the L-enantiomer are generated using the sequence command in tleap within the AmberTools 16 package.[81] The dipeptide sequence is capped with a positively charged amino N-terminus and a negatively charged carboxylate C-terminus. Systems

with 1, 2, 60, and 120 FF dipeptides are simulated using initially dispersed monomers in this linear conformation. Monomer and dimer simulations are performed at concentrations of 4 and 15 mg/mL for 300 and 250 ns respectively. Simulations using 60 or 120 monomers are performed at 50 mg/mL concentrations. As the primary system of investigation, systems of 60 FF dipeptides are simulated in triplicate for 3 μs each. Only one replica of the system with 120 FF dipeptides is simulated for 5 μs. The combined unbiased simulation time for these systems is 14.5 μs.

A fragment of the FF nanotube crystal structure[33] with 168 FF dipeptides, depicted in Figure 2.1, is also simulated for 1 μs. The FF nanotube simulation is created using the Mercury[82] program, translating the unit cell of the crystal to produce seven layers of the representation in Figure 2.1. This crystal structure simulation is performed with periodic boundaries such that the FF nanotube is continuous in the x and y dimensions, while water is allowed to move through the pores in the z dimension. Analysis performed on the FF nanotube system is used as a metric of molecular order within the initially dispersed systems.

2.3.3 Simulation Details

All simulations are performed in explicit TIP3P water in the NPT ensemble using the ff14ipq force field in the GPU accelerated AMBER16 program.[81,83,84] A Langevin thermostat set to 298 K and a Monte Carlo barostat set to 1 bar are used. An integration timestep of 2 fs is used in all simulations with the SHAKE algorithm restraining the bond length of bonds including a hydrogen atom. Direct non-bonding interactions are calculated up to 12 Å for all simulations with the particle mesh Ewald method calculating long range electrostatic interactions.

Prior to production runs of unbiased simulations, the following protocols are performed on all simulations. During initial energy minimization, a steepest descent protocol for 8000 steps followed by the conjugate gradient method for 4000 steps are performed with restraints placed on the peptide atoms with a force constant of 100 kcal/mol/Å^2. The restraints on peptide atoms are then removed and energy minimization is again performed with a steepest decent protocol for 12000 steps followed by the conjugate gradient method for 6000 steps. The systems are then heated

from 0 K to 298 K, at constant volume, with restraints placed on the peptides (force constant of 5 kcal/mol/Å^2) for 50 ps. The volume of each system is then allowed to relax for 40 ps with restraints kept on the peptides. Production runs are then started after 3 ns to remove transient effects due to initial configurations.

2.3.4 Umbrella Sampling Simulations

Umbrella sampling is performed along the monomer center of mass separation between 3.0 Å and 17.5 Å. Harmonic potentials each with a force constant of 20 kcal/mol/Å^2 are used with wells positioned every 0.5 Å and simulated for 15 ns for each window. Initial coordinates for the umbrella sampled windows are culled from 250 ns of unbiased simulation of two FF molecules. Umbrella sampling for FF dimerization results in 450 ns of simulation.

To mimic monomeric addition to a FF nanotube, umbrella sampling is performed between the center of mass of a surface monomer's backbone and the center of mass of the amide bond of the monomer directly below in the FF nanotube simulation. Steered umbrella sampling is used to create initial states ranging from 3.5 to 25.0 Å in 0.5 Å increments. A harmonic potential is placed on each window of 20 kcal/mol/Å^2. Further details on how these simulations were performed using the AMBER16 program are found in the "PMF Comparison" section of the Supporting Information. Each window is simulated for 33 ns. Further simulation, additional windows, and larger force constants are applied in regions of poor collective variable space sampling, discussed further in section "PMF Comparison" of Supporting Information, for a total of 1.55 μs of umbrella sampled simulation.

2.4 Analysis

The VMD software is used to visualize trajectories, generate structural figures, and calculate solvent accessible surface area (SASA).[85] Analysis of MD trajectories is done using python 2.7 and the MDAnalysis module (version 0.18.0) unless stated otherwise.[86,87] Data is then plotted

using Matplotlib.[88] Further details and the scripts themselves are available on Github (https://github.com/mccullaghlab/FFInitialAggregationOrdering).

2.4.1 Free Energy Decomposition

The potentials of mean force (PMF) are generated by integrating the mean force measured in umbrella sampled trajectories. The relative free energy, $\Delta A(r)$, is computed as the integral of the mean force due to all particles $\langle f \rangle$ projected along a given coordinate r. Since the force is pairwise additive between all particles, $\langle f \rangle$ can be separated into solvent–solute (V–U) and solute–solute (U–U) contributions. Furthermore, in classical force fields such as ff14ipq the mean force can be exactly separated further into Lennard-Jones (LJ) and Coulomb (C) components. Integrating the components of the mean force allows us to decompose $\Delta A(r)$ into solvent–solute and solute–solute components, and then further separate these into Lennard-Jones and Coulomb components

$$\begin{aligned} \Delta A(r) &= \int_{\infty}^{r} \langle f \rangle(\alpha) d\alpha \\ &= \delta A_{V-U}(r) + \delta A_{U-U}(r) \\ &= \delta A^{LJ}_{V-U}(r) + \delta A^{C}_{V-U}(r) \ + \delta A^{LJ}_{U-U}(r) + \delta A^{C}_{U-U}(r) \end{aligned} \tag{2.1}$$

where we use δ to denote that $\delta A_x(r)$ are not state functions; the quantities represent the relevant contributions to the total free energy along a chosen path. While computed differently, the PMF's generated by integrating the mean force are shown to agree with PMF's generated using the eigenvector method for umbrella sampling (EMUS)[89] (Figure S4). While it may be appealing to interpret $\delta A^{LJ}_{V-U}(r)$ as the component of the free energy contributed by the hydrophobic force as it is an integral of the non-electrostatic force between the water and the FF molecules, we are cautious to call it thus since the hydrophobic force is an ill defined object. Instead, the solvent–solute Lennard-Jones will be referred to in the context of this paper as the solvent mediated force.

2.4.2 Pair Interaction Type Analysis

Pairs of interacting molecules present in initially dispersed systems are classified through a decision tree analysis. For all possible FF molecule pairs at 0.1 ns increments, two molecules are considered to be interacting if the center of mass distance separation is less than 16 Å. Cutoffs of functional group pair distances are determined from the radial distribution function, $g(r)$, for that pair and are reported in Table A2 of Supporting Information. Backbone interacting pairs are then considered if the two FF molecules have N-terminal–amide, N-terminal–C-terminal, amide–amide, or C-terminal–amide distances within respective cutoffs. Pair interaction types are defined by combinations of cutoffs between atom pair distances described in the "Interaction Pair Type Analysis" section of the Supporting Information. The classification of interacting pairs is then determined in a most–least inclusive stepwise manner. If the pair is not a backbone interacting pair, phenyl–phenyl ring distances are measured between the two molecules and classified as a sidechain interacting pair if these are under cutoffs described in the "Interacting Pair Type Decision Tree Analysis" section within the Supporting Information.

2.4.3 IR Spectra Calculation

We investigate the infrared (IR) spectra of aggregates of FF for three normal modes of the monomer: amide I ($\sim$1630 cm^{-1}), C-terminal carboxylate ($\sim$1580 cm^{-1}) and amide II ($\sim$1440 cm^{-1}). In order to compute the predicted spectra directly from the structures observed in molecular dynamics simulations, we utilize a transition dipole coupling (TDC) scheme.[90] This method has been improved with the use of non-TDC nearest neighbor couplings which has been demonstrated to work well for globular proteins.[91] Additional advancements of ϕ and ψ dependent parameters[92] and transition charge models[93] have been utilized to improve these fits and have even been employed on dipeptides.[93] Here we employ a single-site energy multi-mode transition dipole model motivated by the only non-bonded amide couplings and relative rigidity of the monomers in $\phi - \psi$ space.

The input frequencies, relative infrared intensities and dipolar derivative vectors are taken from electronic structure calculations of the two observed dominant monomer configurations. Geometry optimizations and frequency calculations are performed at the ωB97xD/6-311+G* level of theory in the Gaussian09 software package.[77,78] The PCM implicit solvent model is used to mimic the effects of solvation in water. The geometry optimizations and frequency calculations are performed with all hydrogens bound to a nitrogen being in the deuterated state. The frequencies are scaled by 0.93 in order to match experimental IR frequencies for the FF crystal structure.[37] This scaling factor is within values determined for this basis-set and functional.[94]

The dipole derivative vectors are placed on the center of carbonyl bond of the amide group for amide I, the center of mass of the carboxylate group for the C-terminus, and the nitrogen of the amide for the amide II. These vectors are given their local projection in the basis of the amide group for each molecule. Gaussian (width of 10 cm^{-1}) convolved spectra are averaged over the trajectory.

2.5 Results and Discussion

Previous simulation studies have yielded conflicting conclusions on the driving forces of FF self-assembly. Here we present a thorough set of simulations that investigate the mechanism across multiple scales. We couple these simulations with unique quantitative analyses to paint a consistent picture of the self-assembly process. Specifically, we investigate the dimerization process, the aggregation of many initially dispersed molecules, the expected difference in infrared (IR) spectra of the initial aggregates to the FF nanotube, and the addition of a monomer to the FF nanotube. The results of important molecular interactions within each of these systems suggest a complicated self-assembly process which explains the discrepancies between previous studies.

2.5.1 Driving Forces of Initial Aggregation

As dimerization can be considered to be the smallest aggregation event, the driving forces of this process are critical for the understanding of higher order assembly. The PMF of dimerization

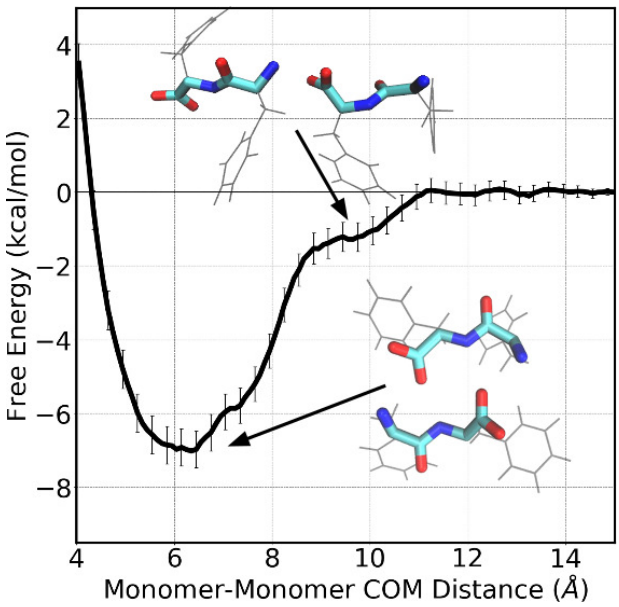


Figure 2.2: The relative free energy of FF dimerization as a function of FF heavy atom center of mass separation. The PMF is computed by integrating the mean force as described in the "Methods" section. The PMF has a global well at 6.25 Å with a depth of 6.31 kcal/mol and a shoulder at 9.75 Å with a depth of 1.03 kcal/mol. Representative dimer structures are also shown and demonstrate the types of interactions sampled between the two dipeptides at the indicated distances.

as a function of heavy-atom center of mass separation is computed using umbrella sampling simulations and is shown in Figure 2.2. The FF dimer PMF has a shoulder at 9.75 Å with a depth of 1.2($\pm$0.4) kcal/mol in which representative structures at this distance are comprised of dimers with one N-terminus–C-terminus interaction. The global minimum of the PMF is at 6.25 Å with a depth of 7.0($\pm$0.5) kcal/mol and contains dimers with both termini of one molecule interacting with the oppositely charged termini on the other molecule. The dominance of these representative structures within the wells of the PMF, namely one N-terminus–C-terminus contact around 9.75 Å and two terminus–terminus contacts around 6.25 Å (Figure 2.2), indicate the importance of backbone–backbone interactions at least for the dimerization of FF.

A quantitative measure of the driving forces for dimerization can be achieved by decomposing the PMF using Equation 3.1. First, the PMF is decomposed into solute–solute and solvent–solute contributions as shown in Figure 2.3a. The solvent–solute curve (green) demonstrates a positive free-energy contribution over center of mass separation distances less than 11 Å. The solute–solute

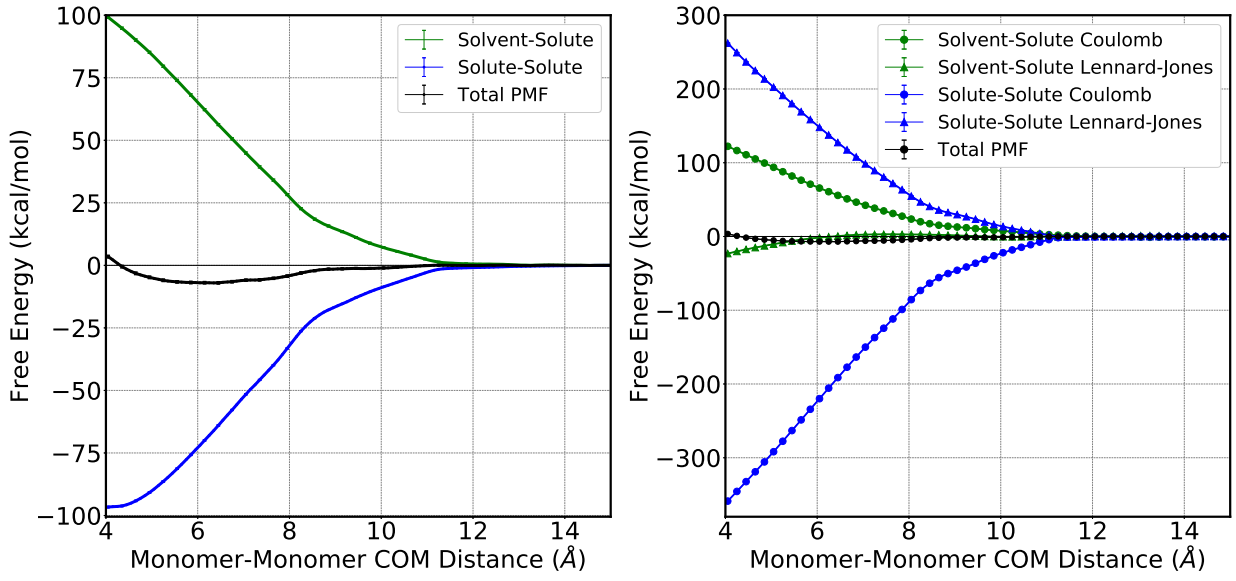


Figure 2.3: The relative free energy of FF dimerization (black curves) as a function of FF heavy atom center of mass separation decomposed into (a) solvent–solute (green curves) and solute–solute (blue curves) components and (b) further decomposition of both solvent–solute and solute–solute components into Lennard-Jones (triangles) and Coulomb (circles) components. Error bars are present but are smaller than the symbols or lines.

curve (blue), however, demonstrates attractive behavior over the same sampled distances. This indicates that the solvent–solute forces hinder dimerization, while solute–solute interactions promote it.

Further separation of the solute–solute and solvent–solute forces into Lennard-Jones (triangles) and Coulomb (circles) forces provides additional differentiation of the underlying driving forces. Figure 2.3b shows that the dimerization of FF is driven predominantly by solute–solute electrostatic forces (blue circles), as this component of the free energy is negative for distances less than 11 Å, while the solute–solute Lennard-Jones forces (blue triangles) are repulsive for the entire PMF. Additionally, the repulsive force between the solvent and the solute is largely dictated by electrostatics. While the solvent–solute Lennard-Jones component of the free energy becomes negative at distances less than 6 Å, this is a shorter distance than the minimum of the well for the total PMF and thus can be interpreted as not a driving force for dimerization. These results indicate that the dimerization of FF in water is driven by solute–solute electrostatic interactions.

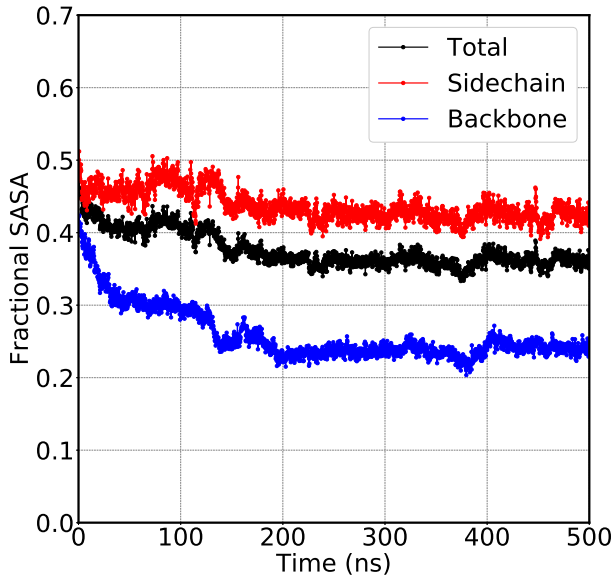
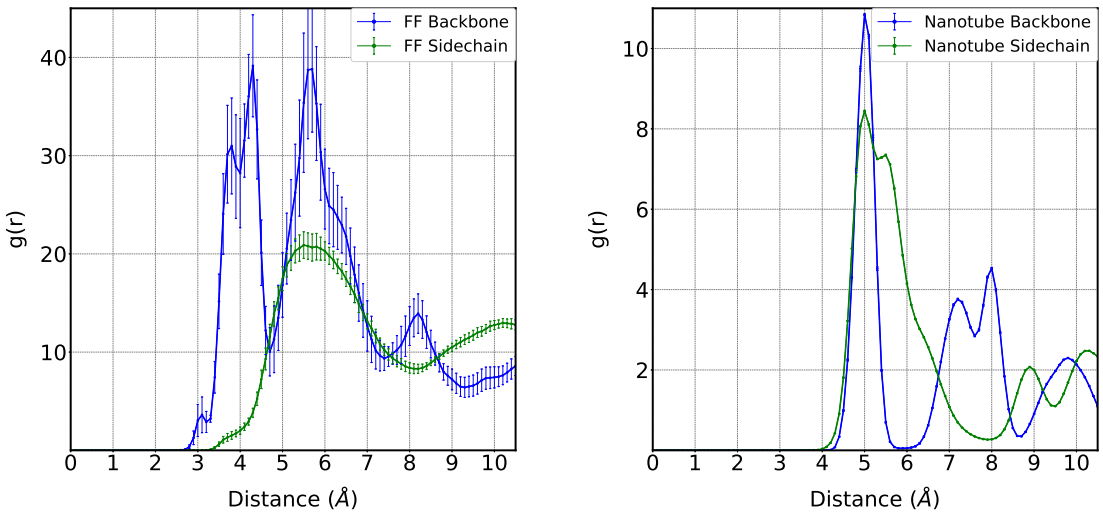


Figure 2.4: The total (black curve), sidechain (red curve), and backbone (blue curve) fractional SASA of a system with 60 initially dispersed FF dipeptides relative to the SASA sampled by a monomer.

Electrostatics continues to drive aggregation beyond dimerization as demonstrated by simulations of many initially dispersed FF molecules. To quantify this behavior, the fraction of total, sidechain, and backbone solvent exposed surface area is shown in Figure 2.4 for the first 500 ns of a 60 FF simulation. The fractional SASA is defined here as the SASA measured in the system normalized by the average SASA of a monomer over 300 ns and the number of molecules in the system. Aggregation occurs during this portion of the simulation as is indicated by the fractional total SASA (black curve) at 0 ns dropping from 0.46 to 0.36 by 500 ns. The sidechain fraction (red curve) has little change over this time frame while the backbone fraction (blue curve) has a significant reduction from 0.42 to 0.23. This behavior is indicative of intermolecular backbone–backbone contacts while sidechains remain similarly solvated during this initial aggregation. In all other initially dispersed systems, the relative total, sidechain, and backbone fractional SASA's are comparable to the values seen in Figure 2.4 with the same conclusions inferred.

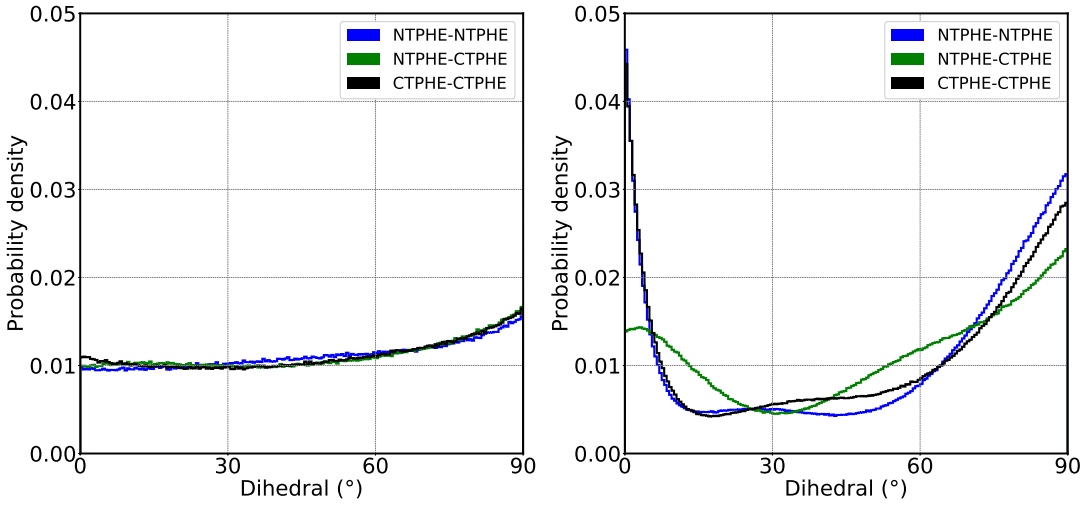
2.5.2 Backbone Electrostatics Drive Further Ordering of Initially Dispersed Systems

Each initially dispersed system forms one aggregate by the last 250 ns of the respective trajectory and is defined here as the stable aggregate. The radial distribution function, $g(r)$, is measured to determine the relative degree of ordering in backbone or sidechain selections to one another for the stable aggregates of initially dispersed systems relative to the FF nanotube. As can be seen in Figure 2.5a, the backbone $g(r)$ (blue) of the stable aggregates has four peaks while the sidechain $g(r)$ (green) has one peak at distances less than 8 Å. The larger number of peaks demonstrates that there is more order between the backbones of FF molecules than the sidechains. Within the backbone center of mass $g(r)$, pairs of molecules from the peaks below 5 Å have both termini interacting while pairs from the peaks between 5 Å and 8 Å have one N-termini–C-termini interaction. These pair interactions are consistent with the representative dimers shown with the dimer PMF of Figure 2.2.



(a) Initially Dispersed **(b)** FF Nanotube

Figure 2.5: The center of mass backbone–backbone (blue curves) and center of mass sidechain–sidechain (green curves) $g(r)$'s over the last 10 ns (a) averaged between the triplicate 60 FF and 120 FF systems and (b) the FF nanotube system. The error bars are included as the standard error of the mean between the triplicate simulations and the individual $g(r)$'s are shown in SI.

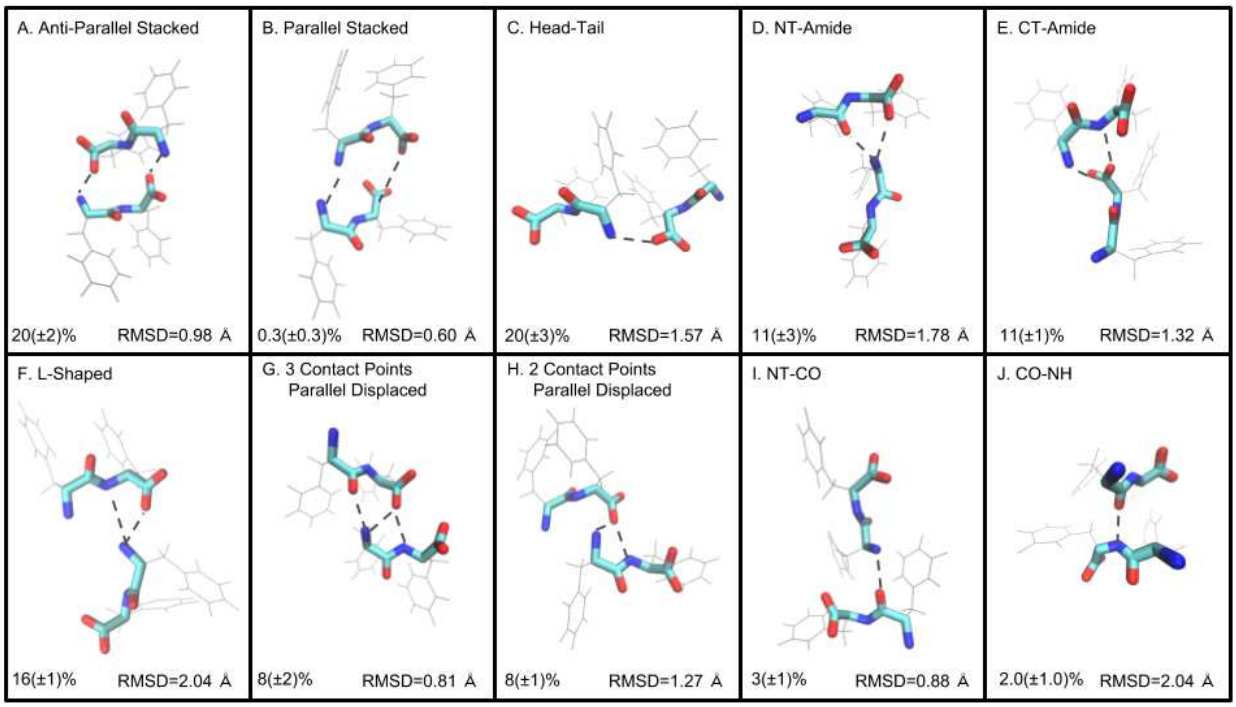


(a) Initially Dispersed **(b)** FF Nanotube

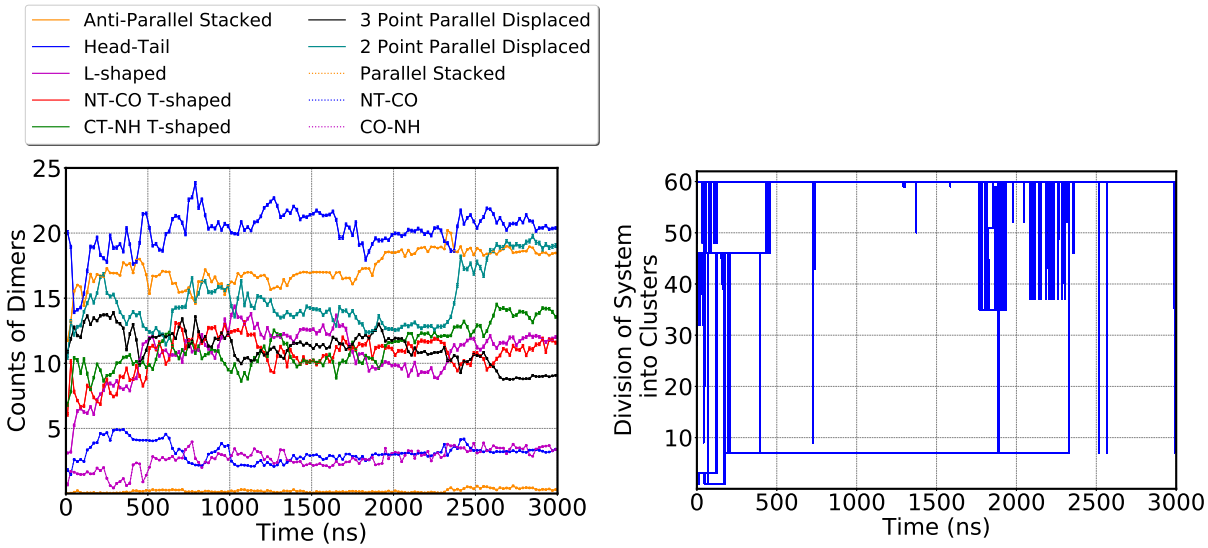
Figure 2.6: The intermolecular phenyl-ring dihedral probability densities in (a) a 60 FF system and (b) the FF nanotube. The dihedral is measured for all combinations of N-terminus and C-terminus phenyl-ring combinations for molecules with phenyl-rings atoms within 6 Å. Values around 0° and 90° are indicative of $\pi - \pi$ stacked ordering occurring as is evident for the FF nanotube system.

Since it has been suggested that $\pi - \pi$ stacking plays a large role in the self-assembly mechanism of FF, the phenyl–phenyl ring dihedral is measured for the stable aggregate of the initially dispersed systems and the FF nanotube. As can be seen in Figure 2.6, values of 0° correspond to face-centered $\pi - \pi$ stacking, face-staggered $\pi - \pi$ stacking, or edge-edge non-$\pi - \pi$ stacking arrangements, while edge-to-face $\pi - \pi$ stacking interactions are represented with values around 90°. While the self-assembly of FF has been proposed to be driven by hydrophobic interactions stabilized by $\pi - \pi$ stacking, the probability densities of $\pi - \pi$ stacking interactions are significantly less ordered in initially dispersed systems (Figure 2.6a) than in the FF nanotube (Figure 2.6b). The lack of order in the sidechain $g(r)$ and in the phenyl-ring dihedral is suggestive that $\pi - \pi$ stacking, while known to be important for stability of FF nanotubes, is not important in the initial aggregation stages of the FF self-assembly mechanism.

Pairwise ordering of aggregates is quantified using a pair type analysis implemented in a decision tree (a scheme for which is shown Figure A5 in Supporting Information). Classification of pair types is based on a set of pairwise distance cutoffs tabulated in Table A2 of Supporting Infor-



(a)



(b) **(c)**

Figure 2.7: (a) Representative structural images of pair interaction types as well as the RMSD and the percentage of the interacting pair type, over the last 250 ns of trajectory, averaged over each initially dispersed system. Contact pair distances with favorable interactions are represented with dashed lines. RMSD's are determined for backbone heavy atoms excluding the C-terminal oxygens. (b) The number of classified pair interaction types over 3 μs of a system with 60 initially dispersed FF dipeptides. (c) The clustering of these backbone interacting pairs as a function of time.

mation that is trained on the pairs observed in the stable aggregate. The initial classification of pairs into two populations yields 68($\pm$5)% backbone–backbone pairs and 32($\pm$5)% sidechain–sidechain pairs for the stable aggregate. The dominance of backbone–backbone pairs in this analysis is consistent with the backbone $g(r)$ having more order than the sidechain $g(r)$ and the phenyl–phenyl ring dihedrals of the stable aggregates containing little evidence of $\pi - \pi$ stacking. Backbone interacting pairs are further classified based on chemical intuition and combinatorics of possible types of pairs, resulting in 10 distinct pairs (Figure 2.7a) through which 99.94($\pm$0.04)% of backbone interacting pairs in the stable aggregates are resolved. The RMSD computed for each interacting pair type (shown in the bottom right of each panel of Figure 2.7a) results in RMSD's below 1 Å for pairs with minimal fluctuations in the angle between the termini–termini vectors of the two molecules. RMSD values above 1 Å are expected if the distance between interacting functional groups as well as the angle between the terminus–terminus vectors fluctuates significantly.

The time evolution of backbone–backbone pairs, shown for one replica of 60 initially dispersed molecules in Figure 2.7b, demonstrates the pairwise ordering of the initially dispersed system. Specifically, the T-shaped and parallel displaced pairs fluctuate significantly over the 3 μs while head–tail and anti-parallel stacked pairs remain relatively constant over the course of the last 1 μs. While number counts varies over other initially dispersed systems, head–tail (solid blue) and anti-parallel stacked (solid orange) are dominant pair types across all initially dispersed systems comprising 20($\pm$3)% and 20($\pm$2)% of backbone interacting pairs in the stable aggregates, respectively. There exists two types of backbone interacting pairs within the FF nanotube, head–tail and parallel stacked pairs. Interestingly, while parallel stacked pairs rarely form, head-tail pairs are the most dominant species present within initially dispersed systems.

Ordering beyond that of the dimers can be assessed by clustering of pairwise contacts from the dimer type decision tree analysis. The time evolution of clusters is measured with the hypothesis that fusion events occur between clusters far more frequently than fission. Clustering is performed by assigning each molecule a unique vertex, labeling edges for backbone interacting vertex pairs, and then determining the vertex count of all connected graphs at each time point. As seen in

Figure 2.7c for one replica of 60 FF molecules, the summation of cluster sizes at a given time is equivalent to the number of molecules in the system such that at 300 ns there are clusters of size 7, 14, and 39. The molecules in the clusters of size 14 and 39 at 300 ns are incorporated into the cluster of size 53 at 500 ns. For this one replica the system begins to form an initial aggregate within the first 250 ns wherein there is a considerable amount of fluctuations. The final stages of clustering into a single grouping of molecules takes a longer time and demonstrates transient fluctuations as the system finds higher order. For example, Figure 2.7c shows that this system forms into two clusters of size 7 and 53 which persist between 500-2300 ns before fusing into a single cluster. However, at time 1700-2300 ns when the system has two stable clusters, the cluster of 53 molecules temporarily breaks into two clusters as seen in the fluctuations. These fluctuations are indicative of a rearrangement of the molecules within the cluster. By the end of all initially dispersed system trajectories, one cluster is formed. The formation of one cluster using only backbone interacting pairs again demonstrates the importance of backbone–backbone interactions. Furthermore, the fluctuations in clusters suggests that initial aggregation is a dynamic process occurring on the microsecond time scale, consistent with the Ostwald ripening seen for N-(t-butoxycarbonyl)-L-PHE-L-PHE-COOH.[95]

2.5.3 Comparison of Computed IR Spectra

The FF amide bond environment within the aggregates formed during the initially dispersed simulations is distinct from the amide bond environment within the pre-ordered nanotube simulations. As such, infrared (IR) spectroscopy should be sensitive to the structural differences in these configurations. It is well established that the position and shape of the amide I band of globular proteins can be correlated to amounts of secondary structure elements in the protein.[91,96] Since IR spectroscopy can probe systems on the picosecond time scale and due to FF nanotubes having a distinct spectrum, we attempt to predict the vibrational frequencies during the initial aggregation of FF relative to the self-assembled FF nanotube using a transition dipole coupling scheme.

IR spectra of the amide I band has been shown to be sensitive to the presence of nearby oscillators as well as the internal configuration of the amide group.[90,91] The FF nanotube structure has been previously observed to contain monomer configurations that are distinct from that of solution phase monomers.[29,33,74] This is consistent with the systems studied here as can be seen in the Ramachandran plot in Figure 2.8. The initially dispersed systems predominantly sample two states between ϕ values of -60°–180° and ψ values of 0°–70° while the FF nanotube samples one states centered around 60° and 150° in ϕ-ψ space. Approximately 5% of the initially dispersed molecules sample the ϕ-ψ space of the FF nanotube suggesting that the systems are not restricted in this dihedral phase space. Since the molecules are in different but relatively fixed orientations, the vibrational frequencies, relative IR intensities, and dipolar derivative vectors for the two systems are taken from two distinct electronic structure calculations.

Using a transition dipole coupling model, the amide I, carboxylate, and amide II frequencies are measured for the last 10 ns of FF nanotube simulation and for the stable aggregate of the initially dispersed systems. The resulting spectra are shown in Figure 2.9. The computed IR spectra were scaled by 0.93 to align the computed amide I peak of the FF nanotube system to the amide

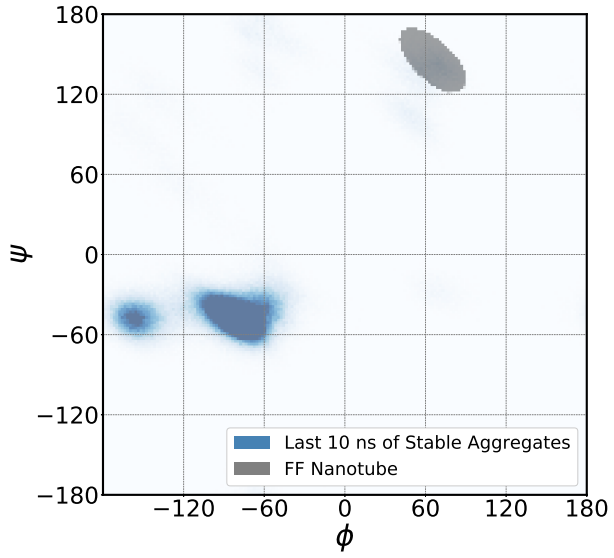


Figure 2.8: The probability density of the ϕ and ψ dihedrals sampled in the last 10 ns of all initially dispersed dipeptide system (light blue) and the FF nanotube system (grey).

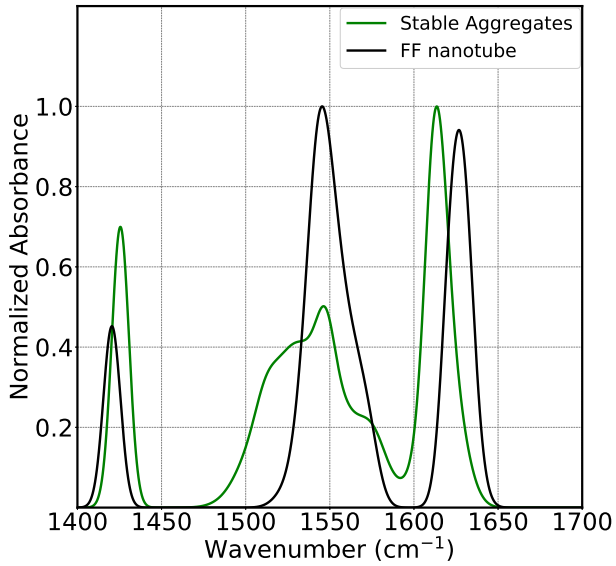


Figure 2.9: The predicted IR spectra of the stable aggregate of the initially dispersed systems (green) and the computed spectra of the FF nanotube (black).

I stretch, centered around 1630 cm^{-1}, in experimental IR spectra of the FF nanotube.[37] The FF nanotube has a narrow amide I peak in agreement with experiment, while the stable aggregates has a blue shifted and slightly broader amide I peak at 1614 cm^{-1} after scaling by the same factor. Although not reported in the literature, the stable aggregate systems and the FF nanotube system both contain significant carboxylate peaks. There is only one sharp peak for frequencies corresponding to the carboxylate modes for the FF nanotube system at 1545 cm^{-1} in stark contrast to the stable aggregates which contain a broad distribution of multiple peaks between 1475–1592 cm^{-1}. The amide II peak is located at 1425 cm^{-1} for the stable aggregates and 1421 cm^{-1} for the FF nanotube. While the amide II peak is sharp in both systems, the stable aggregates have a significantly higher amide II peak intensity relative to the FF nanotube amide II peak.

