

Bridging the Gap Between Optical Spectroscopic Experiments and Computer Simulations for Fast Protein Folding Dynamics

Raymond Z. Cui¹, Daniel-Adriano Silva¹, Jian Song³, Gregory R. Bowman², Wei Zhuang^{3,*} and Xuhui Huang^{1,*}

¹Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; ²Department of Chemistry, Stanford University, Stanford, CA, 94305, USA; ³State Key Lab of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Dalian, Liaoning, China, 116023

Abstract: Fast folding techniques use optical spectroscopic tools to monitor protein folding or unfolding dynamics after a fast triggering such as the laser induced temperature jump. These techniques have greatly improved time resolution of experiments and provide new opportunities for comparison between theory and simulations. However, the direct comparison is still difficult due to two main challenges: a gap between folding relevant timescales (microseconds or above) and length of molecular dynamics simulations (typically tens to hundreds of nanoseconds), and difficulty in directly calculating spectroscopic observables from simulation configurations. This review is focused on recent advances in addressing these two challenges. We describe new methodology that allows simulating folding timescales with an emphasis on Markov State Models. We also review progress on modeling infrared, circular dichroism, and fluorescence spectroscopic signals from protein conformations. At last, we discuss a few studies that directly simulate time-resolved spectroscopy of temperature jump induced unfolding dynamics for a few small proteins. These studies not only provide direct validation of theoretical models, but also greatly improve our understanding of protein folding mechanisms by connecting ensemble averaged spectroscopic observables with atomistic protein conformations.

Keywords: Protein folding, Optical spectroscopy, Molecular dynamics simulations, Markov state model.

INTRODUCTION

Protein native structure is solely determined by its amino acid sequence, and denatured proteins can spontaneously fold into their native conformations under physiologic conditions [1]. Understanding molecular mechanisms of protein folding is fundamental for biochemistry and molecular biology. In the past decades, substantial advances have been made on both theoretical and experimental sides to investigate protein folding mechanisms [2-15]. An ultimate solution of protein folding has to involve a tight combination of theory and experiment. In particular, theoretical approaches can provide a complete and atomistic detailed picture of protein folding, while experiments have to perform a reality-check on theoretical models. However, direct comparison between theory and experiment for protein folding is still difficult due to two major challenges. First, a timescale gap exists between experimental protein folding times (typically microseconds or above [16-20]) and length of molecular dynamics (MD) simulations (typically tens to hundreds of nanoseconds). Second, even with sufficiently long simulations, it is still a difficult task to carry out a direct comparison between MD simulations and experimental results that are often optical spectroscopic observables. In

recent years, a number of research groups have been working toward the direction to bridge the gap between theoretical and experimental investigations in order to achieve a better understanding of protein folding [21-23].

All-atom MD simulation has the potential to provide realistic kinetic information by explicitly modeling protein and solvent atoms, and thus makes it a good candidate for direct comparison with experiments. However, there exists a timescale gap of orders of magnitude between all-atom simulations and experiments. Coarse-grained simulations with simplified representations of proteins are a natural solution to bridge this timescale gap, and such simulations using lattice and off-lattice models [24-30] have provided great insights for protein folding. However, these simplified models are normally incapable of capturing atomistic details and realistic dynamics. Thus they are insufficient to directly simulate the experiments because most optical spectroscopic experiments for protein folding [23, 31] monitor time evolution of protein structure ensembles during its conformational changes. Another solution for the timescale gap is to develop algorithms that can construct models from short simulations to predict long timescale dynamics for protein folding. Markov State Model (MSM) approach [32-37] is one of such striking examples that have recently shown success in investigating protein folding kinetic at microsecond or even millisecond timescales [22]. At last, in recent years, progress in technologies has enable all-atom simulations to rapidly approach experimental protein folding timescales [38-42]. In this review, we will describe recent

*Address correspondence to these authors at the Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; Tel: (852) 23587363; Fax: (852) 23581594; E-mail: xuhuihuang@ust.hk and State Key Lab of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Dalian, Liaoning, China, 116023; Tel: (86) 41184379659; Fax: (86) 41184675584; E-mail: wzhuang@dicp.ac.cn

advances in both simulation methodology and hardware technology that can bridge the timescale gap, and enable direct comparison between all-atom MD simulations and experiments.

In the past twenty years, introduction of fast folding techniques have greatly improved time resolution of experiments [23]. In these techniques, protein folding/unfolding can be triggered by photochemical reaction [43], temperature jump (T-jump) [23, 31, 44] or fast mixing [45], and relaxation dynamics are then be monitored by optical spectroscopic tools. Among different triggers, T-jump is more general and quickly gains popularity because all proteins melt at both low and high temperatures [31]. In the T-jump experiment, a laser is typically used to excite the water OH stretching overtone, and the protein sample can be heated up within 100 ps, which leaves protein backbone structure largely unchanged. Relaxation of protein structures after the T-jump can then be measured by Infrared (IR) spectroscopy, Circular Dichroism (CD), or fluorescence at timescales from tens of nanoseconds to tens of microseconds. Some of these timescales overlap with state-of-the-art all-atom MD simulations, thus providing a good opportunity for direct quantitative comparison between theory and experiment. However, accurate calculation of the spectroscopic observables from MD conformations is difficult, and often requires Quantum Mechanics (QM) to properly treat vibrational and electronic state excitation. Fortunately, extensive research has been performed aiming to calculate different spectroscopy using a wide range of methods [46-51]. In this review, we will survey theoretical studies to model protein IR, CD and fluorescence spectroscopy.

Directly simulating fast protein folding dynamics has recently become possible with advances in both generating long timescale simulations and calculating time resolved spectroscopic observables from protein conformations. A few groups have successfully modeled T-jump triggered optical spectroscopy such as fluorescence for PinWW domain [52], IR and 2DIR for trpzip2 [53], and CD for alanine polypeptide [54]. These studies provide direct validation of theoretical model and could further help refine simulation methodology and force field parameterizations. Moreover, they can greatly help interpretation of experimental data by connecting ensemble averaged spectroscopic observables with atomistic protein conformations. Although direct prediction of spectroscopic observables for fast folding dynamics is still an emerging area, we believe it is one of the most exciting new directions for the protein folding community.

This review is organized as follows. We will first describe recent advances in methodology that allow simulating relevant timescales for folding. Next, we will review algorithms for calculating spectroscopic signals from protein conformations with a focus on IR, CD and fluorescence. We will then describe a few studies that directly simulate fast folding optical spectroscopic experiments. Last, we will discuss some future challenges in directly modeling fast folding experiments.

SIMULATING BIOLOGICAL RELEVANT TIMESCALES FOR FOLDING

Around ten years ago, the length of all-atom MD simulations is still limited to tens or hundreds of nanoseconds for small proteins. Since then, rapid progress in simulation software and hardware technology has greatly extended the length of protein simulations, and they start to reach relevant timescales for protein folding (microsecond or even longer). For example, the Shaw group has reported a single simulation of 1031 μ s for protein BPTI (58 residues) in the explicit solvent with Anton supercomputer, a specialized hardware for MD simulations [40]. Using IBM BlueGene supercomputer, the Zhou and Berne group have performed microsecond simulations on the 115-residue protein lysozyme, and observed its single mutant induced misfolding [55]. MD simulations can also be greatly accelerated in Graphic Processing Units (GPUs) in the implicit solvent using OpenMM [38] or NAMD [56]. In addition to the improvement on hardware, software packages such as GROMACS [42], NAMD [41], and DESMOND [39] have also improved the scaling efficiency and allowed tens of microsecond individual simulations running parallel on many CPU cores. All these advances in simulation technology open up the door of directly simulating protein folding experiments.

In order to directly compare with experiments, one needs to construct a statistical significant model for protein folding. This requires simulating a large number of folding events due to the heterogeneous nature of protein folding. Therefore, it is still an extremely difficult task to obtain enough sampling for protein folding, and may require many simulations. Furthermore, these simulations can easily generate large quantities of data containing protein conformations along many trajectories. Another significant challenge is how to efficiently analyze these massive datasets to gain biological insight, since it is impossible to examine all these heterogeneous trajectories by eye. A number of methods have been developed to provide a statistical description of long timescale dynamics based on stochastic kinetic models constructed from many short simulations [32, 57-59]. Some of them such as milestoning [60] and transition interface sampling [61] require a pre-determined reaction coordinate, and assume relaxation orthogonal to this reaction coordinate is fast. In many cases, it may not be easy to identify such an appropriate reaction coordinate for protein folding.

An alternative approach is to build Markov State Models (MSMs) [32-37]. These models decompose configuration space into a set of metastable states with fast intra-state but slow inter-state transitions. This separation of timescales ensures that the model could be Markovian, i.e. the probability of being in a given state at time $t+\Delta t$ depends only on the state at time t . This property allows MSMs to be built from many short simulations, and then propagated to give global long timescale dynamics using Eq. (1).

$$P(n\Delta t) = [T(\Delta t)]^n P(0) \quad (1)$$

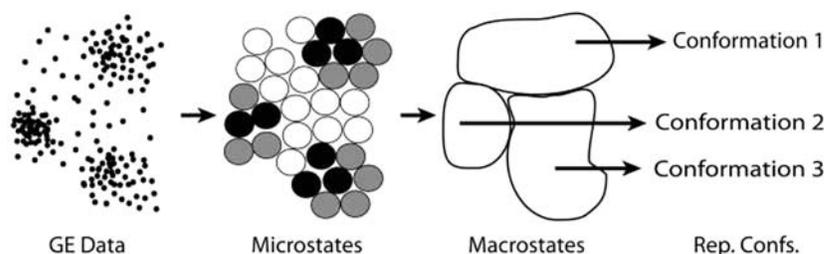


Fig. (1). Schematic diagram of the two-step procedure to build MSMs. First, conformations from MD simulations represented by points are grouped into microstates represented by circles, with darker circles for more highly populated microstates. Kinetically related microstates are then lumped together into metastable states, represented by amorphous shapes. *Reprinted from "Using generalized ensemble simulations and Markov state models to identify conformational states.", Bowman, G. R.; Huang, X.; Pande, V. S., Methods 2009, 49 (2), 197-201, with permission from Elsevier.*

where $P(n\Delta t)$ is a vector of state populations at time $n\Delta t$ and T is the transition probability matrix. Δt , the time interval to count the transitions, is the lag time. T is calculated by normalizing the number of transitions between each pair of states after a lag time in the simulation data. If the model is Markovian and Eq. (1) holds, the exponentiation of T should be identical to an MSM constructed with a longer lag time, and any kinetic properties will then be independent of the lag time. This provides a good way to check if the model is Markovian, and the implied timescales are often used for this check [32, 35, 62].

