MOLECULAR MODELLING AND SIMULATION: FROM SOFT CONDENSED MATTER TO BIOLOGICAL SYSTEMS

Phuong Nguyen

UNIVERSITÉ PARIS-SUD

MÉMOIRE

Présenté pour obtenir

Habilitation à Diriger des Recherches

par

Phuong Hoang NGUYEN

MOLECULAR MODELLING AND SIMULATION: FROM SOFT CONDENSED MATTER TO BIOLOGICAL SYSTEMS

Soutenance 07 September 2016

Jury

Rapporteur :	**Daniel BORGIS**
Rapporteur :	**Tap HA DUONG**
Rapporteur :	**Human REZAEI**
Examinateur:	**Pierre TUFFÉRY**
Examinateur:	**Konrad HINSEN**
Examinateur:	**Luca MONTICELLI**

Contents

1 Introduction

In this $memoire$ I present a summary of my research activity and an overview of my future projects. In the first chapter I present a detailed Curriculum Vitae reporting my education, my teaching experiences as well as my academic achievements. The second chapter is dedicated to my research activity which is divided into three main career periods. A list of publications and selected oral presentations are also included. The future research project is presented in the third chapter. The teaching activity and the administrative activity are presented in the chapters 4 and 5.

Here I briefly introduce the body of my memoire. I had been a PhD student for about three years and about nine years for the post-doc. Then I spent about two years as a temporary principal investigator before being hired as CR1 researcher at CNRS and join the Laboratoire de Biochimie Théorique at IBPC. During this long period I have had opportunities to challenge myself in different fields of research and to acquire deep expertise in computer simulations as well as to be exposed to various scientific environments.

During my PhD work in the field of soft condensed matter simulation, I had been trained in simulation of coarse-grained liquid crystalline systems using classical Monte Carlo technique. This required intensive coding of the simulation programs and analytical derivations related to the liquid state density functional theory. I started to learn biomolecules, specifically protein and RNA during my postdoc. During this time, I had been trained in equilibrium and nonequilibrium molecular dynamics simulations of atomistic protein models. The experience in both coarse-grained and atomistic simulations was the key for acquiring expertise in my current research activity which focuses on the development and application of computational methods for studying the equilibrium and nonequilibrium structure, dynamics, thermodynamics of single proteins and amyloids at all-atom and coarse-grained levels. Personally, my idea in the development is that the developed methods should be as simple as possible, yet efficient. Also, I am more interested in principals governing the system rather than going into details.

During my PhD and postdoc I had been involved in supervising master and PhD students. I had also been regularly involved in teaching activity. In particular, I was in charge of the tutoring associated to the Mathematics course for undergraduate students for eight years of my postdoc. Since this year, I have been involved in giving some lectures on advanced topics of computer simulations such as simulated tempering, metadynamics as well as weighted histogram analysis method and disconnectivity graph.

I would like to conclude this short introduction spending a few words on those extra-academic activities. I review regularly manuscripts for the American Chemical Society, the American Institute of Physics and the Royal Society of Chemistry.

2 Curriculum Vitae

Personal Information

Dr. Phuong Hoang NGUYEN (15.03.1974, Vietnam)
Laboratoire de Biochimie Théorique, UPR 9080, CNRS
IBPC, 13 rue Pierre et Marie Curie, 75005, Paris
Phone: +33 (01) 58 41 51 81
Email: phuong.nguyen@ibpc.fr

University Education

1992-1996: B.Sc student, Physics Department, Hue University of Sciences, Vietnam.
B.Sc. exam: 24 June 1996 (first class).

1997-1998: Diploma student, Condensed Matter Theory Section, International Centre for Theoretical Physics, Trieste, Italy.
Diploma thesis: "Ab-initio Spin Polarized Study of the Ideal α-Sn(111) Surface."
Supervisor: Dr. Sandro Scandolo
Diploma degree: 25 Septembre 1998.

1998-2002: Ph.D student at Physics Department, University of Mainz (Nov 1998-May 1999), Theory Department, Max-Planck Institute for Polymer Research (Jun 1999- Jun 2000), Physics Department, University of Bielefeld (Jun 2000-Jan 2002), Germany.
Ph.D thesis: "Structure and Elasticity of Nematic and Isotropic Liquid Crystals"
Research supervisor: Prof. Friederike Schmid
Ph.D degree: 31 January 2002.