The IR spectra shown in Figure 2.9 indicate that experiments which are able to resolve the early time behavior of an initially dispersed FF system would be sensitive to the formation of different initially ordered structures. Experiments of the initial aggregation of FF remain a significant challenge due to the lack of solubility of FF molecules. It is possible, however, to consider gas-phase cluster growth experiments that may be able to directly test some of these observables. Namely,

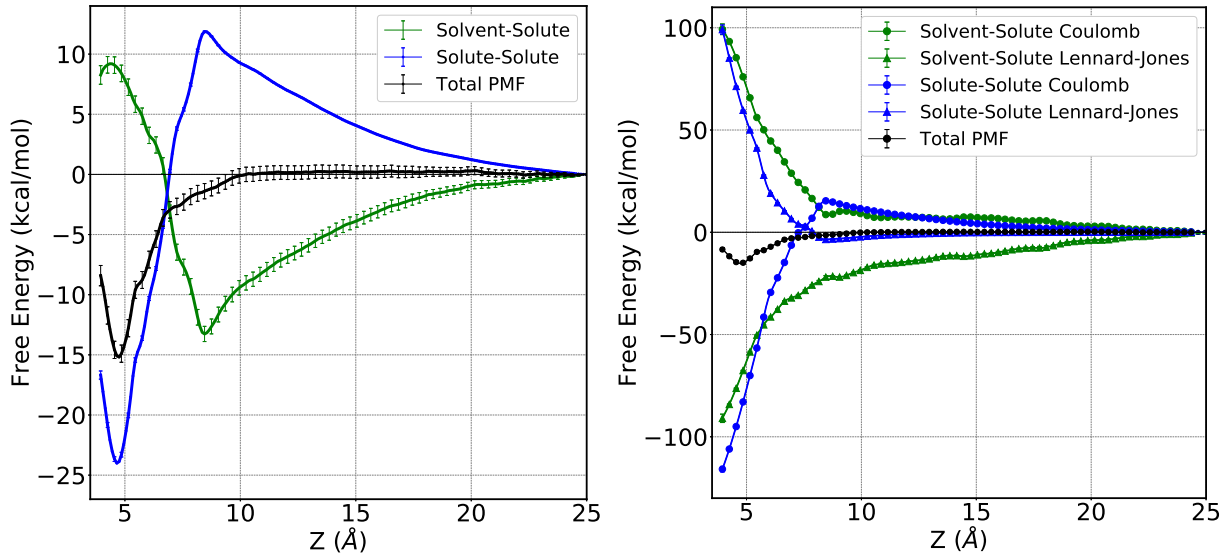


Figure 2.10: The PMF of monomeric addition to the FF nanotube (blue curves) decomposed into (a) solvent–solute (green curves) and solute–solute (blue curves) components which are further separated (b) into Lennard-Jones (triangles) and Coulomb (circles) components. The PMF is measured along the separation distance, labeled here as Z, of the molecule of interest's backbone center of mass and the amide bond center of mass of the monomer in the nanotube layer below.

we anticipate three distinct changes in the linear IR spectra as the FF molecules aggregate and ultimately self-assemble into the nanotubes. First, the amide I band will undergo a blueshift as the monomers reconfigure from the linear solution phase conformation to the distinct state observed in the nanotubes. Furthermore, the carboxylate peak will increase in intensity and the ratio of amide I to carboxylate peaks will approach unity. Lastly, the peaks corresponding to the amide II stretch will redshift and decrease in intensity as the nanotube forms.

2.5.4 Monomeric Addition to the FF Nanotube

Initially dispersed systems push the bounds of time and system size scales for aaMD simulations but do not provide evidence for the initial formation of crystal-like order within the aggregates. This would suggest that this process is incredibly slow and requires alternative methods to study. To understand why FF organizes to form nanotubes with extensive stability, we instead probe the underlying driving forces of FF nanotube growth.

The relative free energy of monomeric addition to the FF nanotube is measured using umbrella sampling along the distance between the backbone center of mass of the monomer of interest and the amide bond center of mass of the monomer in the nanotube layer below, defined as Z. The decomposition of the resulting PMF (Figure 2.10a) demonstrates that monomeric addition to the FF nanotube is initially driven by the solvent–solute component of the free energy (green). The average backbone RMSD of the monomer of interest at 5 Å, relative to the starting structure had a value of 0.17 Å. Inclusion of the sidechain heavy atoms does increase this RMSD due to fluctuations in the $\chi 2$ phenyl ring twisting, but the conformation of the peptide is consistent with molecule conformations found in the crystal. At 8.4 Å the driving forces switch between hindrance/promotion and the solute–solute component of the free energy (blue) then drives the monomeric addition to the nanotube surface. Similar to dimerization, an example structural image at this distance has one terminus of the monomer of interest coordinating to the molecule below it. However, unlike in dimerization, there are additional molecules which cause an increase in the solvent–solute free energy.

Further decomposition of the PMF (Figure 2.10b) suggests that the solvent–solute Lennard-Jones force (green triangles) is promoting the addition of the monomer to the FF nanotube over all distances less than 20 Å. The solvent mediated force driving nanotube growth is in stark contrast to the solvent mediated component of the free energy for FF dimerization which does not demonstrate a promoting force prior to sampling the free energy minimum. Even though this solvent mediated force continues to be attractive in the monomeric addition PMF, the solute–solute Coulomb (blue circles) force begins to decrease and promote monomeric addition within 8.4 Å and is the driving force responsible for the lowest free energy within 6.1 Å.

Features within the decomposed PMF, namely the drastic transition at 8.4 Å, are explored by considering the average orientation of the molecule of interest ($\langle \cos(\theta) \rangle$) relative to the surface of the nanotube. Specifically, $\langle \cos(\theta) \rangle$ is the dot product of the z-vector to the vector between termini, averaged with 0.1 Å bins over the collective variable space as shown in Figure 2.11. Values of $\langle \cos(\theta) \rangle$ greater than zero correspond to the molecule of interest having the N-terminus closer

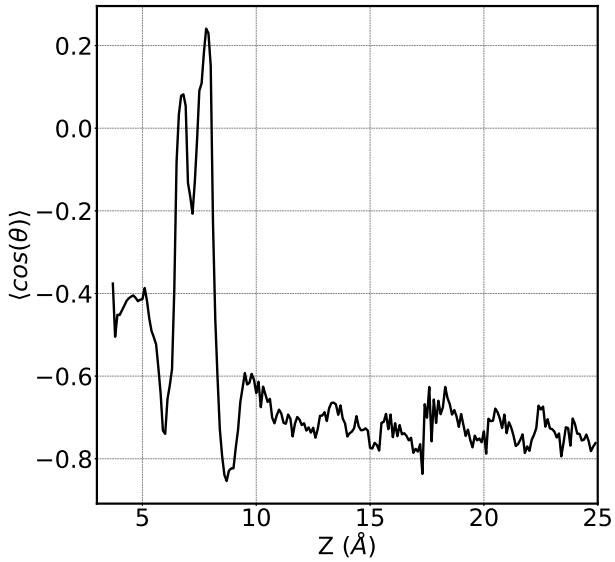


Figure 2.11: The alignment of the monomer of interest $\langle \cos(\theta) \rangle$ measured as the z-component of the vector between both termini relative to the surface of the nanotube as a function of the monomer of interests backbone center of mass separation distance to the amide bond center of mass of the monomer in the nanotube layer below (Z).

to the nanotube surface than the C-terminus and vice versa for values less than zero. Figure 2.11 demonstrates that at distances around 8.4 Å there is a dramatic switch in orientation which is correlated to the switch in the driving forces at 8.4 Å. It is at this transition distance that the molecule of interest's N-terminus coordinates to a C-terminus of an adjacent molecule within the same layer of the nanotube. Additionally, the molecule of interest prefers an orientation with the C-terminus closer to the nanotube than the N-terminus at distances greater than 10 Å. This unidirectional monomeric addition observed here is in agreement with microfluidic studies performed by Arnon *et al.*[97] This orientation preference is correlated with a long-range negative free energy contribution from the solvent mediated forces (green triangles in Figure 2.10b). The correlation between these quantities could be explained by a reduced solvent density on the side of the monomer pointing towards the crystal that is not averaged out due to the orientational preference of the molecule.

2.6 Conclusions

The decomposition of the FF dimerization PMF as well as the sharp decrease in backbone SASA within the first 200 ns relative to the sidechain and total SASA suggest initial aggregation of FF is primarily driven by backbone interactions between solutes. Order within the backbones of dispersed systems, as demonstrated by the backbone $g(r)$ and pair interaction type analysis, further suggests electrostatics promote the initial aggregation process. The sidechains do not demonstrate order during the initial aggregation as seen in the $g(r)$ or $\pi - \pi$ stacking dihedrals. While the order seen in dispersed systems does not resemble the order in the FF nanotube, we postulate that a reorientation event of the individual dipeptides to sample the same dihedral phase space must occur prior to further ordering as is also suggested by Dalgicdir *et al.*[74] Results from the calculated IR spectra further suggests that the reorientation event and subsequent ordering will result in a blue shift of the amide I peak, a decrease in the width of the carboxylate peak, and a decrease in the intensity of the amide II peak. While we do not observe initially dispersed systems forming a FF nanotube, future studies will attempt to investigate the kinetically effective nucleus required to seed crystallization.

Results from decomposition of the monomeric addition to the FF nanotube PMF suggest that initial aggregation and crystal growth have different driving forces. Even though solute–solute electrostatic forces impact the growth at short distances, solvent mediated forces play a dominant role over the total PMF in our modeling of crystal growth. This lends support to the hypothesis that the uniqueness of FF self-assembly is promoted by the solvent and $\pi - \pi$ stacking.

The study presented here attempts to address the disagreement within the literature about the underlying driving forces for FF self-assembly needed for rational design of FF derivatives for a wide variety of potential applications. The multi-step mechanism we propose here highlights the complexity of FF self-assembly and suggests the dominant driving force for FF self-assembly is step dependent. While electrostatics are critical during the initial dimerization and aggregation steps, solvent mediated forces are critical for nanotube growth. It would therefore be interesting

to compare the driving forces seen here, to the driving forces of other hydrophobic dipeptides to investigate the role of $\pi - \pi$ stacking and steric sidechain size.

2.7 Funding

We gratefully acknowledge the Army Research Office for financial support (ARO Fund number W911NF-17-1-0383). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. We would specifically like to acknowledge the San Diego Supercomputer Center Comet and Pittsburgh Super Computing Bridges resources used under XSEDE allocation CHE160008 awarded to MM. This work also used the RMACC Summit supercomputer funded by the National Science Foundation (awards ACI-1532235 and ACI-1532236), which is a joint effort between the University of Colorado Boulder and Colorado State University. We would additionally like to thank Amber Krummel and Anthony Rappé for insightful conversations and guidance.

Chapter 3

Investigating the Sequence Dependence of Valine-Phenylalanine and Phenylalanine-Valine: A Combined IR Spectroscopy and Multiscale Modeling Study[2]

3.1 Overview

Dipeptide self-assembled materials have been shown to have an array of interesting properties which have led to a wide variety of potential applications including use as a piezoelectric material and as a vehicle for drug delivery. These proposed applications, among others, are predicated on the ability to rationally modify a dipeptide to self-assemble in an optimized way for a given application. Previous work has suggested that aggregation properties of certain hydrophobic amino acid combinations are dependent on the primary sequence, an understanding for which is missing in the literature. To elucidate the origins of sequence dependence in dipeptides, we present a joint experimental and computational study on valine-phenylalanine (VF) and phenylalanine-valine (FV). FTIR spectra for VF and FV differ for vibrational frequencies corresponding to the amide I mode. From the second derivative of the linear spectra, VF is shown to have two amide I mode peaks located at 1653 cm^{-1} and 1666 cm^{-1} while FV has peaks located at 1657 cm^{-1} and 1669 cm^{-1}. Using 2D IR, the lifetime for the VF amide I underlying mode located at 1653 cm^{-1} is determined to be 0.25 ps longer lived than the corresponding FV peak at 1657 cm^{-1}. Ramachandran plots determined from all-atom molecular dynamic simulations of monomeric VF and FV demonstrate that while similar backbone comformations are sampled between the two systems,

[2]Jakob Anderson[a,†], Christopher Kuhs[a,†], Martin McCullagh[b], Amber Krummel[a];[a] Department of Chemistry, Colorado State University, Fort Collins, CO, USA, [b] Department of Chemistry, Oklahoma State University, Stillwater, OK, USA, † Contributed equally to this work.

the backbone dihedral states are sampled in differing probabilities. Additionally, from umbrella sampling of the phenylalanine χ_1 for both dipeptides, FV is shown to sample fewer sidechain states relative to VF. Although experimentally the ensemble of conformational states is being observed in the spectra, the lifetimes of underlying amide I modes are far shorter than the interconversion of these states. Through measuring the lifetimes of waters around the amide carbonyl for each of the monomeric conformations using all-atom molecular dynamic simulations restricted to the each of the backbone conformational states, we suggest the differences in experimental lifetimes are caused by water dynamics around the amide carbonyl. Computed linear spectra further suggests VF to be more likely to have a water interacting with the amide carbonyl than in the case of FV in agreement with the amide carbonyl–water lifetime hypothesis. Combined, our results suggest the sequence dependence observed between VF and FV is a combination of monomeric conformations and water dynamics. As a result, we suggest that a phenylalanine on the C-terminus in a dipeptide will result in slower water exchange dynamics suggesting a lower aggregation propensity relative to a phenylalanine residue on the N-terminus.

3.2 Author Contribution

All experimental work was performed by Christopher Kuhs and Professor Amber Krummel. The differences in the IR spectra of VF and FV were explained using computational methodology developed and performed by Jakob Anderson and Professor Martin McCullagh. Descriptions of the experimental methods found in this chapter were influenced by Christopher Kuhs' Ph.D. dissertation. All other text was written by Jakob Anderson.

3.3 Introduction

Naturally occurring protein-based materials have garnered significant interest due to their ease of synthesis and appealing physical properties,[99] such as enhanced strength-to-weight ratio[33] and the combined hardness and flexibility[30,100,101] compared to synthetic materials. While large protein-based molecular motors have a wide variety of potential applications,[102] interesting physical prop-

erties have been demonstrated for self-assembling peptides as short as two amino acids of which diphenylalanine (FF) nanotubes is the exemplar.[60] The stability of FF self-assembled structures has inspired a diverse number of applications including use in drug delivery,[103] antibiotics,[104] hydrogel materials,[105] and piezoelectric materials.[63,64] Despite the relative simplicity of dipeptides, the holy grail of designing a dipeptide to produce an optimized self-assembled structure for a given application remains elusive. The gap in knowledge required for rational dipeptide design is due, in part, to the large number of factors governing the assembly of these materials and a lack of understanding of the fundamental driving forces involved in the self-assembly process.

Primary sequence of a peptide represents an easy synthetic handle that readily affects the aggregation propensity and degree of order in self-assembled structures. Various chemical modifications of FF have been studied due to the potential applications proposed for FF. While non-natural modifications have been studied,[106] natural amino acid substitution has a large impact on the assembled structure and is an easy variable to modify in the optimization of a material for a given activity. Slight sequence modifications can have a large impact on self-assembly as has been observed in both short[107] and long[108] peptides. However, the driving forces behind these differences are poorly understood. Computational approaches have been used to study the self-assembly of the full sequence space of dipeptides[32] and tripeptides in water.[109] The results have demonstrated some important trends, namely hydrophobic residues contribute significantly to aggregation propensity while the reversing of a sequence had little effect on the resulting structure. Important exceptions in which sequence dependence is observed include valine-phenylalanine/phenylalanine-valine (VF/FV) and tryptophan-phenylalanine/phenylalanine-tryptophan (WF/FW) where a difference in aggregation propensity is observed.[32] As neither VF nor FV dipeptides aggregate at 150 mM concentrations in D_2O, this enables isolation of differences between FF substitution of a valine residue on the N- versus the C-terminus. Additionally, this system allows for conservation of overall hydrophobicity of the dipeptide and removes noncovalent intermolecular interactions, simplifying the impact of the modification. Investigation into the termini difference behaviors within a dipeptide sequence would aid in the development of dipeptide self-assembly design principles.

An additional variable to consider for the rational design of dipeptide self-assembled materials is the solvent choice. When dissolved in water, FF will self-assemble into nanowires or nanotubes depending on the FF concentration.[16] While deionized water is often used, varying the salt concentration has been shown to produce long nanotube agglomerates.[13] A recent computational study investigating the underlying driving forces of FF self-assembly found that FF nanotube growth is driven by water mediated forces.[8] Water has further been suggested to be an important trace solvent for sizable and stable fiber formation.[36] Although water is a common solvent, additional solvents such as hexafluoroisopropanol, acetic acid, methanol, and formamide have also been shown to impact the resulting self-assembled morphologies in a variety of ways.[12] These previous studies on FF self-assembly in a range of solvents demonstrate that formation of macroscopic structures is attainable and the properties of these structures will be sensitive to the specific solvent environment chosen. Unfortunately, while the solvent impact on FF self-assembly has been extensively studied, few general rules have been extracted which allow one to predict solvents role on other peptide systems.

Chemical modification and choice of solvent are both variables which have a large impact on the self-assembly of dipeptides and are independent of each other. In order to build predictive design principles, we must first understand how a chemical modification will impact the solvent environment. Here we present a joint computational and experimental study on the sequence dependent behavior of VF and FV in water. Linear infrared (IR) spectroscopy and two-dimensional infrared (2D IR) spectroscopy techniques are ideally suited to reveal differences in solvation dynamics that are promoted by the dipeptide structural fluctuations. Computational and theoretical methods are ideally suited to investigate these systems with atomic resolution at the femtosecond to microsecond timescales. The combined structural sensitivity and high time resolutions of these techniques provide new insights into the way monomer conformations and solvation dynamics promote sequence dependent behaviors between the VF and FV dipeptides. The results herein demonstrate that VF and FV have differences in the linear spectra and differences in the underlying modes for the amide I stretch observed in the second derivative of the spectra. The lifetimes of the

underlying amide I modes of VF and FV measured through 2D IR demonstrates that one of the underlying modes for VF is longer lived than the similar mode for the FV dipeptide. Computational results demonstrate VF and FV sample different populations of similar monomeric conformations described by backbone dihedrals. Using a combination of all-atom molecular dynamics (aaMD) and electronic structure calculations, we hypothesize that the experimental underlying modes of the amide I stretch are caused by the solvent dynamics weighted by the monomeric conformations. These longer lived water interactions would thus imply the water is stabilized in VF relative to FV.

3.4 Methods

3.4.1 Experimental

VF and FV dipeptides were purchased from Bachem without further purification. Each dipeptide was dissolved in D_2O for a final concentration of 150 mM. Aliquots of the dipeptide solutions were placed between two calcium fluoride windows separated by a 25 μm Teflon spacer. Fourier transform infrared spectra (FTIR) were collected using a Vertex 70 with 1 cm^{-1} steps and averaging 64 spectra. Within the OPUS software package (Bruker Optics), second derivative spectra were calculated and smoothed with a Savitzky-Golay filter and scaled by a factor of -4.5 for VF and -1.5 for FV.

2D IR spectra were collected on a home built 2D IR spectrometer with a Ti:Sapphire laser operating at 1 kHz generating pulses centered at 790 nm and sub-50 fs pulse durations with 2.7 mJ of energy per pulse which was used to pump an optical parametric amplifier (OPA). The OPA generates mid-IR light centered at 6000 nm with 9 μJ of energy per pulse and a FWHM pulse duration of 100 fs. A 90:10 beam splitter was used to send 90% of the mid-IR light to a pulse shaper to produce the pump pulses necessary for the 2D IR experiments; the remaining 10% of the mid-IR light was directed into a beam line to produce the necessary probe pulse in the experiments. An acoustic optical modulator (AOM) pulse shaper was used to generate a pulse pair with a specified phase, amplitude, and time delay. The phase, amplitude, and time delays were controlled by driving a series of acoustic waveforms, or masks, across the germanium crystal in the AOM. The time

delays between the two pump pulses were scanned from 0 ps to 2.5 ps with 0.025 ps steps. Masks were generated to employ a four-frame phase cycling scheme and a 1400 cm^{-1} rotating frame to reduce both noise and the number of pulse delays required in the spectrum.[110] A computer-controlled delay stage temporally overlapped the pump and probe beams at the sample position. This stage also delayed the probe pulse from 0 ps to 4 ps with 0.1 ps steps for chemical exchange experiments. A half-wave plate and polarizer were placed in the pump line to rotate the polarization of the pump light to collect data for cross-polarized and co-polarized experiments. The 2D IR signal was frequency resolved by passing the time domain signal through a grating spectrometer (Triax 190, Horiba), which collected the signal on a 64-element mercury cadmium telluride (MCT) array detector. The geometry of the spectrometer taken together with the physical size and spacing of the MCT elements produces 5 cm^{-1} resolution along the probe axis of the 2D IR spectrum. The signal-to-noise ratio was improved by multiplying the signal from each pixel by a Hanning window function prior to taking the fast Fourier transform (FFT).[111] Magic angle spectra were calculated to remove the rotational contribution to the decay rates[112] using spectra collected in the co-polarized (XXXX) configuration and the cross-polarized (XXYY) configuration in accord with the equation XXXX - 2·XXYY. Additionally, a spline function was used to better differentiate the areas of integration in the 2D IR spectrum. Further details of our 2D IR spectrometer are described in previous work.[113,114]

3.4.2 Computational

Electronic Structure Calculations

Geometry optimizations and frequency calculations were performed at the ωB97xD/6-311G* level of theory using the Gaussioan09 software package.[77,78] All calculations used the PCM implicit solvent model in an effort to replicate solvent effects. All exchangeable hydrogens were replaced with deuterium to replicate experimental conditions. Vibrational frequency calculations were scaled by a factor of 0.959 which is the average needed to fit to the carboxylate peak between the two systems and is the standard range for scaling factors using comparable levels of theory.[94]

Molecular Dynamic Simulations

Individual dipeptide molecules in the L-enantiomer zwitterionic conformer were initially generated using the sequence command in tleap within the AmberTools 16 package.[81] All simulations were performed in explicit TIP3P water in the NPT ensemble using the GPU accelerated AMBER16 program.[81,83,84] Simulations, performed in duplicate, with one dipeptide were performed at a concentration of 4 mg/mL for 300 ns for both VF and FV. These systems were then simulated for an additional 10 ns with all degrees of freedom held constant with a 10 fs write frequency in order to capture solvent structure. With just the backbone dihedrals held constant, these systems were again simulated for 20 ns with a 10 fs write frequency to capture short-lived solvent lifetimes. The ff14ipq force field was chosen due to agreement between experiment and quantum calculations relative to other commonly used Amber force fields.[8,79,80] A Monte Carlo barostat set to 1 bar and a Langevin thermostat set to 298 K were used. All simulations used an integration timestep of 2 fs with the SHAKE algorithm restraining the bond lengths of bonds including a hydrogen atom. A non-bonding interaction cutoff of 12 Å was used in all simulations with the particle mesh Ewald method calculating long range electrostatic interactions.

The following protocols were performed prior to production runs of unbiased simulations. A steepest descent energy minimization protocol was first performed for 8000 steps, followed by the conjugate gradient method for 4000 steps. During this initial energy minimization, restraints were placed on the peptide atoms with a force constant of 100 kcal/(mol $Å^2$). The restraints were then removed and an energy minimization procedure was again performed with a steepest descent protocol for 12000 steps followed by a conjugate gradient method for 6000 steps. With restraints again placed on the peptide atoms with a force constant of 5 kcal/(mol $Å^2$), the systems were heated from 0 K to 298 K at constant volume. After heating, the volume was then allowed to relax for 40 ps with restraints kept on the peptides. To avoid transient effects due to initial configurations, production runs were started after 3 ns of unbiased simulation. The unbiased simulations of FV and VF monomer systems each were simulated for 600 ns for a total of 1.2 μs.

Umbrella sampling was performed along the N-terminal χ_1 coordinate space for FV and along the C-terminal χ_1 coordinate space for the VF dipeptide. Starting in the conformations produced through tleap, steered umbrella sampling was used to sample the χ_1 dihedral values with windows ranging from -180° to 180° in 10° increments. A harmonic potential was placed on each window of 0.12184 kcal/(mol deg^2). Each window was simulated for 50 ns for a total of 3.65 μs and a potential of mean force was generated using the weighted histogram method.[115]

Dihedral and hydrogen bonding analyses of simulated trajectories were performed using python 2.7 and the MDAnalysis module (version 0.18.0).[86,87] Hydrogen bonding environment lifetimes are reported with timescales being scaled by 2.729 to reproduce the diffusion of D_2O.[116,117] The VMD software was used to visualize trajectories and generate structural figures.[85]

3.5 Results and Discussion

3.5.1 Monomeric Configuration

Since it has been hypothesized that the VF and FV dipeptides would demonstrate sequence dependent behaviors, IR spectroscopy was used due to the conformational and solvation environment sensitivity of the technique. The FTIR spectrum for VF (black, Figure 3.1a) and for FV (black, Figure 3.1b) reports the amide I mode is located at 1661 cm^{-1} and 1665 cm^{-1} for the respective dipeptides while the carboxylate peak is located at 1597 cm^{-1} for VF and 1590 cm^{-1} for FV. Figures 3.1a and 3.1b also contain second derivative spectra (blue curves) that identify different vibrational modes that lie underneath a single FTIR peak. Each maximum of the second derivative spectrum corresponds to an underlying vibrational mode. From these second derivative spectra, the amide I mode of each dipeptide contains two distinct and separate vibrational frequencies. Peaks within the second derivative spectrum are located at 1653 cm^{-1} and 1666 cm^{-1} for VF while FV has peaks located at 1657 cm^{-1} and 1669 cm^{-1}. The second derivative within the carboxylate mode the FTIR spectra would suggest 2–3 underlying modes for VF and 1–2 modes for FV (Figures 3.1a and 3.1b respectively). We hypothesize that additional underlying modes introduced in VF are due to the relative position of the phenyl ring. While interesting, since there is more con-

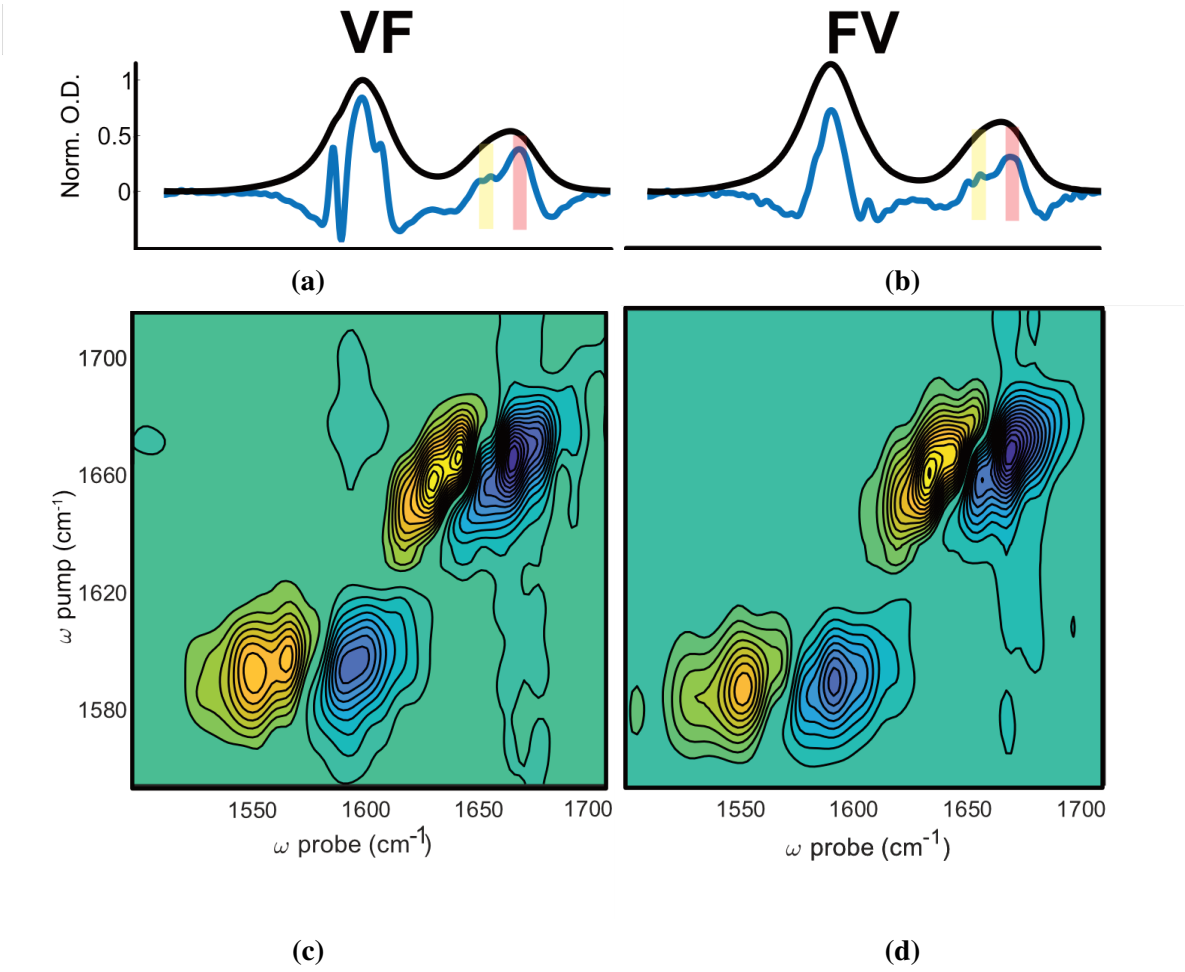
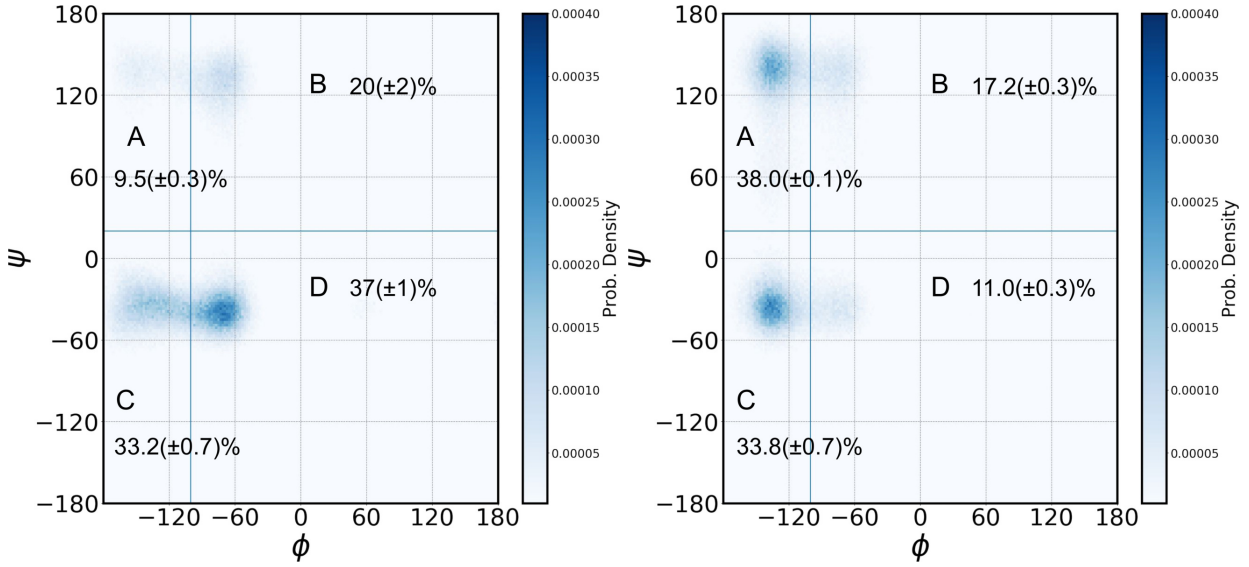


Figure 3.1: Linear infrared spectra and second derivative spectra of the amide I and carboxylate modes of (a) VF and (b) FV. 2D IR spectra for both modes for (c) VF and (d) FV.

sistency between the dipeptides within the amide I mode, for the remainder of this study we focus on the amide I mode. Figures 3.1c and 3.1d show 2D IR spectra of each dipeptide.

Differences in the sampled monomeric conformations between the dipeptides would cause drastic differences in the amide I environment and would thus have a large impact on the resulting experimental spectra.[112] This suggests the differences within VF and FV experimental spectra originate from differing monomeric conformations contributing to different water structure and dynamics around the amide I mode. Monomeric conformations can largely be described by ϕ and ψ backbone dihedrals as well as the phenylalanine residue χ_1 dihedral for each dipeptide, as other degrees of freedom will largely be included in these dihedrals or change minimally. Using



(a) VF **(b)** FV

Figure 3.2: Ramachandran plots of VF (a) and FV (b) with the probability of the dipeptide sampling a given monomeric conformational state as defined by cutoffs which separate peaks in the 1D probability density of each dihedral.

unbiased aaMD simulations of VF and FV monomers, Ramachandran plots defined by the ϕ and ψ dihedrals were generated with cutoffs between states being determined from one dimensional probability densities. As can be seen in Figure 3.2, both VF and FV have two ψ dihedral states existing around -50° and 150°, with FV being evenly distributed (55% and 45% of the respective sampling) in each state whereas VF predominately (70% of the frames) samples values centered around -50° (states labeled C and D in Figure 3.2a). Within the ϕ dihedral, VF samples states centered around -70° and -150° in comparable amounts (43% and 57% respectively). FV, however, predominately (72% of the frames) samples ϕ values centered around -150° (states A and C) with a shoulder around -70°. The Ramachandran plots in Figure 3.2 demonstrate that for both dipeptides, the combined ϕ and ψ dihedral states result in four sampled monomeric conformational states defined by these backbone dihedrals. The probability of sampling these states (ratio of frames sampled in a state and the total number of frames) is largely different between the VF and FV with the exception being state C which is sampled with equal probability for both dipeptides.

With clear backbone conformational states, the contribution of the sidechain to monomeric conformational states must then be addressed. Of the sidechain degrees of freedom, phenyl ring

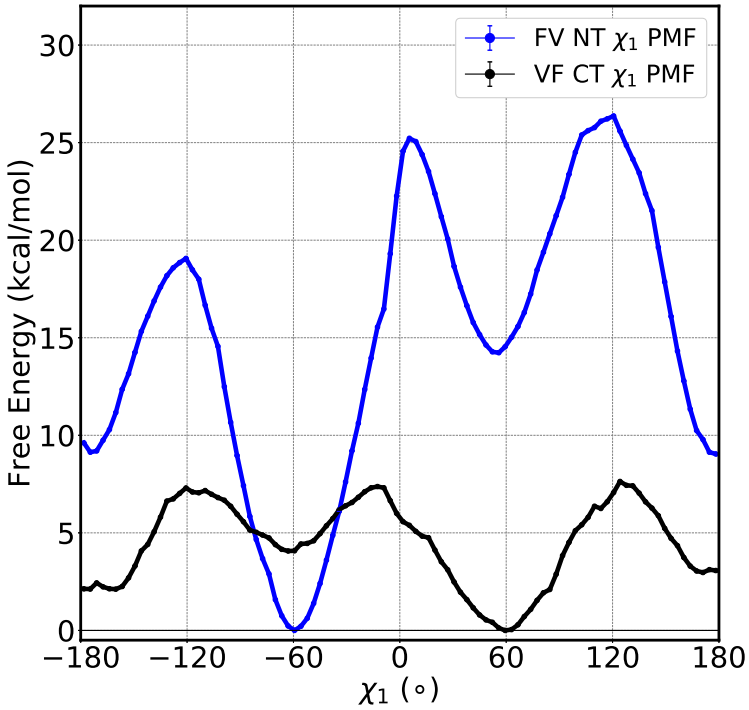
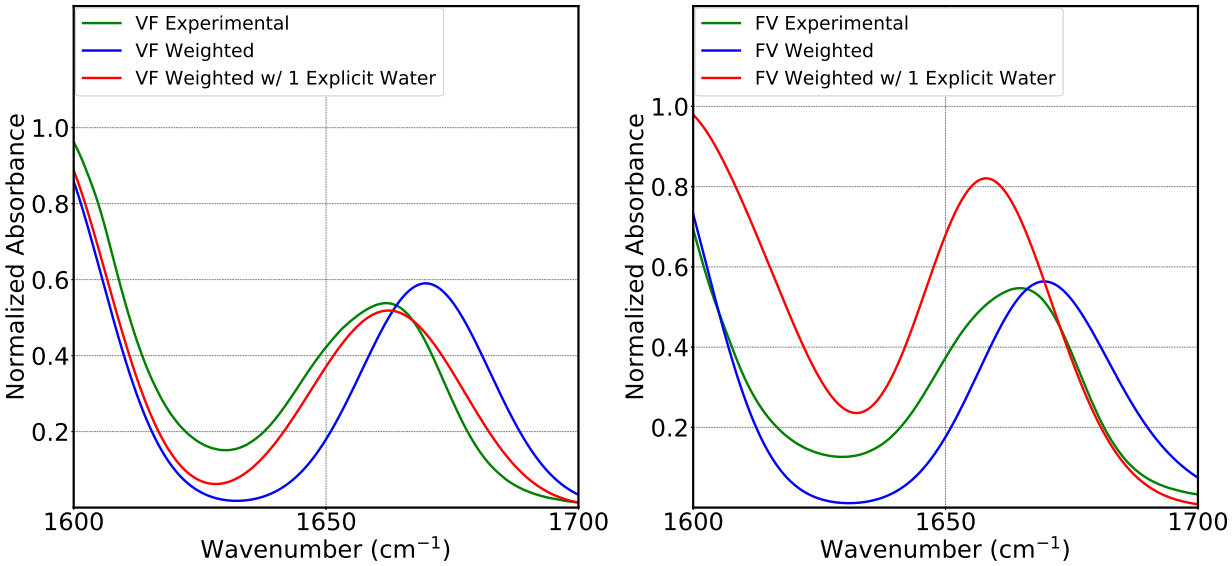


Figure 3.3: Relative free energy of the phenyl ring rotation as is described by the C-terminal χ_1 dihedral for VF (black) and N-terminal χ_1 dihedral for FV. Both dipeptides have three minima centered around -60°, 60°, and 180°.

rotation about the C_α–C_β bond would cause the largest difference in monomeric conformation and is described by the χ_1 dihedral. Umbrella sampling was performed over the phenylalanine residue χ_1 dihedral to compare the amount the phenyl ring interacts with the amide carbonyl between the two systems and to determine the probability of χ_1 dihedral conformational sampling. As is shown in Figure 3.3, the free energy barrier heights for three states of the N-terminal χ_1 for FV (blue) are such that, once the χ_1 dihedral samples about -60°, it is highly improbable for it to exit this conformation. Indeed, FV remains in this global minimum in all systems where this dihedral is unbiased and is initiated in this conformational state. While VF (black) also has three χ_1 states, these states have much lower barriers such that transitions between these states are sampled. Since the χ_1 dihedral sidechain states of both dipeptides are both slow and the number of relevant states are inconsistent (only one relevant state in FV and three in VF), we simplify the monomeric conformational states to the backbone states while we note that one could include



(a) VF **(b)** FV

Figure 3.4: Computed spectra convolved over all four monomeric conformational states of VF (a) and of FV (b) in implicit solvent, with (red) and without (blue) an explicit water hydrogen bonding to the amide carbonyl and compared to experimental FTIR spectra (green).

additional conformational states. Additionally, Figure 3.3 demonstrates that for the experimental systems, the phenyl ring in FV is not likely to impact the solvent environment of the amide bond while the opposite is true for VF.

To test the physical relevance of our chosen monomeric conformations, representative structures from each of the four backbone conformations for both dipeptides were geometry optimized and their linear IR spectra were generated from subsequent frequency calculations. This was done for all eight monomeric backbone conformations in implicit solvent with and without an explicit water hydrogen bonding to the amide carbonyl. As shown in Figure 3.4a, the computed spectrum of VF measured with one explicit water (red) agrees well in both normalized peak height and scaled peak width to the experimental IR spectrum (green). The computed spectrum of VF with only implicit solvent (blue), however, is blue shifted and has a slightly larger peak height relative to the experimental spectrum. The computed spectrum of VF would then suggest that VF in experimental systems is more probable to be found with one water hydrogen bonding to the amide carbonyl. Computed spectrum of VF further demonstrates that, while the monomeric conformations found

from MD simulations are a good approximation to experimental systems, the inclusion of explicit water increases the agreement. In contrast to VF, the computed spectrum of FV (Figure 3.4b) with one explicit water (red) has a much higher normalized absorbance and is red shifted relative to the experimental spectrum (green). The computed spectrum of FV in implicit solvent (blue) agrees better, relative to with an explicit water, to the experimental spectrum (green) in peak height but is also slightly blue shifted. Since FV's computed spectrum is not in as good of agreement with the experimental spectrum relative to VF, we hypothesize that either thermal fluctuations within the system or additional missing waters are giving rise to this discrepancy. The computed spectrum of FV does, however, suggest FV in experimental systems has a solvent environment more like the implicit spectrum than with an explicit water. Interestingly, for both dipeptide systems the addition of an explicit water red shifts the spectra with peak separation distances close to the peak separation distances seen in the second derivative spectra (Figure 3.1), suggesting the experimental amide I underlying modes are dependent on both whether or not there is an explicit water hydrogen bonding to the amide carbonyl and on the ensemble of monomeric conformations. In total, these results suggest VF is more likely to have an explicit water hydrogen bonding to the amide carbonyl while FV does not.