Building MSMs with good state decomposition is challenging. Typically, a two-step procedure is adopted (see Fig. 1) [63]: First, the massive number of MD conformations is divided into a large set of microstates by geometrical similarity. These microstates have to be fine enough so that they do not combine kinetically separated regions of the phase space. Second, a kinetic clustering [64] is performed to group these microstates into a number of metastable states so that transitions between microstates within the same metastable state are much faster than transitions between different metastable states. A number of software packages are already available for automatically constructing MSMs from MD simulations [33, 52, 63, 65].

MSMs have been successfully used to model dynamics of a number of small proteins [33, 52, 66-71]¹. Jayachandran *et al.* [66] constructed an MSM from simulations generated in a distributed computing environment [72] for the Villin headpiece, a 35-residue protein that folds at microsecond timescale. A more recent MSM study [67] on the same system has yielded reasonable agreements with experiments on prediction of the native state and folding rates. MSMs have also been used to study the PinWW domain, another fast folding protein consisted of a three-stranded β -sheet. Using the transition path theory [73], Noe *et al.* [52] identified folding pathways that have most contributions to the folding flux. Morcos *et al.* [68] have investigated the conformational dynamics of a flexible loop in the Pin1 WW domain. Their results from MSMs are consistent with NMR

relaxation data. More recently, researchers have started to apply MSMs on even larger proteins that fold at millisecond such as a 39-residue segment from NTL9 [69] and a 80-residue lambda repressor protein [71]. These MSM studies have stimulated the development of the new protein folding theory, the kinetic hub model [74, 75]. Pande and coworkers discovered from MSMs of various proteins a common feature of protein folding that native states servers as kinetic hubs (see Fig. 2 from NTL9 folding as an example [69]). In their hublike model, many parallel folding pathways co-exist, and folding rates along these pathways are much faster than those connecting different unfolded states. Therefore, the overall folding kinetics aggregated over numerous folding pathways may still be single or bi-exponential as observed in experiments.

MSMs can also be considered as a data-mining tool to generate comprehensive folding models from massive simulation datasets. They cannot provide any information on the regions of the phase space that have not been explored by the simulations. Therefore, one of the major challenges for the MSM is to ensure that all the relevant conformation states have been visited. This issue may be alleviated by Generalized Ensemble (GE) algorithms [76-79], which can enhance conformational sampling by inducing a random walk in Temperature or Hamiltonian space. In recent years, GE algorithms especially Replica Exchange Method (REM) [78, 79] have been widely applied in protein folding studies [14, 80-91]. Huang *et al.* [92] have used non-equilibrium GE simulations to explore the phase space, and then seed short simulations at constant temperature from GE conformations to construct MSMs to obtain both equilibrium thermodynamics and kinetics. Adaptive sampling, allowing one to use an initial MSM to decide where to run new simulations, is another solution to alleviate the sampling issue. In the adaptive sampling, new simulations are started from those states that contribute most to the statistical uncertainty in kinetic properties of interest calculated from the initial MSM. It has been shown that performing adaptive sampling and constructing MSMs iteratively can quickly yield a good model [93].

Another challenge for the MSM is to identify the transition state. MSMs do not give any information on the transition states since conformations at free energy barriers are grouped together with their nearest free energy basins.

¹Silva, D., Bowman, G.R., Sosa-Peinado, A., Huang, X., "A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein", *PLoS. Comp. Bio.*, 2011, 7(5), e1002054

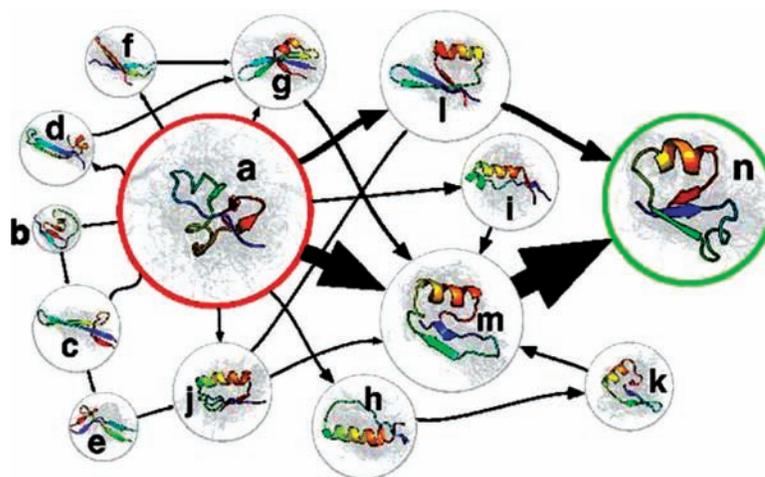


Fig. (2). Top 10 flux folding pathways from a 2000-state Markov State Model (MSM) for NTL9 (1-39). These pathways account for ~25% of the total folding flux. The visual size of each state is proportional to its free energy, and arrow size is proportional to the inter-state flux. The native state (labeled a) acts as a kinetic hub. Reprinted with permission from Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S., "Molecular simulation of *ab initio* protein folding for a millisecond folder NTL9(1-39)." *J Am Chem Soc* **2010**, 132 (5), 1526-8}. Copyright 2011 American Chemical Society.

Yao *et al.* [94, 95] showed how to capture the transition states by constructing a network model using topological methods from MD simulations. More recently, Huang *et al.* [65] developed a new algorithm to construct MSMs by incorporating topological information of the free energy landscape in their model construction.

CALCULATING OPTICAL SPECTROSCOPY OF PROTEINS

Direct calculation of the spectroscopic observables from protein conformations is challenging, but has been extensively studied. In this section, we will review algorithms to model protein fluorescence, CD, and IR spectroscopy.

MODELING FLUORESCENCE SPECTROSCOPY

Fluorescence is a popular optical tool to measure the fast-triggered protein relaxation dynamics [23, 31]. Amino acid tryptophan (Trp) has often been selected as the fluorescence probe for protein dynamics since its fluorescence is sensitive to the polarity of its protein and solvent environment. In this section, we will review recent theoretical and computational advances that aim to model the protein intrinsic Trp fluorescence. There are mainly two categories of methods to model the Trp fluorescence using classical molecular modeling and a hybrid of quantum mechanics and molecular mechanics (QM/MM), respectively [50].

The fluorescence of Trp is due to its side-chain: the 3-methylindole (3MI) moiety. Upon absorbing a photon, 3MI can be excited to 1L_a and 1L_b electronic excited states. Compared with 1L_b , 1L_a transition has a sizable charge transfer (CT) component, and is believed to be the dominating fluorescing state in polar solution. After the excitation to the 1L_a state, the electronic cloud is shifted from

the pyrrole ring to the benzene ring. At this time, placing a positively charged residue near the benzene ring or a negatively charged residue near the pyrrole ring will produce a red shift of the fluorescence and vice versa [49, 50]. Among many environment factors that can alter the Trp fluorescence, three may be dominant in protein systems: the conjugation of the aromatic systems, nearby charged or polarized groups, and the exposure to the solvent molecules. The Barkeley group [96, 97] shows that strong nearby quencher such as histidine, cysteine, or methionine residues could greatly reduce Trp fluorescence lifetime through electron or charge transfer. These Trp-quencher pairs can be used to measure the distance at a few angstroms, and thus provide a good probe for protein folding or unfolding [31]. Residues such as Phe and Tyr may also strongly quench the Trp fluorescence [98]. Without nearby quenching side-chains, the fluorescence will be mostly quenched by backbone amide carbonyl groups at nanosecond timescales. Förster resonance energy transfer (FRET) can also occur between Trp residues or Trp and other chromophores. FRET is often used to measure distance at tens of angstroms [31].

Rotamer model [99] is the simplest classical method to capture the non-exponential decay of the Trp fluorescence. In this model, the multiple-exponential decay of the Trp fluorescence is considered as a result of the ground state heterogeneity, i.e. different stable side-chain conformations of the Trp residues. Excitation of different ground state conformations following the Franck-Condon principle provides the origin of the heterogeneity of the excited state dynamics. Efficient methods such as DEE algorithm [100] have been developed to determine energy minimums of the Trp side-chain attached to a fixed backbone. If strong quenchers exist nearby, fluorescence quenching due to collision for each rotamer can then be modeled by Stern-Volmer relationship [101] (see Eq. (2)).

$$\frac{I_f^0}{I_f} = 1 + k_q \tau_0 [Q] \quad (2)$$

Where I_f^0 and I_f are intensity of fluorescence without and with a quencher respectively. k_q is the quenching rate constant. τ_0 is the fluorescence lifetime without a quencher, and $[Q]$ is the concentration of quencher. Since k_q is highly dependent on the accessibility of the Trp residue, its solvent accessible surface area (SASA) provides a good link between fluorescence lifetimes and rotamer conformations. If no strong quenchers are present nearby, fluorescence is likely to be quenched by the backbone carbonyl groups through charge transfer process, where electron deficient carbon atom in the carbonyl group can serve as the electron acceptor. In this case, the charge transfer rate can be modeled using Marcus-Sutin equation [102]. Rotamer model has been successfully applied to predict Trp fluorescence lifetime in several protein systems [103]. However, algorithms to predict Trp side-chain conformations often treat the backbone fixed. All-atom MD simulations and other sampling algorithms such as REM [78, 79] can take into account the effect of backbone flexibility when computing rotamer state populations, and further predict rotamer transition rates [50, 104].