Professional Career

2002-2009: Postdoct, Institute of Physical and Theoretical Chemistry, Frankfurt University, Germany.
Advisor: Prof. Gerhard Stock
2009-2010: PI, Institute of Physical and Theoretical Chemistry, Frankfurt University, Germany
2011-present: CNRS scientist (CR1), Institut de Biologie et Physico-Chimique, Paris, France

Teaching Experiences

1997: Lecturer in General Physics, Hue University of Sciences, Vietnam (45 hours)
2001: Teaching assistant in Statistical Thermodynamics, Bielefeld University, Germany (one semester)
2002-2008: Teaching assistant in Mathematics for Natural Sciences Students, Frankfurt University, Germany (every semester)
2002: Teaching assistant in Statistical Thermodynamics and Kinetic, Frankfurt University, Germany (one semester)
2003: Teaching assistant in Chemical Binding, Molecular Spectroscopy, Frankfurt University, Germany (one semester)
2007: Teaching assistant in Theoretical Chemistry, Frankfurt University, Germany (one semester)
2015: Lecturer in Advanced Methods in Molecular Dynamics Simulation, University Paris VII, France (10 hours)

Technical Skills

Programming languages: C, C++, Fortran 77/90, shell scripting, Library MPI and OpenMPI for parallel computing.
Operating systems: Unix/Linux, Mac OSX, Window
Simulation techniques: Classical Molecular Dynamics, Monte Carlo, *ab initio* Molecular Dynamics

Software for molecular simulations: Gromacs, Gaussian, Amber
Graphic Software: Rasmol, VMD, Pymol, Pov-Ray, Matlab, Xmgrace, Gnuplot

Working Experiences

Co-supervisor of 2 Ph.D students and 1 B.Sc student (2002-2008, University of Frankfurt).
Co-supervisor of 1 Master student (2015, Universié Paris-Sud)
Co-supervisor of 1 Ph.D student (2013-2016, Universié Paris 7)

Academic achievements

1994-1995: Scholarship of the Australian Embassy in Vietnam for: "In recognition of past academic achievement and with the promise of future education attainment"
1996: 2nd prize "National Student Research Competition" and "Diploma of Merit" awarded by Vietnamese Minister of Education and Training.
2009-2010: Individual grant from the German Research Foundation for the Principal Investigator
2013-2014: Research grant from the Institute of Computational Science and Technology, Vietnam

Reviewer activity

Regular reviewer for the scientific journals including J. Am. Chem. Soc., J. Phys. Chem. B, J. Chem. Phys., Phys. Chem. Chem. Phys, Nanoscale, Chem. Phys, Lett., RSC Advances, Int. J. Mol. Sci., Biochemistry, Biophysics Report, Expert Opinion on Drug Discovery, ACS Chemical Neuroscience.

Conferences and Publications

62 published and 3 submitted papers, 3 books chapters
1 invited talk at a International Conference, 10 talks at Conferences, 4 seminars.

Extra-academic activity

Member de conseil de Laboratoire de Biochimie Theorique, UPR 9080 CNRS

3 Research Activities

3.1 PhD research

My PhD research focuses on the structure and elasticity properties of liquid crystals. Nowadays, the applications of liquid crystals can be found everywhere, from the laptop screen, digital watch, switchable windows, smartphone screen etc. Liquid crystal phases are the phases between the solid and liquid phases. They most often occur in compounds of the rod-like shape or the disc-like shape organic molecules which interact through strongly anisotropic intermolecular forces. There are three main phases of liquid crystals: nematic, smectic and columnar phases. Nematic phase is the simplest phase where the molecules tend to be parallel to a common direction called the director. In this phase, there is no translational order but the orientational order is apparently[1]. This broken symmetry of space leads to many interesting physics properties. My PhD work focuses on two properties that are long range structure and elasticity of the nematic phase.