3.5.2 Solvent Dynamics

The computed spectrum's of VF and FV suggest that the two systems not only sample different monomeric conformations, but that their water environment around the amide is also different. Using the frequency positions of the underlying amide I modes as defined from the second derivative spectra, boxes of integrated intensities in the 2D IR spectra are specified and are shown in Figure 3.5. Figure 3.5a contains 2D IR spectra of the amide I mode of VF and Figure 3.5c the 2D IR spectra of FV. The red circles on the spectra represent the areas that were integrated to determine vibrational lifetimes of the individual modes. Figures 3.5b and 3.5d shows the decay curves for VF and for FV which were fit to single exponentials to determine the vibrational lifetimes for the different transitions and are reported in Table 3.1. Rate constants associated with chemical exchanges

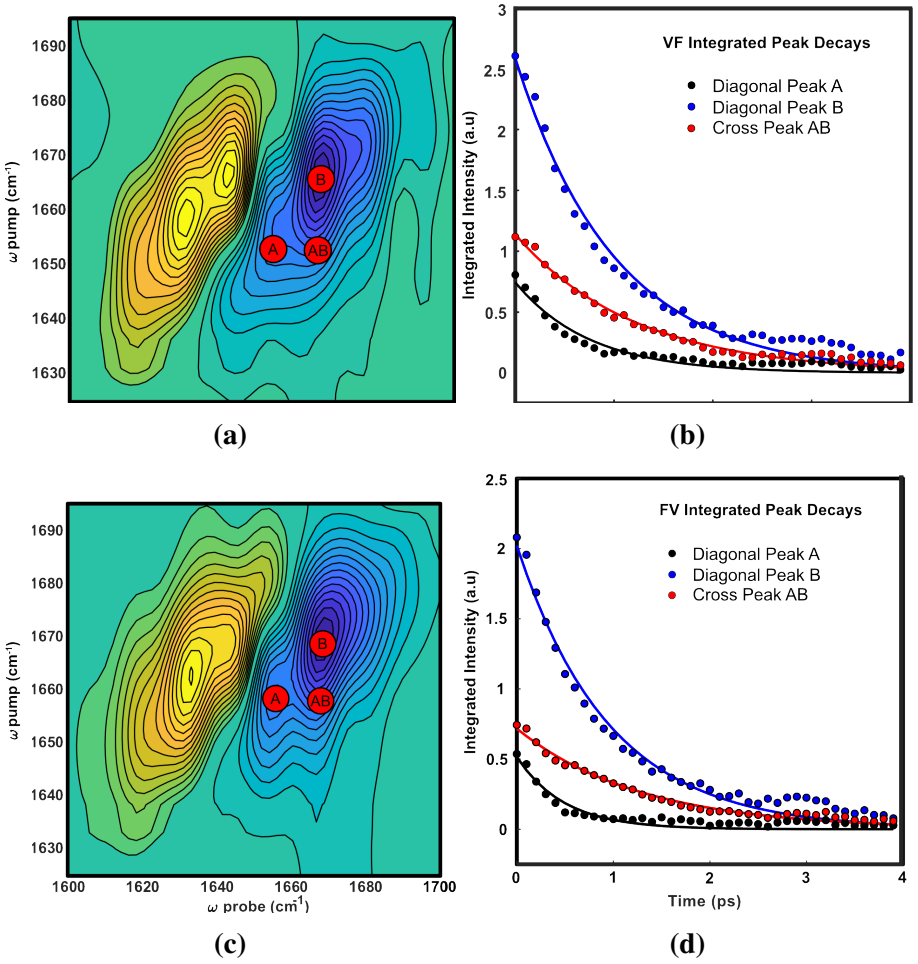


Figure 3.5: Amide I region of (a) VF and (c) FV. Within (a) and (c), the red circles indicate were the integration box was defined for the on diagonal peaks and cross peaks. The integrated peak decays for (b) VF and (d) FV from vibrational lifetimes were calculated for peaks A (black), B (blue), and AB (red) found in (a) and (c).

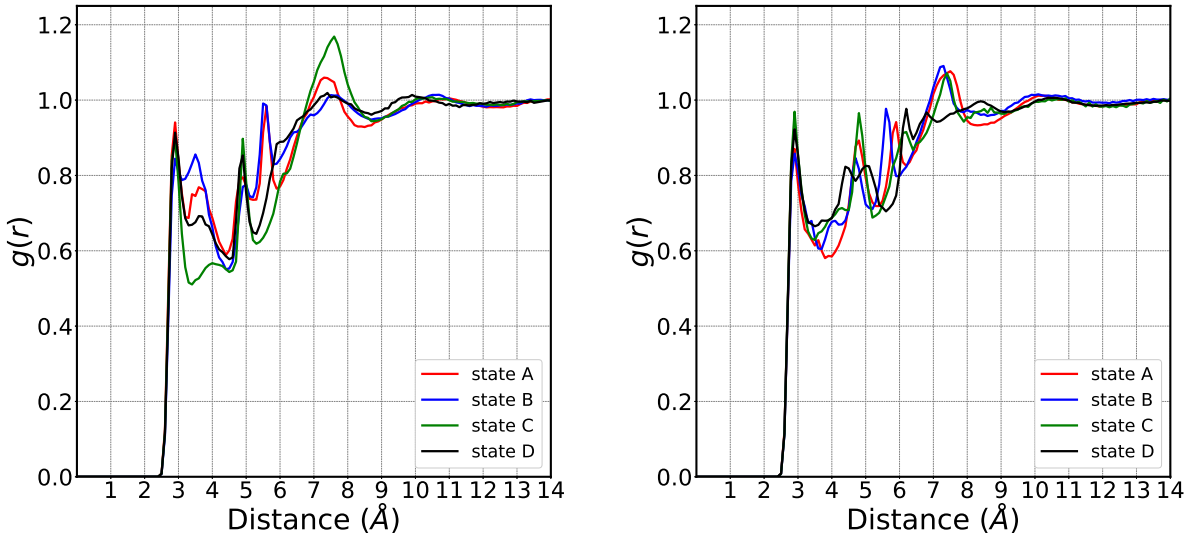
were then determined by plugging the vibrational lifetimes of the different transitions into

$$k_{CP} = \frac{k_B + k_A}{2} - \sqrt{\left(\frac{k_B - k_A}{2}\right)^2 + k_{AB}k_{BA}} \tag{3.1}$$

where k_A and k_B are the respective decay rates for underlying modes A and B and k_{CP} is the cross-peak decay rate. Solving for $k_{AB}k_{BA}$ and assuming k_{AB} is equal to k_{BA}, the exchange rate constant for VF was determined to be $(3.5 \pm 0.6 \text{ ps})^{-1}$ while FV has a faster exchange rate of $(1.7 \pm 0.3 \text{ ps})^{-1}$.

Table 3.1: Vibrational lifetimes determined from 2D IR waiting time experiments of both dipeptide isomers.

	Transition A (ps)	Transition B (ps)	Cross Peak AB (ps)
VF	0.75 ± 0.08	1.00 ± 0.05	1.21 ± 0.05
FV	0.50 ± 0.07	0.97 ± 0.05	1.7 ± 0.3



(a) VF **(b)** FV

Figure 3.6: Radial distribution functions of the amide carbonyl oxygen to all water oxygens in each of the geometry optimized backbone states for both the VF (a) and FV (b) dipeptides.

The experimentally observed transition rate timescales are on the water hydrogen bond dynamic timescales (sub picosecond), and thus, the experimental spectra is hypothesized to be described by the combination of backbone monomeric conformations and water solvation structure dynamics. To study the structure of water around the amide carbonyl, biased monomer simulations were performed on the four states of each dipeptide, forcing the dipeptides to only sample the optimized structures. The water density was then considered by measuring the radial distribution functions ($g(r)$'s) of the oxygen of the amide carbonyl to the water oxygens. As is demonstrated in Figure 3.6a, the water around state A, B, and D of VF are more structured with two peaks while all the states of FV (Figure 3.6b) have only one peak under 4 Å. At distances larger than 9 Å, both dipeptide densities go to a $g(r)$ value of one. While the $g(r)$'s between each of the states

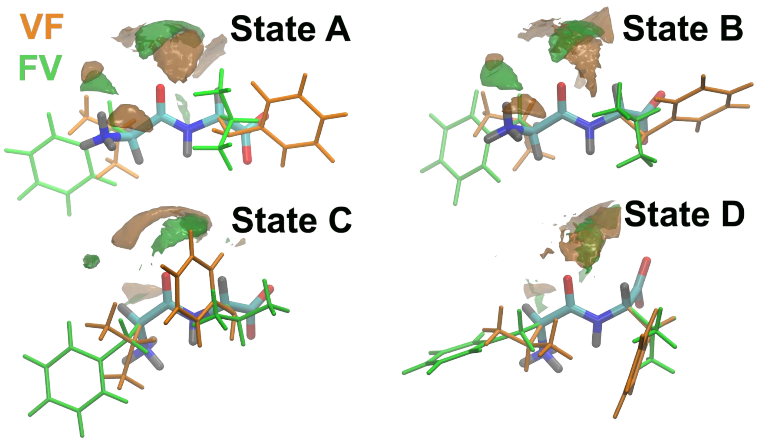


Figure 3.7: Comparison of 3-dimensional water density around the amide carbonyl of each geometry optimized backbone state of both VF (orange) and FV (green) dipeptides.

have differences, within a cutoff of 3.25 Å, both dipeptides have one peak ($\sim$ 2.9 Å) suggesting this to be the first solvation shell. Water density can be further investigated by considering the three-dimensional density. The water density shown in Figure 3.7 is five times that of bulk density within 2.85 Å of the amide carbonyl. The largest difference between the dipeptides is for state C, which is more localized for FV (green) and is a larger band of density for VF (orange). Visual inspection of VF in state C indicates this larger band of water density contains waters within interaction distances of both the amide carbonyl group and the phenyl ring. Potential interactions with the phenyl ring influencing the amide carbonyl environment, motivates non-restrained sampling of the sidechain degrees of freedom.

Since sidechain fluctuations are hypothesized to impact water dynamics, the ϕ and ψ dihedrals were held constant at the optimized values for each of the states, water dynamics were studied with additional simulations with all other degrees of freedom allowed to be freely sampled. Simulations of each system were then performed for 22 ns with frames written every 10 fs for a total of 2.2 million measurements per system. For a given water with a water oxygen to amide oxygen distance within 3.25 Å, the length of time this water remained within the cutoff distance was recorded. Residence times were then averaged for each state, these averaged residence times were then weighted by the conformational sampling probability to determine the overall residence time for the dipep-

Table 3.2: Water residence lifetimes around the amide carbonyl for all four backbone states of both dipeptides, as well as the total, state probability weighted water residence lifetime of each dipeptide.

	Lifetime of water (ps)	
	FV	VF
state A	0.569 ± 0.003	0.656 ± 0.003
state B	0.659 ± 0.003	0.595 ± 0.003
state C	0.653 ± 0.003	0.707 ± 0.003
state D	0.599 ± 0.002	0.684 ± 0.003
State Probability Weighted	0.616 ± 0.001	0.671 ± 0.002

tide. As shown in Table 3.2, for states A, C, and D, VF has longer lived water residence times than FV, consistent with experimental lifetimes. While state B for FV has a longer-lived water residence than in VF, FV only samples this state 17% of the frames such that the overall water residence lifetimes are qualitatively consistent. As the rates seen from MD and experiment agree in trend and sup-picosecond timescale, these results suggests that the dipeptides primary sequence, namely whether the phenylalanine residue is on the N-terminus or the C-terminus, impacts the monomeric conformations sampled as well as a difference in the water dynamics sampled around the dipeptide.

3.6 Conclusion

Linear FTIR spectra of VF and FV dipeptides demonstrates that the amide I environment is significantly different between the two dipeptides with underlying mode differences highlighted by the second derivative of the spectra. Vibrational lifetimes of the underlying modes within the amide I mode measured from 2D IR experiments are longer lived for VF, suggesting a slower exchange rate, than for FV. A combination of QM and aaMD techniques were performed to investigate the origin of the differences observed experimentally. Unbiased simulations suggest the two dipeptides sample discrete backbone conformational states. Additional umbrella sampling simulations of the N-terminal χ_1 for FV suggest it is likely in one conformational state while for VF the C-terminal

χ_1 has lower energy barriers enabling sampling of additional sidechain conformational states. To be consistent between the two dipeptides, we approximated VF and FV monomer conformations by the ϕ and ψ dihedral states while noting additional conformational states are possible. Representative structures from the backbone monomeric states were geometry optimized from which linear IR spectra were then calculated. Computed spectra of each state, weighted by the probability density, and combined to form the total spectrum reproduces experimental spectrum well for VF when there is one explicit water hydrogen bonding to the amide carbonyl. However, computed spectra performed in the same way for FV is in better agreement with experimental spectra when there is only an implicit solvent rather than when there is one explicit water system indicating the amide carbonyl is more solvated for VF than for FV. The agreement between computed and experimental spectra suggests the combination of monomeric states and solvent environment give rise to differences in the environment surrounding the amide I mode between the two dipeptides. Using the $g(r)$'s of the amide carbonyl oxygen to oxygen of water, we were able to determine the first solvation shell to be within 3.25 Å. Visual inspection of the water density demonstrates the phenyl ring in VF influences the water dynamics around the amide oxygen, suggestive of influencing the amide I stretch. Using the first solvation shell as the cutoff, the lifetimes of waters around the amide carbonyl are shown to be longer lived for VF than FV, consistent in magnitude and trend with experimental rates. Together these results suggest the phenylalanine residue on the C-terminus of VF increases the water residence time around the peptide bond. An increase in water residence time around the amide carbonyl suggests a stabilization of water around VF which would then suggest the free energy barriers for aggregation to be higher, and thus the aggregation propensity to be lower than for the FV dipeptide. From our results we hypothesize that chemical modification of the FF N-terminal residue will result in a decrease aggregation propensity and the order of an aggregate.

This work forms the foundation for future studies to investigate the impact sequence dependence may have on other solvents. Future studies will investigate if the water dynamics for other functional groups are also impacted by sequence to more completely study the relationship be-

tween chemical modification and solvent environment. The results presented here demonstrate when a phenylalanine residue is on the C-terminus, the water dynamics will be slower than if there is a valine residue. This result highlights that while the overall hydrophobicity of the molecule remains constant between the two dipeptides, the sequence impacts the monomeric conformations sampled and the solvent dynamics around the dipeptide.

Chapter 4

Investigation of Sequence Dependence in the Dimerization of Valine-Phenylalanine, Phenylalanine-Valine, Isoleucine-Phenylalanine, and Phenylalanine-Isoleucine.[3]

4.1 Overview

The self-assembly of diphenylalanine (FF) has a broad list of potential applications due to the formation of the highly stable and ordered nano-structures. To rationally design a FF derivative optimized for a given application, the impact of chemical modification must be understood. While FF and FF derivatives have been studied extensively, inferred design principles are inconclusive as a lack in experimental design consistency has yielded comparisons in which the multiple variables are changed. Using the free-energy decomposition analysis described in Chapter 2 and by considering the differences in N- and C-terminal modification of FF, this chapter connects Chapter 2 and Chapter 3. Presented here is the dimer potential of mean force (PMF) of valine-phenylalanine (VF), phenylalanine-valine (FV), isoleucine-phenylalanine (IF), and phenylalanine-isoleucine (FI). From the dimer PMFs of FV and FI, it is suggested that FF modification of the C-terminus to a residue smaller than phenylalanine enables the energetic possibility of tighter dipeptide packing. Decomposition of the VF, FV, IF, and FI PMFs demonstrates that backbone electrostatics predominantly drive dimerization of dipeptides. Additionally, it is shown that solvent-mediated forces hinder the dimerization of VF while solvent–solute Lennard-Jones forces promote dimerization of FV, IF, and FI. We postulate that the solvent-mediated force impediment of VF is a result of

[3]Jakob Anderson[a], Martin McCullagh[b]; [a] Department of Chemistry, Colorado State University, Fort Collins, CO, USA, [b] Department of Chemistry, Oklahoma State University, Stillwater, OK, USA

longer lived water residence lifetimes around the carbonyl oxygen of the peptide bond relative to FV as seen in Chapter 3. We further hypothesize that since this hindrance is not seen for IF, the solvent shell sequence dependence seen for VF and FV is also a function of steric sidechain size. The results herein are a foundation to understanding the impact of chemical modification of either residue on dipeptide self-assembly.

4.2 Introduction

The self-assembly of diphenylalanine (FF) produces macroscopically observable nanotubes which are highly ordered and are remarkably stable when exposed to environmental or mechanical stress.[21–25] Due to the chemical properties of FF, there have been a wide variety of proposed applications[21] for FF nanotubes including drug delivery[62], antibacterial[61] use, and hydrogel materials[105] amongst others. FF, however, is not optimized for all of the proposed applications resulting in numerous FF derivatives being studied with the hope of measuring properties which are more suited for a given application.[21,26,118–121] While FF nanotubes and FF derivatives have been extensively studied both experimentally and computationally, there is often a lack of consistency in important factors such as concentration, choice of solvent, zwitterionic state, and methodology resulting in muddled comparisons. In order to rationally design a FF derivative specifically optimized for a given application, FF and perturbations of FF must be studied in a consistent manner so as to isolate the impact of individual variables on the self-assembly process.

The foundation for developing design principles for FF self-assembly has been the focus of previous chapters. In Chapter 2, it was concluded that the initial aggregation steps of FF are driven by the electrostatics of the backbone atoms and that a reorientation event in the dihedral phase space of the N-terminal sidechain must occur prior to further ordering. Solvent mediated forces were then shown to drive monomeric addition to a FF nanotube illustrating the complexity of FF self-assembly. A multitude of analyses were performed on the FF systems in a generalized way which enables analysis and comparison of different dipeptide systems. In particular, the potential of mean force (PMF) decomposition analysis from Chapter 2 is used for this study. In Chap-

ter 3, experimental differences in FTIR and 2D IR spectra between valine-phenylalanine (VF) and phenylalanine-valine (FV) were validated and explained through a combination of QM calculations and MD simulations. Computational results showed a primary difference between VF and FV is the sampling of monomeric conformations described both by the backbone and sidechain dihedral phase spaces. Although the backbone conformational space of VF and FV sampled the same states, the probabilities for sampling the given backbone dihedral states differed between the two dipeptides. Additionally, the N-terminal χ_1 of FV sampled within one free energy minima while the C-terminal χ_1 of VF sampled across the dihedral phase space demonstrating the rigidity of a N-terminal phenylalanine residue. The differences in the sampling of the conformational phase space between VF and FV resulted in longer lived water residence time around the amide bond of VF which was in agreement with the lifetimes determined from the 2D IR spectra. The results from Chapter 3 highlight the fact that while the chemical modification of the functional groups may have been the same, valine substitution of FF at the N-terminus results in drastically different dipeptide properties relative to substitution at the C-terminus.

This chapter connects Chapter 2 and Chapter 3 by considering the sequence dependence of dipeptide dimerization through comparison of the dimer PMF for FF to the dimer PMFs of VF, FV, isoleucine-phenylalanine (IF), and phenylalanine-isoleucine (FI). The isoleucine amino acid was chosen for FF substitution as it is a hydrophobic amino acid which differs from valine by a methyl functional group. As all dimer dipeptide systems studied here are solvated in water at a concentration of 8 mg/mL, as few variables as possible will be altered relative to previous chapters enabling a foundation for future systematic investigations towards developing dipeptide design principles. After discussing the methods used in our simulation study, dimerization of each of the systems are compared, followed by an investigation of the underlying driving forces and trends therein between the systems. Our results suggest that while there are similarities in the dimerization processes of VF, FV, IF, and FI, drastic differences in the underlying driving forces suggests FF modification of the C-terminus will promote dimerization more than FF N-terminal modification.

4.3 Methods

Dipeptide molecules were generated using the sequence command in tleap found within the AmberTools 16 package.[81] Each dipeptide is in the L-enantiomer conformation and zwitterionic charge state. Umbrella sampling simulations were performed in explicit TIP3P water at a concentration of 8 mg/mL using the GPU accelerated Amber16 program.[81,83,84] Simulations were performed in the NPT ensemble with a Monte Carlo barostat set to 1 bar and a Langevin thermostat set to 298 K. The ff14ipq force field was chosen so as to be able to compare to previous work[8] and because of agreement between experiment and quantum mechanical calculations relative to other commonly used Amber force fields.[79,80] All simulations employed an integration timestep of 2 fs with the SHAKE algorithm restraining the bond lengths of bonds which included a hydrogen atom. The particle mesh Ewald method of calculating all long range electrostatics was used for non-covalent bond interaction distances greater than 12 Å.

Prior to starting umbrella sampling windows, the following protocol was used for each dipeptide. Energy minimization was initially performed using the steepest decent protocol for 8000 steps followed by the conjugate gradient method for 4000 steps. Restraints were placed on the peptide atoms during this energy minimization with a force constant of 100 kcal/(mol·Å^2). The energy was minimized again, this time with the restraints removed, using the steepest decent protocol for 12000 steps followed by the conjugate gradient method for an additional 6000 steps. The peptide atoms were then restrained with a force constant of 5 kcal/(mol·Å^2) as the systems were heated from 0 K to 298 K at constant volume. Once heated to room temperature, the volume was allowed to fluctuate for 40 ps with restraints maintained on the peptides. After 3 ns of unbiased simulation, production runs were started and performed for 300 ns for each dimer dipeptide system for a total of 1.2 μs. The umbrella sampling windows were then pulled from these unbiased simulations.

Umbrella sampling was performed along the dipeptide–dipeptide center of mass separation distance for distances between 2.5 Å and 17.5 Å with windows in 0.5 Å increments for each dipeptide system. Each system was simulated for 30 ns per window with a force constant of 20 kcal/(mol·Å^2). Additional windows and higher force constants were used in some instances

to ensure proper overlapping of windows. Umbrella sampling for each dipeptide dimerization resulted in at least 900 ns for a total of more than 3.6 μs of simulation. Using the free energy decomposition analysis described in Chapter 2, the PMF for dimerization of each dipeptide was computed and decomposed into solvent–solute and solute–solute contributions, both of which were then further decomposed into Lennard-Jones and Coulombic components of the free energy. The VMD software was used to visualize trajectories and generate representative structures.[85] All PMFs were plotted using Matplotlib.[88]

4.4 Results and Discussion

What follows is a breakdown of each of the dimer PMFs for VF, FV, IF, and FI as well as a comparison of the driving forces for dimerization through the decomposition of each PMF. While it may be appealing to compare the separation distances of specific features between the PMFs, the dipeptide heavy-atom center of mass chosen as the collective variable prevents direct comparison. Instead, only the general conclusions from the PMFs should be compared (i.e. the driving forces of dimerization). The authors considered choosing the dipeptide backbone heavy-atom center of mass separation distance for the collective variable, however, the PMF decomposition analysis described in Chapter 2 can only be performed when the whole molecule of each monomer is considered. Additionally, choosing the dipeptide center of mass separation distance as the collective variable enables comparison of the dimerization driving forces for the dipeptides considered here to those of FF in Chapter 2. To enable ease of reference, some of the results from the FF dimer PMF found in Figures 2.2 and 2.3b are recreated and shown in Figure 4.1 (see Chapter 2 for full details).

4.4.1 Dipeptide Dimerization Sequence Dependence

The dimer PMFs for VF, FV, IF, and FI are shown in Figure 4.2 while the distance and energy specifics of the common features are detailed in Table 4.1. The dimer PMFs for IF and FI suggest substitution of the FF C-terminus to isoleucine decreases the well depth of the PMF compared to the N-terminus substitution which increases the well depth. While enticing to report this as the

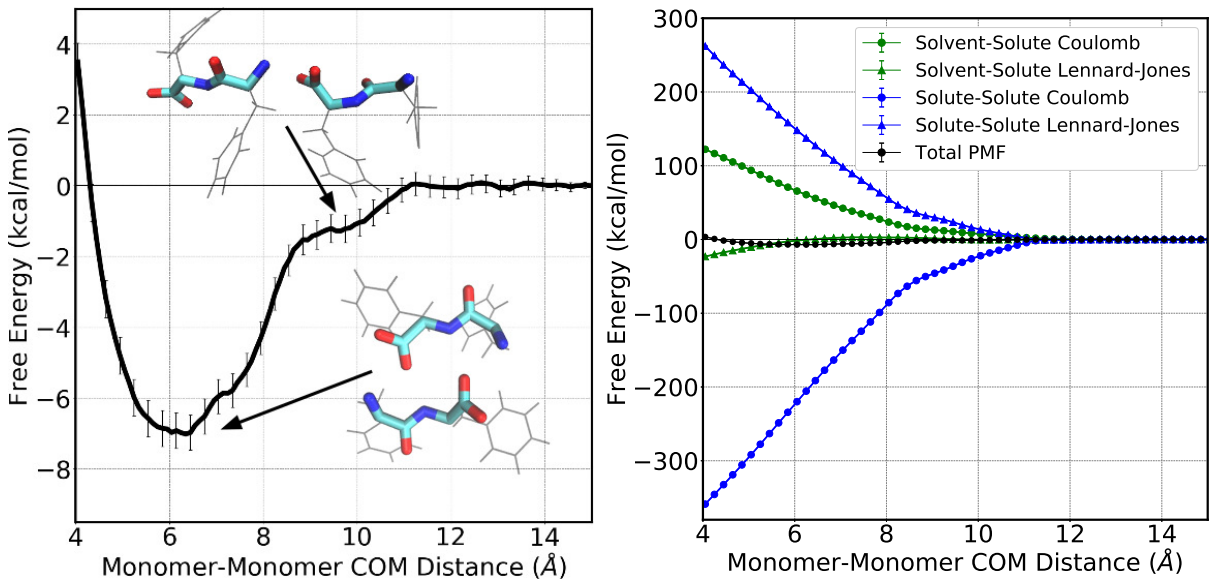
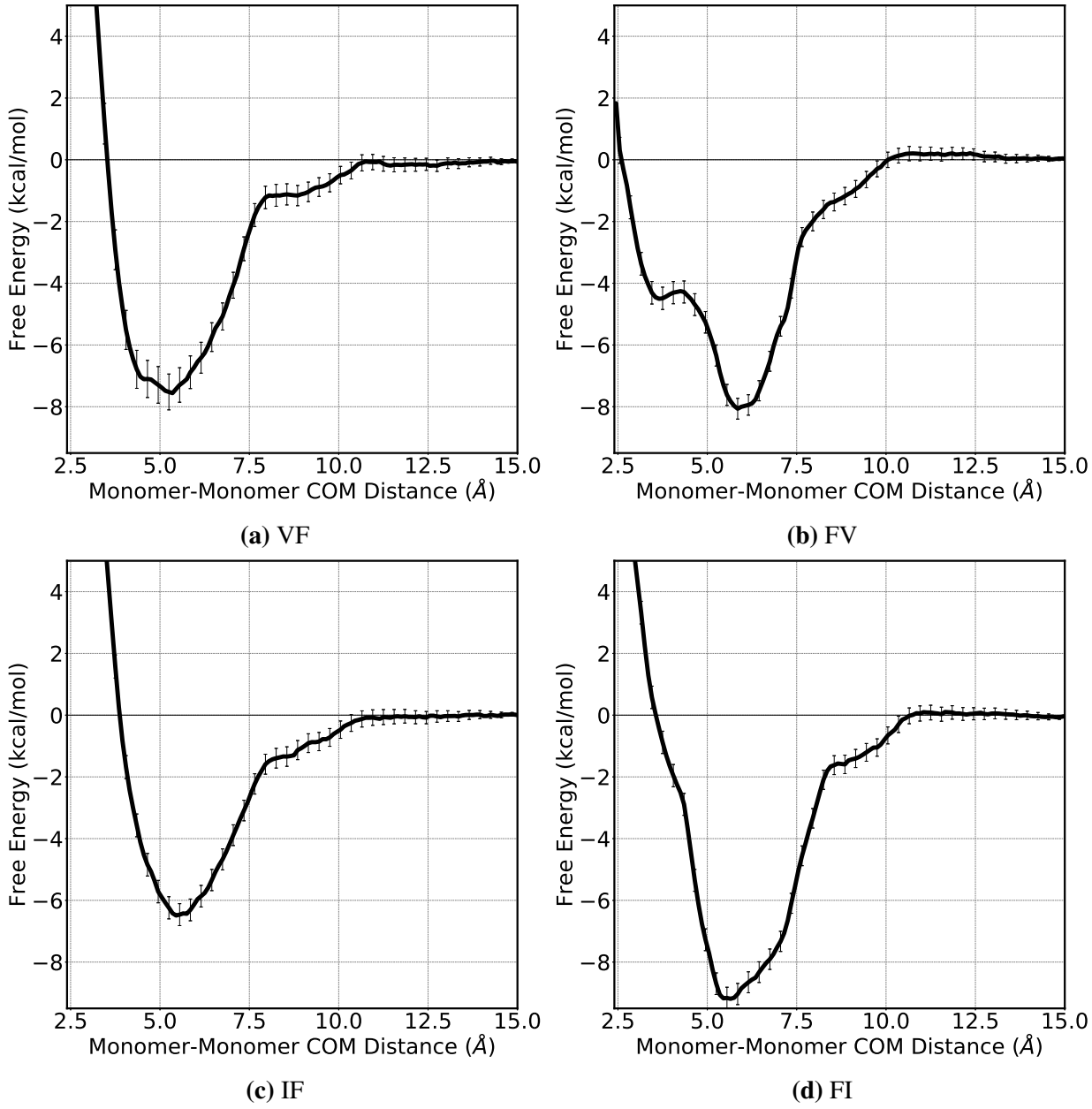


Figure 4.1: (a) The relative free energy of FF dimerization as a function of FF heavy atom center of mass separation. The PMF has a global well at 6.25 Å with a depth of 7.0(±0.5) kcal/mol and a shoulder at 9.75 Å with a depth of 1.03 kcal/mol. (b) The relative free energy of FF dimerization (black curves) as a function of FF heavy atom center of mass separation decomposed into solvent–solute (green curves), solute–solute (blue curves), Lennard-Jones (triangles), and Coulomb (circles) components of the free energy. Error bars are present but are smaller than the symbols or lines. Reproduced with permission from Anderson, J., & McCullagh, M. (2018). Initial Aggregation and Ordering Mechanism of Diphenylalanine from Microsecond All-Atom Molecular Dynamics Simulations. Journal of Physical Chemistry B, 122(51), 12331–12341.https://doi.org/10.1021/acs.jpcb.8b10335. Copyright 2018 American Chemical Society.

case for VF and FV, the global wells are within error of one another. The FV dimer PMF contains a second well located at 3.7 Å with a well depth of -4.5(±0.4) kcal/mol. Additionally, in the FI dimer PMF there is a small bank located around 4 Å demonstrating that there is a significant impact to dimerization for C-terminal substitution. This additional shoulder has been seen previously for dialanine, although, it is unclear if the N- or C-terminus is enabling this shoulder to be present in this case.[31] Since the phenylalanine residue is energetically trapped pointing towards the N-terminus for FV as is demonstrated in Figure 3.3, we assume this to be true for FI. Thus, we hypothesize that the reduction in steric size of the sidechain when modified from phenylalanine to valine or isoleucine enables tighter packing of the FV and FI systems relative to FF, VF, and IF.



(a) VF **(b)** FV

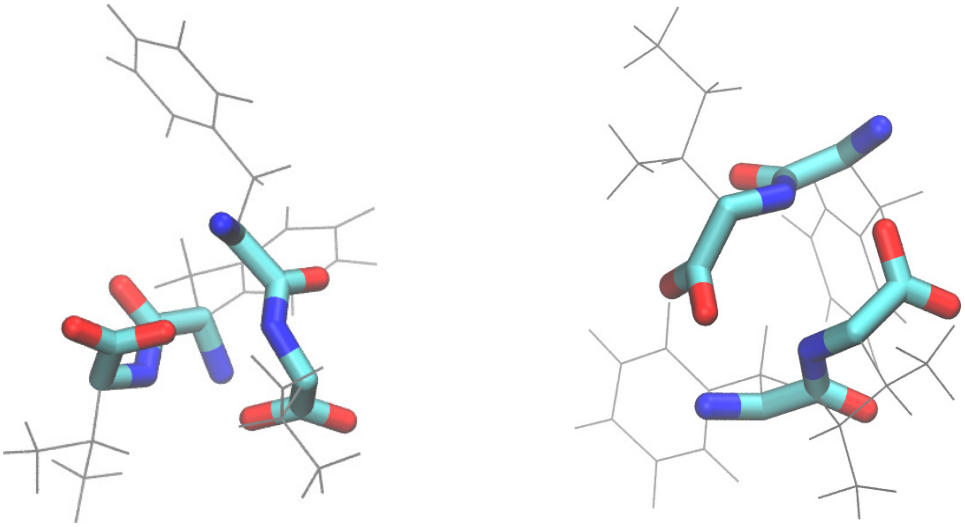
(c) IF **(d)** FI

Figure 4.2: The relative free energy of (a) VF, (b) FV, (c) IF, and (d) FI dimerization as a function of respective dipeptide monomer heavy atom center of mass separation. These PMFs are computed by integrating the mean force as described in Chapter 2. The distance and corresponding relative free energies for the common shoulder and global well of each PMF can be found in Table 4.1.

The shoulders at distances greater than the global minimum seen in each of the dimer PMFs in Figure 4.2 are comparable in shape to the shoulder seen in FF (Figure 4.1a) and are energetically within error. Representative structures of this shoulder contain the N-terminus of one dipeptide

Table 4.1: The dipeptide center of mass separation distance for the global wells and common shoulder range of VF, FV, IF, and FI as well as the corresponding free energies. Distances are reported for clarity, not comparison.

	VF	FV	IF	FI
Global well dist. (Å)	5.35	5.90	5.45	5.65
Global well depth (kcal/mol)	-7.6(±0.6)	-8.1(±0.3)	-6.5(±0.4)	-9.2(±0.3)
Shoulder range (Å)	8–10.5	7.5–10	8–10.5	8.5–10.5
Shoulder depth (kcal/mol)	-1.2(±0.4)	-2.0(±0.3)	-1.6(±0.3)	-1.6(±0.3)



(a) FV **(b)** FI

Figure 4.3: Representative figures of the 4 contact point anti-parallel dimer interaction type for (a) FV and (b) FI hypothesized to be the source of the local minima and shoulder for distances less than the global minima for these dipeptides.

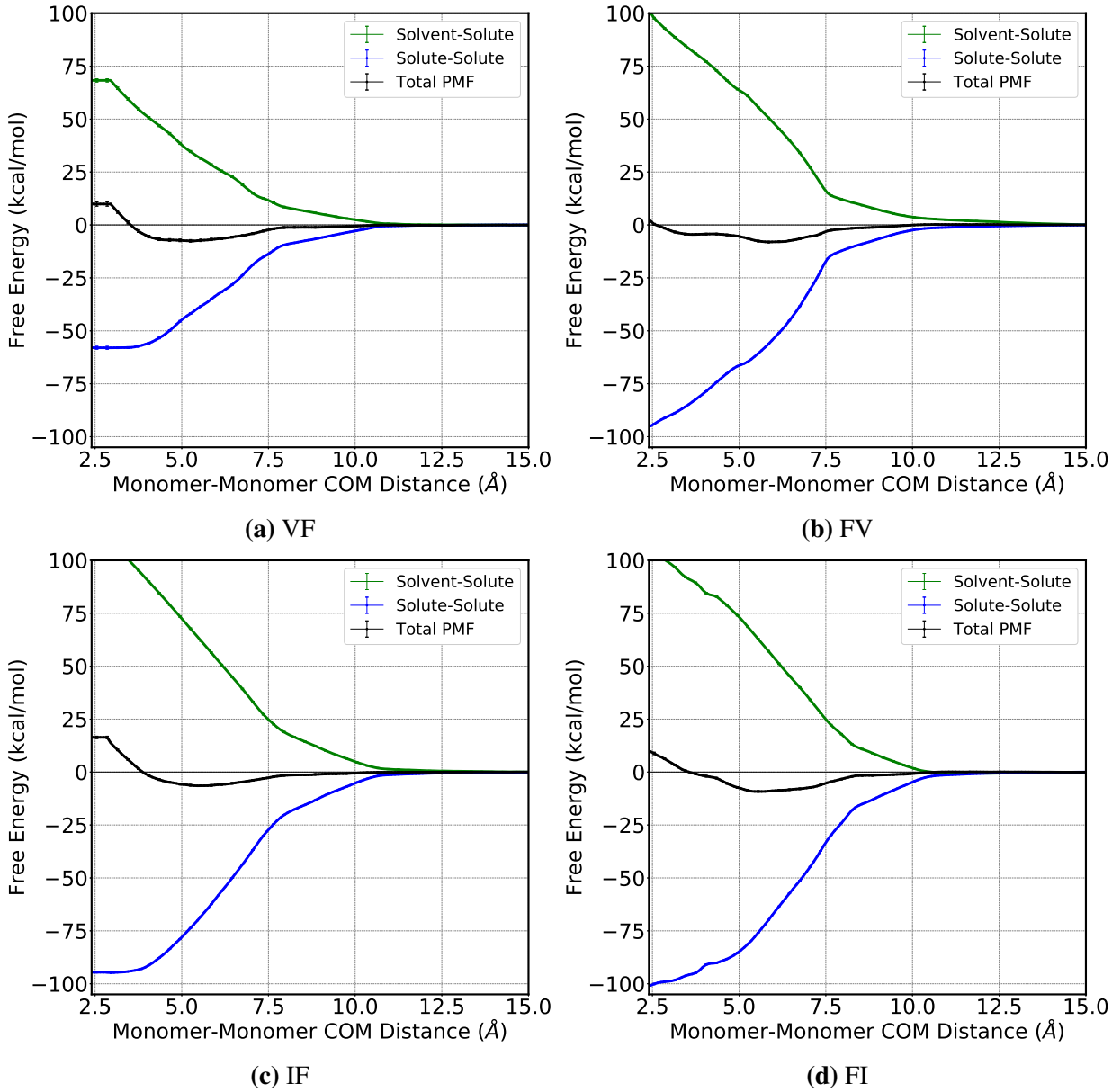
interacting with the C-terminus of the other dipeptide, similar to the dimer interactions seen in the dimer PMF of FF. For distances at the global minimum, representative structures for each of the dipeptides are also comparable to the the anti-parallel stacked dimers seen for FF (Figure 2.7a) in which both termini of one dipeptide are interacting with the oppositely charged termini of the other dipeptide. Neither FF, VF, or IF have an additional well or shoulder at distances lower than the global well, however, there is a local well and a slight shoulder for FV and FI respectively. Upon visual investigation, there are dimer structures which contain the C-terminus of each dipeptide interacting with the N-terminus and the N–H of the amide bond of the alternate dipeptide as is shown

in Figure 4.3 for both FV and FI. While this pair interaction type was not directly observed for FF, it is analogous to the 2 contact points (Figure 2.7a) if instead of parallel, the backbones were anti-parallel resulting in a 4 contact points anti-parallel pair interaction type. The pair interaction type analysis from Chapter 2 should be performed to quantify this. Due to this feature in the PMF, the authors hypothesize that modification of the C-terminus to valine or isoleucine reduces the steric sidechain size, enabling favorable interactions at distances shorter than the global minimum. While backbone dimer interactions for representative structures for the dipeptides studied are similar and demonstrate the importance of backbone electrostatics in dimerization, differences in the general shape of the PMFs suggest there are differences in the underlying physics driving dimerization of the VF, FV, IF, and FI dipeptides.