MD simulations have also been used to compute the Time-dependent Fluorescence Stokes Shift (TDSS), which measures the response of the environment to the new charge distribution of the probe upon the excitation. Nilsson and Halle [105] have obtained the time evolution of the fluorescence maximum wavelength of TDSS from a non-equilibrium solvation response function. This response function is determined by the difference of the probe-solvent electrostatic interactions between the ground and excited states. They suggest that the slow and fast decay of fluorescence relaxation in TDSS are due to protein motion and solvation dynamics separately on protein monellin. In a further study using the same methodology on protein G, Golosov *et al.* [106] found the slow decay may actually be caused by the contributions from both protein conformational and solvent dynamics.

QM is required in order to directly calculate properties of the excited electronic states. Due to the size limitation of quantum calculations, hybrid QM/MM methods are often used to model the protein Trp fluorescence where Trp side-chains are treated quantum mechanically, while the rest of the protein and solvent are treated classically. In a seminal work by Callias and co-workers [49], a QM/MM protocol was developed based on semi-empirical INDO/S-CIS quantum chemical method in combination with CHARMM molecular dynamics simulation to calculate fluorescence of 16 proteins. With a proper Trp ring partial charge rescaling, they can predict the maximum wavelength of the fluorescence emission within an error of 15 nm. In this study, electronic state excitation is simulated by using the 1L_a excited state geometry and charges from 3MI (the Trp fluorophore moiety). Emission maximum wavelength is then determined by averaging many MD snapshots

from 30ps trajectories. More recently, Pistolesi *et al.* [107] have applied another QM/MM algorithm using multi-configurational second-order perturbation theory (CASPT2/CASSCF) and AMBER molecular dynamics to model emission wavelengths of two proteins. Their results show that multiple factors will affect emission energy including the solvent and the specific protein residues. They also argue that multi-configurational quantum chemistry theories may be necessary to properly treat the fluorescence decay due to the crossing between potential energy surfaces. The QM/MM method has also been applied to calculate the FRET probabilities where the protein residue Trp and an antibody Tetracycline (Tc) serve as the donor and acceptor respectively [108]. For each snapshot obtained from classical MD, two QM/MM calculations using the semi-empirical configuration interaction (CI) approach were performed to compute the vertical Trp emission and Tc absorption respectively. The FRET probability is then obtained from the interactions between the transition dipoles of the donor and acceptor. The results support the rotamer model by showing that different rotamer states of Trp display quite different FRET rates.

Both current QM/MM and MM studies focused on modeling the maximum of the fluorescence wavelengths assuming vertical emissions. However, in reality, the fluorescence is always vibrationally resolved in protein-solvent systems. Therefore, it will be important to take into account the coupling between electronic state and vibrational modes. Lin and co-workers [109] have demonstrated that for small multi-atom molecules, such coupling can be calculated using accurate *ab initio* molecular wave functions. For small molecules, the multimode Brownian oscillator (BO) model [110] is often used to compute vibrationally resolved emission spectra. In harmonic approximation, the Huang-Rhys factor can be used to calculate the coupling strength between electronic transitions and nuclear degree of freedoms. These methods developed for small molecules in a uniform environment may fail for complex protein-solvent systems. Thus new algorithms need to be developed in order to model the line-shape of the protein Trp emission fluorescence in addition to the maximum wavelength.

MODELING CIRCULAR DICHROISM (CD)

CD is originated from differential absorption of left and right polarized light at a given wavelength. This absorption difference arises from the molecule's chirality and environment. Far-UV CD (below 250nm) is especially sensitive to changes in the secondary structure of proteins and has proven to be a valuable optical technique to study protein structures at equilibrium as well as their time resolved changes [51]. Due to the difficulty that implies the *ab initio* calculation of protein CD spectra, it has often been empirically interpreted by matching measured CD curves with signature bands for secondary structures motifs [111].

The density of CD bands arises from the rotational strength of the chromophores. In proteins the main contributors to the far-UV spectrum are the amides.

Theoretically, the rotational strength is defined as the imaginary part of the scalar product of the electric (μ) and magnetic (m) dipole transition moments of an electronic transition $R = \text{Im}\{\mu \cdot m\}$. In proteins, the interaction between the amides and the environment can induce perturbations on the CD spectra [51]. Since a first principle calculation of the protein CD spectrum is extremely expensive, two major simplified methods have been developed in the past decades: the dipole interaction method [112, 113], and the matrix method [114].

The dipole interaction method is a classical approach developed by DeVoe [112, 113], in which the electronic transitions are considered as oscillators. The CD spectra at a given wavelength can then be calculated as a summation of contributions from all the oscillators in the molecule.

The matrix method is based on quantum mechanics theory, but has often been implemented semi-empirically [114]. To take a simple dipeptide as an example, the main contributions to its CD in the far-UV region are from the electronic transitions of the backbone amides. If two electronic transitions for the amide at 220 nm ($n\pi^*$) and 193nm ($\pi\pi^*$) are considered, a 4 by 4 Hamiltonian matrix can be constructed for this dipeptide containing two amide groups (see Eq. (3)).

$$H = \begin{bmatrix} E_{n\pi^*}^1 & V_{n\pi^*\pi\pi^*}^{11} & V_{n\pi^*n\pi^*}^{12} & V_{n\pi^*\pi\pi^*}^{12} \\ & E_{\pi\pi^*}^1 & V_{\pi\pi^*n\pi^*}^{21} & V_{\pi\pi^*\pi\pi^*}^{22} \\ & & E_{n\pi^*}^2 & V_{n\pi^*\pi\pi^*}^{22} \\ & & & E_{\pi\pi^*}^2 \end{bmatrix} \quad (3)$$

Diagonal terms correspond to the excitation energies of individual amides, while off-diagonal terms describe the interactions between different localized electronic transitions. Diagonal terms can easily be determined from quantum mechanics. Off-diagonal terms are often determined by coulomb interactions from a charge distribution that can describe the localized transitions (e.g. $\pi\pi^*$ transitions can be represented by dipolar charge distributions). These charges can either be determined from quantum mechanics wave functions or from experiments [51]. The rotational strength R can then be obtained from the eigen-decomposition of the Hamiltonian matrix (H). The eigenvalues correspond to the energies of the transitions, while the eigenvectors describe how localized transitions for different amides are coupled. With these coupling coefficients, one can easily determine the overall transition moments (m and μ) from the localized ones (m^0 and μ^0) to calculate R .

Recent studies show that an accurate prediction of CD spectra requires an ensemble of protein conformations [115-122]. For example, Hirst *et al.* [116] shows that better predictions of the β -I and β -II CD spectra can be obtained when many protein conformations from MD simulations are taken into account.

The Mukamel group [120] have recently improved the CD spectra calculation by incorporating the effect of electrostatic fluctuations. In order to take into account of the effect of the electrostatic potential fluctuations on the transition energy, they calculated the interactions between excited states and environment directly by integrating over the product of the transition charge density and electric field. Their results are in good agreement with experiments (see Fig 3), confirming that an accurate description of the electrostatic environment is an important factor to model the CD spectra. The Mukamel group has also proposed that ultra-fast laser UV sources can extend experimental CD studies to the two-dimensional UV. The authors have simulated the 2DUV of several proteins using QM/MM methods. The results suggest that compared to one-dimensional CD, 2DUV is more sensitive to the protein structure change, and may provide more insights on protein dynamics [119, 123].

MODELING INFRARED (IR) SPECTROSCOPY

IR has been widely used to monitor the fast-triggered protein relaxing dynamics [31]. In protein systems, the backbone amide I is the most intensive absorption band, which is originated from stretching vibration of the C=O (70-85%) and C-N groups (10-20%) [124]. The range of the amide I band lies between 1600 and 1700 cm^{-1} , and is highly sensitive to the protein and solvent environment. IR spectra of the single amide I unit has been extensively studied using the N-methylacetamide (NMA) model system [125]. However, different amide I modes are coupled in protein systems, and the protein IR spectra line shape is highly correlated to its secondary structures. Isotope labeling of a particular carbonyl can shift its frequency, and provides a good way to focus on a specific residue [126].

Normal mode analysis provides a straightforward way to model IR spectra. Normal mode analysis on MD simulations can only calculate low frequency slow motions accurately, while IR spectra are mostly determined by high frequency vibrations. Quantum mechanics is able to precisely compute these high-frequency motions, but is too expensive when applying on complex proteins and many conformations. The vibrational frequency map is thus developed by multiple groups [46, 47, 125] in order to obtain amide I frequencies and avoid repeated quantum computations.

The Cho group [48, 125] parameterize the amide I frequencies as a linear function of the electrostatic potentials at the C, O, N, and H and two methyl sites using a least square regression method for normal mode frequency of NMAD-water clusters. The Skinner group [47, 127] has also constructed a map for the NMAD-water system. Using *ab initio* calculations, they parameterized frequencies as a function of electric fields projected on the C, O, N and D atoms. Both groups have related the amide I frequency with electric field projected at several atoms within NMAD, which is due to the surrounding environment. Cho and Skinner's map can provide the fundamental frequency