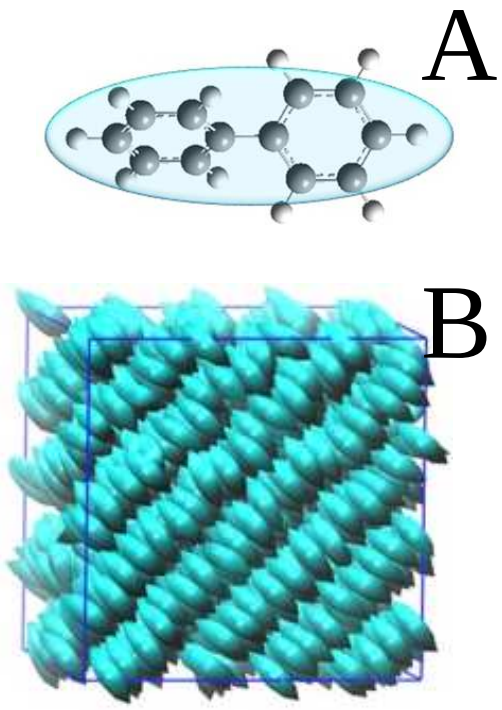

Figure 1: *(A) An example diphenyl crystalline molecule is represented by an ellipsoid particle; (B) a typical equilibrium configuration of the crystalline phase.*

3.1.1 Correlation functions

We construct a coarse-grained model for the liquid crystal, where each molecule is represented by an ellipsoid particle [Fig.1(A)]. The interaction between particles are modeled by a simple repulsive pair potential

$$V_{12} = \begin{cases} 4\epsilon_0(X_{12}^{12} - X_{12}^{6}) + \epsilon_0 : X_{12}^{6} > 1/2, \\ 0 : \text{otherwise} \end{cases} \tag{1}$$

The function $X_{12} = \sigma_0/(r_{12} - \sigma_{12} + \sigma_0)$ depends on the distance r_{12} between particles 1 and 2, and on the shape function

$$\sigma_{12}(\mathbf{u_1}, \mathbf{u_2}, \hat{\mathbf{r}_{12}}) = \sigma_0 \Big\{ \mathbf{1} - \frac{\kappa}{\mathbf{2}} \Big[\frac{(\mathbf{u_1}.\hat{\mathbf{r}}_{12} + \mathbf{u_2}.\hat{\mathbf{r}}_{12})^2}{\mathbf{1} + \kappa \mathbf{u_1}.\mathbf{u_2}} + \frac{(\mathbf{u_1}.\hat{\mathbf{r}}_{12} - \mathbf{u_2}.\hat{\mathbf{r}}_{12})^2}{\mathbf{1} - \kappa \mathbf{u_1}.\mathbf{u_2}} \Big] \Big\} \tag{2}$$

which approximates the contact distance between two ellipsoids of elongation $\kappa = \sqrt{(1+\chi)(1-\chi)}$ with orientations $\mathbf{u_1}$ and $\mathbf{u_2}$ and center - center vector $\mathbf{r}_{12}$ pointing in the direction $\hat{\mathbf{r}}_{12}$ =$\mathbf{r}_{12}$/ r_{12}[2]. Having defined the potential, we carry out Monte Carlo (MC) simulations for the isotropic phase (low density $\rho = 0.24/\sigma_0^3$) and the nematic phase (high density $\rho = 0.30/\sigma_0^3$). Fig.1(B) shows a snapshot of the system of ellipsoids in the nematic phase. Our primary aim is to study the influence of the isotropic/nematic phase transition on the direct correlation functions (DCF) in anisotropic fluids. The DCF is a central quantity in liquid state theories[3], and a study in the nematic phase is therefore interesting in its own right. The DCF $c(\mathbf{u}_1, \mathbf{u}_2, \mathbf{r}_{12})$ is determined from the pair distribution functions by solving the full Ornstein-Zernike equation[3]

$$h(\mathbf{u}_1, \mathbf{u}_2, \mathbf{r}_{12}) = c(\mathbf{u}_1, \mathbf{u}_2, \mathbf{r}_{12}) + \int c(\mathbf{u}_1, \mathbf{u}_3, \mathbf{r}_{13}) \rho^{(1)}(\mathbf{u}_3) h(\mathbf{u}_3, \mathbf{u}_2, \mathbf{r}_{32}) d\mathbf{u}_3 d\mathbf{r}_3, \tag{3}$$

The total correlation function h, which describes the total interaction between two particles, and the one-particle distribution function $\rho^{(1)}$, which is the probability for finding a particle, are determined directly from the simulation data. The DCF has been studied quite intensely in isotropic fluids. In contrast, the determination of this function in the nematic phase is quite challenge. As mentioned above, in a nematic liquid crystal, the particles remain positionally disordered, but align preferentially along the director vector $\mathbf{n}(\mathbf{r})$. Therefore, the isotropy of space is spontaneously broken. Consequently, the pair correlation functions lose their rotational invariance. This means that the pair correlation $F(\mathbf{u}_1, \mathbf{u}_2, \mathbf{r}_{12})$ (F stands for h or c) must depend not only on the orientation of each particle $\mathbf{u}_i$