4.4.2 Noncovalent Forces Dictating Dipeptide Dimerization

For each of the dipeptide systems, the dimer PMFs were decomposed into solvent–solute and solute–solute contributions. As is seen in Figure 4.4, the solute–solute curves (blue) exhibit a negative free-energy contribution for heavy atom center of mass separation distances less than 10.5 Å for each of the dipeptides. The inverse is true for solvent–solute curves (green) which demonstrate a positive free-energy and thus a repulsive contribution over the range of distances for each of the dipeptides. At global well distances for each of the dipeptides, the magnitudes for both the solvent–solute and solute–solute curves are lower for VF than for FV, IF or FI. Additionally, these components of the free energy for IF and FI are larger in magnitude than FV and are more comparable to those seen in FF. Figure 4.4 demonstrates that solute–solute interactions promote dimerization for VF, FV, IF, and FI while solvent–solute forces hinder it, in agreement with what was shown for FF.

The dimer PMFs for VF, FV, IF, and FI are then further separated into solute–solute Lennard-Jones (blue triangles), solute–solute Coulomb (blue circles), solvent–solute Lennard Jones (green triangles), and solvent–solute Coulomb (green circles) contributions and are shown in Figure 4.5. As was shown for FF, the dimerization for each dipeptide is predominately driven by solute–solute

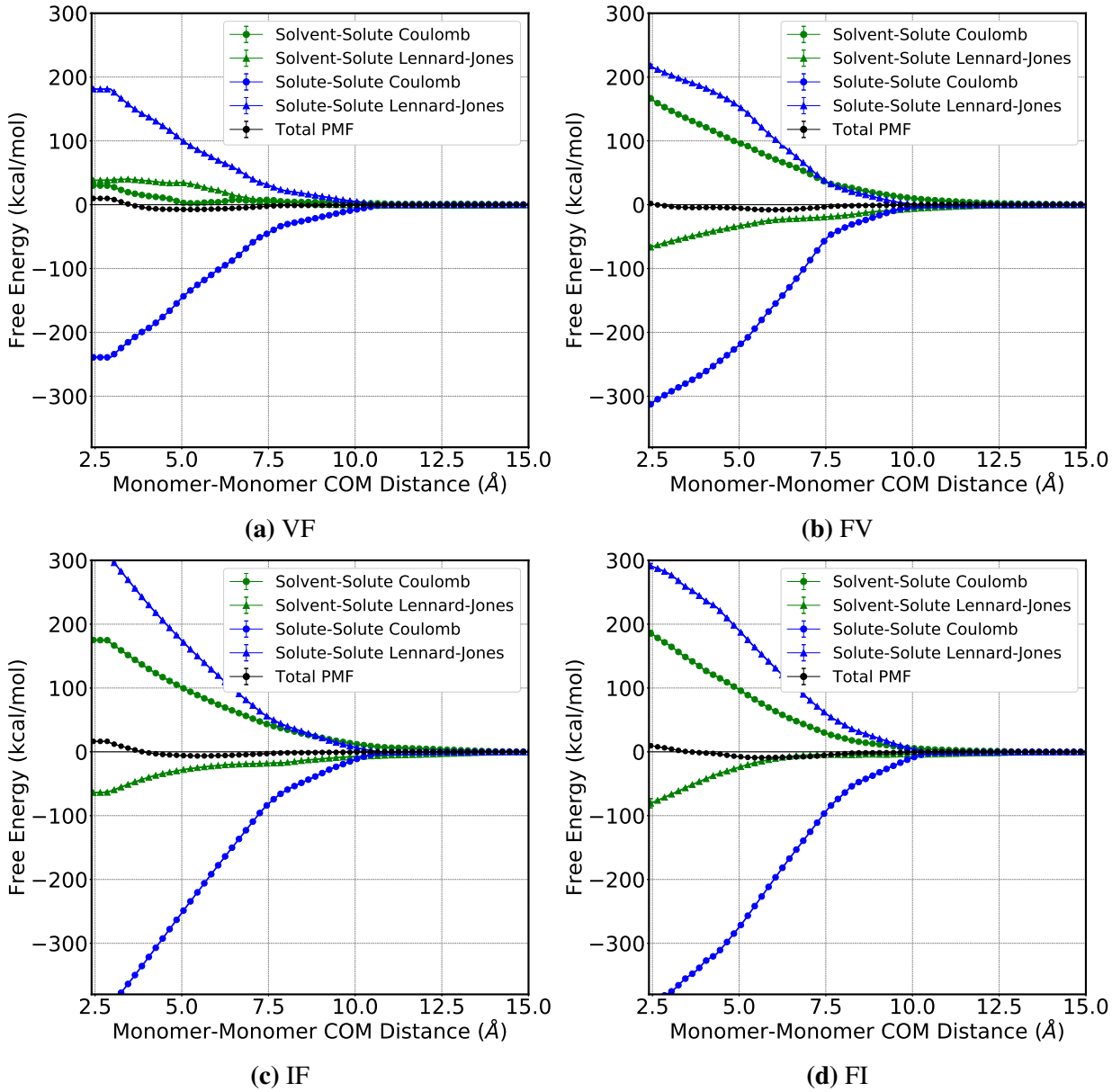


(a) VF **(b)** FV

(c) IF **(d)** FI

Figure 4.4: The relative free energy (black curves) of VF (a), FV (b), IF (c), and FI (d) as a function of respective dipeptide heavy atom center of mass separation decomposed into solvent–solute (green curves) and solute–solute (blue curves) components as described in Chapter 2. Error bars are present but are smaller than the lines.

electrostatic forces (Figure 4.5, blue circles), as this component of the free energy is negative and larger in magnitude than other contributing forces for distances less than 10 Å. The solute–solute Lennard-Jones (Figure 4.5, blue triangles) component of the free energy is the dominant repulsive force for distances less than the common shoulder. Separation of the solvent–solute forces into



(a) VF **(b)** FV

(c) IF **(d)** FI

Figure 4.5: The relative free energy (black curves) of VF (a), FV (b), IF (c), and FI (d) as a function of respective dipeptide heavy atom center of mass separation further decomposed into solvent–solute (green) and solute–solute (blue) Lennard-Jones components (triangles) and solvent–solute (green) and solute–solute (blue) Coulomb (circles) components. Error bars are present but are smaller than the symbols or lines.

Lennard-Jones and Coulombic forces demonstrates differences in the trends of the underlying forces between each of the dipeptides. The solvent–solute Coulombic force contribution to the PMF (Figure 4.5, green circles) are repulsive for the entire PMF of VF, FV, IF, and FI similar to the dimer PMF of FF (Figure 2.3b). However, a significant difference between the dipeptides

exists for VF as the solvent–solute Coulombic contribution is much lower in magnitude relative to the other dipeptides. Furthermore, the differences continue for the solvent–solute Lennard-Jones force contribution for the VF PMF (Figure 4.5a, green triangles) as it is positive for distances less than 9 Å. While the sum of the hydrophobicity of the sidechains is constant between VF and FV, this solvent mediated force promotes dimerization for FV (Figure 4.5b, green triangles) while impeding VF's dimerization. We propose that this significant difference in the dimer PMFs for VF relative to FV originate from the longer lived residence times for water around the carbonyl of the peptide bond for VF relative to FV. Additionally, for both IF and FI the solvent–solute Lennard-Jones force contributes to dimerization. If our hypothesis of VF's solvation shell causing the net repulsive solvent mediated force is true, this result would suggest that the water residence lifetimes around the carbonyl would be comparable between IF and FI, further suggesting a steric size limit to the previously seen sequence dependence. We also postulate that as the steric size of the C-terminal sidechain increases, the solvent–solute Lennard-Jones component of the free energy will increase. As this trend holds true for FV, FI, and FF, it would be of interest to study an additional system with a smaller sidechain size to provide clarity on sequence dependence of FF modification. These results indicate that solvent mediated forces hinder the dimerization of VF in water and only the solute–solute electrostatic interactions enable dimerization. Additionally, while solute–solute electrostatic interactions dominate the driving forces for dimerization of FV, IF, and FI, dimerization is also promoted by the solvent–solute Lenard-Jones force.

4.5 Conclusions

The work presented here provides the foundation for a systematic investigation into the underlying physics for sequence dependence of dipeptide self-assembly. An emphasis has been placed on simultaneously modifying as few variables as possible to enable comparison to previous studies. This chapter is a natural extension of Chapter 3 which considered the sequence dependence in monomers of VF and FV. As the methods and analyses used were consistent to those in Chapter 2, direct comparisons of dimer PMFs of VF, FV, IF, and FI were made to that of FF. Although there

were differences in overall shape of the dimer PMFs for VF, FV, IF, and FI, they contained features and representative structures similar to those of FF. A dimer PMF feature not previously described in the literature occurs for FV and, to a lesser extent, FI centered around a center of mass separation distance of 4 Å. While we hypothesize this is caused by a 4 contact point anti-parallel interaction type, future work would be needed to confirm this. Additionally, due to this feature and a lower global well depth, we hypothesize that modification of the C-terminus would promote dimerization more than substitution of the N-terminus. Decomposition of the dimer PMFs demonstrates that dimerization of VF, FV, IF, and FI is predominantly driven by backbone electrostatics in agreement with FF results. The dimer PMF decomposition also demonstrated that while both solvent–solute Lennard-Jones and Coulomb forces hinder VF dimerization, solvent–solute Lennard-Jones forces promote dimerization for FV, IF, and FI. When considered with the increased water residence lifetimes seen for VF relative to FV in Chapter 3, this would suggest that the longer lived residence times cause an energetic penalty to dimerization. Since this was not seen for IF, this also suggests that the steric size of the isoleucine residue and/or increased hydrophobicity would not increase water residence lifetimes, a hypothesis left to be tested. While these results suggest dimerization would be favorable, studies of larger systems would be required to generalize these conclusions to larger, more ordered systems as kinetic trapping effects may frustrate the self-assembly process.

As stated previously, this work provides a foundation for further investigation. Future work would include investigating the self-assembly process for larger system sizes as was done for FF in Chapter 2 and investigating the monomeric solvent dynamics of IF and FI as was done for VF and FV in Chapter 3. Additionally, the solvent dependence for dimerization for these dipeptides as well as FF will be considered in a non-polar solvent.

Chapter 5

Computing the Entropy of Association with the Rigid-Rotor Gaussian Approximation.[4]

5.1 Overview

Determination of the free energy of association is a vital measure of coordination affinity and in determining reaction mechanisms. While computational methodologies are appealing, computing the free energy of association is challenging due to low energy vibrational modes being anharmonic leading to a large entropic error. Despite significant previous efforts, there is yet a general approach to correcting the entropic error for associating molecules. Here we present a general rigid-rotor-Gaussian-oscillator (RRGO) model for computing the entropy for low lying frequencies. The RRGO approximation uses a Gaussian potential, Gaussian basis set wave functions, and a temperature dependent interpolation function to determine a corrected entropy. The free energy correction using the entropy from the RRGO approximation is reported here for a set of molecules. Specifically, the systems studied here include ethane, methane dimer at a range of carbon–carbon separation distances, benzene dimer in the T-shaped and parallel displaced configurations, and chalcone dimer. Although the entropic impact for using the RRGO is small for ethane, the resulting free energy correction is significantly higher for associating molecules. The impact on the free energy correction is most notable for the chalcone dimer in implicit cyclohexane, the result of which is an experimentally comparable binding free energy. Additionally, results from benzene dihedral umbrella sampling in combination with the RRGO entropy would suggest the T-shaped configuration for benzene dimer to be more energetically stable than the parallel displaced configuration.

[4]Jakob Anderson[a], Jacob Nite[b], Anthony Rappé[a]; [a] Department of Chemistry, Colorado State University, Fort Collins, CO, USA, [b] Gaussian Inc., Wallingford Connecticut

5.2 Introduction

Chemistry, the science of change, dominantly concerns the association and subsequent reaction of two or more molecules. For multistep noncovalent association processes or mechanistic pathways which involve noncovalent association, experimental studies are often able to only determine the free energy of the barrier for the rate limiting step.[122] Whether it be atmospheric chemistry, small molecule synthesis, industrial catalysis or protein-ligand binding events, the inherent experimental difficulties for many association mechanisms make it appealing to utilize computational methodologies. A generalizable computational method, Quantum Mechanical (QM) calculations are a popular computational tool to study relatively large systems with atomic resolution and which accurately compute interaction energies. The minimal error in calculating the interaction energy, however, does not translate to the error in the association free energy. Problems within the accuracy of computational methods in computing the free energy of association arise due to: 1) associating molecules being non-ideal systems[123] and 2) the conversion of the six intermolecular degrees of freedom (translations and rigid rotations) into decidedly anharmonic, low energy vibrational modes during the association process. As illustrated for the rotational barrier of ethane in Figure 5.1, the usual rigid-rotor-harmonic-oscillator (RRHO) statistical thermodynamics model over counts modes above the barrier height (there are only 3 harmonic modes below the 1020 cm^{-1} rotational barrier) and under counts the energy level compression associated with barrier approach. This problem has been addressed extensively for torsions[55–58] as well as for gas phase reactive dynamics[59].

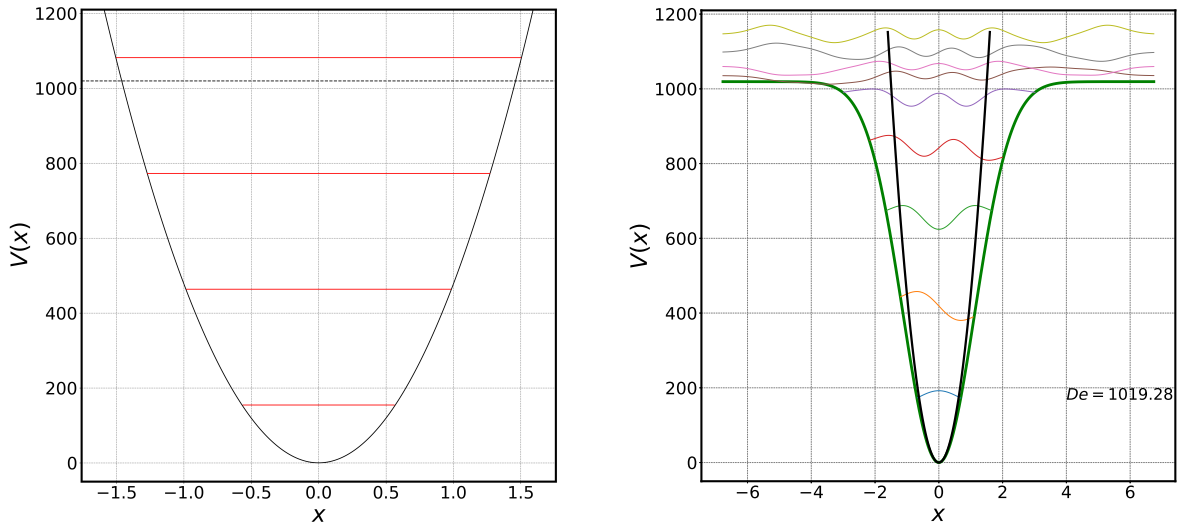


Figure 5.1: (a) A harmonic potential (solid, black curve) with analytical wave functions (red line) which are equally spaced and the barrier height (De) represented as a black dashed line. (b) Energy profile of ethane for the torsional mode where the green curve is a Gaussian potential, the black curve is a harmonic potential, and the all other curves are the energy levels represented as Gaussian oscillators.

The error in the free energy of association in many QM studies has been suggested to originate in an inaccurate calculation of the entropy contribution.[55] While attempts at correcting the entropy have been diverse, improvements can still be made to make the approximation for the entropic term more generalizable and rigorous. One common approximation is to apply a free volume theory correction to the translational entropy, however, the authors note that extension to irregularly shaped molecules may introduce significant errors.[39] It is also common to attempt to compensate for the error in the free energy by scaling the entropy by 50%[40–46] or by 14.3 e.u.[47–52] the latter of which attempts to correct the gas–solution-phase concentration difference. Plata and Singleton found that these empirical scaling factors for the entropy were a "spectacular failure" for the Morita Baylis–Hillman reaction and cautioned against using them in general.[122] Plata and Singleton also attempted to compute the entropy along the Morita Baylis–Hillman reaction pathway using a temperature dependent implicit solvent model (SM8T), however, the computed free energy did not

change and actually shifted the error from the entropy to the enthalpy.[122] Others have found the entropic error is substantially reduced with the use of two or more explicit solvent molecules in addition to the implicit solvent but this does not address the root cause of the entropic error.[53,54] An alternative approach to computing the free energy of association is to utilize all-atom Molecular Dynamics (aaMD) simulations which enables the study of thermodynamic variable and explicit solvent dependence of the association mechanism. A major disadvantage to this classical treatment of the system is the inability to accurately simulate most transition metals and the obvious dependence on the parameters within the force field. A popular approximation is that of Grimme in which a free-rotor correction is applied to low-frequency modes, although, Grimme notes that a "prerequisite is a certain conformational rigidity" an often overlooked requirement.[55] As will be shown in this paper, this approximation requires an explicit temperature dependence, though it is generalizable for non-torsional degrees of freedom.

A general Gaussian model is presented in this paper as well as the application of the model to the hindered rotation of ethane, the formation of methane and benzene dimers, and the dimerization of Chalcone, a larger complex which is a privileged scaffold in medicinal chemistry[124] and is of interest to us[125]. In this paper, we attempt to build off of the RRHO approximation by first considering the free energy PMFs determined with aaMD of several example systems demonstrating the significant entropy of associating molecules and the ability to reproduce the enthalpy of association for the systems studied. A novel approximation is then described which uses a Gaussian oscillator to compute the entropy for low frequency modes and the correction terms associated with this approximation for this set of test systems.

5.3 Methods

Geometry optimization and frequency calculations were performed using the Gaussian16 (G16) software package.[126] The range separated-dispersion corrected ωB97xD functional in conjunction with a 6-311+G* basis set was chosen due to the implemented dispersion corrections which are needed for supramolecular interactions.[77] Coordinates for the optimized structures are provided in

the Supporting Information section. The GaussView software was used to view structures generated from QM calculations.[127]

Umbrella sampling simulations were performed using the Amber18 program. All simulations were performed in the NPT ensemble using a Monte Carlo barostat set to 1 bar and a Langevin thermostat set to the designated temperature. The SHAKE algorithm was used to restrain the bond lengths of bonds which included a hydrogen atom. A integration timestep of 2 fs was used and all long range electrostatics were calculated using the particle mesh Ewald method with a cutoff distance of 12 Å. The force field for chalcone was developed using atom types found in the generalized amber force field (GAFF) coupled with implementation of the RESP and ANTECHAMBER charge fitting procedure to fit atom charges.[128–130] The OPLS-AA force field was used for benzene as this force field was developed to reproduce properties of small organic molecules.[131]

Using steered umbrella sampling, *in vacuo* simulations were performed biased along the monomer center of mass separation between 3.5 Å and 25 Å for chalcone and 1.5 Å and 20 Å for benzene. Harmonic potentials with a force constant of 20 kcal/(mol·Å^2) were used for windows positioned every 0.5 Å. Each window of the chalcone umbrella sampling was simulated at a temperature of 208 K to 398 K with a 10 K increment for 3 ns for a total of 2.6 μs of simulation. Benzene dimer umbrella sampling was performed at 160 K, 180 K, 200 K, and 220 K for 30 ns per window for a total simulation time of 4.5 μs.

The rotation of one benzene relative to another, also described as the dihedral or the angle between the planes of the two benzene's was also simulated using steered umbrella sampling. The dihedral is described by the fifth carbon of one benzene to the first, second, and third carbon atoms of the second benzene. A restraint of 5 kcal/(mol·Å^2) was placed on the carbons of one benzene and the first carbon of the second benzene. Each degree between -180° and 180° was simulated using a harmonic potential with a force constant of 0.06092 kcal/(mol·deg^2) for 50 ns for a total of 1.8 μs. While an imperfect description of the angle between the vectors perpendicular to the planes of the benzene rings, visual analysis confirms rotation of one benzene relative to the other.

In house python 2.7 scripts were developed to analyze QM data and are described in more detail below. The center of mass separation distance during the benzene dimer dihedral umbrella sampling made use of the MDAnalysis module for python.[87,132] All data from MD simulations was plotted using Matplotlib.[88]

5.4 Results and Discussion

5.4.1 Thermodynamic Quantities from aaMD

Chalcones are a group of compounds found in a wide variety of plants (Toxicodendron vernicifluum, Piper methysticum, and Lophira alata to name a few) and plant products such as beer (Humulus lupus), tea (e.g. Piper methysticum, Angelica keiskei), spices (e.g. Carthamus tinctorius, Boesenbergia rotunda), fruits (e.g. Dalbergia odorifera, Semecarpus anacardium, Alpinia rafflesiana, Alpinia conchigera), and vegetables (e.g. Angelica keiskei, Allium ampeloprasum).[133–135] A significant amount of research has been done on chalcones due to their of structural diversity, ease of synthesis and chemical manipulation, as well as their potential therapeutic applications.[136–153] The broad list of potential chalcone therapeutics include applications for anticarcinogenic,[154–158] antioxidant,[154,159–161] antidiabetic,[162–164] and antimalarial[144,165–167] drugs. In order to develop an optimized chalcone derivative for a given therapeutic application, understanding how the chalcone derived drug would bind to a receptor relative to other drugs is a critical point of comparison.

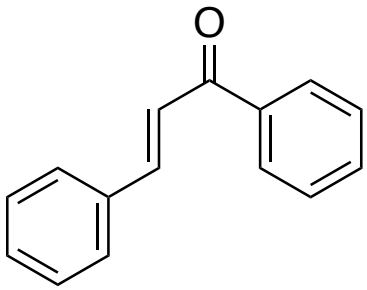


Figure 5.2: The unsubstituted chalcone considered in the context of this paper.

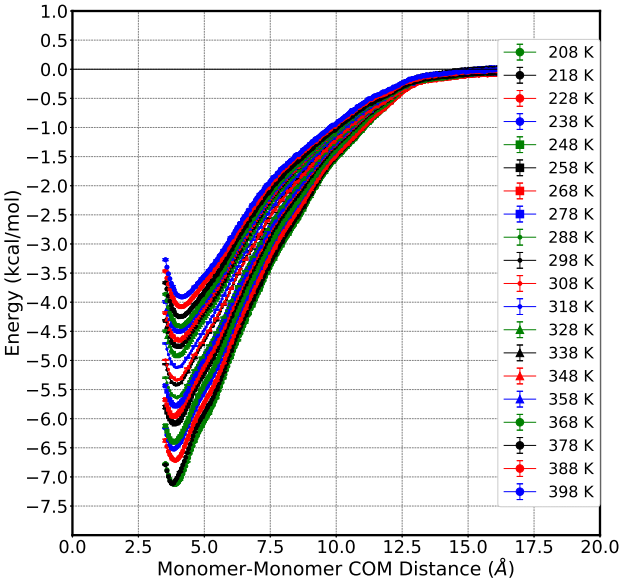


Figure 5.3: Chalcone dimer PMF's performed *in vacuo* at a range of temperatures from 208 K to 398 K with a 10 K increment. For reference, lower temperature PMF's have lower energies with the umbrella sampling performed at 208 K resulting in a PMF with a global minimum of -7.13 kcal/mol while the PMF from umbrella sampling performed at 398 K has a global minimum of -3.91 kcal/mol.

As discussed in the Introduction, determining the binding energies for associating molecules is a non-trivial task. In order to compute the binding energy for a chalcone derivative to a receptor, being able to compute the binding energy for chalcone dimer is a good first step. For the purposes of this paper, we will be referring to the unsubstituted chalcone as depicted in Figure 5.2. While the chalcone monomer has been computed using DFT computations,[168] to the authors knowledge DFT results for the unsubstituted chalcone dimer have not previously been reported. All energies can be seen in Table A4. Experimental studies suggest a binding free energy for chalcone dimer is -5.1 kcal/mol at 298 K when solvated in cyclohexane.[169] The binding free energy computed using the ωB97xD functional in an implicit cyclohexane solvent, however, is -12.2 kcal/mol (See Supplemental Information (SI)). To demonstrate the cause of the experimental and QM discrepancy, aaMD simulations of the chalcone dimer were first simulated in *in vacuo*. Visual inspection of the simulations demonstrates that the chalcone dimer at room temperature does not form a rigid dimer complex and instead samples a variety of different dimer local free energy minima in very little simulation time. To probe the chalcone dimer association energy, umbrella sampling was

Table 5.1: The free energy in kcal/mol of chalcone dimer at the equilibrium distance at the given temperature in K.

Temperature (K)	E (kcal/mol)	Temperature (K)	E (kcal/mol)
208	-7.132±0.004	308	-5.338±0.007
218	-7.120±0.005	318	-5.122±0.007
228	-6.716±0.006	328	-4.920±0.008
238	-6.523±0.005	338	-4.756±0.009
248	-6.402±0.008	348	-4.651±0.009
258	-6.09±0.02	358	-4.50±0.01
268	-5.963±0.008	368	-4.42±0.01
278	-5.780±0.006	378	-4.25±0.01
288	-5.643±0.006	388	-4.08±0.01
298	-5.424±0.007	398	-3.91±0.01

performed for chalcone dimer biased along the chalcone monomer heavy atom center of mass separation distance for distances between 3.5 Å and 25 Å with 0.5 Å window increments. For this range of distances, umbrella sampling simulations were performed at 20 different temperatures ranging from 208 K to 398 K with a 10 K increment. As can be seen in Figure 5.3, there is a single well located at 3.525 Å for all PMF's. For temperatures above 298 K, the PMF's have a broader and more shallow well as compared to those temperatures below 298 K. Additionally, as the temperature is decreased, an ill-defined shoulder appears between 5 Å and 5.6 Å. The relative free energy at the bottom of the well for each temperature is reported in Table 5.1. Fitting a line to the free energy at the bottom of the well vs. the temperature, results in an enthalpy of -10.6 kcal/mol, an entropy of 0.017 kcal/(mol·K), and a R^2 value of 0.992. As shown in the Table A4, the chalcone dimer computed *in vacuo* has an interaction energy of -13.7 kcal/mol and an enthalpy of -11.9 kcal/mol. The agreement between the enthalpy computed using QM and aaMD suggests that the error between the computed and experimental binding energies for the previous systems solvated in cyclohexane is predominately resulting from an error in the entropy. While there is a increase in entropy caused by the lack of rigidity within the individual molecules, the magnitude of the translational and rotational entropy for one monomer relative to another is significantly larger as is the focus of this study.

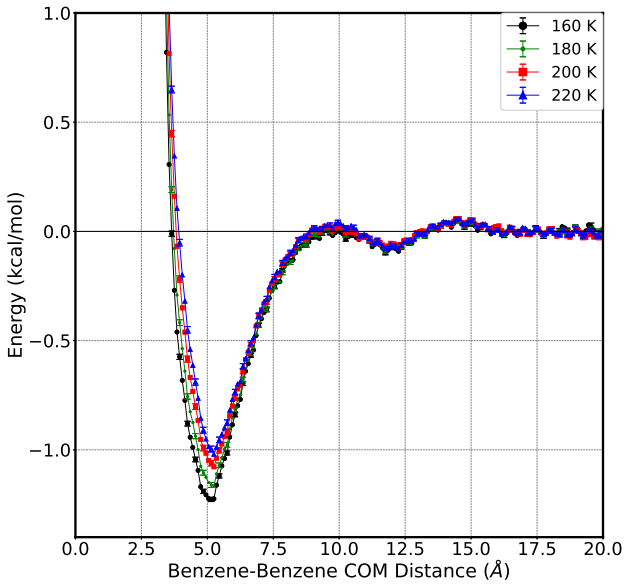


Figure 5.4: Benzene dimer PMF's performed *in vacuo* at temperatures of 160 K (black circles), 180 K (green dots), 200 K (red squares), and 220 K (blue triangles) along the benzene–benzene center of mass separation distance. Error bars are present but are often smaller than the symbols, demonstrating the PMF's are outside of error of one another for distances around the global minimum.

To demonstrate that the discrepancy between experimental and QM results is a general problem and not just an issue with large molecules or molecules with many intramolecular degrees of freedom (flexible molecules), the dimerization of benzene was also considered in a similar manner. Umbrella sampling was performed for two benzene's *in vacuo* between center of mass separation distances of 2.0 Å and 20 Å with a 0.5 Å increment at temperatures of 160 K, 180 K, 200 K, and 220 K. A PMF for each temperature was subsequently generated and is shown in Figure 5.4. At distances below 10 Å, the PMF's begin to differ with global well depths of -1.227($\pm$0.010) kcal/mol, -1.17($\pm$0.01) kcal/mol, -1.08($\pm$0.01) kcal/mol, and -1.02($\pm$0.01) kcal/mol for the respective PMF's simulated at temperatures of 160 K, 180 K, 200 K, and 220 K. The global well of the benzene dimer PMF's is located at 5.15 Å for the 160 K PMF and 5.25 Å for all others, however, the free energy for the 160 K PMF at 5.25 Å is within error of the free energy at 5.15 Å. Similar to the chalcone dimer PMF's, the global well of the benzene dimer PMF's broadens with decreasing temperature. Fitting a line to the free energy at the global minima vs. temperature

results in an enthalpy of -1.8 kcal/mol, an entropy of 0.00354 kcal/(mol·K), and a R^2 value of 0.994. The enthalpic energy is in agreement with the experimentally determined binding energies (D0= 1.6±0.2 kcal/mol or D0=2.4±0.4 kcal/mol) at the limit of 0 K demonstrating the validity of the analysis.[170,171] Since benzene is a rigid molecule, the enthalpic agreement to experiment further indicates that the disagreement between free energies from experiment and QM calculations is predominately dependent on rotational and translational entropy.

Interestingly, the interaction energy computed from the force field for the geometry optimized structures using the ωB97xD functional is -2.2288 kcal/mol for the parallel displaced and -2.1733 kcal/mol for the T-shaped benzene dimer conformations. This would suggest that, for the parameters in the used force field, the parallel displaced structure would be modestly favored. Instead, the free energy shown by the PMF in Figure 5.5a, generated using umbrella sampling biased along the angle between the planes of the benzene's or the dihedral, suggests the T-shaped geometry is actually favored. The dihedral PMF for benzene dimer has maxima centered around 90° when the dimer is in a parallel displaced configuration and has an average free energy of 0.342 kcal/mol. The minimas are centered around -180°, 0°, and 180° with an average free energy of 0.07 kcal/mol corresponding to a T-shaped configuration. As can be seen in Figure 5.5b, the center of mass separation distances for the corresponding angles are consistent with parallel displaced conformation around ±90° and T-shaped around 0° and ±180°. The relative equivalence between the two configurations in interaction energies (difference of 0.05 kcal/mol) suggests that the difference in free energy results from the entropy of the system. While not conclusive, the restriction of the system to fewer possible conformations for the parallel displaced relative to the T-shaped configuration would cause an decrease in the entropy and thus an increase in the free energy as is supported by previous studies.[172]

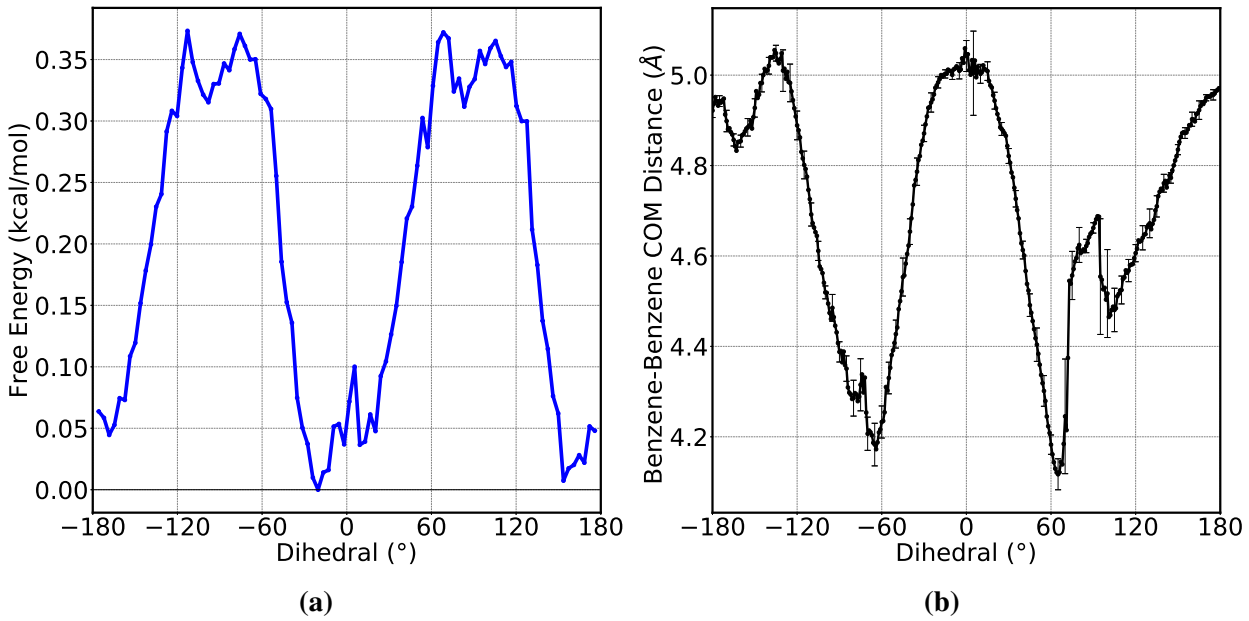


Figure 5.5: (a) The PMF of the angle between the planes of each benzene molecule or the dihedral of the two benzene's. This PMF has a periodicity of 180°. The benzene dimer is in a parallel displaced configuration for dihedral values of ±90° and in a T-shaped configuration for dihedral values around -180°, 0°, and 180°. (b) The benzene–benzene center of mass separation distance measured from the umbrella sampled trajectory used to generate (a). The separation distance is lowest around ±90° while highest around -180°, 0°, and 180° corresponding to parallel displaced and T-shaped configurations respectively. The error bars included in (b) are the standard deviation of the distances for the frames that were biased to sample at the corresponding dihedral angle.

5.4.2 Rigid-Rotor-Gaussian-Oscillator (RRGO) Approximation

In order to account for the entropy of low-lying modes of associating molecules, a python script was developed which utilizes the information in a standard G16 checkpoint file and produces a corrected entropic term. This script initially uses the Hessian matrix computed from G16 to compute the force constants and reduced masses of all modes including the imaginary and low energy modes. Using the force constants, reduced masses, and the partition functions described below, the energy and entropy of low lying modes are determined using a rigid-rotor-Gaussian-oscillator (RRGO) approximation. Considerable care was taken to ensure that units were being consciously followed and as such, and especially where it was a hurdle for us in the development of the script, we explicitly include the units within the script. The following is a more detailed summary of the script used to compute the correction to the entropy term using the RRGO approximation.

Loosely following "Vibrational Analysis in *Gaussian*" by Ochterski (1999),[173] the following quantities are determined. In the entropy.py script found in the SI, the Cartesian Hessian matrix is read in using fchk_ext.py (also in SI), after which the Hessian is weighted by the atomic mass of each atom. This mass weighted Cartesian Hessian matrix is then diagonalized to determine an eigenvalue (ρ), scaled to account for the weight of an electron relative to a neutron, after which this scaled mass weighted Hessian matrix is diagonalized to determine a different eigenvalue (λ) and a eigenvector (*L*). The diagonalization of the mass weighted Cartesian Hessian matrix is performed prior to and after scaling for ease of unit conversions. The reduced mass (μ) for each mode including the low-lying modes is then computed by

$$\mu = \sum_{i}^{3N} \frac{L^2}{m_i} \tag{5.1}$$

where m_i is a matrix of the atomic masses in the same size and shape as the hessian matrix. The mass weighted Cartesian Hessian prior to scaling is in units of $E_h/r_{bohr}{}^2$ and thus ρ is in units of $E_h/r_{bohr}{}^2$·amu). Once converted to mdyne/(Å^2·amu), the force constant (*k*) is then computed by

$$k = \rho\mu \tag{5.2}$$

The vibrational energy for all modes (ν) including translational and rotational modes is computed using Equation 5.3 after λ is converted to cm^{-1}.

$$\nu = \sqrt{\lambda} \tag{5.3}$$

To account for imaginary frequencies, the absolute value of λ is taken prior to the square root, then multiplied by the sign of λ.

As is shown by Ochterski, a transformation matrix (*D*) is then determined which includes the translation and rotation vectors.[173] The D matrices are found by first computing the moment of

inertia of the atom selections. The eigenvectors of the diagonalized moment of inertia matrices are then used to manually populate the D matrix which is then normalized.

The translational, rotational, and vibrational entropic contributions for all modes are then calculated in a similar fashion as detailed in "Thermochemistry in *Gaussian*" by Ochterski (2000).[123] The fraction of translational and rotational entropy in the x, y, and z dimensions for low-lying modes is then determined with the corresponding vibrational energy and entropy using a harmonic oscillator. Using the μ's and k's determined from a transformation of the Cartesian data to the normal mode space, the vibrational frequencies are computed again using Equation 5.4

$$\nu = \hbar\sqrt{\frac{k}{\mu}} \tag{5.4}$$

where $\hbar$ is the reduced Planck's constant.

For a number of vibrational modes such as bond dissociation, complexation, low frequency bends, and torsional degrees of freedom, there are a small, finite number of bound states as is illustrated in Figure 5.1b for ethane. Modeling such systems with a harmonic oscillator is problematic. As described in the introduction, there are a multitude of approaches that attempt to correct for this. Here we suggest a general alternative. Consider a separation of the general partition function (Q) into two terms, the first involving the n bound states, and second the set of continuum of states starting at n+1. We can treat the bound states conventionally and the continuum states in terms of translational or rotational partition functions as done in Equations 5.5 – 5.8

$$\begin{aligned} Q = \sum_{i}^{\infty} e^{-\beta E_i} &= \sum_{i}^{n} e^{-\beta E_i} + e^{-\beta E_{n+1}} \sum_{i=n+1}^{\infty} e^{-\beta (E_i - E_{n+1})} \qquad (5.5) \\ &= \sum_{i}^{n} e^{-\beta E_i} + e^{-\beta E_{n+1}} Q_R \\ B_f &= e^{-\beta E_{n+1}} \qquad (5.6) \end{aligned}$$

$$\Big(\frac{\delta Q}{\delta T}\Big)_V = \Bigg[\sum_i^n \frac{\beta E_i}{T} e^{-\beta E_i} - \frac{\beta E_{n+1}}{T} e^{-\beta E_{n+1}} \sum_{i=n+1}^{\infty} e^{-\beta(E_i - E_{n+1})} + e^{-\beta E_{n+1}} \sum_{i=n+1}^{\infty} \frac{\beta(E_i - E_{n+1})}{T} e^{-\beta(E_i - E_{n+1})}\Bigg]$$

$$= \frac{\beta}{T}\Big[\sum_i^n E_i e^{-\beta E_i} + B_f(E_R + E_{n+1} Q_R)\Big] \tag{5.7}$$

$$\Big(\frac{\delta ln Q}{\delta T}\Big)_V = \frac{1}{Q}\Big(\frac{\delta Q}{\delta T}\Big)_V \tag{5.8}$$

where β is thermodynamic beta, E is the energy, Q_R is the rotational partition function, B_f is a Boltzmann factor, V is volume, and T is temperature. The bound state energies could be obtained from the quantum mechanics of a numerical potential, a Fourier expansion, a Morse potential or other choices. Here we illustrate the approach with a Gaussian potential. A α_h (Equation 5.9) and a α_g (Equation 5.10) are computed and used in the Gaussian exponent in the basis set expansion and for the Gaussian potential (Equation 5.12) respectively.

$$\alpha_h = \frac{\nu\mu}{2} \tag{5.9}$$

$$\alpha_g = \frac{k}{2D_e} \tag{5.10}$$

$$D_e = \frac{4k}{\pi} \tag{5.11}$$

$$V = D_e(1 - e^{-\alpha_g x^2}) \tag{5.12}$$

$$fw90m = 2\sqrt{\frac{log(90)}{\alpha_g}} \tag{5.13}$$

where D_e is the dissociation energy and is computed in Equation 5.11 and the ν is that of Equation 5.4. The range of values for which the potential and the basis set expansion is dependent on is equal to $\pm$*fw90m* divided by a user defined number of grid points. For the results provided in the next section, the number of grid points used is 70. The continuum of states are described using the rotational model provided by Grimme.[55] In Ref. 55, an interpolation formula was used to smoothly transition from a low frequency rotational model to a high frequency vibrational model

using Equations 5.14 – 5.15. A ω_0 of 100 cm^{-1} was selected to describe room temperature thermodynamics. In the current work a Boltzmann factor (Equation 5.6) associated with the barrier height provides the interpolation.

$$S = \mathrm{w}(\omega)S_V + [1 - \mathrm{w}(\omega)]S_r \tag{5.14}$$

$$\mathrm{w}(\omega) = \frac{1}{1 + (\frac{\omega_0}{\omega})^4}, \ \omega_0 = 100cm^{-1} \tag{5.15}$$

The even spacing of Gaussians over the range of distances is determined by dividing De by the frequency (Equation 5.4) rounding to the nearest integer. For the present study 70 gaussians are used. The spacing is illustrated in Figure 5.6 for a Gaussian potential (solid curve) with a maximum of one, the placement of 21 Gaussians (black dots), and a normalized Gaussian function which is offset by one grid point from the middle (dashed curve). The variational principle is used to determine the basis set solution for the kinetic and potential energy where the potential is given by Equation 5.12 and the basis functions are Gaussians (Equation 5.16) for which N is a normalization constant.