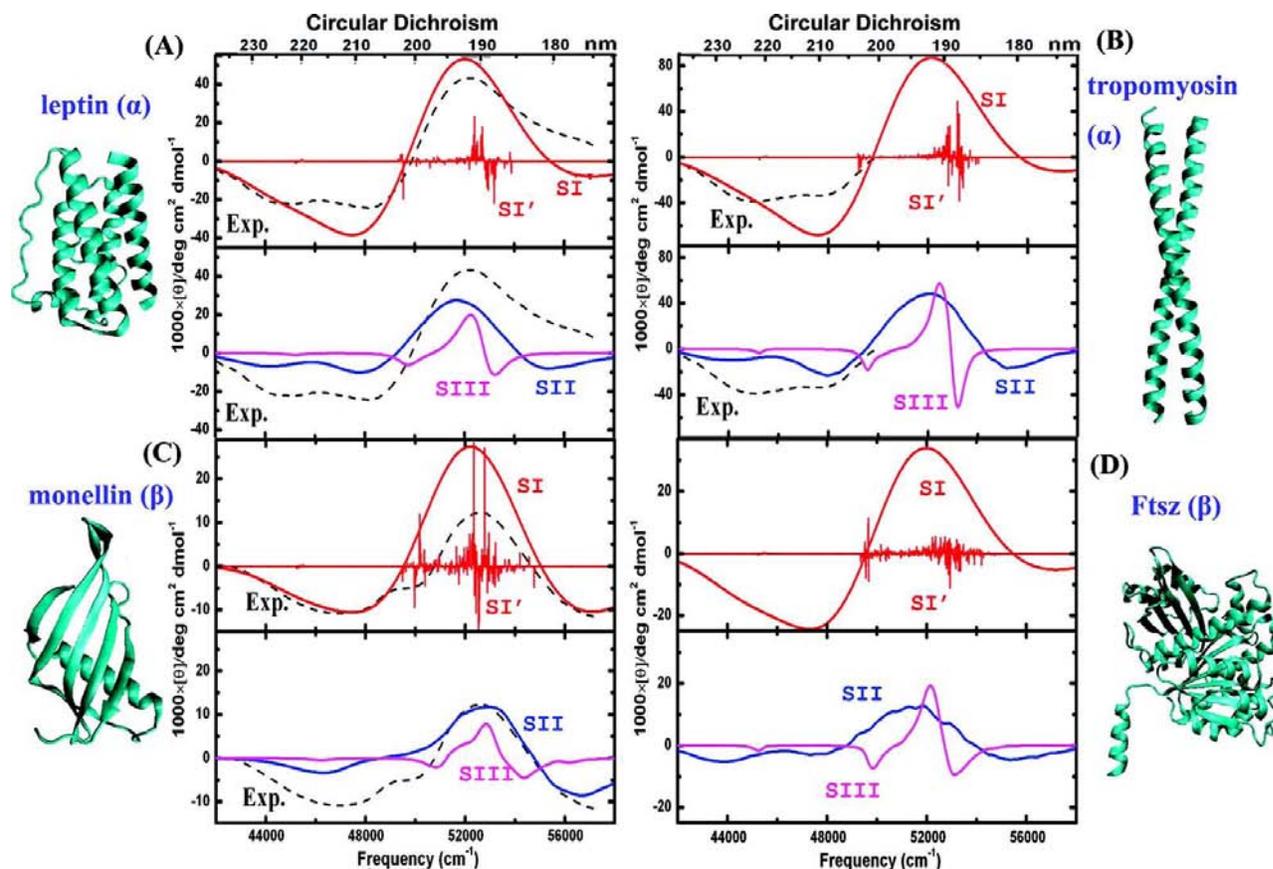


Fig. (3). Simulated and experimental CD spectra are compared for α -helix proteins: leptin (PDB code: 1ax8) and tropomyosin (PDB code: 2d3e), sheet-containing protein: monellin (PDB code: 1mol), and $\alpha\beta$ -protein: FtsZ (PDB code: 1fsz). SI' represents the simulated line spectra based on a single conformation scaled by a factor 1/1000, and SI indicates SI' convoluted with a Gaussian envelope. SII are simulated spectra based on 2000 MD snapshots that consider the electrostatic potential from all surroundings, and SIII takes account of only peptide groups scaled by a factor 1/15. The experimental CD spectrum is shown as dashed black lines. X-ray crystal structures for different proteins are shown in green. Reprinted with permission from Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S., "Ultrafast spectroscopy of protein backbone transitions in aqueous solution: combined QM and MM simulations." *J Phys Chem B* **2010**, *114* (24), 8270. Copyright 2011 American Chemical Society.

fluctuations of amide I, and they used frequency time-correlation function to produce the IR line shape. However, these maps focus on individual amide units, and do not take into account coupling between different amides.

In order to model the coupling between amide units, the exciton model was used to construct a vibrational Hamiltonian operator by various groups [46, 128]. Diagonal terms of the Hamiltonian matrix provide the local vibrational transitions, while off-diagonal terms represent their couplings. Coupling between neighboring amides can be obtained from *ab initio* calculations [129], while coupling between non-neighboring amides was treated with the transition dipole coupling model [124]. Based on this model, the Mukamel group [130] developed a complete map which parameterized the fundamental, overtone, combination frequencies, and transition dipoles of all the amide modes as a quadratic equation from DFT calculations. Jansen and

Knoester [128] later adopted the similar approach to construct a simpler map only for amide I frequencies.

In recent years, two-dimensional infrared spectroscopy (2DIR) has emerged as a powerful tool in the investigation of protein ultrafast dynamics [131-138]. In 2DIR experiment, three laser pulses interact with the protein system to generate a coherent nonlinear signal. 2D correlation plots can be obtained from these experiments, and significantly enhance the spectral resolution. Several methods have been developed to model the 2DIR spectra by numerically integration of the Schrodinger equation (NISE) [139-143] or performing the integration using direct Nonlinear Exciton Propagation (NEP) [144]. Using NISE protocol, Smith *et al.* [138] have modeled isotope edited 2DIR for the trpzip2 peptide using structures from different conformation states generated from a MSM [33]. More recently, Zhuang *et al.* [53] simulated equilibrium 2DIR spectra for the same

peptide by combining the NEP method and MSMs. They have first obtained equilibrium Fourier Transform Infrared (FTIR) spectra at 300K, which are consistent with the experimental results of a folded hairpin [145-147] with a major peak at $\sim 1640\text{cm}^{-1}$ and a minor one at $\sim 1670\text{cm}^{-1}$ (see Fig. 4). More strikingly, they found from simulating 2DIR spectra that MSMs can capture the asymmetrical feature elongated along the diagonal line observed experimentally, while nanosecond MD simulations started from the NMR structure failed to reproduce this inhomogeneous broadening (see Fig. 5). These results indicated that a sufficient sampling of important relevant conformational states is crucial for obtaining the accurate spectroscopic observables.

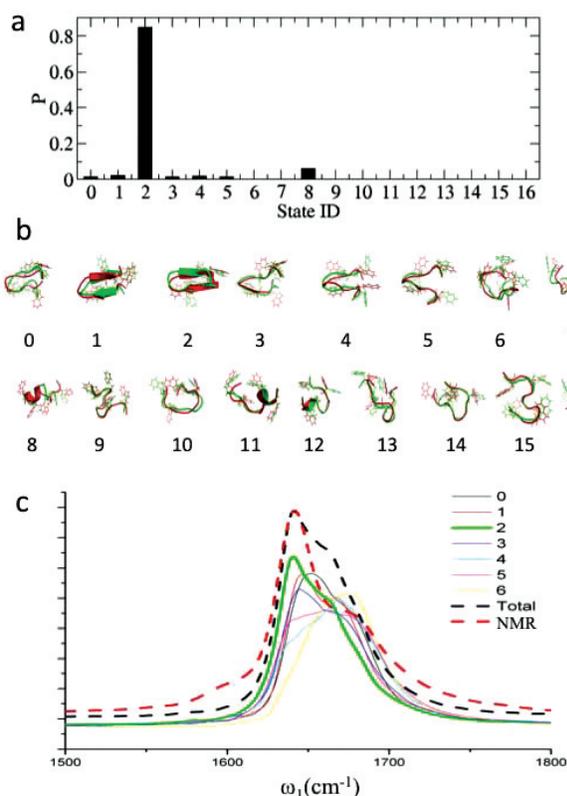


Fig. (4). (a) Equilibrium populations of metastable states at 300K computed from a 17-state MSM are displayed for the trpzip2 peptide. (b) Representative structures from metastable states are shown. (c) The overall 300K equilibrium linear absorption FTIR signal (black dash) and the contributions (solid lines) from states with population $> 1\%$, simulated using MSM data. The overall signal is scaled by 1.2 for a better presentation. Green solid line gives the contribution from the most significant macro state (#2). Red dash line gives the simulated absorption signal based on the NMR structure, and it is scaled to have the same maximum as the MSM overall signal for a better comparison. *Reprinted with permission from Zhuang, W.; Cui, R. Z.; Silva, D.-A.; Huang, X., "Simulating the T-Jump-Triggered Unfolding Dynamics of trpzip2 Peptide and Its Time-Resolved IR and Two-Dimensional IR Signals Using the Markov State Model Approach." The Journal of Physical Chemistry B 2011, 115(18), 5415-5424. Copyright 2011 American Chemical Society.*

DIRECTLY SIMULATING FAST FOLDING SPECTROSCOPY

Atomistic MD simulations can model the dynamics of protein folding through the physical interactions that drive them. Thus, these simulations may provide atomistic protein folding dynamics that is presently inaccessible through experimental observations. However, the great challenge for simulations is the reality check through quantitative experimental validation. Many simulation studies have focused on quantitatively predictions of folding rates, free energies, and structures [21]. For example, Bolhuis [148] obtained a $5\ \mu\text{s}$ folding time for a β -hairpin peptide using transition interface sampling [61], which is in good agreement of the experimental value of $6\ \mu\text{s}$ [149]. Using a deviation of the AMBER-99 force field, Sorin *et al.* [150] have calculated the helix formation rate for a 21-residue helical peptide to be 0.05 to $0.06\ \text{ns}^{-1}$, which agrees well with experimental value of $0.06\ \text{ns}^{-1}$. More recently, Bowman *et al.* [67] have successfully predicted experimental native structure of the villin headpiece as well as the average folding rate. However, the accurate prediction of experimental folding rates or other overall properties does not guarantee that simulations can predict correct atomistic folding mechanisms.

The introduction of fast folding techniques provides new opportunities for quantitative comparison between experiment and theory. Fast triggered relaxation dynamics monitored by optical spectroscopy can capture protein folding or unfolding dynamics from nanoseconds to microseconds. Simulating the relaxation spectroscopy will thus provide theory a quantitative experimental validation not only on overall folding rates, but also on the whole course of folding or unfolding dynamics. In order to connect simulated atomistic structure and dynamics with experimental ensemble averaged spectroscopic observables, two challenges need to be overcome: simulating experimental relevant timescales, and calculating spectroscopic signals. In the previous two sections, we have reviewed recent advances in addressing the above two challenges. In this section, we will review recent studies to directly model the fast folding experiment by combining algorithms to simulate long timescale dynamics with those to calculate spectra signals [21-23].