(like in the isotropic phase) but also on the center - center vector $\mathbf{r}_{12}$. The latter makes the calculation of the function F to be much more difficult. One of the first tasks of my thesis is to solve this issue, and below I briefly describe the method which I developed to calculate pair correlation functions in the simulated nematic phase. Details of the method can be found in Ref.[4]. First, a Cartesian coordinate frame where the z axis points in the direction of the director vector $\mathbf{n}(\mathbf{r})$ is constructed, and all orientation-dependent functions are expanded in spherical harmonics $Y_{lm}(\mathbf{u})$. This yields

$$\rho^{(1)} = \rho \sum_{l\text{even}} f_l Y_{l0}, \tag{4}$$

with the bulk number density ρ, and

$$F(\mathbf{u}_1, \mathbf{u}_2, \mathbf{r}) = \sum_{l_1,l_2,l,m_1,m_2,m} F_{l_1 m_1 l_2 m_2 lm}(r) Y_{l_1 m_1}(\mathbf{u_1}) Y_{l_2 m_2}(\mathbf{u_2}) Y_{lm}(\hat{\mathbf{r}}). \tag{5}$$

In uniaxially symmetric phases, only real coefficients with $m_1 + m_2 + m = 0$ and even $l_1 + l_2 + l$ enter the expansion. Since our particles have uniaxial symmetry, every single l_i is even as well. The total correlation function h is defined as

$$h(\mathbf{u}_1, \mathbf{u}_2, \mathbf{r}) = \frac{\rho^(2)(\mathbf{u}_1, \mathbf{u}_2, \mathbf{r})}{\rho^{(1)}(\mathbf{u_1})\rho^{(1)}(\mathbf{u_2})} - 1, \tag{6}$$

where $\rho^(2)(\mathbf{u}_1, \mathbf{u}_2, \mathbf{r})$ is the pair distribution function. In terms of the spherical harmonic expansion, Eq.6 is rewritten as

$$\rho^{(2)}_{l_1,m_1,l_2,m_2,l,m}(r) = \rho^2 \Big(\sqrt{4\pi} f_{l_1} f_{l_2} \delta_{m_1 0} \delta_{m_2 0} \delta_{l0} \delta_{m0} + \sum_{l_1',l_1'',l,l_2',l_2''} h_{l_1' m_1 l_2' m_2 lm}(r) f_{l_1''} f_{l_2''} \Gamma^{l_1 l_1' l_1''}_{m_1 m_1 0} \Gamma^{l_2 l_2' l_2''}_{m_2 m_2 0} \Big), \tag{7}$$

with

$$\Gamma^{ll'l''}_{mm'm''} = \int d\mathbf{u} Y^*_{lm}(\mathbf{u}) Y^*_{l'm'}(\mathbf{u}) Y^*_{l''m''}(\mathbf{u}), \tag{8}$$

and the coefficients of the pair distribution are determined directly from the simulation data

$$\rho^{(2)}_{l_1,m_1,l_2,m_2,l,m}(r) = 4\pi\rho^2 g(r) \langle Y^*_{l_1 m_1}(\mathbf{u_1}) Y^*_{l_2 m_2}(\mathbf{u_2}) Y^*_{lm}(\mathbf{u}) \rangle_{\delta r}, \tag{9}$$

where the average $\langle \cdots \rangle_{\delta r}$ is performed over all molecule pairs at distances $|r_1 - r_2| = [r, r + \delta r]$, and $g(r)$ is the radial distribution function, i.e., the average total number of molecule pairs divided by $N 4\pi\rho^2 \delta r$. Then, the coefficients of the total correlation function h are determined by solving the matrix Eq.7. The DCF c is most easily calculated in Fourier representation. Therefore, we perform a Hankel transformation of all functions in Eq.3 and obtain a new representation in the k-space

$$\begin{aligned} h_{l_1,m_1,l_2,m_2,l,m}(k) &= c_{l_1,m_1,l_2,m_2,l,m}(k) + \rho \sum_{\substack{l_3,l_3',l_3'',m_3,\\ l'm'l''m''l_3}} c_{l_1,m_1,l_3,m_3,l',m'}(k) h_{l_3',m_3,l_2,m_2,l'',m''}(k) \\ &\times \; f_{l_3''} (-1)^{m_3} \Gamma^{ll'l''}_{mm'm''} \Gamma^{l_3 l_3' l_3''}_{m_3 m_3 0}. \end{aligned} \tag{10}$$