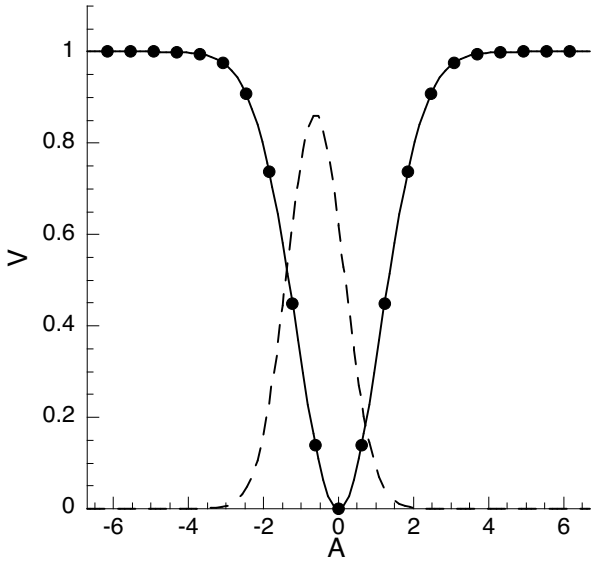


Figure 5.6: A depiction of the placement (dots) of Gaussian modes (dashed curve) on a Gaussian potential (solid curve) as an example of what is done for low energy vibrational modes in the RRGO approximation.

$$f(x_i) = Ne^{-\alpha_g x_i^2} \tag{5.16}$$

The resulting bound energy spectrum is used in Equations 5.5 – 5.8, to obtain the potential energy (U, Equation 5.17) and entropy (S) of a given mode. As shown in Equations 5.18 – 5.19, the total vibrational entropy is then taken as the sum of entropies over all modes. The derivation of the following equations is described in further detail in "Molecular Thermodynamics" by McQuarrie and Simon (1999).

$$U = k_B T^2 \Big(\frac{\delta ln Q}{\delta T}\Big)_V \tag{5.17}$$

$$S_i = k_B T \Big(\frac{\delta ln Q}{\delta T}\Big)_V + e^{\frac{-De}{k_B T}} k_B \Big(\frac{1}{2} + ln\Big(\sqrt{\frac{2\pi^2 k_B T}{\hbar \nu}}\Big)\Big) \tag{5.18}$$

$$S_{total} = \sum_{n=i}^{N\ modes} S_n \tag{5.19}$$

To illustrate the difference between RRHO and RRGO models as well as a literature[55] approach, the entropy as a function of Harmonic vibrational frequency for a range of conditions is provided in Figure 7. As described above, the Grimme model interpolates between the vibrational entropy and a rotational entropy using the weighting procedure given in Equation 5.14.

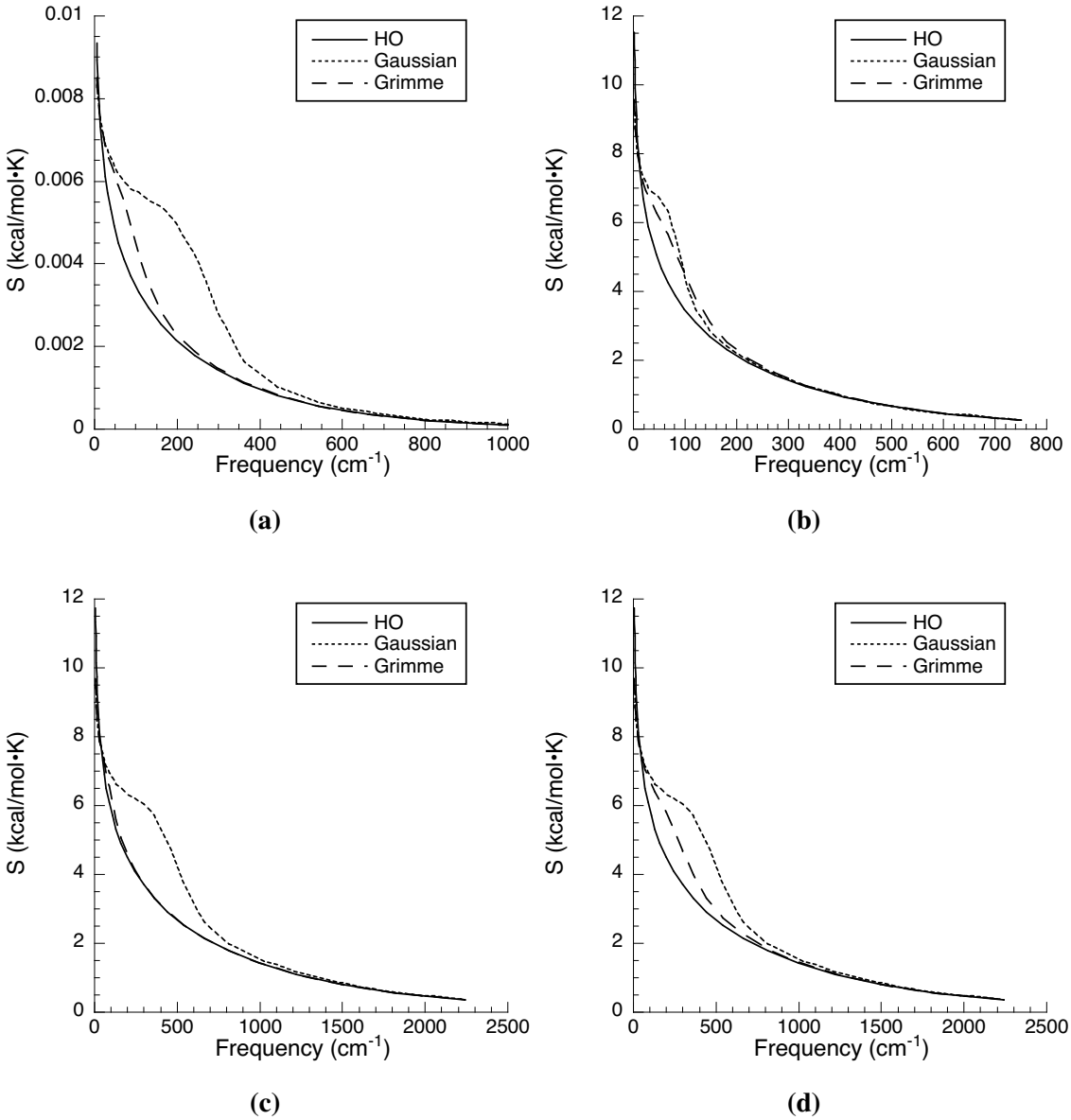


Figure 5.7: A comparison of computing the entropy as a function of frequency using a harmonic oscillator (solid line, HO), a Gaussian oscillator (dotted line, Gaussian), and the correction term used by Grimme[55] (dashed line, Grimme). For a), the reduced mass (μ) is set to 1, the ω_0=100 cm^{-1}, and the temperature is 298 K. In (b), μ=9, ω_0=100 cm^{-1}, and the temperature is set to 298 K. For (c) relative to (b) the temperature was increased to 1000 K. For (d), μ=9, ω_0=300 cm^{-1}, and temperature=1000 K.

As shown in Figure 5.7a, for low frequencies both the Gaussian and Grimme models plateau at finite values while the harmonic oscillator continues to rise. At intermediate frequencies the Gaussian and Grimme models provide a larger entropy than the harmonic oscillator model, accounting for anharmonicity effects, and both approach the harmonic oscillator model at high frequencies. The impact of reduced mass is illustrated in Figure 5.7b and temperature in Figures 5.7c and 5.7d where a ω_0 of 300 cm^{-1} recovers the intermediate frequency anharmonicity impact.

5.5 RRGO Approximation Results

Using the RRGO approximation detailed in the previous section, the entropy of five systems were studied. The systems chosen represent a set of molecules with a diverse number of possible configurations and degrees of freedom. Ethane was chosen as it is a single molecule and thus previous approximations treating it as an ideal gas is still appropriate. Methane dimers computed at the geometry optimized equilibrium distance as well as carbon–carbon separation distances of 4.5 Å, 6.5 Å, and 8.5 Å were considered for the molecular simplicity and the additional degree of freedom (intramolecular stretch). As a point of comparison to MD results, benzene dimer was computed in both the parallel displaced and T-shaped conformations. Additionally, the chalcone dimer, which was also studied in the "Thermodynamic Quantities from aaMD" section, offers a system with significantly more degrees of freedom compared to other systems studied here.

Table 5.2: The lowest mode frequency using the G16 code as well as the frequency used in the RRGO approximation script, number of bound states (n), barrier height (De), partition function (Q), and the derivative with respect to temperature of the natural log of Q (dlnQ/dT) for each system using the developed script and the RRGO approximation.

System	Frequency$_{G16}$ (cm^{-1})	Frequency$_{RRGO}$ (cm^{-1})	n	De (cm^{-1})	Q	dlnQ/dT
Ethane	309.2535	291.6437	5	1019.28	1.33	2.73E-3
Methane dimer at equilibrium	58.3916	42.0237	1	37.11	20.219	3.55E-4
Methane dimer at 4.5 Å	14.2187	7.6210	1	5.44	42.960	8.16E-5
Methane dimer at 6.5 Å	3.6602	0.7114	1	0.45	84.359	3.71E-5
Methane dimer at 8.5 Å	15.3362	3.8953	1	2.51	41.219	9.79E-5
Benzene dimer T-shaped	26.1999	21.6509	1	28.14	30.587	3.37E-4
Benzene dimer parallel displaced	33.9250	29.0638	2	44.59	25.623	5.30E-4
Chalcone dimer	9.2607	6.5897	1	5.71	50.707	8.62E-5

Prior to discussing the entropic results from using the RRGO approximation, and in an attempt at transparency, the first mode for all systems is reported. As can be seen in Table 5.2, the frequencies for all systems using the RRGO approximation do not agree with G16 output frequencies because a Gaussian potential is decidedly non-harmonic. While initially alarming, the frequencies are relatively close with the largest disagreement off by 17.6098 cm^{-1}. Additionally, it is common practice to project the translations and rotations out from the hessian prior to, and thus excluding them from, the frequency determination contributing to this disagreement. For all systems the number of Gaussian basis set functions below the mode barrier is equal to one except for ethane and the parallel displaced conformation of the benzene dimer. Figure 5.1 and Figures 5.8 and 5.10 depict the energy profiles for the lowest modes for all of the systems quantitatively described by Table 5.2.

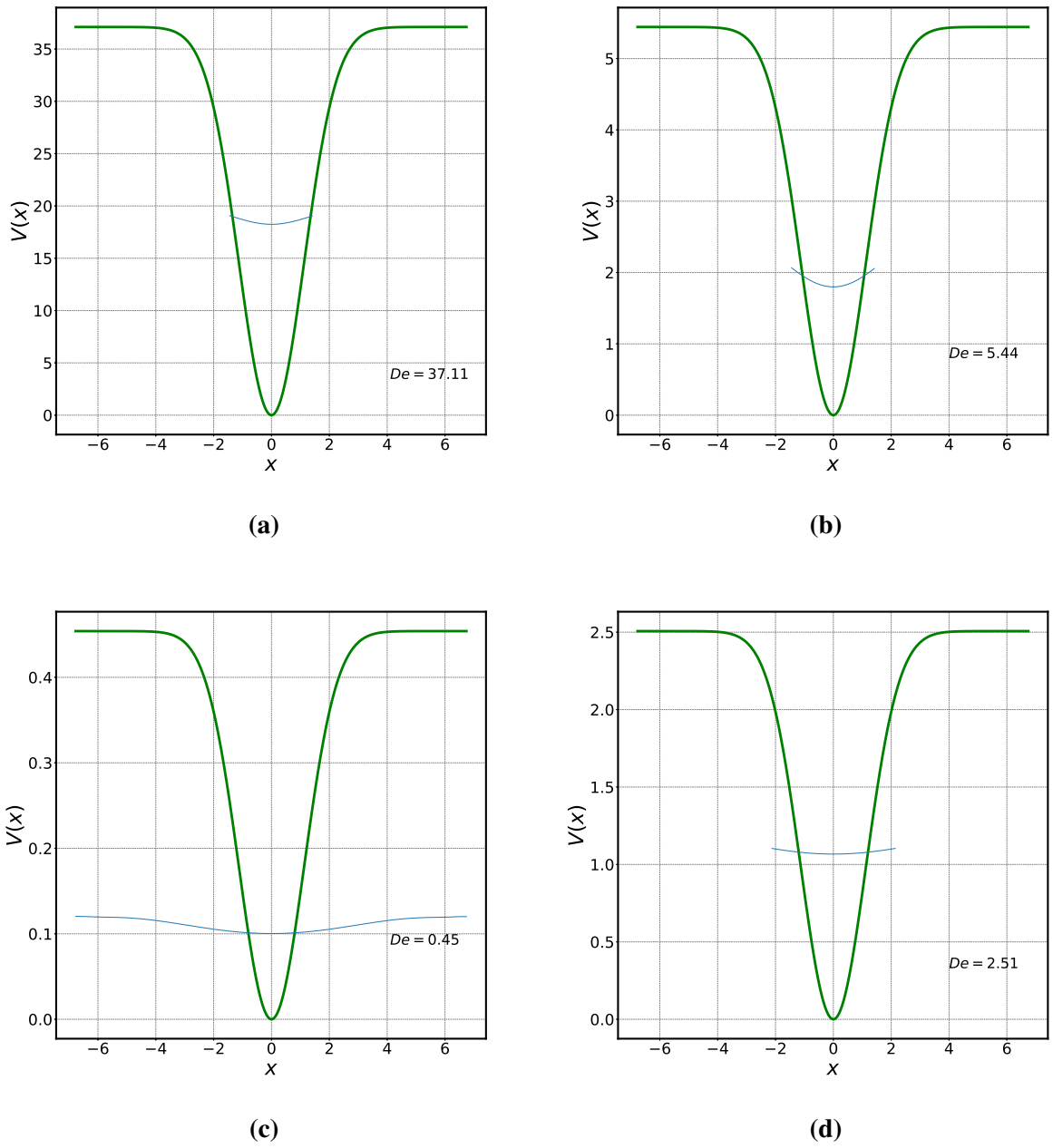
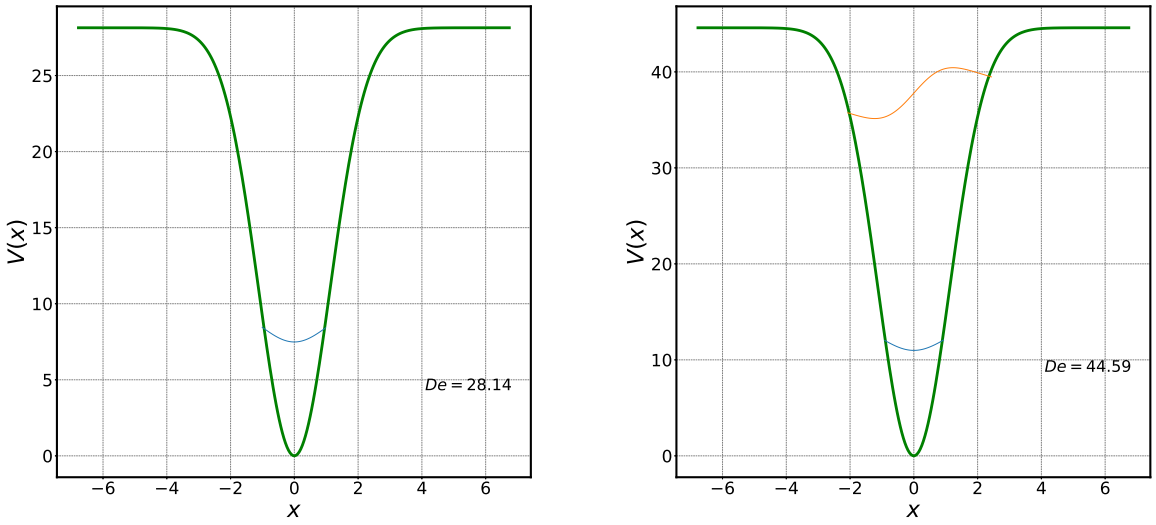


Figure 5.8: Energy profile of methane dimer at an (a) equilibrium monomer separation distance and carbon–carbon distances of (b) 4.5 Å, (c) 6.5 Å, and (d) 8.5 Åeach for the first real mode where the green curve is a Gaussian potential and the blue curve ground state energy level.



(a) **(b)**

Figure 5.9: Energy profile of benzene dimer in the (a) T-shaped configuration and (b) in the parallel displaced configuration for the first real mode where the green curve is a Gaussian potential and the blue/orange curve are the energy levels.

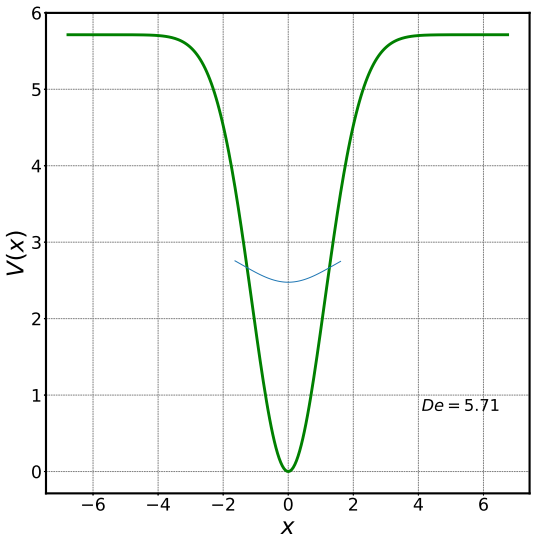


Figure 5.10: Energy profile of chalcone dimer for the first real mode where the green curve is a Gaussian potential and the blue curve is the energy level.

The RRGO approximation was used for all modes and the total entropy was determined for each of the systems. As shown in Table 5.3, the entropy of each system is larger for the RRGO approximation than the entropy determined from the RRHO approximation. This is to be expected as the RRHO approximation will ignore these low energy modes. The difference between the entropy for two approaches is relatively small for ethane (0.820 cal/(mol·K)), as is to be expected since it is a single molecule. This minimal entropic difference results in a -0.24 kcal/mol impact on the free energy at 298 K for ethane. When two molecules are considered, however, the difference between the two approximations becomes substantially larger. For the methane dimer at equilibrium, the difference in the RRGO and RRHO entropies would decrease the free energy by -3.92 kcal/mol. As the methane dimer carbon separation distance is increased, the entropic impact of the low frequency modes also increases and as such the difference between the RRHO and RRGO approximations becomes larger. For methane dimer with a carbon separation distance of 4.5 Å, using the RRGO approximation would correct the free energy by -4.38 kcal/mol. While the methane dimer at 6.5 Å has an entropy (41.774 cal/(mol·K)) which is larger than at 8.5 Å (40.536 cal/(mol·K)), we hypothesize this is caused by long range noncovalent forces and is likely physical. Using the RRGO approximation, the free energies for the methane dimers at 6.5 Å and 8.5 Å would be corrected by -3.62 kcal/mol and -3.85 kcal/mol respectively. From the benzene dimer dihedral umbrella sampling PMF (Figure 5.5), it was suggested that the free energy for the T-shaped configuration will be slightly more stable than the parallel displaced configuration due to entropic accessibility of additional states. This hypothesis is further supported in Table 5.3 for both the RRHO and RRGO approximations as the entropy for the T-shaped configuration is larger in magnitude than the parallel displaced configuration for both approximations. Using the RRGO approximation would correct the free energy by -2.81 kcal/mol for the T-shaped configuration and by -2.71 kcal/mol for the parallel displaced configuration, further stabilizing the T-shaped configuration. The above systems are relatively simple in the intramolecular degrees of freedom and have fewer intermolecular sources for noncovalent interactions relative to the chalcone dimer system. As a result, the chalcone dimer contains significantly more low energy modes which would be

ignored by the RRHO approximation but are considered for the RRGO approximation leading to a free energy correction of -6.74 kcal/mol when *in vacuo*. This substantial free energy correction translates to a -6.60 kcal/mol when in cyclohexane resulting in a corrected dimerization free energy for chalcone in cyclohexane of 5.56 kcal/mol which is in close agreement with the experimental value of 5.1 kcal/mol.[169]

Table 5.3: For each of the eight systems, the entropy computed using the RRHO approximation (S_{HO}), the entropy computed using the RRGO approximation (S_{GO}), and the impact on the free energy ($\Delta\Delta G^{\circ}_{Corr}$) is reported.

System	S_{HO} (cal/(mol·K))	S_{GO} (cal/(mol·K))	$\Delta\Delta G^{\circ}_{Corr}$ (kcal/mol)
Ethane	1.958	2.778	-0.24
Methane dimer at equilibrium	22.205	35.340	-3.92
Methane dimer at 4.5 Å	24.384	39.082	-4.38
Methane dimer at 6.5 Å	29.623	41.774	-3.62
Methane dimer at 8.5 Å	27.633	40.536	-3.85
Benzene dimer T-shaped	37.709	47.141	-2.81
Benzene dimer parallel displaced	36.250	45.346	-2.71
Chalcone dimer *in vacuo*	111.134	133.743	-6.74
Chalcone dimer in cyclohexane PCM	112.821	134.960	-6.60

5.6 Conclusion

As computing the free energy of association is valuable to a wide range of disciplines, a novel approach to nonempirically correcting the QM computed free energy of association using a RRGO approximation is proposed. The chalcone and benzene dimer PMFs demonstrate that the low-mode frequencies for the intermolecular rotation and translation will contribute significantly to the entropy of the free energy and thus can not be ignored. Using the PMFs of these systems, it is shown that thermodynamic quantities for association can be determined from aaMD umbrella sampling for a range of temperatures by considering the bottom of the well vs. temperature. Doing so,

results in good enthalpic agreement for chalcone dimer to QM results and an enthalpy which is in line with experimental results for the benzene dimer. While the entropy for benzene dimer is almost entirely composed of rotational and translational entropy as the molecule is fairly rigid, the entropy in the chalcone system is also composed of intramolecular degrees of freedom. Additionally, the dihedral for benzene dimer PMF suggests that while the potential energy for the system would favor a parallel displaced configuration, the T-shaped configuration is more favorable when considering the free energy of the system.

To correct the entropic component from QM calculations, the RRGO approximation is developed in a general way so as to be applicable to any system. Using a Gaussian potential, Gaussians for the basis set expansion, and applying a temperature dependence to an interpolation within the partition function, the RRGO approximation is able to decrease the entropic error for all the systems studied. While the ethane entropy correction using the RRGO approximation was minimal compared to the RRHO approximation, this was not the case for systems with associating pairs of molecules. As the carbon–carbon separation distance in the methane dimer increased from equilibrium up to 6.5 Å, so did the entropy contribution. Although the entropic contribution decreased slightly between 6.5 Å and 8.5Å carbon–carbon separation distance, we hypothesize this is a result of long range noncovalent interactions. Using the RRGO approximation, benzene dimer in the T-shaped configuration is suggested to have a higher entropic contribution to the free energy than the parallel displaced conformation, in agreement with the umbrella sampling results for the dimer dihedral. The impact of using the RRGO approximation was most pronounced for the chalcone dimer. The free energy correction for the chalcone dimer in cyclohexane is in good agreement with the literature value.

The work presented here enables broad employment of the RRGO approximation to a wide variety of systems. While the RRGO approximation script associated with our results was developed specifically for G16, the script can be readily adapted for any QM software package provided a hessian is able to be extracted. Future work will consider a variety of other systems of interest to us in which a rigorous understanding of the free energy is necessary. These future studies will

include reaction mechanisms as well as comparisons of binding free energies between associating species.

Chapter 6

Conclusions

Noncovalent interactions are found throughout nature and are a fundamental component to the field of chemistry. Aggregation is the process through which a finite number of molecules associate through noncovalent interactions. The previous four chapters have focused on the driving forces for small molecule aggregation by detailing dipeptide self-assembly properties (Chapters 2, 3, and 4) and the development of a more rigorous approach to computing the entropy of association (Chapter 5). Detailed here are the conclusions from these chapters.

In Chapter 2, FF was studied at multiple points in the self-assembly pathway. The decomposition of the FF dimer PMF and the significant decrease in the backbone SASA within the first 200 ns of initially dispersed systems demonstrated that backbone–backbone interactions promote the earliest stages of aggregation. Backbone interactions continued to promote the aggregation of initially dispersed systems as was evident by the pair interaction type analysis and the backbone $g(r)$ relative to the sidechain $g(r)$. Although we did not observe nanotube formation, the backbone Ramachandran plot and the calculated IR spectra suggested a reorientation event must occur prior to nanotube ordering. Decomposition of the monomeric addition to the FF nanotube PMF demonstrated that long range solvent mediated forces promote the final step of the FF self-assembly mechanism. From this study, we determined that the self-assembly mechanism is multi-step consisting of backbone interactions driving initial aggregation after which a conformational reorientation must occur prior to nanotube formation at which point solvent mediated forces drive nanotube growth. Finally, this study developed a protocol of generalized analyses for which the aggregation of other dipeptide systems could easily be compared, enabling the genesis for a set of dipeptide self-assembly design principles.

Chapter 3 considers the sequence dependence of VF and FV. The IR and 2DIR results from Christopher Kuhs and Professor Amber Krummel demonstrate a difference in spectra between these two dipeptides. Although expected since they are two different molecules, the origin of the

spectroscopic perturbation was unclear. Unbiased aaMD simulations demonstrated that spectra were likely capturing the ensemble of monomeric conformations as was supported from calculated spectra weighted by the probabilities of conformers. Umbrella sampling simulations demonstrate that the phenylalanine residue is significantly more restricted in the sampling of the χ_1 dihedral phase space for FV than in VF. Therefore, this suggests the sidechain of the phenylalanine residue would not play a role in the water dynamics around the amide bond for FV but would for VF. Solvent dynamics were then measured using aaMD biased to sample the geometry optimized structures for backbone states. The $g(r)$ of the amide carbonyl oxygen to all water oxygens in each of the backbone states demonstrates that water is more ordered around the amide carbonyl in VF than in FV. The water residence lifetimes from these simulations were then measured and weighted by the backbone state probability for each dipeptide. The lifetimes of waters around the amide carbonyl were shown to be longer lived for VF than FV, in agreement in magnitude and trend with rates measured in 2D IR. These results suggested that the combination of differences in conformer sampling and water residence lifetimes would cause the water-amide carbonyl to be more stabilized for VF than FV, resulting in a larger free energy penalty for aggregation.

As the analyses in the above studies were performed in a generalized way, in Chapter 4 the dimerization of VF, FV, IF, and FI was compared to that of FF connecting the studies in Chapter 2 to those in Chapter 3. While not a completed study, the work from this chapter is a foundation from which a study about sequence dependence or solvent effects can be further explored. Although the dimer PMFs of these dipeptides were distinct from one another, there were common features with similar representative structures between them. The common representative structures demonstrate the importance of backbone electrostatic interactions as these were the primary interactions. It is tempting to conclude C-terminal modification of FF decreases the global well depth more than N-terminal modification based on the dimer PMFs, especially those of IF and FI, in this chapter. However, we are reticent to fully put forth this hypothesis as the global well depths for the VF and FV dimer PMFs are within error of each other and instead suggest additional study of comparably sized or smaller sidechains. An additional feature at distances less than the global well was

observed for both FV and FI with representative structures characterized as a 4 contact point anti-parallel dimer interaction type. This feature suggests that, as a function of sidechain steric size, modification of FF's C-terminus to a smaller amino acid enables the possibility for tighter packing. Decomposition of the solute–solute and solvent–solute contributions both into Lennard-Jones and Coulombic components of the free energy demonstrated that backbone electrostatics principally drives dimerization of all dipeptides studied. Additionally, the solvent–solute Lennard-Jones component of the free energy hinder dimerization of VF while promoted it for FV, IF, and FI. When the longer water residence times around the carbonyl of the peptide bond for VF relative to FV seen in Chapter 3 are considered, we hypothesize this causes solvent mediated forces to impede dimerization of VF. Furthermore, we hypothesize that since solvent mediated forces are not hindering the dimerization of IF, the sequence dependence seen between VF and FV has a steric sidechain size dependence.

In Chapter 5, a novel approach to computing the entropy for associating molecules called the rigid-rotor-Gaussian-oscillator (RRGO) approximation is developed. By applying a Gaussian potential, Gaussians in the basis set expansion for the wave function, and a temperature dependence to the interpolation within the partition function, the RRGO approximation is able to compute an expectedly higher entropy for associating molecules after summation over all modes. Using the entropic difference, the free energy of binding can be corrected resulting in a more accurate value for associating complexes. The impact of the RRGO approximation was shown to be most notable for the chalcone dimer system in an implicit cyclohexane solvent, reducing the free energy difference between calculated and experimental values from 7.1 kcal/mol to 0.5 kcal/mol. As the chalcone dimer was seen in MD simulations to sample many intramolecular and intermolecular degrees of freedom, the agreement to experimental results using the RRGO approximation motivates its usage to all associating molecules of interest. Additionally, from umbrella sampling simulations as well as RRGO entropic results give further evidence to suggest that the benzene dimer is more stable when in the T-shaped configuration than in the parallel displaced configuration.

6.1 Continued Work

The work presented in the dipeptides chapters are a general foundation for further study of dipeptides which is needed in order develop generalized design principles. Additionally, many pressing studies within the Rappé group would benefit from computing the entropy in a more rigorous way. Ideas of continued research on the self-assembly of dipeptides as well as the extension of the rigid-rotor-Gaussian-oscillator approximation are presented here.

6.1.1 Dipeptide Self-Assembly

FF is described as an exemplar of small molecule self-assembly[21,24,28,29,62,65–67] with a broad range of potential applications.[21,61–64] To optimize a FF-derivative for a given application, one would need to develop a comprehensive set of design principles. Chapters 2, 3, and 4 provide a foundation for achieving this, however, there are still a significant amount of additional studies to be performed prior to rationally designing FF derivatives. Fortunately, these studies provide broadly applicable analyses and points of comparison with a conscious goal of systematically extending the studies towards development of a set design principles.

An immediately apparent extension of these studies is studying the water residence time around the amide bond of IF and FI. In Chapter 3, it was shown that N-terminal modification of FF to a valine residue increased the water residence time relative to C-terminal modification. To perform this study, simulations of monomeric IF and FI would need to be performed in order to determine backbone dihedral states. After which, belly simulations would be performed of these states and the water residence times measured. This study will test if the sequence dependence of dipeptides is dependent on steric size or a general design principle.

An additional extension of Chapters 3 and 4 would be to consider the impact of solvent on dipeptide self-assembly. All studies presented in this dissertation were solvated in water although experimental studies have shown that solvent impacts FF self-assembly in a variety of ways.[12] To measure the impact solvent would have on dipeptide self-assembly, future studies should start with the decomposition of the dimer PMF for FF in a nonpolar solvent. Once the dimerization

of FF is understood, similar studies to those seen in Chapters 3 and 4 should be considered. As solvent hindered VF and FF dimerization while solvent–solute Lennard-Jones forces promoted dimerization for FV, IF, and FI, it would be of interest to see if these results are caused by the polarity of the solvent or some other factor.

Additionally, characteristics of larger system sizes should be measured while studying sequence dependence and the impact of choice in solvent. The SASA, backbone and sidechain $g(r)$'s, pair interaction type, and calculated IR spectra seen in Chapter 2 would provide quantification of differences for different dipeptide systems relative to FF. While somewhat arbitrary, systems with 60 FF dipeptides performed in triplicate did not demonstrate significant differences relative to the system with 120 FF dipeptides. Therefore, for future work, 60 dipeptides is a recommended system size and is substantially larger than most other aaMD dipeptide systems in the literature.

Two additional amino acids are suggested for further study of dipeptide aggregation and self-assembly. The amino acids studied here include phenylalanine, isoleucine, and valine all of which are hydrophobic to varying degrees. By considering the substitution of a phenylalanine residue to an alanine residue, the impact of sidechain hydrophobicity and steric size could be further studied. Furthermore, the impact of a hydrophilic sidechain could be considered while steric sidechain size would be comparable to the valine sidechain size by considering the threonine amino acid. Apart from these, there 15 other natural amino acids that one could choose from prior to considering non-natural amino acids or termini capping groups which would have significant impacts to dipeptide aggregation and self-assembly.

Dipeptide self-assembly has a plethora of potential applications which are dependent on the rational design of a dipeptide for a given application. Previous chapters have made a concerted effort to provide a foundation from which future studies could be built out of in a consistent manner so as to enable concise comparisons and isolation of specific modification impacts. That said, there are still many research avenues needed to be explored in order to begin to have design principles necessary for control of the self-assembled structures.

6.1.2 Development of a Rigid-Rotor-Gaussian-Oscillator (RRGO) Approximation

The RRGO approximation is a generalizable model which can be implemented for any system and has been shown to significantly improve the computed free energy of association with respect to experiment. While there is likely to be small changes to the chapter prior to publication in a peer reviewed journal, the methodology in the application of the RRGO approximation is complete. As such, immediate application of the RRGO approximation is of interest for several studies currently being investigated within the Rappé group. Specifically, these include the study of photo-redox complexes and perfluoroalkyl substances (PFAS).

The resulting free energy correction from the RRGO approximation being in significantly better agreement with experimental results for the chalcone dimer motivates the study of other associating complexes. In particular, the Rappé group has several studies involving chromium-ligand complexes and the resulting photoredox catalysis. Due to the metal center, QM calculations in combination with the RRGO approximation will give significant insight into ligand selection, guiding experimental studies further.

As described in Chapter 5, low energy modes are often ignored in computing the free energy for a system using common QM software packages. For the PFAS family of molecules, many of the modes involving a fluorine stretches are energetically under this threshold and would not be considered. Therefore, the error in the entropy of independent PFAS molecules would be significant. As it is of interest to study the temperature dependence of the activation energy, the slope of which is the entropy, of various PFAS molecules during the incineration process, the RRGO approximation will be of great use.

Bibliography

(1) Zhang, M.; Cao, S.; Liu, A.; Cornelissen, J. J.; Lemay, S. G. Self-Assembly of Viral Capsid Proteins Driven by Compressible Nanobubbles. *J. Phys. Chem.* **2020**, *11*, 10421–10424.

(2) Garmann, R. F.; Goldfain, A. M.; Manoharan, V. N. Measurements of the self-assembly kinetics of individual viral capsids around their RNA genome. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 22485–22490.

(3) Huang, C.; Ghanati, E.; Schmit, J. D. Theory of Sequence Effects in Amyloid Aggregation. *J. Phys. Chem. B* **2018**, *122*, 5567–5578.

(4) Spillantini, M. G.; Schmidt, M. L.; Lee, V. M.-Y.; Trojanowski, J. Q.; Jakes, R.; Goedert, M. α-Synuclein in Lewy bodies. *Nature* **1997**, *388*, 839–840.

(5) Hardy, J.; Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **2002**, *297*, 353–356.

(6) Gupta, S.; Singh, I.; Sharma, A. K.; Kumar, P. Ultrashort Peptide Self-Assembly: Front-Runners to Transport Drug and Gene Cargos. *Front. Bioeng. Biotechnol.* **2020**, *8*.

(7) Mandal, D.; Nasrolahi Shirazi, A.; Parang, K. Self-assembly of peptides to nanostructures. *Org. Biomol. Chem.* **2014**, *12*, 3544–3561.

(8) Anderson, J.; Lake, P. T.; McCullagh, M. Initial Aggregation and Ordering Mechanism of Diphenylalanine from Microsecond All-Atom Molecular Dynamics Simulations. *J. Phys. Chem. B* **2018**, *122*, 12331–12341.

(9) Gale, C. D.; Derakhshani-Molayousefi, M.; Levinger, N. E. How to Characterize Amorphous Shapes: The Tale of a Reverse Micelle. *J. Phys. Chem. B* **2022**, *126*, 953–963.

(10) Ghosh, S.; Ray, A.; Pramanik, N. Self-assembly of surfactants: An overview on general aspects of amphiphiles. *Biophys. Chem.* **2020**, *265*, 106429.

(11) Fenyves, R.; Schmutz, M.; Horner, I. J.; Bright, F. V.; Rzayev, J. Aqueous self-assembly of giant bottlebrush block copolymer surfactants as shape-tunable building blocks. *J. Am. Chem. Soc.* **2014**, *136*, 7762–7770.

(12) Mason, T. O.; Chirgadze, D. Y.; Levin, A.; Adler-Abramovich, L.; Gazit, E.; Knowles, T. P. J.; Buell, A. K. Expanding the Solvent Chemical Space for Self-Assembly of Dipeptide Nanostructures. *ACS Nano* **2014**, *8*, 1243–1253.

(13) Acuña, S. M.; Veloso, M. C.; Toledo, P. G. Self-Assembly of Diphenylalanine-Based Nanostructures in Water and Electrolyte Solutions. *J. Nanomater.* **2018**, *2018*, 1–7.

(14) Bock, H.; Gubbins, K. E. Anomalous temperature dependence of surfactant self-assembly from aqueous solution. *Phys. Rev. Lett.* **2004**, *92*, 135701.

(15) Burz, D. S.; Ackers, G. K. Cooperativity Mutants of Bacteriophage λ cI Repressor: Temperature Dependence of Self-Assembly. *Biochemistry* **1996**, *35*, 3341–3350.

(16) Kim, J.; Han, T. H.; Kim, Y.-I.; Park, J. S.; Choi, J.; Churchill, D. G.; Kim, S. O.; Ihee, H. Role of Water in Directing Diphenylalanine Assembly into Nanotubes and Nanowires. *Adv. Mater.* **2010**, *22*, 583–587.

(17) Lu, X.; Li, X.; Guo, K.; Xie, T. Z.; Moorefield, C. N.; Wesdemiotis, C.; Newkome, G. R. Probing a hidden world of molecular self-assembly: Concentration-dependent, three-dimensional supramolecular interconversions. *J. Am. Chem. Soc.* **2014**, *136*, 18149–18155.

(18) Xue, W. F.; Homans, S. W.; Radford, S. E. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *PNAS* **2008**, *105*, 8926–8931.

(19) Davey, R. J.; Schroeder, S. L.; Ter Horst, J. H. Nucleation of Organic Crystals–A Molecular Perspective. *Angew. Chem. Int. Ed.* **2013**, *52*, 2166–2179.