Kim *et al.* [54] have simulated temperature jump triggered unfolding dynamics of a helical peptide (Alanine 20). In order to simulate the T-jump process from 181K to 214K, they instantly heated conformations extracted from the equilibrium ensemble at 181K generated from REM simulations, and started MD simulations at 214K from these conformations. These 214K MD simulations were then used to simulate non-equilibrium relaxation dynamics after the T-jump. At each time slice of individual 214K MD simulations, CD spectra were calculated using the matrix method [51] and averaged over many conformations. Finally, the authors computed the folding rate by assuming a simple two state kinetics model with folded and unfolded states defined from a cluster analysis. Their calculated CD spectra clearly

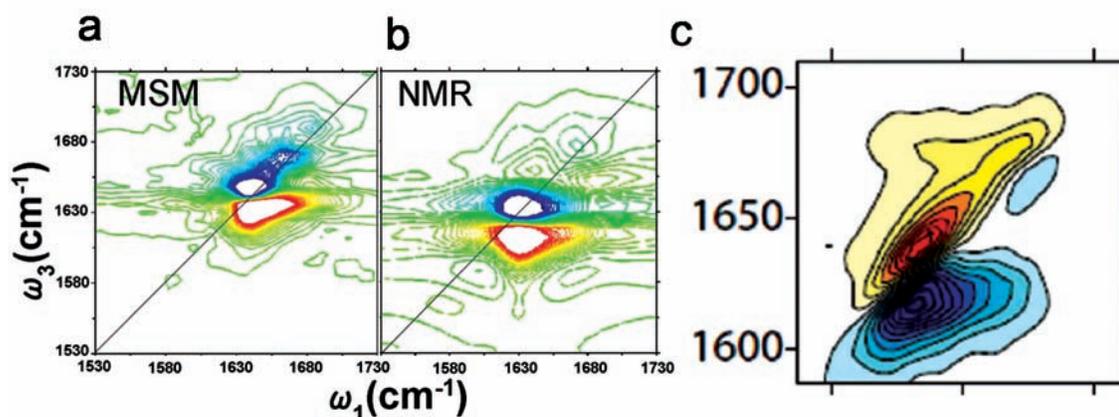


Fig. (5). Simulated absorptive 2DIR spectra for the trpzip2 peptide at 300K [53], computed from the MSM and nanosecond MD simulations starting from the NMR structure are shown in (a) and (b) respectively. In (c), the experimental results [138] at 298K are shown. Part (a) and (b) were reprinted with permission from Zhuang, W.; Cui, R. Z.; Silva, D.-A.; Huang, X., "Simulating the T-Jump-Triggered Unfolding Dynamics of trpzip2 Peptide and Its Time-Resolved IR and Two-Dimensional IR Signals Using the Markov State Model Approach." *The Journal of Physical Chemistry B* **2011**, *115*(18), 5415-5424. Copyright 2011 American Chemical Society. Part (c) was reprinted with permission from Smith, A. W.; Lessing, J.; Ganim, Z.; Peng, C. S.; Tokmakoff, A.; Roy, S.; Jansen, T. L. C.; Knoester, J., "Melting of a beta-Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations." *Journal of Physical Chemistry B* **2010**, *114* (34), 10913-10924. Copyright 2011 American Chemical Society.

showed the melting of the α -helix structure with the increase of the temperature, but did not capture the relatively intensity between the two α -helix signature minimums at 208 and 222 nm. Their predicted folding time (47.5ns) is in reasonable agreement with experimental value for a 21-residue alanine-based peptide (16ns) [151]. However, the experimental temperatures (Jumping from 282.4K to 300.5K) are much higher than the simulation temperatures (Jumping from 181K to 214K). The authors stated that this difference may mainly due to the force field and the implicit solvent model they used.

MSMs provide a natural formalism to make connection between all-atom MD simulations and time-resolved optical spectroscopy. If one can calculate the spectroscopic observable q_i for state i , the overall spectroscopic signal at time $n\Delta t$ is just the weighted average over states ($q = \sum P_i(n\Delta t)q_i$). Therefore, one can obtain the time-resolved spectroscopy from the relaxation dynamics of the state populations $P_i(n\Delta t)$ as defined in Eq. (1).

Noe *et al.* [52] have used the MSM approach to model the T-jump triggered intrinsic Trp fluorescence for the protein PinWW domain. Without involving quantum mechanics, they simply relate the fluorescence intensity to the SASA. Furthermore, they define an observable ($\chi(t)$) to describe the percentage of the ensemble being relaxed to the temperate after the T-jump. $\chi(t)$ can be calculated from the time evolution of the averaged SASA. Finally, they compared $\chi(t)$ with the experimental fluorescence relaxation curve [150] to show good qualitative agreement. Both plots display an initial non-exponential decay followed by a single exponential relaxation at the timescale of tens of μ s (simulation value is 26 μ s and the experimental value is 13 μ s [150]). This is a seminal study by applying MSM to model

fast folding experiments, but has several limitations. Since simulations are only performed at the temperate after the T-jump (360K), the authors have selected a uniform distribution as their initial state populations, while these initial values should be determined by the temperate before the T-jump. Although SASA is a good indicator for the collision fluorescence quenching, it may not be sufficient to describe other processes such as charge transfer to the carbonyl groups of the backbone. Therefore, more sophisticated algorithms are necessary to calculate the fluorescence.

Zhuang *et al.* [53] have recently simulated the T-jump triggered time resolved IR and 2DIR spectroscopy for a 12 residue β -hairpin trpzip2 peptide. They modeled protein unfolding upon the T-jump using MSMs, and calculate IR and 2DIR spectra signal from protein conformations using the NEP method. In this study, they simulated dynamics at temperatures both before (300K) and after (350K) the T-jump with a total of around two thousand 70ns MD simulations. They then simulated the relaxation dynamics at 350K by taking equilibrium state populations at 300K as the initial point. By propagating Eq. (1), they were able to obtain spectra signals at different time points after the T-jump. From the simulated FTIR spectra, they observed a weaker lower frequency peak at 1640 cm^{-1} corresponding to the hairpin structure, while a stronger peak at 1660 cm^{-1} corresponding to the coil structure after the T-jump. This is consistent with the protein unfolding dynamics at the higher temperature (see Fig. (6a)). Moreover, the simulated 2DIR spectra also indicated an unfolding event with more elongated diagonal peak (see Fig. (6b)). This study nicely demonstrated the power of the MSM in modeling fast-triggered optical experiments, and the same protocol may be applied to study other systems in the future.

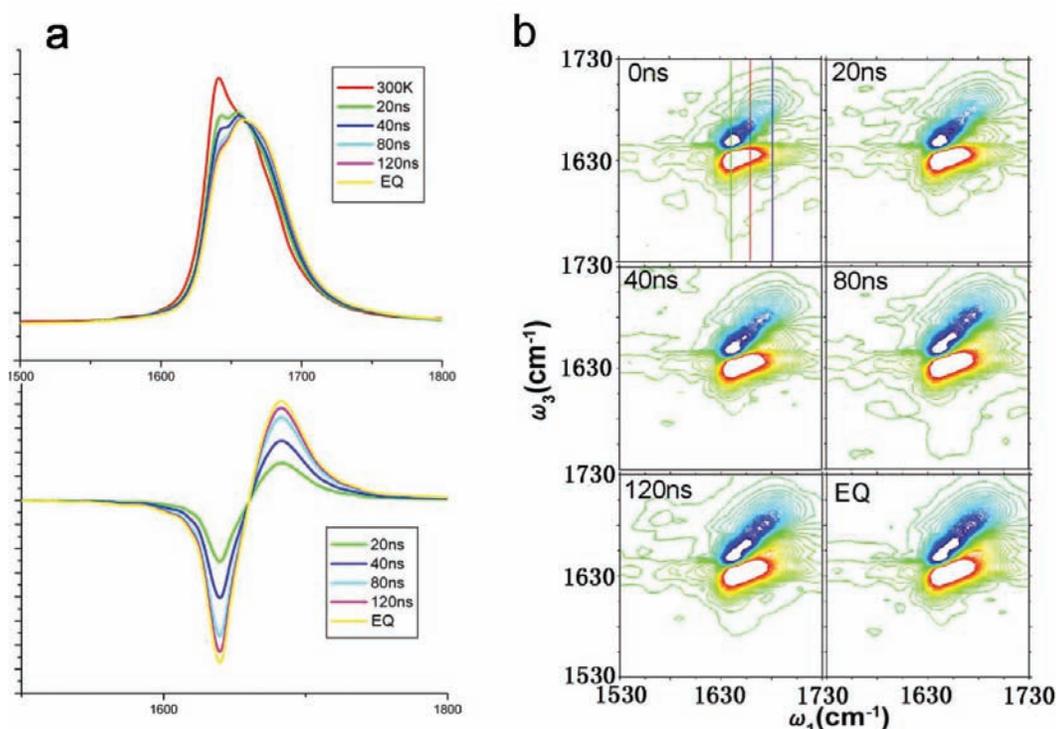


Fig. (6). Simulated Transient FTIR and 2DIR at the different time points after the T-jump are shown in (a) and (b) respectively. For (a), the plots on the lower panel are the difference spectra with the signal before the T-jump at 300K. Reprinted with permission from Zhuang, W.; Cui, R. Z.; Silva, D.-A.; Huang, X., "Simulating the T-Jump-Triggered Unfolding Dynamics of trpzip2 Peptide and Its Time-Resolved IR and Two-Dimensional IR Signals Using the Markov State Model Approach." *The Journal of Physical Chemistry B* **2011**, 115(18), 5415-5424. Copyright 2011 American Chemical Society.

CONCLUSIONS AND FUTURE PERSPECTIVES

With the ability to simulate experimentally relevant timescales and calculate experimental spectroscopic signals, theoreticians begin to have the power to predict not only the overall experimental properties such as folding rate, but also the whole time-resolved relaxation optical spectra. This will not only greatly assist the interpretation of experimental data, but also help to refine the theoretical model. Moreover, linking ensemble averaged spectroscopic observables with atomistic protein dynamics will provide new opportunities for directly assessing different protein folding models. For example, folding dynamics following the native-centric model [152, 153] and the hub-like model [74, 75] may both produce overall folding rate in reasonable agreement with experiments, but quite different spectroscopic relaxation curves.