The coefficients of the DCF are obtained by solving the matrix equation 10. Fig.2 shows an example of an expansion coefficient, c_{202020} , which vanishes in the case of rotational invariance. In the isotropic phase ($\rho \leq 0.283\sigma_0^{-3}$), the values are indeed close to zero. In the nematic phase, a nonzero DCF coefficient not only emerges, it may even grow quite large. At $\rho = 0.3\sigma_0^{-3}$ the values are comparable to those of a symmetry preserving coefficient of similar order. Hence the DCF really reflects the broken symmetry in the nematic phase. To summarize, I have developed a method to calculate DCF without any approximations, and applied it to study the local structure of fluids of uniaxial ellipsoids at different densities in the vicinity of the nematic/isotropic phase transition. We find that the DCF suitably characterizes the local structure, in the sense that it is short ranged both in the isotropic and the nematic phase.

3.1.2 Elastic constants

We then use the DCF to calculate the elastic constants of the nematic phase. Due to the long range orientational order breaks a continuous symmetry, the isotropy of space, there exist soft fluctuation modes-spatial variations of the director $\mathbf{n}(\mathbf{r})$ - which cost no energy in the infinite wavelength limit (i.e., the limit where n is rotated uniformly) and are otherwise penalized by elastic restoring forces[5]. For symmetry reasons, the latter depend on only three material parameters at large finite wavelengths[1], which are described by an elastic free energy functional[6]

$$F\{\mathbf{n}(\mathbf{r})\} = \frac{1}{2}\int d\mathbf{r}\{K_{11}[\bigtriangledown.\mathbf{n}]^2 + K_{22}[\mathbf{n}.(\bigtriangledown\times\mathbf{n}]^2 + K_{33}[\mathbf{n}\times(\bigtriangledown\times\mathbf{n})]^2\}, \quad (11)$$

which has three contributions: the splay, twist, and bend modes. The parameters $K_{\alpha\alpha}$ ($\alpha = 1, 2, 3$), called Frank elastic constants, control almost exclusively the structure and the properties of nematic liquid crystals at mesoscopic length scales. It has been derived in Ref.[7], these elastic constants are expressed in terms of the DCF in the k-space as the following

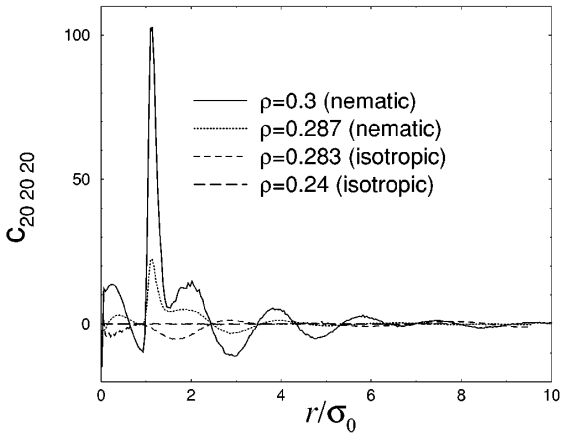

Figure 2: *Expansion coefficient with $l_i = l = 2$ and $m_i = m = 0$ of the direct correlation function in the molecular frame vs r for different densities. In the isotropic phase, this coefficient must vanish for symmetry reasons. At $\rho = 0.283\sigma_0^{-3}$, a small nonzero remnants are observed due to finite size effects*