(20) Anwar, J.; Zahn, D. Uncovering Molecular Processes in Crystal Nucleation and Growth by Using Molecular Simulation. *Angew. Chem. Int. Ed.* **2011**, *50*, 1996–2013.

(21) Yan, X.; Zhu, P.; Li, J. Self-Assembly and Application of Diphenylalanine-Based Nanostructures. *Chem. Soc. Rev.* **2010**, *39*, 1877–1890.

(22) Rosenman, G.; Beker, P.; Koren, I.; Yevnin, M.; Bank-Srour, B.; Mishina, E.; Semin, S. Bioinspired Peptide Nanotubes: Deposition Technology, Basic Physics and Nanotechnology Applications. *J. Pept. Sci.* **2011**, *17*, 75–87.

(23) Kol, N.; Adler-Abramovich, L.; Barlam, D.; Shneck, R. Z.; Gazit, E.; Rousso, I. Self-Assembled Peptide Nanotubes Are Uniquely Rigid Bioinspired Supramolecular Structures. *Nano Lett.* **2005**, *5*, 1343–1346.

(24) Adler-Abramovich, L.; Reches, M.; Sedman, V. L.; Allen, S.; Tendler, S. J. B.; Gazit, E. Thermal and Chemical Stability of Diphenylalanine Peptide Nanotubes: Implications for Nanotechnological Applications. *Langmuir* **2006**, *22*, 1313–1320.

(25) Andersen, K. B.; Castillo-Leon, J.; Hedström, M.; Svendsen, W. E. Stability of diphenylalanine peptide nanotubes in solution. *Nanoscale* **2011**, *3*, 994–998.

(26) De Groot, N. S.; Parella, T.; Aviles, F. X.; Vendrell, J.; Ventura, S. Ile-Phe Dipeptide Self-Assembly: Clues to Amyloid Formation. *Biophys. J.* **2007**, *92*, 1732–1741.

(27) Mayans, E.; Casanovas, J.; Gil, A. M.; Jiménez, A. I.; Cativiela, C.; Puiggalí, J.; Alemán, C. Diversity and Hierarchy in Supramolecular Assemblies of Triphenylalanine: From Laminated Helical Ribbons to Toroids. *Langmuir* **2017**, *33*, 4036–4048.

(28) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. Probing the Self-Assembly Mechanism of Diphenylalanine-Based Peptide Nanovesicles and Nanotubes. *ACS Nano* **2012**, *6*, 3907–3918.

(29) Jeon, J.; Mills, C. E.; Shell, M. S. Molecular Insights into Diphenylalanine Nanotube Assembly: All-Atom Simulations of Oligomerization. *J. Phys. Chem. B* **2013**, *117*, 3935–3943.

(30) Azuri, I.; Adler-Abramovich, L.; Gazit, E.; Hod, O.; Kronik, L. Why Are Diphenylalanine-Based Peptide Nanostructures so Rigid? Insights from First Principles Calculations. *J. Am. Chem. Soc.* **2014**, *136*, 963–969.

(31) Rissanou, A. N.; Georgilis, E.; Kasotakis, E.; Mitraki, A.; Harmandaris, V. Effect of Solvent on the Self-Assembly of Dialanine and Diphenylalanine Peptides. *J. Phys. Chem. B* **2013**, *117*, 3962–3975.

(32) Frederix, P. W. J. M.; Ulijn, R. V.; Hunt, N. T.; Tuttle, T. Virtual Screening for Dipeptide Aggregation: Toward Predictive Tools for Peptide Self-Assembly. *J. Phys. Chem. Lett.* **2011**, *2*, 2380–2384.

(33) Görbitz, C. H. Nanotube Formation by Hydrophobic Dipeptides. *Chemistry* **2001**, *7*, 5153–5159.

(34) Mayans, E.; Ballano, G.; Casanovas, J.; Díaz, A.; Pérez-Madrigal, M. M.; Estrany, F.; Puiggalí, J.; Cativiela, C.; Alemán, C. Self-Assembly of Tetraphenylalanine Peptides. *Chem. Eur. J.* **2015**, *21*, 16895–16905.

(35) Li, Q.; Ma, H.; Jia, Y.; Li, J.; Zhu, B. Facile fabrication of diphenylalanine peptide hollow spheres using ultrasound-assisted emulsion templates. *Chem. Commun.* **2015**, *51*, 7219–7221.

(36) Wang, J.; Liu, K.; Yan, L.; Wang, A.; Bai, S.; Yan, X. Trace Solvent as a Predominant Factor To Tune Dipeptide Self-Assembly. *ACS Nano* **2016**, *10*, 2138–2143.

(37) Reches, M.; Gazit, E. Casting Metal Nanowires Within Discrete Self-Assembled Peptide Nanotubes. *Science* **2003**, *300*, 625–627.

(38) Amdursky, N.; Molotskii, M.; Gazit, E.; Rosenman, G. Elementary Building Blocks of Self-Assembled Peptide Nanotubes. *J. Am. Chem. Soc.* **2010**, *132*, 15632–15636.

(39) Mammen, M.; Shakhnovich, E. I.; Deutch, J. M.; Whitesides, G. M. Estimating the Entropic Cost of Self-Assembly of Multiparticle Hydrogen-Bonded Aggregates Based on the Cyanuric Acid·Melamine Lattice. *J. Org. Chem.* **1998**, *63*, 3821–3830.

(40) Wang, T.; Liang, Y.; Yu, Z.-X. Density Functional Theory Study of the Mechanism and Origins of Stereoselectivity in the Asymmetric Simmons-Smith Cyclopropanation with Charette Chiral Dioxaborolane Ligand. *J. Am. Chem. Soc.* **2011**, *133*, 13762–13763.

(41) Kua, J.; Krizner, H. E.; De Haan, D. O. Thermodynamics and Kinetics of Imidazole Formation from Glyoxal, Methylamine, and Formaldehyde: A Computational Study. *J. Phys. Chem. A* **2011**, *115*, 1667–1675.

(42) Deubel, D. V.; Lau, J. K. In silico evolution of substrate selectivity: comparison of organometallic ruthenium complexes with the anticancer drug cisplatin. *Chem. Commun.* **2006**, 2451–2453.

(43) Huang, F.; Lu, G.; Zhao, L.; Li, H.; Wang, Z.-X. The Catalytic Role of N-Heterocyclic Carbene in a Metal-Free Conversion of Carbon Dioxide into Methanol: A Computational Mechanism Study. *J. Am. Chem. Soc.* **2010**, *132*, 12388–12396.

(44) Zhao, L.; Wen, M.; Wang, Z.-X. Reaction Mechanism of Phosphane-Catalyzed [4+2] Annulations between α-Alkylallenoates and Activated Alkenes: A Computational Study. *Eur. J. Org. Chem.* **2012**, 3587–3597.

(45) Li, H.; Lu, G.; Jiang, J.; Huang, F.; Wang, Z.-X. Computational Mechanistic Study on C_p*Ir Complex-Mediated Acceptorless Alcohol Dehydrogenation: Bifunctional Hydrogen Transfer vs β-H Elimination. *Organometallics* **2011**, *30*, 2349–2363.

(46) Deubel, D. V. Mechanism and Control of Rare Tautomer Trapping at a Metal–Metal Bond: Adenine Binding to Dirhodium Antitumor Agents[1]. *J. Am. Chem. Soc.* **2008**, *130*, 665–675.

(47) Mazzone, G.; Russo, N.; Sicilia, E. Homogeneous Gold Catalysis: Hydration of 1,2-Diphenylacetylene with Methanol in Aqueous Media. A Theoretical Viewpoint. *Organometallics* **2012**, *31*, 3074–3080.

(48) Nasiri, R.; Field, M. J.; Zahedi, M.; Moosavi-Movahedi, A. A. Comparative DFT Study To Determine if α-Oxoaldehydes are Precursors for Pentosidine Formation. *J. Phys. Chem. A* **2012**, *116*, 2986–2996.

(49) Monnat, F.; Vogel, P.; Rayón, V. M.; Sordo, J. A. Ab Initio and Experimental Studies on the Hetero-Diels-Alder and Cheletropic Additions of Sulfur Dioxide to (E)-1-Methoxybutadiene: A Mechanism Involving Three Molecules of SO2. *J. Org. Chem.* **2002**, *67*, 1882–1889.

(50) Hölscher, M.; Leitner, W. Heterolytic Outer-Sphere Cleavage of H2 for the Reduction of N2 in the Coordination Sphere of Transition Metals–A DFT Study. *Angew. Chem. Int. Ed.* **2012**, *51*, 8225–8229.

(51) Sieffert, N.; Bühl, M. Hydrogen Generation from Alcohols Catalyzed by Ruthenium-Triphenylphosphine Complexes: Multiple Reaction Pathways. *J. Am. Chem. Soc.* **2010**, *132*, 8056–8070.

(52) Martin, R. L.; Hay, P. J.; Pratt, L. R. Hydrolysis of Ferric Ion in Water and Conformational Equilibrium. *J. Phys. Chem. A* **1998**, *102*, 3565–3573.

(53) Zhao, L.; Li, S.-J.; Fang, D.-C. A Theoretical Study of Ene Reactions in Solution: A Solution-Phase Translational Entropy Model. *ChemPhysChem* **2015**, *16*, 3711–3718.

(54) Han, L.-L.; Li, S.-J.; Fang, D.-C. Theoretical estimation of kinetic parameters for nucleophilic substitution reactions in solution: an application of a solution translational entropy model. *Phys. Chem. Chem. Phys.* **2016**, *18*, 6182–6190.

(55) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. Eur. J.* **2012**, *18*, 9955–9964.

(56) Ayala, P. Y.; Schlegel, H. B. Identification and treatment of internal rotation in normal mode vibrational analysis. *J. Chem. Phys.* **1998**, *108*, 2314–2325.

(57) Ellingson, B. A.; Lynch, V. A.; Mielke, S. L.; Truhlar, D. G. Statistical thermodynamics of bond torsional modes: Tests of separable, almost-separable, and improved Pitzer–Gwinn approximations. *J. Chem. Phys.* **2006**, *125*, 084305.

(58) Li, Y.-P.; Gomes, J.; Sharada, S. M.; Bell, A. T.; Head-Gordon, M. Improved Force-Field Parameters for QM/MM Simulations of the Energies of Adsorption for Molecules in Zeolites and a Free Rotor Correction to the Rigid Rotor Harmonic Oscillator Model for Adsorption Enthalpies. *J. Phys. Chem. C* **2015**, *119*, 1840–1850.

(59) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J. Phys. Chem. B* **2011**, *115*, 14556–14562.

(60) Reches, M.; Gazit, E. Formation of Closed-Cage Nanostructures by Self-Assembly of Aromatic Dipeptides. *Nano Lett.* **2004**, *4*, 581–585.

(61) Schnaider, L.; Brahmachari, S.; Schmidt, N. W.; Mensa, B.; Shaham-Niv, S.; Bychenko, D.; Adler-Abramovich, L.; Shimon, L. J. W.; Kolusheva, S.; DeGrado, W. F.; Gazit, E. Self-Assembling Dipeptide Antibacterial Nanostructures with Membrane Disrupting Activity. *Nat. Commun.* **2017**, *8*, 1365.

(62) Silva, R. F.; Araújo, D. R.; Silva, E. R.; Ando, R. A.; Alves, W. A. L-Diphenylalanine microtubes as a Potential Drug-Delivery System: Characterization, Release Kinetics, and Cytotoxicity. *Langmuir* **2013**, *29*, 10205–10212.

(63) Nguyen, V.; Zhu, R.; Jenkins, K.; Yang, R. Self-Assembly of Diphenylalanine Peptide with Controlled Polarization for Power Generation. *Nat. Commun.* **2016**, *7*, 13566.

(64) Khanra, S.; Ghosh, K.; Ferreira, F. F.; Alves, W. A.; Punzo, F.; Yu, P.; Guha, S. Probing Nonlinear Optical Coefficients in Self-Assembled Peptide Nanotubes. *Phys. Chem. Chem. Phys.* **2017**, *19*, 3084–3093.

(65) Castillo-León, J., *Chapter 2 - Fabrication of Nanostructures Using Self-Assembled Peptides as Templates: The Diphenylalanine Case*; Castillo-León, J., Svendsen, W. E., Eds.; Micro and Nano Technologies; William Andrew Publishing: Oxford, 2015, pp 21–31.

(66) Han, T. H.; Oh, J. K.; Lee, G.-J.; Pyun, S.-I.; Kim, S. O. Hierarchical Assembly of Diphenylalanine into Dendritic Nanoarchitectures. *Colloids Surf. B* **2010**, *79*, 440–445.

(67) Görbitz, C. H. The Structure of Nanotubes Formed by Diphenylalanine, the Core Recognition Motif of Alzheimer's Beta-Amyloid Polypeptide. *Chem. Comm.* **2006**, *22*, 2332–2334.

(68) Görbitz, C. H. Microporous Organic Materials from Hydrophobic Dipeptides. *Chem. Eur. J.* **2007**, *13*, 1022–1031.

(69) Gazit, E. A Possible Role for Pi-Stacking in the Self-Assembly of Amyloid Fibrils. *FASEB J.* **2002**, *16*, 77–83.

(70) Azriel, R.; Gazit, E. Analysis of the Minimal Amyloid-forming Fragment of the Islet Amyloid Polypeptide an Experimental Support for the Key Role of the Phenylalanine Residue in Amyloid Formation. *J. Biol. Chem.* **2001**, *276*, 34156–34161.

(71) Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C. Molecular Basis for Amyloid Fibril Formation and Stability. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 315–320.

(72) Tamamis, P.; Adler-Abramovich, L.; Reches, M.; Marshall, K.; Sikorski, P.; Serpell, L.; Gazit, E.; Archontis, G. Self-Assembly of Phenylalanine Oligopeptides: Insights from Experiments and Simulations. *Biophys. J.* **2009**, *96*, 5020–5029.

(73) 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* **2006**, *65*, 712–725.

(74) Dalgicdir, C.; Sensoy, O.; Peter, C.; Sayar, M. A Transferable Coarse-Grained Model for Diphenylalanine: How to Represent an Environment Driven Conformational Transition. *J.Chem. Phys.* **2013**, *139*, 234115.

(75) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comp. Chem.* **2004**, *25*, 1656–1676.

(76) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-

Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(77) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

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

(79) Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A. ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 4515–4534.

(80) Koes, D. R.; Vries, J. K. Evaluating Amber Force Fields Using Computed NMR Chemical Shifts. *Proteins* **2017**, *85*, 1944–1956.

(81) Case, D.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Simmerling,

C.; Botello-Smith, W.; Swails, J.; Walker, R.; Wang, J.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P., *AMBER 2017*, University of California, San Francisco, 2017.

(82) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A., *Mercury CSD 2.0 – New Features for the Visualization and Investigation of Crystal Structures*; 2; International Union of Crystallography: 2008; Vol. 41, pp 466–470.

(83) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed Without Compromise–A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.

(84) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

(85) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(86) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations., ed. by Benthall, S.; Rostrup, S., 2016.

(87) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

(88) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.

(89) Thiede, E. H.; Van Koten, B.; Weare, J.; Dinner, A. R. Eigenvector Method for Umbrella Sampling Enables Error Analysis. *J. Chem. Phys.* **2016**, *145*, 084115.

(90) Krimm, S. Vibrational Analysis of Conformation in Peptides, Polypeptides, and Proteins. *Biopolymers* **1983**, *22*, 217–225.

(91) Torii, H.; Tasumi, M. Model Calculations on the Amide-I Infrared Bands of Globular Proteins. *J. Chem. Phys.* **1992**, *96*, 3379–3387.

(92) Małolepsza, E.; Straub, J. E. Empirical Maps For The Calculation of Amide I Vibrational Spectra of Proteins From Classical Molecular Dynamics Simulations. *J. Phys. Chem. B* **2014**, *118*, 7848–7855.

(93) Hamm, P.; Woutersen, S. Coupling of the Amide I Modes of the Glycine Dipeptide. *Bull. Chem. Soc. Jpn.* **2002**, *75*, 985–988.

(94) Katari, M.; Nicol, E.; Steinmetz, V.; van der Rest, G.; Carmichael, D.; Frison, G. Improved Infrared Spectra Prediction by DFT from a New Experimental Database. *Chem. Eur. J.* **2017**, *23*, 8414–8423.

(95) Levin, A.; Mason, T. O.; Adler-Abramovich, L.; Buell, A. K.; Meisl, G.; Galvagnion, C.; Bram, Y.; Stratford, S. A.; Dobson, C. M.; Knowles, T. P. J.; Gazit, E. Ostwald's Rule of Stages Governs Structural Transitions and Morphology of Dipeptide Supramolecular Polymers. *Nat. Commum.* **2014**, *5*, 5219.

(96) Torii, H.; Tasumi, M. Ab Initio Molecular Orbital Study of the Amide I Vibrational Interactions Between the Peptide Groups in Di- and Tripeptides and Considerations on the Conformation of the Extended Helix. *J. Raman Spectrosc.* **1998**, *29*, 81–86.

(97) Arnon, Z. A.; Vitalis, A.; Levin, A.; Michaels, T. C. T.; Caflisch, A.; Knowles, T. P. J.; Adler-Abramovich, L.; Gazit, E. Dynamic Microfluidic Control of Supramolecular Peptide Self-Assembly. *Nat. Commun.* **2016**, *7*, 13190.

(98) McGaughey, G. B.; Gagné, M; Rappé, A. K. pi-Stacking Interactions. Alive and Well in Proteins. *J. Biol. Chem.* **1998**, *273*, 15458–15463.

(99) Colombo, G.; Soto, P.; Gazit, E. Peptide Self-Assembly at the Nanoscale: a Challenging Target for Computational and Experimental Biotechnology. *Trends Biotechnol.* **2007**, *25*, 211–218.

(100) Mahler, A.; Reches, M.; Rechter, M.; Cohen, S.; Gazit, E. Rigid, Self-Assembled Hydrogel Composed of a Modified Aromatic Dipeptide. *Adv. Mater.* **2006**, *18*, 1365–1370.

(101) Zelenovskiy, P.; Kornev, I.; Vasilev, S.; Kholkin, A. On the Origin of the Great Rigidity of Self-Assembled Diphenylalanine Nanotubes. *Phys. Chem. Chem. Phys.* **2016**, *18*, 29681–29685.

(102) Saper, G.; Hess, H. Synthetic Systems Powered by Biological Molecular Motors. *Chem. Rev.* **2020**, *120*, 288–309.

(103) Silva, R. F.; Araújo, D. R.; Silva, E. R.; Ando, R. A.; Alves, W. A. L -Diphenylalanine Microtubes As a Potential Drug-Delivery System: Characterization, Release Kinetics, and Cytotoxicity. *Langmuir* **2013**, *29*, 10205–10212.

(104) Schnaider, L.; Brahmachari, S.; Schmidt, N. W.; Mensa, B.; Shaham-Niv, S.; Bychenko, D.; Adler-Abramovich, L.; Shimon, L. J.; Kolusheva, S.; Degrado, W. F.; Gazit, E. Self-Assembling Dipeptide Antibacterial Nanostructures with Membrane Disrupting Activity. *Nat. Commun.* **2017**, *8*, No. 1365.

(105) Yan, X.; Cui, Y.; He, Q.; Wang, K.; Li, J. Organogels Based on Self-Assembly of Diphenylalanine Peptide and Their Application To Immobilize Quantum Dots. *Chem. Mater.* **2008**, *20*, 1522–1526.

(106) Wijerathne, N. K.; Kumar, M.; Ulijn, R. V. Fmoc-Dipeptide/Porphyrin Molar Ratio Dictates Energy Transfer Efficiency in Nanostructures Produced by Biocatalytic Co-Assembly. *Chem.: Eur. J.* **2019**, *25*, 11847–11851.

(107) Sahoo, J. K.; Nazareth, C.; VandenBerg, M. A.; Webber, M. J. Self-Assembly of Amphiphilic Tripeptides with Sequence-Dependent Nanostructure. *Biomater. Sci.* **2017**, *5*, 1526–1530.

(108) Kar, K.; Wang, Y.-H.; Brodsky, B. Sequence Dependence of Kinetics and Morphology of Collagen Model Peptide Self-Assembly into Higher Order Structures. *Protein Sci.* **2008**, *17*, 1086–1095.

(109) Frederix, P. W. J. M.; Scott, G. G.; Abul-Haija, Y. M.; Kalafatovic, D.; Pappas, C. G.; Javid, N.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. Exploring the Sequence Space for (Tri-)Peptide Self-Assembly to Design and Discover New Hydrogels. *Nat. Chem.* **2015**, *7*, 30–37.

(110) Shim, S.-H.; Zanni, M. T. How to Turn Your Pump–Probe Instrument into a Multidimensional Spectrometer: 2D IR and Vis Spectroscopies Via Pulse Shaping. *Phys. Chem. Chem. Phys.* **2009**, *11*, 748–761.

(111) MacKenzie, K. J. D.; Smith, M. E., *Multinuclear solid-state NMR of inorganic materials*; Pergamon: 2002, p 727.

(112) Feng, C.-J.; Tokmakoff, A. The Dynamics of Peptide-Water Interactions in Dialanine: An Ultrafast Amide I 2D IR and Computational Spectroscopy Study. *J. Chem. Phys.* **2017**, *147*, 085101.

(113) Cyran, J. D.; Nite, J. M.; Krummel, A. T. Characterizing Anharmonic Vibrational Modes of Quinones with Two-Dimensional Infrared Spectroscopy. *J. Phys. Chem. B* **2015**, *119*, 8917–8925.

(114) Cyran, J. D.; Krummel, A. T. Probing structural features of self-assembled violanthrone-79 using two dimensional infrared spectroscopy. *J. Chem. Phys.* **2015**, *142*, 212435.

(115) 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. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(116) Mills, R. Self-Diffusion in Normal and Heavy Water in the Range 1-45°. *J. Phys. Chem.* **1973**, *77*, 685–688.

(117) Mahoney, M. W.; Jorgensen, W. L. Diffusion Constant of the TIP5P Model of Liquid Water. *J. Chem. Phys.* **2001**, *114*, 363–366.

(118) Creasey, R. C. G.; Louzao, I.; Arnon, Z. A.; Marco, P.; Adler-Abramovich, L.; Roberts, C. J.; Gazit, E.; Tendler, S. J. B. Disruption of Diphenylalanine Assembly by a Boc-Modified Variant. *Soft Matter* **2016**, *12*, 9451–9457.

(119) Adler-Abramovich, L.; Gazit, E. The Physical Properties of Supramolecular Peptide Assemblies: from Building Block Association to Technological Applications. *Chem. Soc. Rev.* **2014**, *43*, 6881–6893.

(120) Mu, X.; Eckes, K. M.; Nguyen, M. M.; Suggs, L. J.; Ren, P. Experimental and Computational Studies Reveal an Alternative Supramolecular Structure for Fmoc-Dipeptide Self-Assembly. *Biomacromolecules* **2012**, *13*, 3562–3571.

(121) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. Triphenylalanine Peptides Self-Assemble into Nanospheres and Nanorods that are Different from the Nanovesicles and Nanotubes Formed by Diphenylalanine Peptides. *Nanoscale* **2014**, *6*, 2800–2811.

(122) Plata, R. E.; Singleton, D. A. A Case Study of the Mechanism of Alcohol-Mediated Morita Baylis–Hillman Reactions. The Importance of Experimental Observations. *J. Am. Chem. Soc.* **2015**, *137*, 3811–3826.

(123) Ochterski, J. W., *Thermochemistry in Gaussian*; Gaussian Inc.: 2000, pp 1–19.

(124) Zhuang, C.; Zhang, W.; Sheng, C.; Zhang, W.; Xing, C.; Miao, Z. Chalcone: A Privileged Structure in Medicinal Chemistry. *Chem. Rev.* **2017**, *117*, 7762–7810.

(125) Higgins, R. F.; Fatur, S. M.; Shepard, S. G.; Stevenson, S. M.; Boston, D. J.; Ferreira, E. M.; Damrauer, N. H.; Rappé, A. K.; Shores, M. P. Uncovering the Roles of Oxygen in Cr(III) Photoredox Catalysis. *J. Am. Chem. Soc.* **2016**, *138*, 5451–5464.

(126) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.;

Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian16 Revision C.01., Gaussian Inc. Wallingford CT, 2016.

(127) Dennington, R.; Keith, T. A.; Millam, J. M. GaussView Version 6., Semichem Inc. Shawnee Mission KS, 2019.

(128) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(129) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph.* **2006**, *25*, 247–260.

(130) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP Charges To Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.

(131) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(132) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. In *Proceedings of the 15th Python in Science Conference*, ed. by Benthall, S.; Rostrup, S., 2016, pp 98–105.

(133) Yadav, V. R.; Prasad, S.; Sung, B.; Aggarwal, B. B. The Role of Chalcones in Suppression of NF-κB-Mediated Inflammation and Cancer. *Int. Immunopharmacol.* **2011**, *11*, 295–309.

(134) Rozmer, Z.; Perjési, P. Naturally occurring chalcones and their biological activities. *Phytochem. Rev.* **2016**, *15*, 87–120.

(135) Hsu, Y. L.; Chia, C. C.; Chen, P. J.; Huang, S. E.; Huang, S. C.; Kuo, P. L. Shallot and licorice constituent isoliquiritigenin arrests cell cycle progression and induces apoptosis through the induction of ATM/p53 and initiation of the mitochondrial system in human cervical carcinoma HeLa cells. *Mol. Nutr. Food Res.* **2009**, *53*, 826–835.

(136) Singh, P.; Anand, A.; Kumar, V. Recent developments in biological activities of chalcones: A mini review. *Eur. J. Med. Chem.* **2014**, *85*, 758–777.

(137) Sahu, N. K.; Balbhadra, S. S.; Choudhary, J.; Kohli, D. V. Exploring pharmacological significance of chalcone scaffold: a review. *Curr. Med. Chem.* **2012**, *19*, 209–225.

(138) Zhou, B.; Xing, C. Diverse Molecular Targets for Chalcones with Varied Bioactivities. *Med. Chem.* **2015**, *5*, 388–404.

(139) Dimmock, J. R.; Elias, D. W.; Beazely, M. A.; Kandepu, N. M. Bioactivities of chalcones. *Curr. Med. Chem.* **1999**, *6*, 1125–1149.

(140) Sharma, V.; Kumar, V.; Kumar, P. Heterocyclic Chalcone Analogues as Potential Anti-cancer Agents. *Anti-Cancer Agents Med. Chem.* **2013**, *13*, 422–432.

(141) Boumendjel, A.; Ronot, X.; Boutonnat, J. Chalcones derivatives acting as cell cycle blockers: potential anti cancer drugs? *Curr. Drug Targets* **2009**, *10*, 363–371.

(142) León-González, A. J.; Acero, N.; Muñoz-Mingarro, D.; Navarro, I.; Martín-Cordero, C. Chalcones as Promising Lead Compounds on Cancer Therapy. *Curr. Med. Chem.* **2015**, *22*, 3407–3425.

(143) Mahapatra, D. K.; Asati, V.; Bharti, S. K. Chalcones and their therapeutic targets for the management of diabetes: Structural and pharmacological perspectives. *Eur. J. Med. Chem.* **2015**, *92*, 839–865.

(144) Mahapatra, D. K.; Bharti, S. K.; Asati, V. Chalcone scaffolds as anti-infective agents: Structural and molecular target perspectives. *Eur. J. Med. Chem.* **2015**, *101*, 496–524.

(145) Das, M.; Manna, K. Chalcone Scaffold in Anticancer Armamentarium: A Molecular Insight. *J. Toxicol.* **2016**, *2016*, 7651047.

(146) Bukhari, S. N. A.; Jasamai, M.; Jantan, I. Synthesis and biological evaluation of chalcone derivatives (mini review). *Mini. Rev. Med. Chem.* **2012**, *12*, 1394–1403.

(147) Kumar, D.; Kumar, M.; Kumar, A.; Singh, S. K. Chalcone and curcumin derivatives: a way ahead for malarial treatment. *Mini. Rev. Med. Chem.* **2013**, *13*, 2116–2133.

(148) Kontogiorgis, C.; Mantzanidou, M.; Hadjipavlou-Litina, D. Chalcones and their Potential Role in Inflammation. *Mini. Rev. Med. Chem.* **2008**, *8*, 1224–1242.

(149) Bukhari, S. N. A.; Franzblau, S. G.; Jantan, I.; Jasamai, M. Current prospects of synthetic curcumin analogs and chalcone derivatives against mycobacterium tuberculosis. *Med. Chem.* **2013**, *9*, 897–903.

(150) Mahapatra, D. K.; Bharti, S. K. Therapeutic potential of chalcones as cardiovascular agents. *Life Sci.* **2016**, *148*, 154–172.

(151) Matos, M. J.; Vazquez-Rodriguez, S.; Uriarte, E.; Santana, L. Potential pharmacological uses of chalcones: a patent review (from June 2011 - 2014). *Expert Opin. Ther. Pat.* **2015**, *25*, 351–366.

(152) Ahmed, K.; Methuku, K. R.; Arutla, V. The design and development of imidazothiazole-chalcone derivatives as potential anticancer drugs. *Expert Opin. Drug Discov.* **2013**, *8*, 289–304.

(153) Mahapatra, D. K.; Bharti, S. K.; Asati, V. Anti-cancer chalcones: Structural and molecular target perspectives. *Eur. J. Med. Chem.* **2015**, *98*, 69–114.

(154) Shenvi, S.; Kumar, K.; Hatti, K. S.; Rijesh, K.; Diwakar, L.; Reddy, G. C. Synthesis, anticancer and antioxidant activities of 2,4,5-trimethoxy chalcones and analogues from asaronaldehyde: Structure–activity relationship. *Eur. J. Med. Chem.* **2013**, *62*, 435–442.

(155) Ghosh, C.; Gupta, N.; Mallick, A.; Santra, M. K.; Basu, S. Self-assembled glycosylated chalcone-boronic acid nanodrug exhibits anticancer activity through mitochondrial impairment. *ACS Appl. Bio. Mater.* **2018**, *1*, 347–355.

(156) Saito, Y.; Mizokami, A.; Izumi, K.; Naito, R.; Goto, M.; Nakagawa-Goto, K. α-Trifluoromethyl Chalcones as Potent Anticancer Agents for Androgen Receptor-Independent Prostate Cancer. *Molecules* **2021**, *26*, 2812.

(157) Djemoui, A.; Naouri, A.; Ouahrani, M. R.; Djemoui, D.; Lahcene, S.; Lahrech, M. B.; Boukenna, L.; Albuquerque, H. M.; Saher, L.; Rocha, D. H.; Monteiro, F. L.; Helguero, L. A.; Bachari, K.; Talhi, O.; Silva, A. M. A step-by-step synthesis of triazole-benzimidazole-chalcone hybrids: Anticancer activity in human cells^{+}. *J. Mol. Struct.* **2020**, *1204*, 127487.

(158) Hseu, Y.-C.; Huang, Y.-C.; Thiyagarajan, V.; Mathew, D. C.; Lin, K.-Y.; Chen, S.-C.; Liu, J.-Y.; Hsu, L.-S.; Li, M.-L.; Yang, H.-L. Anticancer activities of chalcone flavokawain B from Alpinia pricei Hayata in human lung adenocarcinoma (A549) cells via induction of reactive oxygen species-mediated apoptotic and autophagic cell death. *J. Cell. Physiol.* **2019**, *234*, 17514–17526.

(159) Mazzone, G.; Malaj, N.; Galano, A.; Russo, N.; Toscano, M. Antioxidant properties of several coumarin–chalcone hybrids from theoretical insights. *RSC Adv.* **2015**, *5*, 565–575.

(160) Kostopoulou, I.; Tzani, A.; Polyzos, N.-I.; Karadendrou, M.-A.; Kritsi, E.; Pontiki, E.; Liargkova, T.; Hadjipavlou-Litina, D.; Zoumpoulakis, P.; Detsi, A. Exploring the 2'-Hydroxy-Chalcone Framework for the Development of Dual Antioxidant and Soybean Lipoxygenase Inhibitory Agents. *Molecules* **2021**, *26*, 2777.

(161) Serifi, O.; Tsopelas, F.; Kypreou, A.-M.; Ochsenkühn-Petropoulou, M.; Kefalas, P.; Detsi, A. Antioxidant behaviour of 2'-hydroxy-chalcones: a study of their electrochemical properties. *J. Phys. Org. Chem.* **2013**, *26*, 226–231.

(162) Enoki, T.; Ohnogi, H.; Nagamine, K.; Kudo, Y.; Sugiyama, K.; Tanabe, M.; Kobayashi, E.; Sagawa, H.; Kato, I. Antidiabetic Activities of Chalcones Isolated from a Japanese Herb, Angelica keiskei. *J. Agric. Food Chem.* **2007**, *55*, 6013–6017.

(163) Rammohan, A.; Bhaskar, B. V.; Venkateswarlu, N.; Gu, W.; Zyryanov, G. V. Design, synthesis, docking and biological evaluation of chalcones as promising antidiabetic agents. *Bioorg. Chem.* **2020**, *95*, 103527.

(164) Kaur, M.; Kaushal, R. Synthesis, characterization and α-amylase and α-glucosidase inhibition studies of novel vanadyl chalcone complexes. *Appl. Organomet. Chem.* **2021**, *35*, e6042.

(165) Vinindwa, B.; Dziwornu, G. A.; Masamba, W. Synthesis and Evaluation of Chalcone-Quinoline Based Molecular Hybrids as Potential Anti-Malarial Agents. *Molecules* **2021**, *26*, 4093.

(166) Hameed, A.; Masood, S.; Hameed, A.; Ahmed, E.; Sharif, A.; Abdullah, M. I. Antimalarial, cytotoxicity and molecular docking studies of quinolinyl chalcones as potential anti-malarial agent. *J. Comput. Aided Mol. Des.* **2019**, *33*, 677–688.

(167) Tomar, V.; Bhattacharjee, G.; Rajakumar, S.; Srivastava, K.; Puri, S. K. Synthesis of new chalcone derivatives containing acridinyl moiety with potential antimalarial activity. *Eur. J. Med. Chem.* **2009**, *98*.

(168) Bakarić, D.; Baranović, G. The conformational equilibrium and vibrational properties of chalcone. *J. Mol. Struct.* **2019**, *1196*, 429–438.

(169) Blanco, S. E.; Ferretti, F. H. Determination of absorptivity and formation constant of a chalcone association complex. *Talanta* **1998**, *45*, 1103–1109.

(170) Krause, H; Ernstberger, B; Neusser, H. J. Binding energies of small benzene clusters. *Chem. Phys. Lett.* **1991**, *184*, 411–417.

(171) Grover, J. R.; Walters, E. A.; Hui, E. T. Dissociation energies of the benzene dimer and dimer cation. *J. Phys. Chem.* **1987**, *91*, 3233–3237.

(172) Rappé, A. K.; Bernstein, E. R. Ab Initio Calculation of Nonbonded Interactions: Are We There Yet? *J. Phys. Chem. A* **2000**, *104*, 6117–6128.

(173) Ochterski, J. W., *Vibrational Analysis in Gaussian*; Gaussian Inc.: 1999, pp 1–10.

Appendix A

Supporting Information

A.1 Chapter 2 SI: The Initial Aggregation and Ordering Mechanism of Diphenylalanine From Microsecond All-Atom Molecular Dynamics Simulations

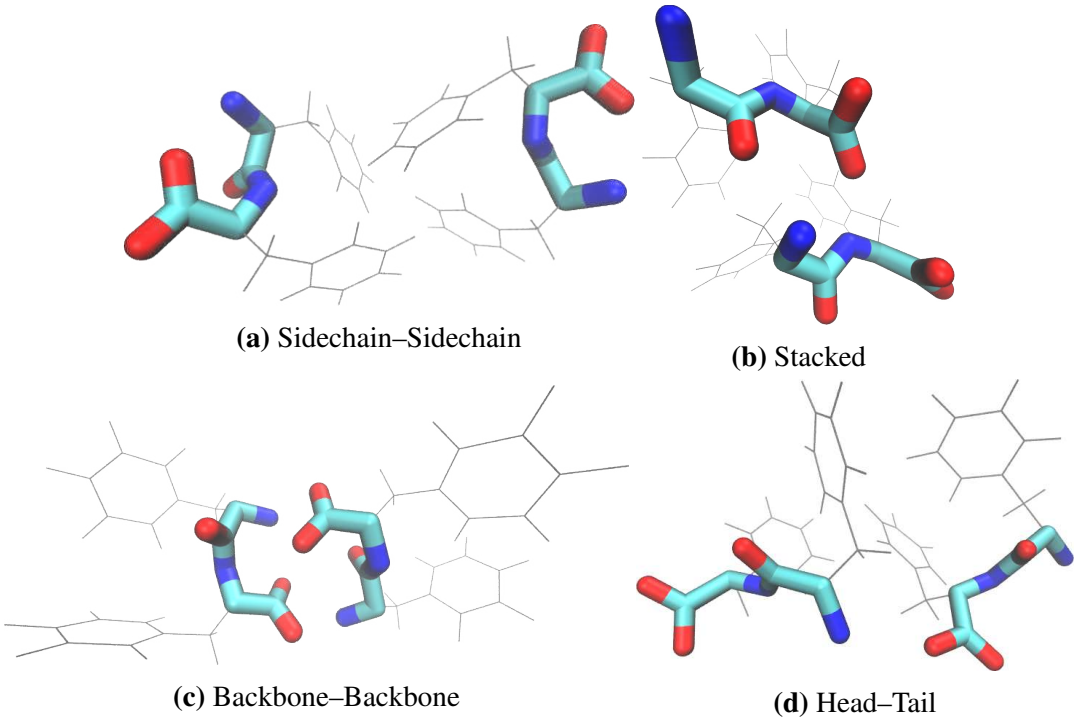
A.1.1 Force Field Verification

Table A1: Binding energies in kcal/mol for FF dimers (see Figure A1) computed in a dielectric of one for commonly used force fields and compared to *ab initio* calculations with a basis set superposition correction.