MSMs provide a good way to simulate the time resolved spectroscopy from the relaxation of the metastable state populations. However, obtaining spectroscopic signals of each state still requires an ensemble average over many protein conformations. Accurately modeling these spectroscopic signals often requires the quantum mechanics in order to properly treat the vibrational or electronic excitations. Therefore, good algorithms should avoid repeated quantum calculations, but still keep a certain level of accuracy. The amide I frequency map approach [47, 48, 125, 128, 130] that has been widely used in modeling IR spectra is one such

example. In the future, these approaches to simulate time resolved spectroscopy for fast protein folding dynamics have a great potential to be widely applied in studying protein misfolding, protein-ligand binding and other related problems.

ACKNOWLEDGEMENTS

X.H. acknowledges the support from Hong Kong Research Grants Council HKUST2/CRF/10, 661011, DAG09/10.SC03, and University Grants Council RPC10SC03. W.Z. acknowledges the support from the support of the NSFC QingNian Grant 21003117 and NSFC Key Grant 21033008.

REFERENCES

- [1] Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H., Jr., The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA*, **1961**, 47, 1309-1314.
- [2] Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R., The protein folding problem. *Annu. Rev. Biophys.*, **2008**, 37, 289-316.
- [3] Gruebele, M., The fast protein folding problem. *Annu. Rev. Phys. Chem.*, **1999**, 50, 485-516.
- [4] Anfinsen, C. B.; Scheraga, H. A., Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.*, **1975**, 29, 205-300.
- [5] Kubelka, J.; Hofrichter, J.; Eaton, W. A., The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.*, **2004**, 14 (1), 76-88.
- [6] Xu, Y.; Purkayastha, P.; Gai, F., Nanosecond folding dynamics of a three-stranded beta-sheet. *J. Am. Chem. Soc.*, **2006**, 128 (49), 15836-15842.

- [7] Maity, H.; Maity, M.; Krishna, M. M.; Mayne, L.; Englander, S. W., Protein folding: the stepwise assembly of foldon units. *Proc. Natl. Acad. Sci. USA*, **2005**, *102* (13), 4741-4746.
- [8] Leopold, P. E.; Montal, M.; Onuchic, J. N., Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*, **1992**, *89* (18), 8721-8725.
- [9] Ceconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S., Direct observation of the three-state folding of a single protein molecule. *Science*, **2005**, *309* (5743), 2057-2060.
- [10] Lei, H.; Wu, C.; Liu, H.; Duan, Y., Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA*, **2007**, *104* (12), 4925-4930.
- [11] Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G., A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.*, **2005**, *346* (4), 1121-1145.
- [12] Matouschek, A.; Kellis, J. T., Jr.; Serrano, L.; Fersht, A. R., Mapping the transition state and pathway of protein folding by protein engineering. *Nature*, **1989**, *340* (6229), 122-126.
- [13] Sosnick, T. R.; Dothager, R. S.; Krantz, B. A., Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *Proc. Natl. Acad. Sci. USA*, **2004**, *101* (50), 17377-17382.
- [14] Zhou, R.; Berne, B. J.; Germain, R., The free energy landscape for beta hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA*, **2001**, *98* (26), 14931-14936.
- [15] Zhang, J.; Li, W.; Wang, J.; Qin, M.; Wu, L.; Yan, Z.; Xu, W.; Zuo, G.; Wang, W., Protein folding simulations: from coarse-grained model to all-atom model. *IUBMB Life* **2009**, *61* (6), 627-43.
- [16] Bunagan, M. R.; Yang, X.; Saven, J. G.; Gai, F., Ultrafast folding of a computationally designed Trp-cage mutant: Trp(2)-cage. *J. Phys. Chem. B*, **2006**, *110* (8), 3759-3763.
- [17] Du, D. G.; Zhu, Y. J.; Huang, C. Y.; Gai, F., Understanding the key factors that control the rate of beta-hairpin folding. *Proc. Natl. Acad. Sci. USA*, **2004**, *101* (45), 15915-15920.
- [18] Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C. Y.; Lahr, S. J.; Daggett, V.; Roder, H.; DeGrado, W. F.; Gai, F., Ultrafast folding of alpha(3): A de novo designed three-helix bundle protein. *Proc. Natl. Acad. Sci. USA*, **2003**, *100* (26), 15486-15491.
- [19] Kubelka, J.; Henry, E. R.; Cellmer, T.; Hofrichter, J.; Eaton, W. A., Chemical, physical, and theoretical kinetics of an ultrafast folding protein. *Proc. Natl. Acad. Sci. USA*, **2008**, *105* (48), 18655-18662.
- [20] Chung, H. S.; Tokmakoff, A., Temperature-dependent downhill unfolding of ubiquitin. I. Nanosecond-to-millisecond resolved nonlinear infrared spectroscopy. *Proteins-Structure Function Bioinformatics*, **2008**, *72* (1), 474-487.
- [21] Snow, C. D.; Sorin, E. J.; Rhee, Y. M.; Pande, V. S., How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.*, **2005**, *34*, 43-69.
- [22] Bowman, G. R.; Voelz, V. A.; Pande, V. S., Taming the complexity of protein folding. *Curr. Opin. Struct. Biol.*, **2011**, *21* (1), 4-11.
- [23] Eaton, W. A.; Munoz, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J., Fast kinetics and mechanisms in protein folding. *Ann. Rev. Biophys. Biomol. Struct.*, **2000**, *29*, 327-359.
- [24] Levitt, M.; Warshel, A., Computer simulation of protein folding. *Nature*, **1975**, *253* (5494), 694-698.
- [25] Chan, H. S.; Dill, K. A., Polymer principles in protein structure and stability. *Annu. Rev. Biophys. Chem.*, **1991**, *20*, 447-490.
- [26] Karplus, M.; Sali, A., Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.*, **1995**, *5* (1), 58-73.
- [27] Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G., Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.*, **1997**, *48*, 545-600.
- [28] Pande, V. S.; Grosberg, A.; Tanaka, T.; Rokhsar, D. S., Pathways for protein folding: is a new view needed? *Curr. Opin. Struct. Biol.*, **1998**, *8* (1), 68-79.
- [29] Shakhnovich, E. I., Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.*, **1997**, *7* (1), 29-40.
- [30] Thirumalai, D.; Klimov, D. K., Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr. Opin. Struct. Biol.*, **1999**, *9* (2), 197-207.
- [31] Gruebele, M.; Sabelko, J.; Ballew, R.; Ervin, J., Laser temperature jump induced protein refolding. *Accounts Chem. Res.*, **1998**, *31* (11), 699-707.
- [32] Noe, F.; Fischer, S., Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.*, **2008**, *18* (2), 154-162.
- [33] Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C., Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.*, **2007**, *126* (15), 155101.
- [34] Buchete, N. V.; Hummer, G., Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B*, **2008**, *112* (19), 6057-6069.
- [35] Swope, W. C.; Pitera, J. W.; Suits, F., Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory. *J. Phys. Chem. B*, **2004**, *108*, 6571-6581.
- [36] Sriraman, S.; Kevrekidis, I. G.; Hummer, G., Coarse master equation from Bayesian analysis of replica molecular dynamics simulations. *J. Phys. Chem. B*, **2005**, *109* (14), 6479-84.
- [37] Bowman, G. R.; Huang, X.; Pande, V. S., Network models for molecular kinetics and their initial applications to human health. *Cell Res.*, **2010**, *20* (6), 622-630.
- [38] Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S., Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.*, **2009**, *30* (6), 864-872.
- [39] Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E., Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*, ACM: Tampa, Florida, 2006; p 84.
- [40] Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y.; Towles, B., Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, ACM: Portland, Oregon, 2009; pp 1-11.
- [41] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **2005**, *26* (16), 1781-802.
- [42] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, **2008**, *4* (3), 435-447.
- [43] Lu, H. S. M.; Volk, M.; Kholodenko, Y.; Gooding, E.; Hochstrasser, R. M.; DeGrado, W. F., Aminothiotyrosine disulfide, an optical trigger for initiation of protein folding. *J. Am. Chem. Soc.*, **1997**, *119* (31), 7173-7180.
- [44] Callender, R.; Dyer, R. B., Probing protein dynamics using temperature jump relaxation spectroscopy. *Curr. Opin. Struct. Biol.*, **2002**, *12* (5), 628-633.
- [45] Roder, H.; Maki, K.; Cheng, H.; Shastry, M. C. R., Rapid mixing methods for exploring the kinetics of protein folding. *Methods*, **2004**, *34* (1), 15-27.
- [46] Zhuang, W.; Hayashi, T.; Mukamel, S., Coherent Multidimensional Vibrational Spectroscopy of Biomolecules: Concepts, Simulations, and Challenges. *Angew. Chem. Int. Edit.*, **2009**, *48* (21), 3750-3781.
- [47] Lin, Y. S.; Shorb, J. M.; Mukherjee, P.; Zanni, M. T.; Skinner, J. L., Empirical Amide I Vibrational Frequency Map: Application to 2D-IR Line Shapes for Isotope-Edited Membrane Peptide Bundles. *J. Phys. Chem. B*, **2009**, *113* (3), 592-602.
- [48] Ham, S.; Cho, M., Amide I modes in the N-methylacetamide dimer and glycine dipeptide analog: Diagonal force constants. *J. Chem. Phys.*, **2003**, *118* (15), 6915-6922.
- [49] Vivian, J. T.; Callis, P. R., Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys. J.*, **2001**, *80* (5), 2093-2109.
- [50] Ceulemans, A.; Moors, S. L. C.; Jonckheer, A.; De Maeyer, M.; Engelborghs, Y., How Do Rotameric Conformations Influence the Time-Resolved Fluorescence of Tryptophan in Proteins? A Perspective Based on Molecular Modeling and Quantum Chemistry. *Curr. Protein Pept. Sc.*, **2008**, *9* (5), 427-446.