$$K_{11} = -\frac{k_BT}{2}\int \frac{\partial^2 c(\mathbf{k},\mathbf{u_1},\mathbf{u_2})}{\partial k_x^2}|_{\mathbf{k}=0}\times\rho^{(1)'}(u_{1z})\rho^{(1)'}(u_{2z})u_{1x}u_{2x}d\mathbf{u_1}d\mathbf{u_2}$$
$$K_{22} = -\frac{k_BT}{2}\int \frac{\partial^2 c(\mathbf{k},\mathbf{u_1},\mathbf{u_2})}{\partial k_x^2}|_{\mathbf{k}=0}\times\rho^{(1)'}(u_{1z})\rho^{(1)'}(u_{2z})u_{1y}u_{2y}d\mathbf{u_1}d\mathbf{u_2}$$
$$K_{33} = -\frac{k_BT}{2}\int \frac{\partial^2 c(\mathbf{k},\mathbf{u_1},\mathbf{u_2})}{\partial k_z^2}|_{\mathbf{k}=0}\times\rho^{(1)'}(u_{1z})\rho^{(1)'}(u_{2z})u_{1x}u_{2x}d\mathbf{u_1}d\mathbf{u_2}. \quad (12)$$

We calculate the DCF from the pair distribution function $\rho^{(2)}$ using an upper cutoff $l_{\max} = 2, 4$, and 6, respectively, in the matrix equations 7 and 10, and the elastic constants are calculated from Eq.12. Already the lowest order calculation with $l_{\max} = 2$ gave elastic constants of the correct order of magnitude. Quantitatively reliable results were obtained with $l_{\max} \geq 6$. In summary, we have presented a reliable method to calculate elastic constants from pair distributions in computer simulations.

3.2 Postdoc research

After my PhD in condensed matter physics, I changed my research field to biophysical chemistry with a focus on proteins. With this, I started to learn molecular dynamics (MD) simulation techniques using protein all-atom models. My first projects were the development and application of computational strategies that allows us to extend well-established MD techniques to the description of photoinduced nonequilibrium dynamics simulations using standard MD program packages. This allows us to study the energy transport[8, 9, 10] and ultrafast conformational dynamics[11, 12, 13] of large systems, which are otherwise very difficult to study by means of semiclassical or quantum molecular dynamics methods. To quantitatively compare with picosecond pump-probe Infrared (IR) experiments, I am also involved in the development of accurate methods to simulate 1D and 2D IR spectra of peptides[14, 15, 16]. I am also interested in equilibrium fast dynamics such as folding of small peptides. Along this direction, I develop methods to reduce the complexity of peptides in order to capture essential physics of folding pathways[17, 18, 19, 20, 21, 22], and collaborate with NMR experiment to determine accurately the structure of small peptides[23]. Finally, I start to work on the protein aggregation to understand the fibril growth mechanism[24], and on RNA systems aiming to characterize the structure RNA hairpins[25] and long-range correlations in ligand-binding riboswitch[26]. Below, I summary some main results in details.

3.2.1 Energy transfer in peptides

(a) Vibrational energy relaxation. In recent years, femtosecond time-resolved infrared spectroscopy has made it possible to watch the flow of vibrational energy within and between molecules with unprecedented time resolution[27]. Particular interest has been payed to biomolecular systems, since the celebrated concept of the relation between protein structure, dynamics, and function ultimately requires an microscopic understanding of the energy flow in the system. Employing the amide I modes (which mainly involve the stretching of the peptide C= O bond) as a conformational probe, experiments have provided new insight into the structure, fluctuations, and conformational transitions of peptides in aqueous solution. Interestingly, the experiments have shown that the $\nu = 1 \rightarrow 0$ population decay time T_1 of the amide I mode is consistently about 1 ps for all systems considered. This includes the model systems N-methylacetamid[28], and trialanine[29], various small globular peptides such as apamin, sylla-toxin, and bovine pancreatic tripsin inhibitor[28], and the protein myoglobin[30]. These findings indicate that the amide I population relaxation in peptides (i) is a highly efficient and unusually fast example of vibrational energy redistribution and (ii) represents a generic, sequence-independent feature of peptides.

To describe the vibrational relaxation dynamics, usually a system-bath approach is employed in which the system includes the initially excited vibrational mode (and possibly further strongly coupling modes) and the bath comprises all other degrees of freedom[31]. Second-order time-dependent perturbation theory with respect to the systembath interaction leads to a reduced density-matrix formulation in which the system is treated numerically exactly, while the dynamics of the bath enters via equilibrium autocorrelation functions. Being mostly interested in the vibrational energy relaxation rate $1/T_1$, well-established assumptions then lead to the widely used Golden Rule expression. To obtain a realistic modeling of the solvent, furthermore, many workers have pursued a mixed quantum-classical strategy and calculate the bath correlation functions from a classical molecular dynamics simulation