Method	sidechain–sidechain	stacked	backbone–backbone	head–tail	MAPE
ωB97xD/	-7.22	-55.53	-91.74	-46.22	NA
6-311+G* ff14SB	12.20	-50.47	-99.64	-30.23	80.32
ff99SB	12.20	-50.47	-99.64	-30.23	80.32
ff14ipq	-7.18	-46.23	-95.42	-49.82	7.28
ff15ipq	-3.91	-43.95	-84.96	-38.49	22.68
CHARMM27	-6.72	-69.58	-113.77	-135.92	62.56
CHARMM36	-6.76	-70.02	-113.99	-136.44	62.98
OPLS	-6.76	-47.35	-58.06	-29.92	23.28
GROMOS53a6	-5.98	-36.58	-23.60	-18.50	46.38

A.1.2 PMF Comparison

The monomer is pulled perpendicular to the surface of the nanotube along the z-axis via anchor dummy atoms 900 Å away in the $\pm x$ and $\pm y$ directions with a RMSD restraint of 100 kcal/mol/Å^2. A harmonic force constant between the monomer of interest and each anchor dummy atom of 20 kcal/mol/Å^2 is used. These anchor dummy atoms are necessary to pull the monomer of interest away from the surface in a specific direction, in this case the z-dimension. Within the topology file, the dummy atoms are defined to have zero Lennard-Jones and Coulombic forces and therefore do



(a) Sidechain–Sidechain

(b) Stacked

(c) Backbone–Backbone

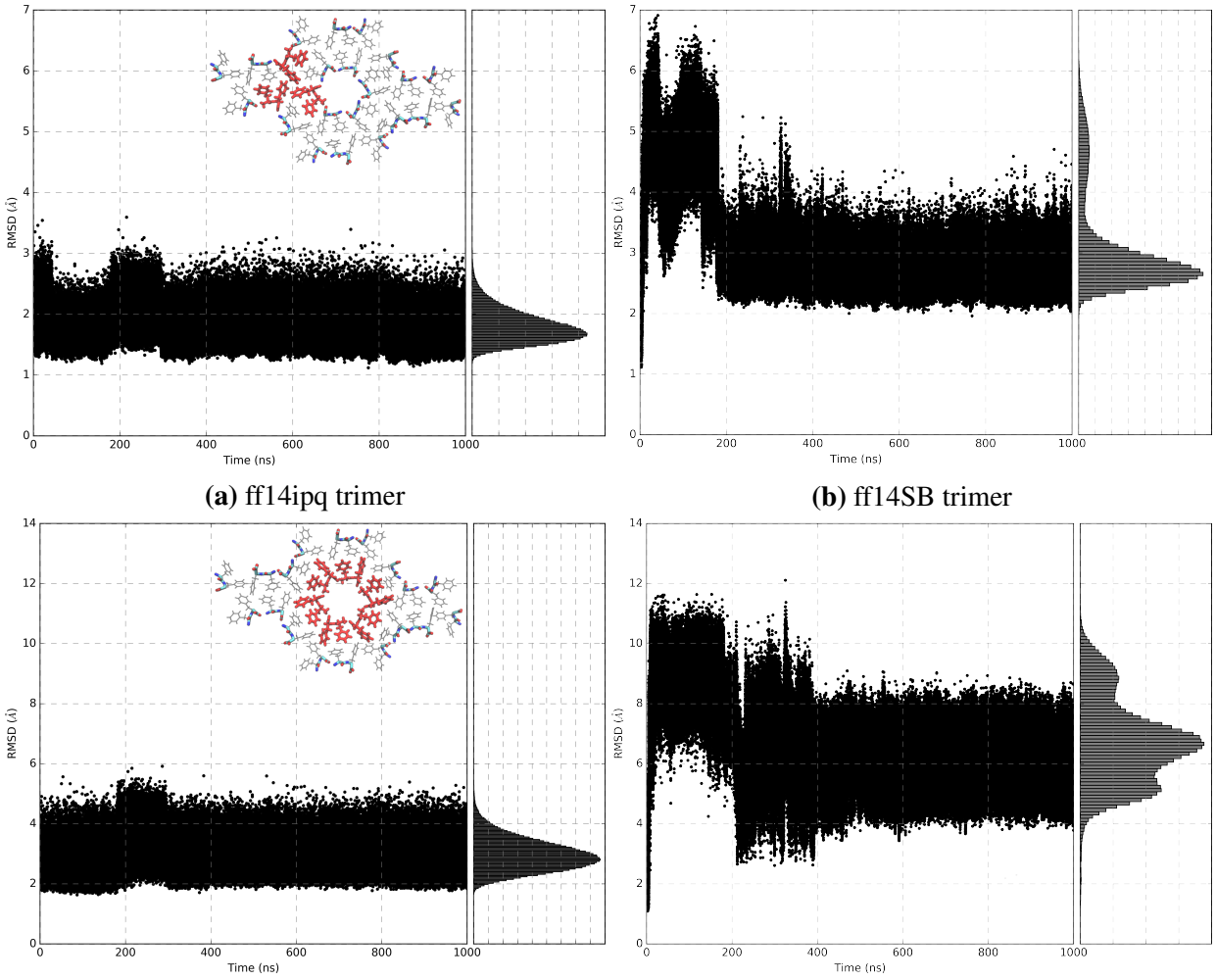
(d) Head–Tail

Figure A1: Interaction energies were computed for dimers in the above four configurations with the sidechain–sidechain, stacked, and head–tail dimer configurations seen in the crystal structure and the backbone–backbone interactions seen in unbiased aaMD.

not impact the force decomposition analyzed. The impact of dummy atom involved restraints does result in a non-zero force at large distances when analyzed using EMUS. This additional energy is removed through the following equation.

$$E = E_{EMUS} - (k[(900^2 + x^2)^{1/2} - 900]^2/6) \tag{A.1}$$

where k is the monomer–dummy atom force constant, E_{EMUS} is the uncorrected energy, and x is the distance between the two. An additional 30 ns of umbrella sampling was performed at distances of 5.25, 5.5, and 6.25 Å for the monomeric addition to the FF nanotube to have well sampled overlap between windows. A larger force constant of 40 kcal/mol/Å^2 was applied to windows centered at distances of 5.0, 6.0, and 6.5 Å for the monomeric addition to the FF nanotube.



(a) ff14ipq trimer

(b) ff14SB trimer

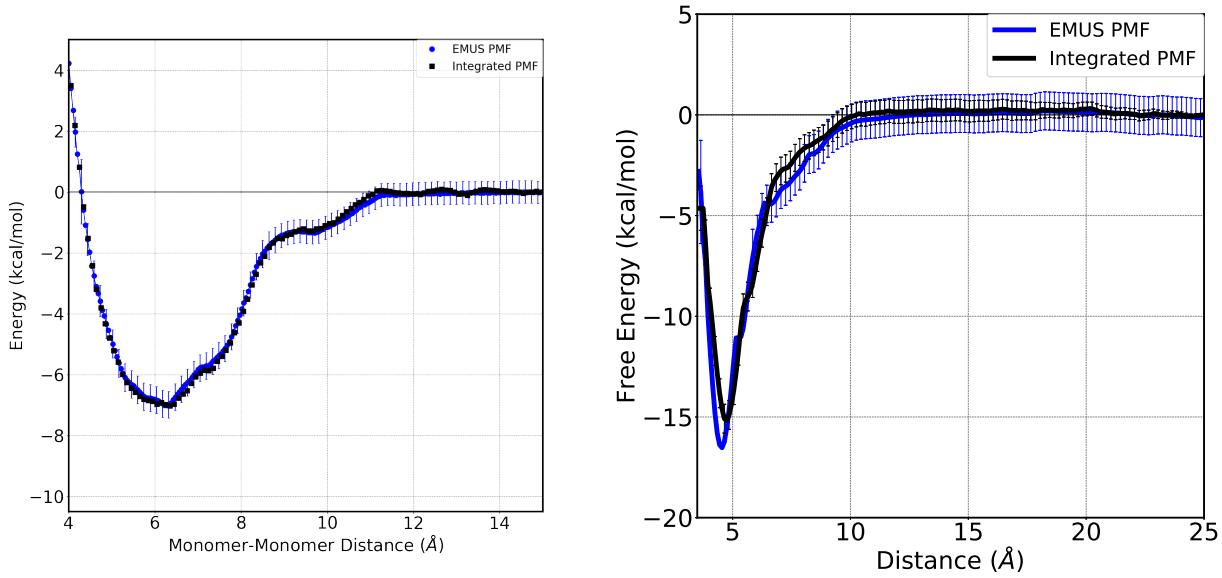
(c) ff14ipq hexamer

(d) ff14SB hexamer

Figure A2: The RMSD of a trimer selection highlighted in red within the nanotube (a) over 1 μs of simulation using the (a) ff14ipq and (b) the ff14SB forcefields. The RMSD of a hexamer selection highlighted in red within the nanotube (c) over 1 μs of simulation using the (c) ff14ipq and (d) the ff14SB forcefields.

A.1.3 Interacting Pair Type Decision Tree Analysis

Classification of pairs becomes more inclusive as pair interaction types are classified. Cutoffs (Table S2) include NT (center-of-mass of the N-terminus), NH (center-of-mass of N and H in amide bond), CO (center-of-mass of the carbonyl group in amide bond), and CT (center-of-mass of the C-terminus) with the order of functional group denoting relative molecule. Interacting pairs are determined based on the combinations of applied cutoffs, shown in Table S2, where 0 means the interacting pair has a functional group distance outside the cutoff, 1 means the distance is less



(a) FF Dimerization

(b) Monomeric Addition to FF Nanotube

Figure A3: The PMF of (a) dimerization and (b) monomeric addition to the FF nanotube, computed by integrating the mean force (black) and generated from EMUS (blue) which are shown here to be within error of each other.

than the cutoff, and NM is denoted as "not measured" for the respective pair of molecules. While redundant in some interacting pair types, each pair is reported twice as there are pairs requiring the degenerate cutoffs (i.e. CT–NH and NH–CT). Percentages for individual stable aggregate systems is shown in Table S3 and the time evolution for these systems is seen in Figure S6. A pair of molecules is classified as sidechain interacting if it is not a backbone interacting pair and had a intermolecular phenyl ring center-of-mass to phenyl ring center-of-mass distance less than 5.5 Å if both phenyl rings were on the N-terminal residue or 5.8 Å otherwise.

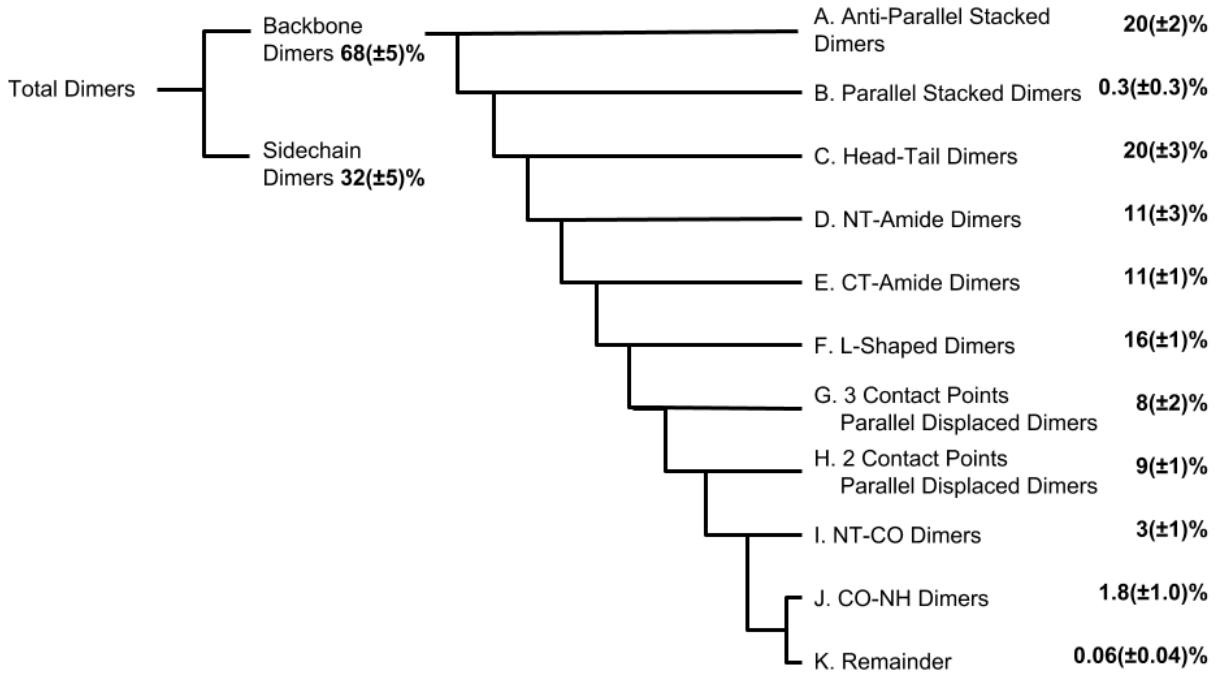


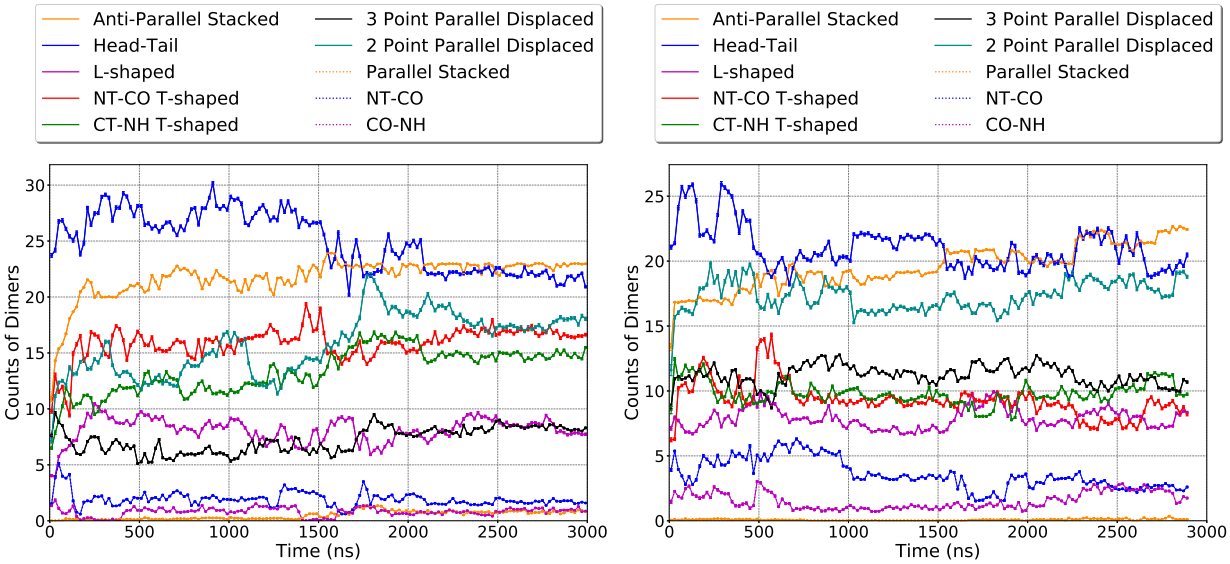
Figure A4: A scheme of the decision tree analysis for classifying backbone interacting pair types performed on initially dispersed systems with the relative percentages, averaged over initially dispersed systems, on the right.

Table A2: The percentages of stable aggregates of initially dispersed systems.

	60 FF Replica 1	60 FF Replica 2	60 FF Replica 3	120 FF	Mean
Anti-Parallel Stacked	16.6	20.1	21.7	19.9	19.6
Parallel Stacked	0.3	0.7	0.1	0.1	0.3
Head-Tail	18.5	19.1	19.0	24.8	20.3
NT-Amide	10.2	14.6	8.7	9.4	10.7
CT-Amide	12.4	13.0	10.5	10.1	11.5
L-Shaped	17.2	15.6	17.7	14.9	16.3
3 Contact Parallel Displaced	10.7	7.2	7.4	7.1	8.1
2 Contact Parallel Displaced	8.0	7.3	10.2	8.0	8.4
NT-CO	2.9	1.5	2.6	4.5	2.9
CO-NH	3.1	0.8	2.1	4.5	1.8
Remainder	0.02	0.1	0.04	0.08	0.06

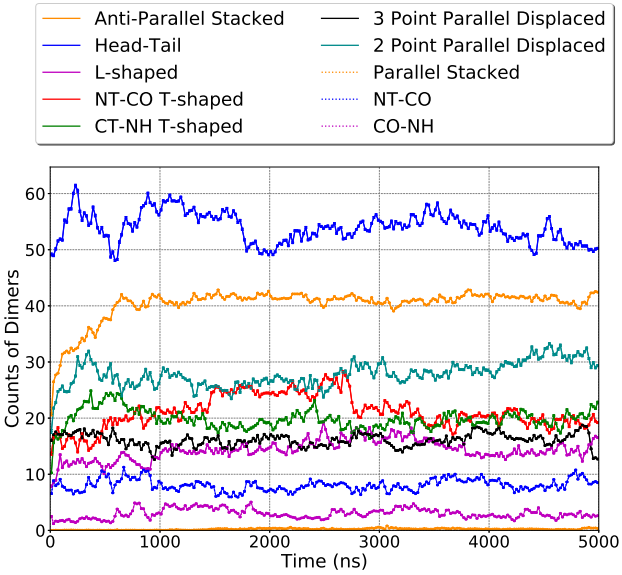
Table A3: The cutoffs and the combinations of cutoffs used to classify backbone interacting pairs.

	NT–NT	NT–CO	NT–NH	NT–CT	CO–NT	CO–NH or NH–CO	NH–NT	NH–CT	CT–NT	CT–NH	CT–CT
Cutoffs (Å)	3.8	4.6	5.0	4.3	4.6	5.8	5.0	4.2	4.3	4.2	4.8
Anti-Parallel Stacked	NM	NM	NM	1	NM	NM	NM	NM	1	NM	NM
Parallel Stacked	1	NM	NM	NM	NM	NM	NM	NM	NM	NM	1
Head-Tail	0	0	0	1	NM	NM	NM	0	NM	0	NM
NT-Amide	NM	1	NM	1	NM	NM	NM	0	NM	0	NM
CT-Amide	0	NM	NM	NM	0	NM	1	NM	1	1	NM
L-Shaped	NM	0	1	1	NM	NM	NM	NM	NM	0	NM
3 Contact Parallel Displaced	NM	NM	NM	NM	1	NM	NM	NM	1	1	NM
2 Contact Parallel Displaced	NM	NM	NM	1	NM	NM	NM	1	NM	NM	NM
NT-CO	NM	1	NM	NM	NM	NM	NM	NM	NM	NM	NM
CT-NH	NM	NM	NM	NM	NM	1	NM	NM	NM	NM	NM



(a) 60 FF rep2

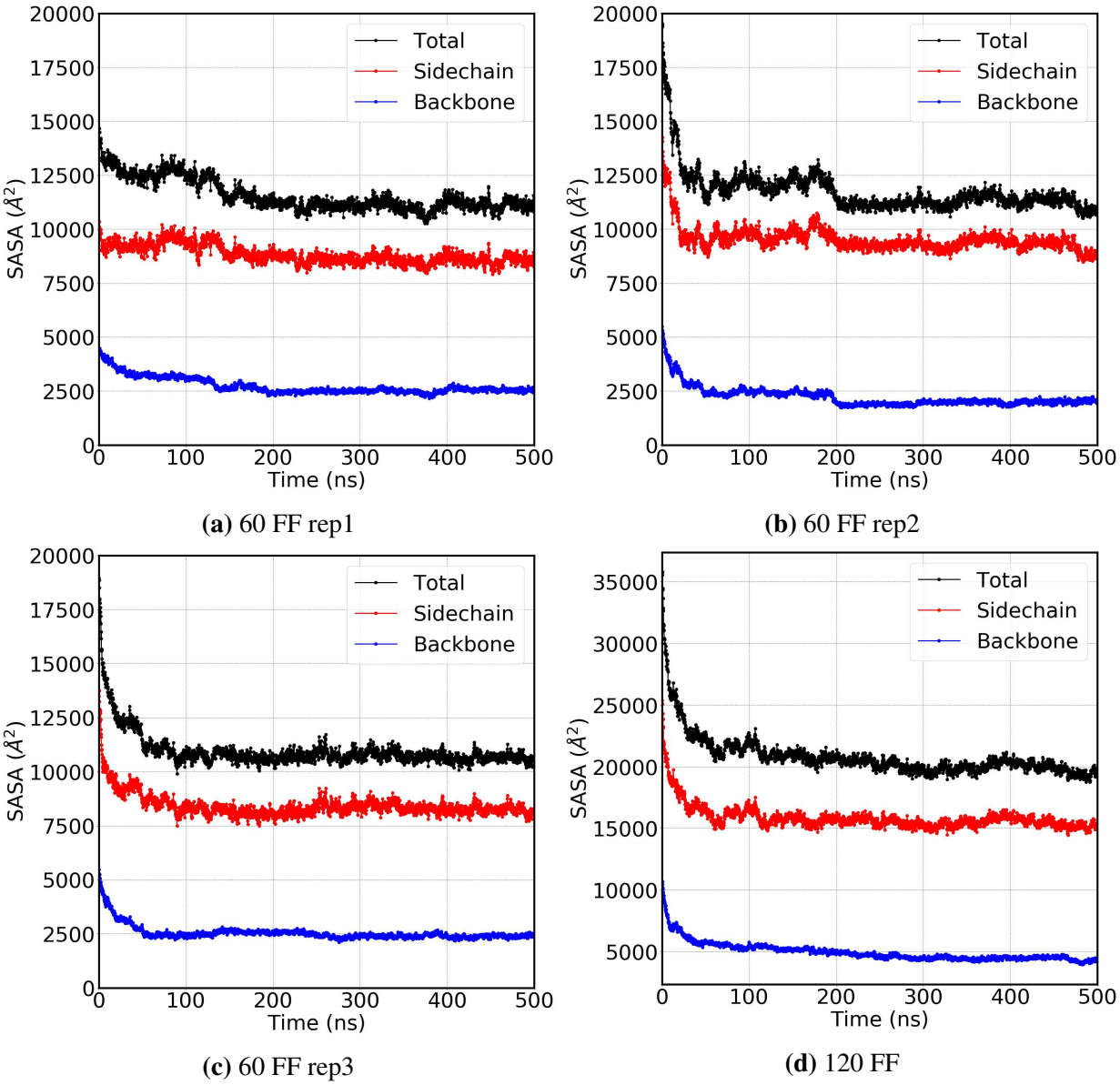
(b) 60 FF rep3



(c) 120 FF

Figure A5: The time evolutions of backbone interacting pairs in the initially dispersed systems with one replica reported in the main document.

A.1.4 SASA



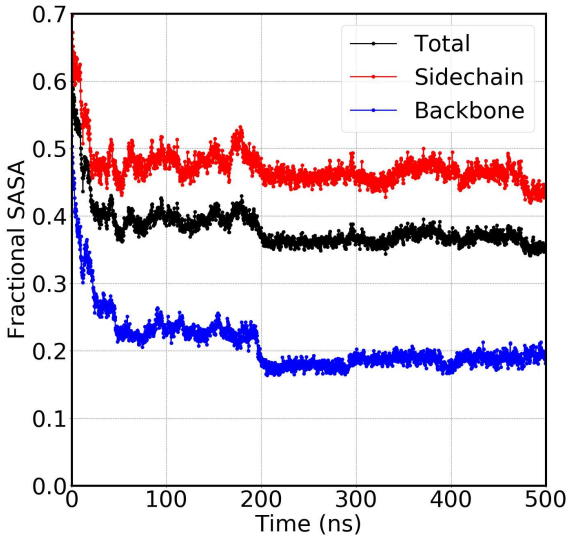
(a) 60 FF rep1

(b) 60 FF rep2

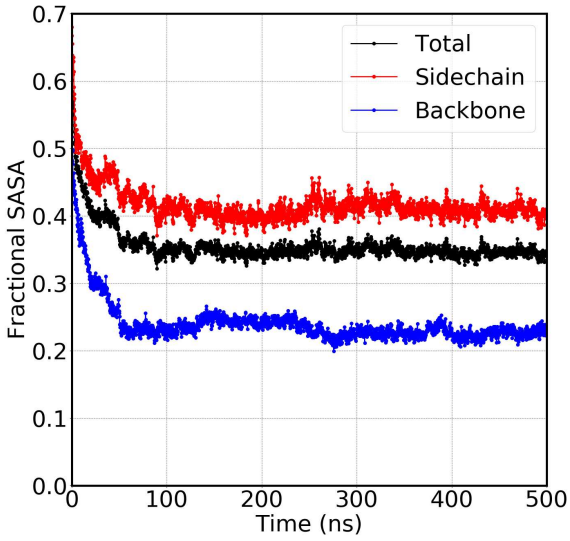
(c) 60 FF rep3

(d) 120 FF

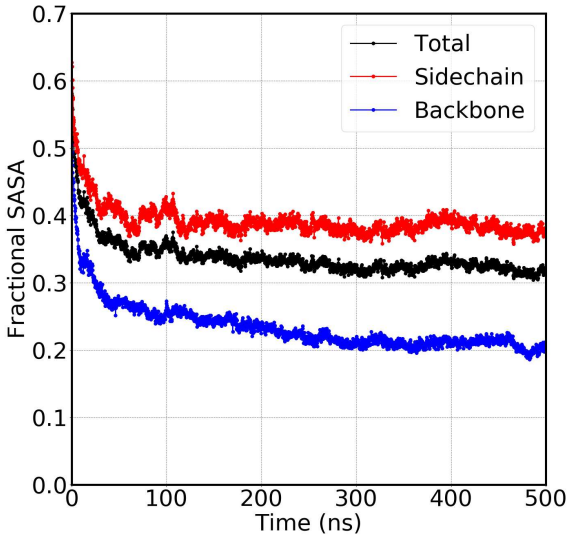
Figure A6: The SASA time evolution of the total SASA (black), sidechain in the presence of backbone SASA (red), and backbone in the presence of the sidechain (blue) SASA for all initially dispersed systems. Quantitatively the sum of the blue and red values equal the black curve.



(a) 60 FF rep2



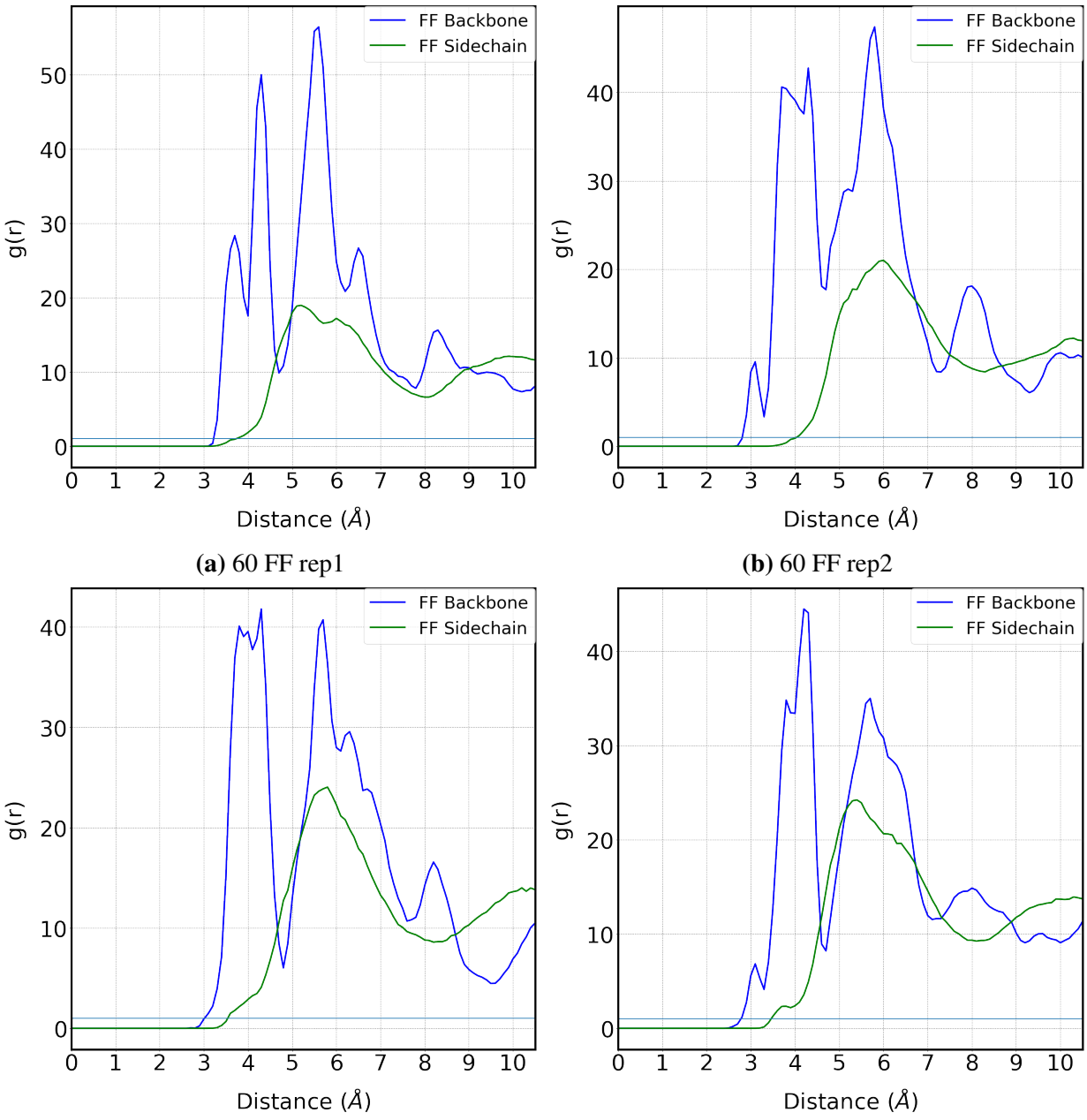
(b) 60 FF rep3



(c) 120 FF

Figure A7: The fractional SASA time evolution, described in the main document, for the initially dispersed systems.

A.1.5 $g(r)$'s



(a) 60 FF rep1

(b) 60 FF rep2

(c) 60 FF rep3

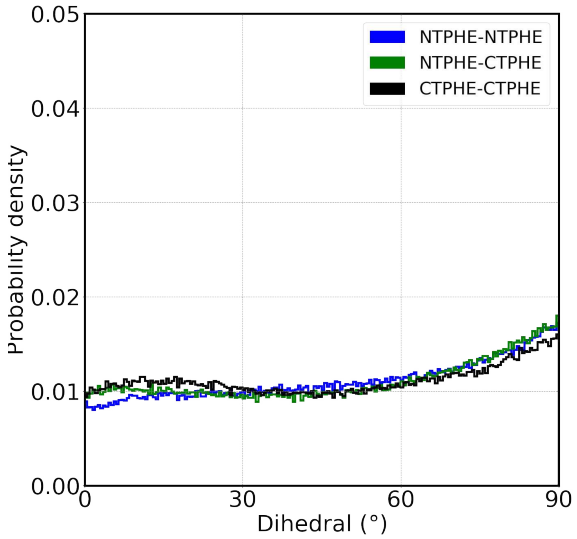
(d) 120 FF

Figure A8: The individual backbone (blue) and sidechain (green) $g(r)$'s for all stable aggregates.

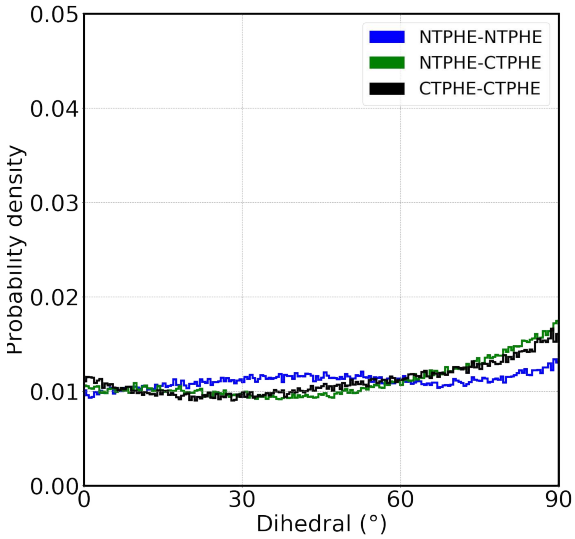
A.1.6 Dihedral Probability Densities

A.1.7 Phenyl Ring Orientation

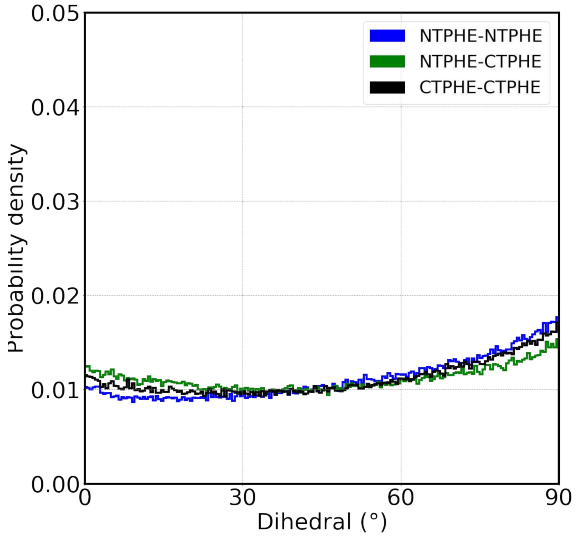
The dihedral between two phenyl rings of different molecules was measured to determine the relative amount of π–π stacking between the initially dispersed systems and the crystal structure. The plane of a phenyl ring is defined here by the plane formed by the C_{ζ}-$C_{\epsilon 1}$-$C_{\delta 1}$ atoms for each phenyl group. The dihedral is calculated as the angle formed between the vectors orthogonal to the planes of the two phenyl rings of different FF molecules that contained any atoms within 6 Å. For more details on this analysis, see the work done by McGaughey *et al.*[98] The dihedral was then mapped to the first quadrant since the dihedral for this system is functionally symmetric about 90°.



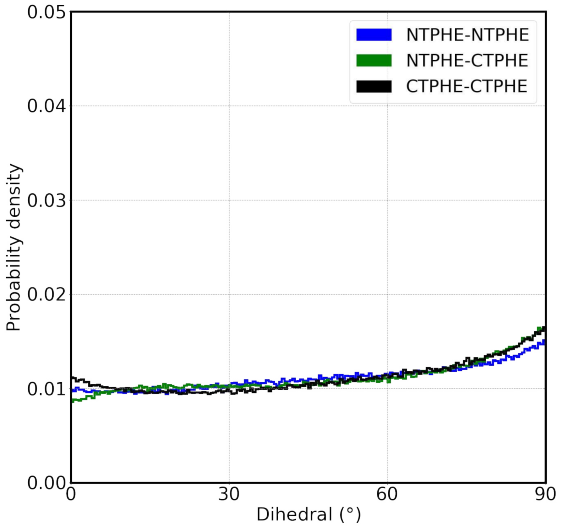
(a) 60 FF rep1



(b) 60 FF rep2



(c) 60 FF rep3



(d) 120 FF

Figure A9: The phenyl ring dihedral dihedral for individual stable aggregate systems.

A.2 Chapter 5 SI: Computing the Entropy of Association with the Rigid-Rotor Gaussian Approximation

A.2.1 Geometry Optimized Energies

Table A4: All energies determined using the ωB97xD functional in conjunction with the 6-311G* basis set. Barrier heights are used in the python script analysis and is the energy of association for all dimer systems the torsional rotation energy for ethane. A positive binding energy is reflective of a stable bond formation. The chalcone dimer in cyclohexane binding energy reported is BSSE corrected.

System	Total E (Hartrees)	Barrier Height/Binding Energy (kcal/mol)
Ethane staggered	-79.817170	
Ethane eclipsed	-79.812868	2.699570
Methane monomer	-40.50971	
Methane dimer at equilibrium	-81.020726	0.819519
Methane dimer at equilibrium with BSSE		0.63
Methane dimer at 4.5 Å	-81.020075	0.411014
Methane dimer at 6.5 Å	-81.019493	0.045808
Methane dimer at 8.5 Å	-81.019433	0.008158
Benzene monomer	-232.210998	
Benzene dimer T-shaped	-464.427933	3.725336
Benzene dimer T-shaped with BSSE		3.25
Benzene dimer parallel displaced	-464.428605	4.1471391
Benzene dimer parallel displaced with BSSE		3.36
Chalcone monomer *in vacuo*	-653.946815	
Chalcone dimer *in vacuo*	-1307.919231	16.06523141
Chalcone dimer with BSSE *in vacuo*		13.72
Chalcone monomer PME cyclohexane	653.9508075	
Chalcone dimer PME cyclohexane	1307.925626	12.15700465