- [51] Sreerama, N.; Woody, R. W., Computation and analysis of protein circular dichroism spectra. *Methods Enzymol.*, **2004**, *383*, 318-351.
- [52] Noe, F.; Schutte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R., Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA*, **2009**, *106* (45), 19011-19016.
- [53] Zhuang, W.; Cui, R. Z.; Silva, D.-A.; Huang, X., Simulating the T-Jump-Triggered Unfolding Dynamics of trpzip2 Peptide and Its Time-Resolved IR and Two-Dimensional IR Signals Using the Markov State Model Approach. *J. Phys. Chem. B*, **2011**, *115* (18), 5415-5424.
- [54] Kim, S.; Roitberg, A. E., Simulating temperature jumps for protein folding. *J. Phys. Chem. B*, **2008**, *112* (5), 1525-1532.
- [55] Zhou, R.; Eleftheriou, M.; Royyuru, A. K.; Berne, B. J., Destruction of long-range interactions by a single mutation in lysozyme (vol 104, pg 5824, 2007). *P Natl Acad Sci USA* **2007**, *104* (27), 11507-11507.
- [56] Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K., Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.*, **2007**, *28* (16), 2618-2640.
- [57] Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M., Simulating replica exchange simulations of protein folding with a kinetic network model. *Proc. Natl. Acad. Sci. USA*, **2007**, *104* (39), 15340-15345.
- [58] Rao, F.; Cafilisch, A., The protein folding network. *J. Mol. Biol.*, **2004**, *342* (1), 299-306.
- [59] Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L., Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Ann. Rev. Phys. Chem.*, **2002**, *53*, 291-318.
- [60] Faradjian, A. K.; Elber, R., Computing time scales from reaction coordinates by milestoning. *J Chem Phys* **2004**, *120* (23), 10880-10889.
- [61] Moroni, D.; van Erp, T. S.; Bolhuis, P. G., Investigating rare events by transition interface sampling. *Physica A*, **2004**, *340* (1-3), 395-401.
- [62] Chodera, J. D., Long-time protein folding dynamics from short-time molecular dynamics simulations
- [63] Bowman, G. R.; Huang, X.; Pande, V. S., Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **2009**, *49* (2), 197-201.
- [64] Deuffhard, P.; Weber, M., Robust Perron cluster analysis in conformational dynamics. *Linear Algebra and Its Applications*, **2005**, *398*, 161-184.
- [65] Huang, X.; Yao, Y.; Bowman, G. R.; Sun, J.; Guibas, L. J.; Carlsson, G.; Pande, V. S., Constructing multi-resolution markov state models (msms) to elucidate RNA hairpin folding mechanisms. *Pac. Symp. Biocomput.*, **2010**, 228-239.
- [66] Jayachandran, G.; Vishal, V.; Pande, V. S., Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece. *J. Chem. Phys.*, **2006**, *124* (16), -.
- [67] Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S., Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.*, **2009**, *131* (12), -.
- [68] Morcos, F.; Chatterjee, S.; McClendon, C. L.; Brenner, P. R.; Lopez-Rendon, R.; Zintsmaster, J.; Ercsey-Ravasz, M.; Sweet, C. R.; Jacobson, M. P.; Peng, J. W.; Izaguirre, J. A., Modeling Conformational Ensembles of Slow Functional Motions in Pin1-WW. *Plos Comput. Biol.*, **2010**, *6* (12), -.
- [69] Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S., Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J. Am. Chem. Soc.*, **2010**, *132* (5), 1526-1528.
- [70] Bowman, G. R.; Voelz, V. A.; Pande, V. S., Atomistic folding simulations of the five helix bundle protein lambda6-85. *Submitted for Publication* **2011**.
- [71] Bowman, G. R.; Voelz, V. A.; Pande, V. S., Atomistic folding simulations of the five-helix bundle protein lambda(6-85). *J. Am. Chem. Soc.*, **2011**, *133* (4), 664-667.
- [72] Shirts, M.; Pande, V. S., COMPUTING: Screen Savers of the World Unite! *Science*, **2000**, *290* (5498), 1903-1904.
- [73] E, W.; Vanden-Eijnden, E., Towards a Theory of Transition Paths. *J. Statist. Phys.*, **2006**, *123* (3), 503-523.
- [74] Bowman, G. R.; Pande, V. S., Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. USA*, **2010**, *107* (24), 10890-10895.
- [75] Pande, V. S., Simple theory of protein folding kinetics. *Phys. Rev. Lett.*, **2010**, *105* (19), 198101.
- [76] Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N., New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.*, **1992**, *96* (3), 1776-1783.
- [77] Marinari, E.; Parisi, G., Simulated Tempering: a New Monte Carlo Scheme. *Europhys. Lett.*, **1992**, *19*, 451-458.
- [78] Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, **1999**, *314* (1-2), 141-151.
- [79] Hansmann, U. H.; Okamoto, Y., New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.*, **1999**, *9* (2), 177-183.
- [80] Mohanty, S.; Hansmann, U. H., Folding of proteins with diverse folds. *Biophys. J.*, **2006**, *91* (10), 3573-3578.
- [81] Nymeyer, H.; Garcia, A. E., Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. USA*, **2003**, *100* (24), 13934-13939.
- [82] Liu, P.; Huang, X.; Zhou, R.; Berne, B. J., Hydrophobic aided replica exchange: an efficient algorithm for protein folding in explicit solvent. *J. Phys. Chem. B*, **2006**, *110* (38), 19018-19022.
- [83] Im, W.; Brooks, C. L., 3rd, De novo folding of membrane proteins: an exploration of the structure and NMR properties of the fd coat protein. *J. Mol. Biol.*, **2004**, *337* (3), 513-519.
- [84] Roitberg, A. E.; Okur, A.; Simmerling, C., Coupling of replica exchange simulations to a non-Boltzmann structure reservoir. *J. Phys. Chem. B*, **2007**, *111* (10), 2415-2418.
- [85] Huang, X.; Bowman, G. R.; Pande, V. S., Convergence of folding free energy landscapes via application of enhanced sampling methods in a distributed computing environment. *J. Chem. Phys.*, **2008**, *128* (20), 205106.
- [86] Rhee, Y. M.; Pande, V. S., Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.*, **2003**, *84* (2 Pt 1), 775-786.
- [87] Pitera, J. W.; Swope, W. C.; Abraham, F. F., Observation of noncooperative folding thermodynamics in simulations of 1BBL. *Biophys. J.*, **2008**, *94* (12), 4837-4846.
- [88] Zhang, W.; Wu, C.; Duan, Y., Convergence of replica exchange molecular dynamics. *J Chem Phys.*, **2005**, *123* (15), 154105.
- [89] Periole, X.; Mark, A. E., Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent. *J. Chem. Phys.*, **2007**, *126* (1), 014903.
- [90] Yang, L. J.; Shao, Q.; Gao, Y. Q., Thermodynamics and Folding Pathways of Trpzip2: An Accelerated Molecular Dynamics Simulation Study. *J. Phys. Chem. B*, **2009**, *113* (3), 803-808.
- [91] Zhou, R. H., Trp-cage: Folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. USA*, **2003**, *100* (23), 13280-13285.
- [92] Huang, X.; Bowman, G. R.; Bacallado, S.; Pande, V. S., Rapid equilibrium sampling initiated from nonequilibrium data. *Proc. Natl. Acad. Sci. USA*, **2009**, *106* (47), 19765-9.
- [93] Hinrichs, N. S.; Pande, V. S., Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *J. Chem. Phys.*, **2007**, *126* (24), -.
- [94] Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S., Structural insight into RNA hairpin folding intermediates. *J. Am. Chem. Soc.*, **2008**, *130* (30), 9676-9678.
- [95] Yao, Y.; Sun, J.; Huang, X.; Bowman, G. R.; Singh, G.; Lesnick, M.; Guibas, L. J.; Pande, V. S.; Carlsson, G., Topological methods for exploring low-density states in biomolecular folding pathways. *J. Chem. Phys.*, **2009**, *130* (14), 144115.
- [96] Chen, Y.; Barkley, M. D., Toward understanding tryptophan fluorescence in proteins. *Biochemistry*, **1998**, *37* (28), 9976-9982.
- [97] Chen, Y.; Liu, B.; Yu, H. T.; Barkley, M. D., The peptide bond quenches indole fluorescence. *J. Am. Chem. Soc.*, **1996**, *118* (39), 9271-9278.
- [98] Nanda, V.; Brand, L., Aromatic interactions in homeodomains contribute to the low quantum yield of a conserved, buried tryptophan. *Proteins-Structure Function and Genetics* **2000**, *40* (1), 112-125.
- [99] Gauduchon, P.; Wahl, P., Pulsefluorimetry of Tyrosyl Peptides. *Biophys. Chem.*, **1978**, *8* (1), 87-104.
- [100] Desmet, J.; Demaeyer, M.; Hazes, B.; Lasters, I., The Dead-End Elimination Theorem and Its Use in Protein Side-Chain Positioning. *Nature*, **1992**, *356* (6369), 539-542.