A.2.2 RRGO entropy.py

```
""" @author: Jakob Anderson and Anthony Rapp$\'{e}$, parts
being modified from a script written by Jacob Nite\\
```

```
%usage: python2.7 entropy.py -i fchk your_file.fchk -t
298.15  -p 1 -trs 1 -f1rs 1 -f2rs 1 > your_file_entropy.dat
"""
import argparse
import numpy as np
from numpy.linalg import eig
import math
from scipy import linalg
from scipy import integrate
from fchk_ext import MoleculeData, TYPE_ACCEPT
import sys
import matplotlib.pyplot as plt


consts = {'pi': 3.141592653589793, 'toAng': 0.5291772085899999,
'avog': 6.0221417899999999E+023, 'sLight': 29979245800.00000,
'hartree': 4.3597439400000000E-018, 'elPerN': 1822.888485005929,
'scaleFC': 1e6, 'plancks': 6.62606957e-34, 'boltzmann':
1.3806488E-23, 'amu_to_kg': 1.66053886E-27, 'bohr_to_m':
5.29177249E-11, 'plancks_amubohr': 1.4249883E+14,
'boltzmann_amubohr': 2.9691744E+24, 'atm_to_Pa': 101325.0}
consts['scale'] = consts['scaleFC'] / consts['elPerN']
consts['factor'] = (10 ** 23 / consts['scale']) *
(consts['avog'] * consts['hartree'] / (4. * (np.pi *
consts['sLight'] * consts['toAng']) ** 2))
T = float(sys.argv[5]) # K
P = float(sys.argv[7]) # ATM
```

```
tot_rot_sym = float(sys.argv[9])

def num_potential(ng,a,b,v):
    # a and b are the basis functions over the distance grid
    # v is the potential over the grid
    sum=0.0
    for i in range(ng):
        sum=sum+a[i]*b[i]*v[i]
    return sum

def a_grid(ng,a,x,zeta,xi):
    for n in range(ng):
        a[n]=math.exp(-zeta*(x[n]-xi)**2)
    return a



# constants
#P. J. Mohr, B. N. Taylor, and D. B. Newell, "CODATA
Recommended Values of the Fundamental Physical Constants:
2010," Rev. Mod. Phys., 84 (2012) 1527-1605. DOI: 10.1103
/RevModPhys.84.1527
#P. J. Mohr, B. N. Taylor, and D. B. Newell, "CODATA
RecommendedValues of the Fundamental Physical Constants:
2010," Chem. Ref. Data, 41 (2012) 043109. DOI: 10.1063
/1.4724320

hbar = 1.0545718E-34 # J*s
```

```
hbar_Eh = hbar * 2.2937124810E17 # from J*s to Eh*s
avog_num = 6.0221417899999999E+023
##########routines##############

def Stat_Thermo(E,T,De,vib_en,hbar):
    #E needs to be in kcal/mol
    #De is the barrier, in kcal/mol
    vib_en*=6.57966E15 #from hartrees to 1/s
    R = 0.0019872 #kcal/K/mol
    n = E.size
    exp_e = np.zeros(n,dtype=float)
    nEleU = 0 # number of energy values the En is less
than or equal to the U model
    for i in range(n):
        if E[i] < De:
            nEleU += 1
    #print "number of bound states: ",nEleU
    #computing Q
    Qt=0.0
    ezero=E[0]
    if n == nEleU:
        nEleU = nEleU-2
    for i in range(nEleU):
        exp_e[i] = math.exp(-(E[i]-ezero)/(R*T))
    Qt = np.sum(exp_e)
    #Bfact = math.exp(-De/(R*T))
    barrier=E[nEleU]-ezero
```

```
    Bfact = math.exp(-(barrier)/(R*T))
    rotQ = (2.0*np.pi**2*consts['boltzmann']*T/hbar/
vib_en)**0.5
    #print "Q,Bfact*rotQ,vib_en=", Q,Bfact*rotQ,vib_en
    # add above the barrier terms
    Q=Qt+Bfact*rotQ
    print "Q=",Q
    # energy terms
    # dQ.dT
    dqdt = 0.0
    Erot=0.5*R*T
    for j in range(nEleU):
        dqdt = dqdt+(E[j]-ezero)*exp_e[j]
    dqdt=dqdt+Bfact*(Erot + barrier*rotQ)
    dqdt = dqdt/(R*T*T)
    # dlnQ/dT
    dlnqdt = dqdt/Q
    # U and S
    U = R*T*T*dlnqdt
    Se = R*T*dlnqdt
    Sq = R*math.log(Q)
    Sg = Se + Sq
    # add above the barrier terms
    # Harmonic entropy
    edelta=E[1]-E[0]
    #print edelta*349.75
    eu = math.exp(-edelta/(R*T))
```

```
    qho=1.0-eu
    dlnq=(edelta/(R*T))*eu/qho
    Sh = R*(dlnq-math.log(qho))
    #print "Harmonic S=",St
    print "dlnQ/dT: ", dlnqdt
    print "U= ", U," kcal*K/mol "
    print "Sg,Sh,Sq,Se=", Sg,Sh,Sq,Se," kcal/(mol K)"
    return U, Sg

def gauss(x,De,alpha):
    return De*(1-np.exp(-alpha*x*x))
#####The basis_V2 routine was adapted from https://
github.com/mccullaghlab/CHEM571A-FALL2018/blob/master
/notes/Morse%20Oscillator.ipynb #####
def basis_V2(Nzeta,alpha,mu,V,Xg,dgrid,Xgaus):
    #N = 3                # half the number of basis
functions
    Ngaus=Xgaus.size
    Ngrid=Xg.size

    #Nzeta: Exponents
    K=Ngaus*Nzeta
    #K = 2*N+1      # total number of basis functions #keeps K
above zero, which keeps i and j greater than zero
    #dx = 0.75                # spacing between basis
functions
    #zstart=1.0
```

```
    #dzeta = 0.5                # 1/spread of basis
functions

    S = np.zeros((K,K),dtype=float)#basis function overlap
matrix
    H = np.zeros((K,K),dtype=float)#Hamiltonian matrix,
Hij = <gi|H|gj>
    Zeta = np.zeros(K,dtype=float)
    Xp = np.zeros(K,dtype=float)
    Norm = np.zeros(K,dtype=float)

    beta=0.75
    alpha0=alpha/(beta**((Nzeta+1)/2))

    dzeta=0.5*alpha

    a = np.zeros(Ngrid,dtype=float)
    b = np.zeros(Ngrid,dtype=float)
    da = np.zeros(Ngrid,dtype=float)
    db = np.zeros(Ngrid,dtype=float)

    #generate the gaussians
    i=0

    for n in range(Ngaus):
        zstep=alpha
```

```
        for m in range(Nzeta):
            Zeta[i]=alpha0*beta**(m+1)
            Xp[i]=Xgaus[n]
            aa=a_grid(Ngrid,a,Xg,Zeta[i],Xp[i])
            Norm[i]=1.0/(math.sqrt(np.dot(aa,aa)*dgrid))
            i=i+1

    # compute S, and Hamiltonian, H, matrices
    for i in range(K):
        xi = Xp[i]
        zetai = Zeta[i]
        # evaluate the basis on the grid
        for n in range(Ngrid):
            dxi=Xg[n]-xi
            a[n]=Norm[i]*math.exp(-zetai*(dxi*dxi))
            da[n]=-2.0*zetai*a[n]*dxi

        for j in range(K):
            xj = Xp[j]
            zetaj = Zeta[j]
            for m in range(Ngrid):
                dxj=Xg[m]-xj
                b[m]=Norm[j]*math.exp(-zetaj*(dxj*dxj))
                db[m]=-2.0*zetaj*b[m]*dxj
            snum=np.dot(a,b)*dgrid
            tnum=0.5*np.dot(da,db)*dgrid/mu #Note: /mu is in
the kinetic energy # da and db are the derivatives of the
```

```
basis functions over the distance grid
            vnum=num_potential(Ngrid,a,b,V)*dgrid
            # basis function value:
            S[i,j]=snum
            H[i,j]=tnum+vnum
    Ssvd = S
    Hsvd = H
    # finalize the S^-1*H*S matrix
    SinvH = np.dot(np.linalg.inv(S),H)
    # compute eigenvalues and eigenvectors
    H_eig_val, H_eig_vec = np.linalg.eig(SinvH)
    #alternative diagonalization procedure
    # svd: single value decomposition
    # print 'Two step Eigenvalue Solver with Reduced Space'
    svd_u, svd_s, svd_v = np.linalg.svd(Ssvd)
    eig_zero = []
    for i in range(svd_s.size):  # Check for zero eigenvalues
        if svd_s[i] < 1e-5: # changed from -2 to -5
            eig_zero.append(i)
    svd_u = np.delete(svd_u, eig_zero, axis=1)
    svd_v = np.delete(svd_v, eig_zero, axis=0)
    svd_s = np.delete(svd_s, eig_zero)
    nsvd = svd_s.size
    for i in range(svd_s.size):
        svd_u[:, i] = svd_u[:, i] / np.sqrt(svd_s[i])
        svd_v[i, :] = svd_v[i, :] / np.sqrt(svd_s[i])
```

```
    h2 = svd_v.dot(Hsvd).dot(svd_u)

    svd_eig_val, svd_eig_vect = np.linalg.eigh(h2)
    svd_eig_vect = svd_u.dot(svd_eig_vect)
    svd_psi = np.zeros((Ngrid,K),dtype=float)
    svd_psiNorm = np.empty(Xg.size,dtype=float)
    # generate psis from coefficients
    for A in range(nsvd):
        count = K-A-1
        psiNorm = 0.0
        for i in range(K):
            xi = Xp[i]
            normi=Norm[i]
            svd_psi[:,A] = svd_psi[:,A] + svd_eig_vect[i,A]*
normi*np.exp(-Zeta[i]*(Xg-xi)**2)
            svd_psiNorm = svd_psiNorm + svd_eig_vect[i,A]*
normi*np.exp(-Zeta[i]*(Xg-xi)**2)

    # normalize the wave functions
    normt=np.dot(svd_psi[:,A],svd_psi[:,A])
    svd_psi[:,A] /= np.sqrt(normt)

    # reorder these so largest eigenvalue is first
    idx = H_eig_val.argsort()
    H_eig_val = H_eig_val[idx]
    H_eig_vec = H_eig_vec[:,idx]
    psi = np.zeros((Ngrid,K),dtype=float)
```

```
    psiNorm = np.empty(Xg.size,dtype=float)
    # generate psis from coefficients
    for A in range(K):
        count = K-A-1
        psiNorm = 0.0
        for i in range(K):
            xi = Xp[i]
            psi[:,A] = psi[:,A] + H_eig_vec[i,A]*
np.exp(-Zeta[i]*(Xg-xi)**2)
            psiNorm = psiNorm + H_eig_vec[i,A]*
np.exp(-Zeta[i]*(Xg-xi)**2)

        # normalize the wave functions
        psi2 = np.power(psiNorm,2)
        norm = float(integrate.simps(psi2,Xg))
        psi[:,A] /= np.sqrt(norm)
        # return normalized wave functions and energies
    return svd_psi, svd_eig_val, nsvd

def extract(obj, full_diag):
    """Determines the translational and rotational vectors of
    two fragments, both individually and together
    """
    """Weights the Hessian by the atomic masses of each atom
    Uses the formula Hmw(i,j) = H(i,j) / sqrt(M(i) * M(j))
    """
    R = consts['boltzmann']*consts['avog'] # J/K/mol
```

```
    cart_hessian = obj.cart_hessian
    fragment_index = obj.fragment_index
    i_mass = obj.atom_mass
    mass_array = np.empty(cart_hessian.shape)
    masses_expanded = np.empty(cart_hessian.shape)

    for i in range(cart_hessian.shape[0] / 3):
            for j in range(cart_hessian.shape[0] / 3):
                mass_array[i * 3:i * 3 + 3, j * 3:j * 3 + 3
] = 1 / np.sqrt(i_mass[i] * i_mass[j])
                masses_expanded[i * 3:i * 3 + 3, j * 3:j *
3 + 3] = i_mass[i]

    mw_cart_hessian = cart_hessian * mass_array
    e1, e2 = linalg.eigh(mw_cart_hessian) #needed to determine
the force constant in the correct unit

    """Scale the mass weighted cartesian Hessian by a given
    scale factor (s_factor)
    """
    mw_cart_hessian = mw_cart_hessian * consts['scale']

    """Do a quick vibrational energy value calculation without
    altering hessian
       Need to first diagonalize the vibrational hessian
    """
    eigen_val, eig_vector = linalg.eigh(mw_cart_hessian)
```

```
    """follow procedure given in Vibrational Analysis in
Gaussian by Joseph W. Ochterski
    """
    lmwc2 = eig_vector**2
    summed = np.sum((lmwc2/masses_expanded),axis=0)
    RM = summed**(-1)
    np.savetxt("reduced_masses.txt", RM)
    np.savetxt("eigen_val.txt", eigen_val)
    """ mw_cart_hessian prior to scale is in units of hartree/
bohr^2 and thus the eigenvalue is in units of hartree/
(bohr^2*amu)
need to convert to mdyne/(ang^2*amu) then multiply by
reduced mass (10^-18 because /10^-20(ang^2)/10^2(J/m^2
to mdyne/ang))
    """
    e1 = e1/consts['toAng']**2/10**-18*consts['hartree']
    FC = e1 * RM
    np.savetxt("force_constants.txt", FC)

    """ Scale the vibrational eigenvalues to the units cm-1
    """
    eigen_val = eigen_val * consts['factor']
    vib_energies = np.sqrt(np.abs(eigen_val)) *
np.sign(eigen_val)
    """ Print out the vibrational energies from the unaltered
mass weighted cartesian hessian
```

```
    """
    vib_energies = np.asarray(vib_energies)


    """Compute trans. and rot. vectors by:
            compute center of mass
            compute moment of inertia (m.o.i)
            diag. m.o.i.
            generate d matrix
    """

    n_trans_rot = 6
    n_vib = mw_cart_hessian.shape[0] - n_trans_rot

    """ Calculates the center of mass for the molecule
    """
    atom_mass=obj.atom_mass
    total_mass = atom_mass.sum()
    mass_array = np.tile(atom_mass, (3, 1)).transpose()
    cart_coords = obj.cart_coords
    weighted_cart = cart_coords * mass_array /
total_mass
    center_mass = weighted_cart.sum(0)


    """ Calculate the moment of inertia
    """
```

```
    center_mass_mat = np.tile(center_mass,
(cart_coords.shape[0], 1))

    d_cart = cart_coords - center_mass_mat

    tensor = np.zeros((3, 3))

    tensor[0, 0] = np.sum(atom_mass * (d_cart[:, 1] **
2 + d_cart[:, 2] ** 2))
    tensor[1, 1] = np.sum(atom_mass * (d_cart[:, 0] **
2 + d_cart[:, 2] ** 2))
    tensor[2, 2] = np.sum(atom_mass * (d_cart[:, 0] **
2 + d_cart[:, 1] ** 2))

    tensor[1, 0] = -1 * np.sum(atom_mass *
d_cart[:, 0] * d_cart[:, 1])
    tensor[2, 0] = -1 * np.sum(atom_mass *
d_cart[:, 0] * d_cart[:, 2])
    tensor[2, 1] = -1 * np.sum(atom_mass *
d_cart[:, 1] * d_cart[:, 2])

    tensor[0, 1] = tensor[1, 0]
    tensor[0, 2] = tensor[2, 0]
    tensor[1, 2] = tensor[2, 1]

    """ Diagonalize tensor
    """
```

```
    moi_eig_val, eig_vect = linalg.eigh(tensor)

    """ Build the d-matrix
    """
    num_atom = atom_mass.size
    num_atom_3 = num_atom * 3
    p_cart = np.zeros(cart_coords.shape)
    t_eig_vec = eig_vect
    for i in range(num_atom):
        p_cart[i, :] = np.dot(t_eig_vec.transpose(),
d_cart[i, :])
    sq_mass = np.sqrt(atom_mass)
    d_matrix = np.zeros((num_atom_3, num_atom_3))

    n1=0
    n2=0
    n3=0
    for i in range(atom_mass.size):
        if i < 3:
            if i\%3 == 0:
                for j in range(len(d_matrix[i])):
                    if j\%3 == 0:
                        d_matrix[i,j] = 1.0 * sq_mass[n1]
                        n1 += 1
            if i\%3 == 1:
                for j in range(len(d_matrix[i])):
                    if j\%3 == 1:
```

```
                    d_matrix[i,j] = 1.0 * sq_mass[n2]
                    n2 += 1
            if i\%3 == 2:
                for j in range(len(d_matrix[i])):
                    if j\%3 == 2:
                        d_matrix[i,j] = 1.0 * sq_mass[n3]
                        n3 += 1
        m = i * 3
        d_matrix[3, m + 0] = (p_cart[i, 1] * t_eig_vec[0, 2]
- p_cart[i, 2] * t_eig_vec[0, 1]) * sq_mass[i]
        d_matrix[3, m + 1] = (p_cart[i, 1] * t_eig_vec[1, 2]
- p_cart[i, 2] * t_eig_vec[1, 1]) * sq_mass[i]
        d_matrix[3, m + 2] = (p_cart[i, 1] * t_eig_vec[2, 2]
- p_cart[i, 2] * t_eig_vec[2, 1]) * sq_mass[i]
        d_matrix[4, m + 0] = (p_cart[i, 2] * t_eig_vec[0, 0]
- p_cart[i, 0] * t_eig_vec[0, 2]) * sq_mass[i]
        d_matrix[4, m + 1] = (p_cart[i, 2] * t_eig_vec[1, 0]
- p_cart[i, 0] * t_eig_vec[1, 2]) * sq_mass[i]
        d_matrix[4, m + 2] = (p_cart[i, 2] * t_eig_vec[2, 0]
- p_cart[i, 0] * t_eig_vec[2, 2]) * sq_mass[i]
        d_matrix[5, m + 0] = (p_cart[i, 0] * t_eig_vec[0, 1]
- p_cart[i, 1] * t_eig_vec[0, 0]) * sq_mass[i]
        d_matrix[5, m + 1] = (p_cart[i, 0] * t_eig_vec[1, 1]
- p_cart[i, 1] * t_eig_vec[1, 0]) * sq_mass[i]
        d_matrix[5, m + 2] = (p_cart[i, 0] * t_eig_vec[2, 1]
- p_cart[i, 1] * t_eig_vec[2, 0]) * sq_mass[i]
```

```
    for i in range(6):
        d_matrix[i, :] = d_matrix[i, :] / np.linalg.norm(
d_matrix[i, :])

    eigen_val, eig_vector = linalg.eigh(mw_cart_hessian)
    norm_mode_dotted_w_dimertransrot = np.dot(d_matrix,
eig_vector)

    """Begin computing entropy. Source: Joseph W. Ochterski;
    /textit{Thermochemistry in Gaussian}; June 2, 2000.
    """

    #### Need to convert ATM to Pa. Pa = kg/m/s^2 so units
    should all cancel out
    qt = (2 * consts['pi'] * total_mass *
    consts['amu_to_kg'] * consts['boltzmann'] * T )**1.5 /
    consts['plancks']**3 * consts['boltzmann'] * T / P /
    consts['atm_to_Pa']
    Trans_entropy = R * (math.log(qt) + 2.5) / 4.184
    print 'Full Translational entropy:', Trans_entropy

    theta_rx = consts['plancks_amubohr']**2 / (8 *
consts['pi']**2 * moi_eig_val[0] *
consts['boltzmann_amubohr'])
    theta_ry = consts['plancks_amubohr']**2 / (8 *
    consts['pi']**2 * moi_eig_val[1] *
    consts['boltzmann_amubohr'])
```

```
    theta_rz = consts['plancks_amubohr']**2 / (8 *
    consts['pi']**2 * moi_eig_val[2] *
    consts['boltzmann_amubohr'])#
    theta_r = theta_rx * theta_ry * theta_rz
    print 'theta_r=',theta_rx,theta_ry,theta_rz,theta_r**
0.333333333
    qr = ((consts['pi'] * T**3) / theta_r)**0.5/
    tot_rot_sym
#Rot_entropy = R * (math.log(qr) + 1.5) / 4.184#
print 'Full Rotational entropy:', Rot_entropy


    theta_vK = []
    HO_S = np.zeros(len(vib_energies),float)
    S_diff = np.zeros(len(vib_energies),float)
    for i in range(len(vib_energies)):
        #if vib_energies[i] > 0.0:
            theta_vK.append(consts['plancks'] *
    vib_energies[i] * consts['sLight'] /
    consts['boltzmann'])
    theta_vK = np.asarray(theta_vK)
    Sv_summation = 0
    Sv_components = []
    for i in range(6,len(theta_vK)):
        if theta_vK[i] > 0.0:
            tvKT = theta_vK[i] / T
            A = tvKT / math.expm1(tvKT) - math.log(abs(1.0 -
math.exp(-tvKT)))
```

```
            Sv_components.append(R * A / 4.184)
            Sv_summation += A
            #print "i= ",i,"Freq= ",vib_energies[i]," Sv= ",
A* R/4.184
            HO_S[i]=0.001*R * A / 4.184
    Vib_entropy = 0.001*R * Sv_summation / 4.184
    print 'Full Harmonic Vibrational entropy:', Vib_entropy

    mu_tot = RM
    mu_tot /= avog_num * 1000 * 9.10938291E-31  # Conversion
from amu to atomic units (Me based)
    FC_tot = FC
    FC_tot /= 15.569141 # mdyne/angstrum to Hartree/bohr^2
    S_tot = np.zeros(len(FC))
    De_tot = np.zeros(len(FC))
    n_low = 0
    for i in range(6,len(FC)): #all mode minus the first 6
(trans/rot of system)
        mu = mu_tot[i]
        FC = FC_tot[i]
        De = FC*4/np.pi
        De_tot[i]=(De*627.51)
#       only include low frequencye modes
        if math.exp(-De_tot[i]/(0.0019872*T))>0.0001:
            n_low=n_low+1
            vib_en =  hbar * (np.abs(FC*15.569141*100) / (
mu*9.10938291E-31))**.5
```

```
            vib_en *= 5.0341170122E22 # from J to cm-1
            print "mode", i-6, "vib_en ", vib_en
            vib_en *= 4.55633525275523E-6  # from cm-1
 to Hartrees
            spring_const = FC
            alpha_h = vib_en * mu / 2.0
            alpha_g = spring_const / (2.0 * De)
            c = np.sqrt(1.0 / (2.0 * alpha_g))
            fwhm = 2.0 * np.sqrt(2.0 * np.log(2.0)) * c
            fwtm = 2.0 * np.sqrt(2.0*np.log(10.0)) * c
            fw90m = 2.0 * np.sqrt(2.0 * np.log(90.0)) * c
            Nx=70 # number of grid points used in the
potential expansion
            xvals = np.zeros(2*Nx+1,dtype=float)
            dgrid=fw90m/Nx
            dstep=-fw90m
            for l in range(2*Nx+1):
                xvals[l]=dstep
                dstep=dstep+dgrid

            U_model= gauss(xvals,De,alpha_g)

            # compute wave functions and energies for the
potential functions
            Ngaus=int(De/vib_en)
            if Ngaus < 5:
            # use 5 gaussians spread over the entire xval
```

```
range
                Ngaus =5
                Xgaus = np.zeros(2*Ngaus+1,dtype=float)
                xstep=xvals[0]
                xdelta=-xstep/Ngaus
                for k in range(Xgaus.size):
                    Xgaus[k]=xstep
                    xstep=xstep+xdelta
            elif Ngaus > 50:
            # use 50 gaussians spread over a compressed
range
                Ngaus=50
                Xgaus = np.zeros(2*Ngaus+1,dtype=float)
                xstep=xvals[0]*100.0*vib_en/De
                xdelta=-xstep/Ngaus
                for k in range(Xgaus.size):
                    Xgaus[k]=xstep
                    xstep=xstep+xdelta
            else :
            # use Ngaus gaussians spread over the entire
xval range
                Xgaus = np.zeros(2*Ngaus+1,dtype=float)
                xstep=xvals[0]
                xdelta=-xstep/Ngaus
                for k in range(Xgaus.size):
                    Xgaus[k]=xstep
                    xstep=xstep+xdelta
```

```
            # compute wave functions and energies for the
potential functions
            psi_model, E_model, n_E = basis_V2(1,alpha_h,mu,
U_model,xvals,dgrid,Xgaus) #uses the above code
            e_conv=219474.6 # from Hartrees to cm-1
            U_model *= e_conv
            E_kcal=627.51*E_model # from Hartrees to kcal/mol
            E_cm=e_conv*E_model
            print "Vibrational Energy: ", E_cm[1],E_cm[0],
"cm-1"
            De_cm = De*e_conv
            U, S = Stat_Thermo(E_kcal,298.15, De_tot[i],
vib_en, hbar)
            S_tot[i] = S
            S_diff[i]=HO_S[i]-S_tot[i]
            print "HO_S,S_tot,S_diff=",HO_S[i],S_tot[i],
S_diff[i]
#############at end of gaussian calc ####################
    # Harmonic Oscillator entropy
    #SHO = S_HO(E_kcal,T)

    print "S vib for low lying modes: ", np.sum(S_tot),
"kcal/mol/K"
    print "S vib corrected: ", np.sum(HO_S)-np.sum(S_diff),
"kcal/mol/K"
    print "S vib difference total: ", np.sum(S_diff),
"kcal/mol/K"
```

```
    return vib_energies, d_matrix


def main():
    np.set_printoptions(linewidth=10000)
    parser = argparse.ArgumentParser()
    parser.add_argument('input_file', help='The input
file to analyze (.fchk or previous extracted)')
    parser.add_argument('-i', '--input_type', type=str,
choices=TYPE_ACCEPT, default="", help='Style of
input_type')
    parser.add_argument('-t', '--Temp', type=str,
default="", help='temp')
    parser.add_argument('-p', '--pressure', type=str,
default="", help='pressure in atm')
    parser.add_argument('-trs', '--tot_rot_sym', type=str,
default="", help='total rotational symmetry from .log file')
    parser.add_argument('-f1rs', '--frag1_rot_sym', type=str,
default="", help='rotational symmetry of only frag1 from
separate, individual .log file')
    parser.add_argument('-f2rs', '--frag2_rot_sym',
type=str, default="", help='rotational symmetry of only frag2
from separate, individual .log file')
    parser.add_argument('-f', '--full_diag', action=
'store_true', help='Diagonalizes full hessian and can
produce cartesian displacements for all 3n modes')

    args = parser.parse_args()
```

```
    mole_dat = MoleculeData(args.input_file)
    mole_dat.extract_file(file_type=args.input_type)

    vib_energies, d_matrix = extract(mole_dat,
args.full_diag)

if __name__ == '__main__':
    main()}
```

A.2.3 Geometry Optimized Coordinates

Table A5: Ethane coordinates in the staggered confirmation.

	x	y	z
C	-0.7626390000	-0.0000070000	-0.0001210000
C	0.7626390000	0.0000070000	0.0001210000
H	-1.1611720000	0.6571530000	-0.7780060000
H	-1.1614200000	0.3450530000	0.9578450000
H	-1.1612660000	-1.0022060000	-0.1803460000
H	1.1612660000	1.0022060000	0.1803460000
H	1.1614200000	-0.3450530000	-0.9578450000
H	1.1611720000	-0.6571530000	0.7780060000

Table A6: Ethane coordinates in the eclipsed confirmation.

	x	y	z
C	0.7688610000	0.0000040000	-0.0000030000
C	-0.7688610000	0.0000040000	-0.0000030000
H	1.1764110000	-0.5913290000	-0.8232550000
H	1.1763950000	-0.4173160000	0.9237320000
H	1.1764310000	1.0086240000	-0.1004600000
H	-1.1763930000	-0.4173000000	0.9237400000
H	-1.1764100000	-0.5913460000	-0.8232430000
H	-1.1764340000	1.0086210000	-0.1004790000

Table A7: Methane monomer coordinates.

	x	y	z
C	0.0000010000	0.0000020000	-0.0000020000
H	0.0536820000	-0.1286530000	1.0816670000
H	0.9598630000	0.3609370000	-0.3712490000
H	-0.7788700000	0.7241230000	-0.2417610000
H	-0.2346810000	-0.9564220000	-0.4686450000

Table A8: Methane dimer at equilibrium distance.

	x	y	z
C	-1.7410970000	0.0000190000	-0.0000230000
H	-2.8313940000	0.0132380000	-0.0265730000
H	-1.3507390000	-0.0047300000	-1.0182040000
H	-1.3776220000	0.8854990000	0.5224580000
H	-1.3993720000	-0.8941220000	0.5224190000
C	1.7410970000	0.0000190000	0.0000230000
H	2.8313940000	0.0132420000	0.0265860000
H	1.3507230000	-0.0043520000	1.0181990000
H	1.3993770000	-0.8943140000	-0.5220940000
H	1.3776350000	0.8853080000	-0.5227910000

Table A9: Methane dimer at a separation distance of 4.5 Å.

	x	y	z
C	2.2500030000	-0.0000010000	-0.0000010000
H	3.3405060000	-0.0120180000	-0.0016070000
H	1.8883960000	0.0337660000	1.0284630000
H	1.8954050000	0.8792870000	-0.5392120000
H	1.8758540000	-0.9010280000	-0.4876350000
C	-2.2499990000	0.0000050000	0.0000020000
H	-2.6258160000	0.9293370000	0.4296540000
H	-2.6087100000	-0.0975040000	-1.0253390000
H	-2.6064780000	-0.8441100000	0.5915290000
H	-1.1591820000	0.0122410000	0.0041370000

Table A10: Methane dimer at a separation distance of 6.5 Å.

	x	y	z
C	-3.2499960000	0.0000030000	0.0000010000
H	-2.8236520000	-0.4458720000	-0.8993830000
H	-2.8490990000	1.0055910000	0.1324290000
H	-4.3348930000	0.0513140000	-0.0989290000
H	-2.9924370000	-0.6110580000	0.8658770000
C	3.2500010000	-0.0000010000	0.0000030000
H	3.5977930000	1.0336510000	-0.0074030000
H	3.6799650000	-0.5221070000	0.8555590000
H	3.5603360000	-0.4945750000	-0.9211590000
H	2.1619530000	-0.0169600000	0.0729860000

Table A11: Methane dimer at a separation distance of 8.5 Å.

	x	y	z
C	-4.2500000000	0.0000020000	-0.0000010000
H	-5.3391870000	0.0027920000	0.0557750000
H	-3.8623360000	-0.8852840000	0.5053960000
H	-3.9395250000	-0.0130570000	-1.0454070000
H	-3.8588860000	0.8955380000	0.4842380000
C	4.2499980000	0.0000040000	-0.0000030000
H	5.3399500000	-0.0372220000	0.0079270000
H	3.9213700000	0.9299150000	-0.4655400000
H	3.8772110000	-0.0455040000	1.0239140000
H	3.8614150000	-0.8472140000	-0.5662800000

Table A12: Chalcone monomer coordinates *in vacuo*.

	x	y	z
C	-0.1487819065	-0.4121964813	0.1681251418
H	-0.0405497403	1.6013613616	0.5296183850
C	-2.3720879958	-1.5825548086	-0.1150994305
H	-2.3029221349	-3.6049660177	-0.4850114031
C	-4.8980185843	-0.4520678099	-0.0030897020
C	-5.3264251351	2.0593761392	0.7140235162
C	-6.9821879285	-1.9489103168	-0.6360194068
C	-7.7551293215	3.0365326609	0.7765488799
H	-3.7533491921	3.2608355769	1.2478559901
C	-9.4173814339	-0.9713323862	-0.5793125093
H	-6.6829044648	-3.9036466589	-1.1847580394
C	-9.8101590909	1.5262693200	0.1259880317
H	-8.0533054058	4.9846814182	1.3421722142
H	-11.0078163999	-2.1631409932	-1.0823386702
C	2.2439494560	-1.8718019376	0.0279225912
O	2.2497490250	-4.1693345739	-0.0182264072
C	4.6900353420	-0.4405952834	-0.0186742722
C	4.8564994146	2.0737040416	-0.7989855966
C	6.8890319962	-1.7151228661	0.6921385995
C	7.1818450381	3.2930308385	-0.8634573782
H	3.1848931585	3.0868862996	-1.4156315326
C	9.2043167185	-0.4890327400	0.6590362693
H	6.7442015067	-3.6762767176	1.2666927788
C	9.3545952880	2.0188963188	-0.1209443530
H	7.2958125227	5.2395412033	-1.4970862816
H	10.8991930604	-1.4892344701	1.2339419377
H	11.1676040706	2.9766850389	-0.1567886755
H	-11.7094565417	2.2967861946	0.1779403785

Table A13: Chalcone dimer coordinates *in vacuo*.

	x	y	z		x	y	z
C	1.3223187530	2.8826919579	-1.6948574446	H	-0.8157701428	-3.1223583461	0.4198064078
H	0.7199553660	1.6194214729	-3.1920741173	C	-4.0653316144	-1.6969777172	2.2647685672
C	-0.3700215766	4.0993788076	-0.2622051501	H	-4.6669258829	-0.6023558280	3.8975447336
H	0.3623908631	5.2582116961	1.2698089369	C	-6.0782469505	-2.5943594563	0.5925651577
C	-3.1240250845	3.9575718797	-0.4885792057	C	-5.6121157998	-3.4968604631	-1.8487094860
C	-4.3342698582	3.1517454750	-2.6975007008	C	-8.5807543854	-2.4522180469	1.4220943954
C	-4.6065283506	4.6227384201	1.5950988097	C	-7.5855963339	-4.2433267950	-3.3998759459
C	-6.9463889594	2.9909316831	-2.8094461786	H	-3.6921938796	-3.5672376384	-2.5651896459
H	-3.2252671543	2.6722747485	-4.3549151147	C	-10.5620924002	-3.2064001300	-0.1239490173
C	-7.2229560270	4.4527348907	1.4866795606	H	-8.9716479856	-1.7213325057	3.2991327637
H	-3.6861580948	5.2240813550	3.3265035550	C	-10.0697507500	-4.1009831849	-2.5428058416
C	-8.3988550309	3.6356476089	-0.7150874319	H	-7.1919853077	-4.9247903324	-5.2936043741
H	-7.8568154770	2.3459568648	-4.5290741515	H	-12.4926572571	-3.0837436853	0.5560102982
H	-8.3446935943	4.9578435297	3.1274757241	C	0.2261813036	-0.7868611147	3.7268363365
C	4.0674575562	3.2174493579	-1.2398397725	O	-0.4704661821	0.7874337016	5.2534004537
O	4.8680720966	4.9695220363	0.0151385949	C	2.9641731631	-1.4448523655	3.4801552854
C	5.8639312364	1.3337818308	-2.3609934145	C	3.7733897362	-3.7813908360	2.5608187097
C	5.0935712091	-1.0321532172	-3.2274781337	C	4.7554198612	0.3545957434	4.1906487934
C	8.4262844036	1.9606700816	-2.4514583770	C	6.3358827431	-4.3008330466	2.3389610992
C	6.8493949935	-2.7391124670	-4.1764267156	H	2.4082970355	-5.2232763317	2.0466488347
H	3.1265894426	-1.5926931885	-3.1105570073	C	7.3123835299	-0.1467088771	3.9191631989
C	10.1771571060	0.2672091445	-3.4124898669	H	4.1141583545	2.1664971466	4.8968129391
H	9.0090097581	3.7932431974	-1.7449521909	C	8.1065125307	-2.4742409136	2.9907672769
C	9.3898046626	-2.0902372543	-4.2778501991	H	6.9521280572	-6.1223531551	1.6279724830
H	6.2316341193	-4.5846698901	-4.8222367927	H	8.6875654777	1.2932509878	4.4045903417
H	12.1616320659	0.7783724655	-3.4864348448	H	10.1071484273	-2.8588511748	2.7594799337
H	10.7599826160	-3.4206402551	-5.0259435851	H	-10.4420475567	3.4917827694	-0.8018919952
C	-1.5772654655	-2.0316368514	1.9791799477	H	-11.6131655539	-4.6802484511	-3.7623140522

Table A14: Benzene monomer coordinates.

	x	y	z
C	2.2804154984	1.3073426745	-0.0000226767
C	0.0079803128	2.6284293261	-0.0000094486
C	-2.2724011704	1.3211811379	-0.0000453534
C	-2.2804457340	-1.3073313362	-0.0000188973
C	-0.0079935409	-2.6284425542	-0.0000548021
C	2.2724540828	-1.3211811379	0.0001492884
H	4.0596227522	2.3275736104	-0.0005083363
H	0.0141691655	4.6793753977	-0.0000623610
H	-4.0453968950	2.3521476072	0.0008182514
H	-4.0596662159	-2.3275225878	-0.0004781007
H	-0.0142447545	-4.6793905155	0.0000472431
H	4.0454516970	-2.3521702839	0.0002078699

Table A15: Benzene dimer coordinates in the T-shaped configuration.

	x	y	z
C	2.3367270000	0.2580540000	1.4097240000
C	2.6505310000	1.2545450000	0.4911890000
C	2.7490110000	0.9474400000	-0.8622310000
C	2.5353580000	-0.3568960000	-1.2967610000
C	2.2266480000	-1.3539250000	-0.3771500000
C	2.1258240000	-1.0462940000	0.9756380000
H	2.2516910000	0.4989630000	2.4642270000
H	2.8147070000	2.2723680000	0.8300400000
H	2.9890430000	1.7255590000	-1.5795960000
H	2.6074690000	-0.5959340000	-2.3527470000
H	2.0619150000	-2.3716780000	-0.7166630000
H	1.8768960000	-1.8223220000	1.6921420000
C	-3.7855440000	-0.2610020000	0.0916640000
C	-3.3806200000	1.0680330000	0.0239940000
C	-2.0322830000	1.3776330000	-0.1244400000
C	-1.0877070000	0.3599240000	-0.2049520000
C	-1.4938990000	-0.9684640000	-0.1370950000
C	-2.8414620000	-1.2797670000	0.0106000000
H	-4.8373530000	-0.5029980000	0.2073470000
H	-4.1168670000	1.8632040000	0.0871820000
H	-1.7162980000	2.4147840000	-0.1772310000
H	-0.0361090000	0.5984110000	-0.3196330000
H	-0.7533000000	-1.7589950000	-0.1991340000
H	-3.1572960000	-2.3170470000	0.0629980000

Table A16: Benzene dimer coordinates in the parallel displaced configuration.

	x	y	z
C	-4.6242010680	1.7933386248	-0.9562089091
C	-3.1484044394	2.5582430132	1.0777806511
C	-2.0351347342	0.7667279740	2.6448472655
C	-2.4048539547	-1.7907515899	2.1822234463
C	-3.8855260763	-2.5592388988	0.1510401291
C	-4.9909023961	-0.7684891986	-1.4219640044
H	-5.4867438188	3.1927884337	-2.1834895627
H	-2.8467002367	4.5560046365	1.4315033411
H	-0.8498872558	1.3640722487	4.2071423218
H	-1.5163482586	-3.1873573612	3.3922074379
H	-4.1702964445	-4.5581627036	-0.2114830148
H	-6.1436976101	-1.3681238212	-3.0094793437
C	2.4055002410	-1.7912032344	-2.1819966792
C	2.0353388246	0.7660816877	-2.6453820580
C	3.1479206696	2.5581560858	-1.0784288272
C	4.6234130523	1.7939792419	0.9560652899
C	4.9907909023	-0.7677087418	1.4224156489
C	3.8860684277	-2.5589592194	-0.1504089606
H	1.5175633524	-3.1883305701	-3.3917879187
H	0.8501196921	1.3628477063	-4.2079265581
H	2.8459481258	4.5558194434	-1.4326031617
H	5.4853265243	3.1938920337	2.1832495675
H	6.1435124170	-1.3667121959	3.0102220060
H	4.1711600492	-4.5577715303	0.2125563792

Table A17: Chalcone monomer coordinates in cyclohexane.

	x	y	z
C	-0.1440878272	-0.4239959303	0.1457299991
H	-0.0283799049	1.5794518785	0.5549501619
C	-2.3729780567	-1.5805951627	-0.1623425802
H	-2.3167265832	-3.5965000453	-0.5667533904
C	-4.8949950227	-0.4451721998	-0.0268208810
C	-5.3053924849	2.0987391315	0.5820242661
C	-6.9942727262	-1.9729551903	-0.5250187920
C	-7.7342856439	3.0736298718	0.6802427744
H	-3.7203016640	3.3323881617	0.9918359132
C	-9.4298592947	-0.9977998884	-0.4292531480
H	-6.7084800163	-3.9501414771	-0.9940884666
C	-9.8058920897	1.5298068871	0.1732028355
H	-8.0191693955	5.0470688319	1.1574155940
H	-11.0324263015	-2.2137516347	-0.8234953427
C	2.2439494560	-1.8811882066	-0.0230206420
O	2.2518522900	-4.1813985847	-0.1251433241
C	4.6880983729	-0.4496300634	-0.0272914227
C	4.8543262297	2.0733053095	-0.7825865545
C	6.8837388737	-1.7237929290	0.6970726740
C	7.1771849738	3.2992952802	-0.8127409121
H	3.1863406887	3.0884320955	-1.4050396184
C	9.1960737337	-0.4904292475	0.6997749822
H	6.7440786745	-3.6888339468	1.2590431680
C	9.3466111957	2.0247771461	-0.0583887536
H	7.2914944988	5.2514483668	-1.4278372727
H	10.8872197565	-1.4898372927	1.2862174277
H	11.1567532640	2.9881311092	-0.0667829164
H	-11.7045300260	2.2993600014	0.2521782640

Table A18: Chalcone dimer coordinates in cyclohexane.

	x	y	z		x	y	z
C	1.3123769045	3.0045962917	-1.4551891667	H	-0.8182456839	-3.1430300587	0.1812303915
H	0.7298820967	1.8543219722	-3.0476555881	C	-4.0556392098	-1.8650934103	2.1502322750
C	-0.3994351617	4.0993145569	0.0524852496	H	-4.6539566934	-0.9068851711	3.8674905315
H	0.3099906512	5.1445957005	1.6744313964	C	-6.0757789683	-2.6233573016	0.4197364880
C	-3.1504869175	3.9688592130	-0.2098219457	C	-5.6170404257	-3.3680983137	-2.0764101293
C	-4.3385122931	3.3065159232	-2.4782887066	C	-8.5766555697	-2.5142860970	1.2608799816
C	-4.6542684982	4.5069738060	1.8958562597	C	-7.5983123001	-3.9924165372	-3.6713615403
C	-6.9502194339	3.1650037925	-2.6275128090	H	-3.6985395795	-3.4139506251	-2.7983515973
H	-3.2148906689	2.9252731536	-4.1510609236	C	-10.5655902830	-3.1474463484	-0.3300104083
C	-7.2700366602	4.3524603606	1.7506062510	H	-8.9613716556	-1.9063536876	3.1824591915
H	-3.7527236928	5.0075773382	3.6689256835	C	-10.0813328806	-3.8846492435	-2.8034708650
C	-8.4238334290	3.6812051230	-0.5117850409	H	-7.2119200272	-4.5513729182	-5.6060837944
H	-7.8436705430	2.6357066609	-4.3947788844	H	-12.4946169029	-3.0523118729	0.3581333105
H	-8.4079011492	4.7561663030	3.4079734218	C	0.2471515929	-1.1179996894	3.6737879485
C	4.0512115818	3.3281702933	-0.9665891740	O	-0.4317097918	0.2995612534	5.3591079461
O	4.8319273076	4.9916015947	0.4213824393	C	2.9813526620	-1.7569443920	3.3487324018
C	5.8675576206	1.5601294303	-2.2295327364	C	3.7752775724	-4.0074493074	2.2229394824
C	5.1168715305	-0.7338561904	-3.2882611701	C	4.7854532764	-0.0359898315	4.2080852951
C	8.4248916756	2.2157830901	-2.2709876554	C	6.3347829226	-4.5188318366	1.9493260565
C	6.8868229065	-2.3402253259	-4.3765373594	H	2.4027847048	-5.3918096543	1.5864419749
H	3.1557592530	-1.3209789372	-3.2192521565	C	7.3395540102	-0.5252852433	3.8870718722
C	10.1896784304	0.6236057968	-3.3704396842	H	4.1586614016	1.7105969724	5.0742978937
H	8.9981854076	3.9895158078	-1.4215633824	C	8.1180285209	-2.7670539555	2.7553508824
C	9.4215596181	-1.6617418864	-4.4269968228	H	6.9381592027	-6.2744061773	1.0795267580
H	6.2844670784	-4.1314550165	-5.1718398801	H	8.7250708694	0.8563142139	4.4968637720
H	12.1690303432	1.1574949625	-3.4038500397	H	10.1159564401	-3.1423119628	2.4886595229
H	10.8022822425	-2.9131713484	-5.2835378038	H	-10.4663154179	3.5516247222	-0.6292522980
C	-1.5683667459	-2.1856060558	1.8304547329	H	-11.6301636392	-4.3681375274	-4.0572133516