- [101] Permyakov, E. A., *Lumin. Spec. Proteins*, CRC Press: 1993; p 164.
- [102] Marcus, R. A.; Sutin, N., Electron Transfers in Chemistry and Biology. *Biochim. Biophys. Acta*, **1985**, *811* (3), 265-322.
- [103] Hellings, M.; De Maeyer, M.; Verheyden, S.; Hao, Q.; Van Damme, E. J. M.; Peumans, W. J.; Engelborghs, Y., The Dead-End Elimination method, tryptophan rotamers, and fluorescence lifetimes. *Biophys J.*, **2003**, *85* (3), 1894-1902.
- [104] Moors, S. L. C.; Jonckheer, A.; De Maeyer, M.; Engelborghs, Y.; Ceulemans, A., Tryptophan Conformations Associated with Partial Unfolding in Ribonuclease T1. *Biophys. J.*, **2009**, *97* (6), 1778-1786.
- [105] Nilsson, L.; Halle, B., Molecular origin of time-dependent fluorescence shifts in proteins. *Proc. Natl. Acad. Sci. USA*, **2005**, *102* (39), 13867-13872.
- [106] Golosov, A. A.; Karplus, M., Probing polar solvation dynamics in proteins: a molecular dynamics simulation analysis. *J. Phys. Chem. B*, **2007**, *111* (6), 1482-1490.
- [107] Olivucci, M.; Pistolesi, S.; Sinicropi, A.; Pogni, R.; Basosi, R.; Ferre, N., Modeling the Fluorescence of Protein-Embedded Tryptophans with ab Initio Multiconfigurational Quantum Chemistry: The Limiting Cases of Parvalbumin and Monellin. *J. Phys. Chem. B*, **2009**, *113* (49), 16082-16090.
- [108] Beierlein, F. R.; Othersen, O. G.; Lanig, H.; Schneider, S.; Clark, T., Simulating FRET from tryptophan: is the rotamer model correct? *J. Am. Chem. Soc.*, **2006**, *128* (15), 5142-52.
- [109] Hayashi, M.; Mebel, A. M.; Liang, K. K.; Lin, S. H., Ab initio calculations of radiationless transitions between excited and ground singlet electronic states of ethylene. *J Chem Phys* **1998**, *108* (5), 2044-2055.
- [110] Zhao, Y.; Knox, R. S., A Brownian oscillator approach to the Kennard-Stepanov relation. *J. Phys. Chem. A*, **2000**, *104* (33), 7751-7761.
- [111] Johnson, W. C., Jr., Secondary structure of proteins through circular dichroism spectroscopy. *Annu. Rev. Biophys. Biophys. Chem.*, **1988**, *17*, 145-166.
- [112] DeVoe, H., Optical Properties of Molecular Aggregates. I. Classical Model of Electronic Absorption and Refraction. *J. Chem. Phys.*, **1964**, *41* (2), 393-400.
- [113] DeVoe, H., Optical Properties of Molecular Aggregates. II. Classical Theory of the Refraction, Absorption, and Optical Activity of Solutions and Crystals. *J. Chem. Phys.*, **1965**, *43* (9), 3199-3208.
- [114] Bayley, P. M.; Nielsen, E. B.; Schellman, J. A., The rotatory properties of molecules containing two peptide groups: theory. *J. Phys. Chem.*, **1969**, *73* (1), 228-243.
- [115] Daura, X.; Bakowies, D.; Seebach, D.; Fleischhauer, J.; van Gunsteren, W. F.; Kruger, P., Circular dichroism spectra of beta-peptides: sensitivity to molecular structure and effects of motional averaging. *Eur. Biophys. J.*, **2003**, *32* (8), 661-670.
- [116] Hirst, J. D.; Bhattacharjee, S.; Onufriev, A. V., Theoretical studies of time-resolved spectroscopy of protein folding. *Faraday Discuss* **2003**, *122*, 253-67; discussion 269-82.
- [117] Bringmann, G.; Mühlbacher, J.; Regges, C.; Fleischhauer, J., MD-based CD calculations for the assignment of the absolute axial configuration of the naphthylisoquinoline alkaloid dioncophylline A. *J. Comput. Chem.*, **2001**, *22* (12), 1273-1278.
- [118] Kruschel, D.; Zagrovic, B., Conformational averaging in structural biology: issues, challenges and computational solutions. *Mol. Biosys.*, **2009**, *5* (12), 1606-1616.
- [119] Abramavicius, D.; Jiang, J.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S., Simulation study of chiral two-dimensional ultraviolet spectroscopy of the protein backbone. *J Am Chem Soc* **2010**, *132* (22), 7769-75.
- [120] Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S., Ultraviolet spectroscopy of protein backbone transitions in aqueous solution: combined QM and MM simulations. *J Phys Chem B* **2010**, *114* (24), 8270-8277.
- [121] Kaminsky, J.; Kubelka, J.; Bour, P., Theoretical Modeling of Peptide alpha-Helical Circular Dichroism in Aqueous Solution. *J. Phys. Chem. A*, **2011**, *115* (9), 1734-1742.
- [122] Glattli, A.; van Gunsteren, W. F., Are NMR-derived model structures for beta-peptides representative for the ensemble of structures adopted in solution? *Angew. Chem. Int. Ed. Engl.*, **2004**, *43* (46), 6312-6.
- [123] Jiang, J.; Mukamel, S., Two-dimensional near-ultraviolet spectroscopy of aromatic residues in amyloid fibrils: a first principles study. *Phys. Chem. Chem. Phys.* **2011**, *13* (6), 2394-400.
- [124] Krimm, S.; Bandekar, J., vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv. Protein Chem.* **1986**, *38*, 181.
- [125] Kwac, K.; Cho, M. H., Molecular dynamics simulation study of N-methylacetamide in water. I. Amide I mode frequency fluctuation. *J. Chem. Phys.* **2003**, *119* (4), 2247-2255.
- [126] Mukherjee, P.; Krummel, A. T.; Fulmer, E. C.; Kass, I.; Arkin, I. T.; Zanni, M. T., Site-specific vibrational dynamics of the CD3 zeta membrane peptide using heterodyned two-dimensional infrared photon echo spectroscopy. *J. Chem. Phys.*, **2004**, *120* (21), 10215-10224.
- [127] Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L., Ultrafast vibrational spectroscopy of water and aqueous N-methylacetamide: Comparison of different electronic structure/molecular dynamics approaches. *J. Chem. Phys.*, **2004**, *121* (18), 8887-8896.
- [128] Jansen, T. L.; Knoester, J., A transferable electrostatic map for solvation effects on amide I vibrations and its application to linear and two-dimensional spectroscopy. *J. Chem. Phys.*, **2006**, *124* (4).
- [129] 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 Spec.*, **1998**, *29* (1), 81-86.
- [130] Hayashi, T.; Zhuang, W.; Mukamel, S., Electrostatic DFT map for the complete vibrational amide band of NMA. *J. Phys. Chem. A*, **2005**, *109* (43), 9747-9759.
- [131] Scheurer, C.; Pryatinski, A.; Mukamel, S., Signatures of β -peptide Unfolding in Two-Dimensional Vibrational Echo Spectroscopy: A Simulation Study. *J. Am. Chem. Soc.*, **2001**, *123*, 3114.
- [132] Chung, H. S.; Khalil, M.; Smith, A. W.; Ganim, Z.; Tokmakoff, A., Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin. *Proc. Natl. Acad. Sci. USA*, **2005**, *102* (3), 612-617.
- [133] Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P., Watching hydrogen-bond dynamics in a beta-turn by transient two-dimensional infrared spectroscopy. *Nature*, **2006**, *444* (7118), 469-472.
- [134] Shim, S. H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T., Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution. *Proc. Natl. Acad. Sci. USA*, **2009**, *106* (16), 6614-6619.
- [135] Bagchi, S.; Falvo, C.; Mukamel, S.; Hochstrasser, R. M., 2D-IR Experiments and Simulations of the Coupling between Amide-I and Ionizable Side Chains in Proteins: Application to the Villin Headpiece. *J. Phys. Chem. B*, **2009**, *113* (32), 11260-11273.
- [136] Jansen, T. L. C.; Knoester, J., Two-dimensional infrared population transfer spectroscopy for enhancing structural markers of proteins. *Biophys. J.*, **2008**, *94* (5), 1818-1825.
- [137] Roy, S.; Jansen, T. L. C.; Knoester, J., Structural classification of the amide I sites of a beta-hairpin with isotope label 2DIR spectroscopy. *Phys. Chem. Chem. Phys.*, **2010**, *12* (32), 9347-9357.
- [138] Smith, A. W.; Lessing, J.; Ganim, Z.; Peng, C. S.; Tokmakoff, A.; Roy, S.; Jansen, T. L. C.; Knoester, J., Melting of a beta-Hairpin Peptide Using Isotope-Edited 2D IR Spectroscopy and Simulations. *J. Phys. Chem. B*, **2010**, *114* (34), 10913-10924.
- [139] Jansen, T. L.; Knoester, J., Nonadiabatic effects in the two-dimensional infrared spectra of peptides: Application to alanine dipeptide. *J. Phys. Chem. B*, **2006**, *110* (45), 22910-22916.
- [140] Jansen, T. L.; Zhuang, W.; Mukamel, S., Stochastic Liouville equation simulation of multidimensional vibrational line shapes of trialanine. *J. Chem. Phys.*, **2004**, *121* (21), 10577-10598.
- [141] Torii, H., Effects of intermolecular vibrational coupling and liquid dynamics on the polarized Raman and two-dimensional infrared spectral profiles of liquid N,N-dimethylformamide analyzed with a time-domain computational method. *J. Phys. Chem. A*, **2006**, *110* (14), 4822-4832.
- [142] Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.*, **2008**, *128* (22), 224511.
- [143] Jansen, T. L. C.; Ruszel, W. M., Motional narrowing in the time-averaging approximation for simulating two-dimensional nonlinear infrared spectra. *J. Chem. Phys.*, **2008**, *128* (21), 214501.

- [144] Falvo, C.; Palmieri, B.; Mukamel, S., Coherent infrared multidimensional spectra of the OH stretching band in liquid water simulated by direct nonlinear exciton propagation. *J. Chem. Phys.*, **2009**, *130* (18).
- [145] Wang, J. P.; Chen, J. X.; Hochstrasser, R. M., Local structure of beta-hairpin isotopomers by FTIR, 2D IR, and ab initio theory. *J. Phys. Chem. B*, **2006**, *110* (14), 7545-7555.
- [146] Smith, A. W.; Tokmakoff, A., Amide I two-dimensional infrared spectroscopy of beta-hairpin peptides. *J. Chem. Phys.*, **2007**, *126* (4).
- [147] Wang, T.; Xu, Y.; Du, D. G.; Gai, F., Determining beta-sheet stability by Fourier transform infrared difference spectra. *Biopolymers* **2004**, *75* (2), 163-172.
- [148] Bolhuis, P. G., Transition-path sampling of beta-hairpin folding. *Proc. Natl. Acad. Sci. USA*, **2003**, *100* (21), 12129-12134.
- [149] Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A., Folding dynamics and mechanism of beta-hairpin formation. *Nature*, **1997**, *390* (6656), 196-199.
- [150] Sorin, E. J.; Pande, V. S., Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.*, **2005**, *88* (4), 2472-2493.
- [151] Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B., Fast events in protein folding: Helix melting and formation in a small peptide. *Biochemistry*, **1996**, *35* (3), 691-697.
- [152] Brooks, C. L.; Hills, R. D., Insights from Coarse-Grained Go Models for Protein Folding and Dynamics. *Int. J. Mol. Sci.*, **2009**, *10* (3), 889-905.
- [153] Onuchic, J. N.; Wolynes, P. G., Theory of protein folding. *Curr. Opin. Struc. Biol.*, **2004**, *14* (1), 70-75.