Here, my primary aim is to develop an all-atom nonequilibrium molecular dynamics simulation (NEMD) method to directly simulate the vibrational relaxation process using a standard MD program package such as GROMACS[32]. Our method includes the following main steps[8]. First, a quasiclassical sampling of the initial atom positions and velocities is performed to account for the nonequilibrium initial preparation of the system. To this end, we represent the solute normal modes $\{p_k, q_k\}$ in terms of classical action-angle variables $\{n_k, \phi_k\}$[33]

$$\begin{aligned} q_k &= \sqrt{(2n_k + \gamma)} \sin \phi_k \\ p_k &= \sqrt{(2n_k + \gamma)} \cos \phi_k. \end{aligned} \tag{13}$$

where the factor $\gamma = 1$ accounts for the zero-point energy (ZPE) of the mode. To obtain the initial positions and momenta of the initially excited amide I normal mode, we associate the action n_k with the initial quantum state of the amide I mode, e.g., $n_k = 1$ for the first excited state. The initial actions of the remaining solute modes (which are at thermal equilibrium) may be sampled from the Boltzmann distribution $P(n_k) \propto e^{n_k \hbar \omega_k / kT}$. In all cases, the vibrational phases ϕ_k are picked randomly from the interval $[0, 2\pi]$. This way an ensemble of normal-mode positions and momenta are calculated, which present a quasiclassical representation of the quantum initial state of the solute molecule. Next, employing these initial conditions, classical MD simulations are performed. To monitor the vibrational dynamics, the overall rotation and internal motion are separated and a normal-mode analysis is performed, which yields the time-dependent energy content of the normal modes. It turns out, however, that these procedures are well established only in the case that the vibrational dynamics is well described by small-amplitude motion around a single equilibrium structure. Considering flexible biophysical systems far from equilibrium (e.g., a folding peptide) , the single-reference normal-mode approximation must break down. To extend the nonequilibrium description of vibrational-energy redistribution to the treatment of molecular systems undergoing large amplitude motion, we adopt the ideas underlying the instantaneous normal-mode theory[34]. In this formulation at every time step a normal-mode calculation is performed, which employs the instantaneous position of the trajectory as reference structure, and the instantaneous vibrational energy of the system at time t_{n+1} is

$$E_{\mathrm{vib}} = V_0 + \frac{1}{2} \sum_{k_1}^{3N-6} \hbar \omega_k (p_k^2 + q_k^2), \tag{14}$$

where $q_k(t_n)$, $p_k(t_n)$, and $\omega_k(t_n)$ represent the positions, momenta, and frequencies of the instantaneous normal modes. $V_0 = V[\mathbf{r}(t_n)] - \frac{1}{2} \sum_k \hbar \omega_k a_k^2$ with $a_k = (m_k \hbar \omega_k^3)^{-1/2} \sum_{i=1}^{3N} F_i U_{ik}$ being the coordinate shift accounts

for the non-zero force term F_i, and U_{ik} is the eigenvector of the normal modes. To demonstrate the performance of the method, we study the vibrational energy relaxation process following the excitation $\nu = 0 \rightarrow 1$ of the amide I mode of the model system N-methylacetamid peptide (H_3C-COND-CH_3).

Figure 3(a) shows the time-dependent energy content of the amide I mode. In the purely classical case, where the ZPE is not included in modes ($\gamma = 0$), the decay can be well fitted to a biexponential function with the decay times $t_1 = 1.9$ ps (80%) and $t_2 = 13.3$ (20%) ps. At longer times, the amide I energy converges to $k_BT \approx 200$ cm^{-1} which is consistent with the equipartition theorem. Including ZPE in the amide I mode ($\gamma = 1$) is seen to lead to a considerable enhancement of the initial relaxation process ($t_1 = 1.5$ ps). There is only little change, though, if the ZPE is also included in the remaining solute modes. If we include only a fraction of ZEP with $\gamma = 0.35$ to the amide I such that the amide I energy remains larger than the ZPE for all times under consideration, we obtain $t_1 = 1.8$ ps (75%) and $t_2 = 7.5$ ps (25%) [Fig.3(b)]. In summary, we show that the vibrational energy relaxation rate obtained from the nonequilibrium simulations is in qualitative agreement with experiment, and our approach therefore may represent a reasonable and direct description of the quantum-mechanical relaxation process.

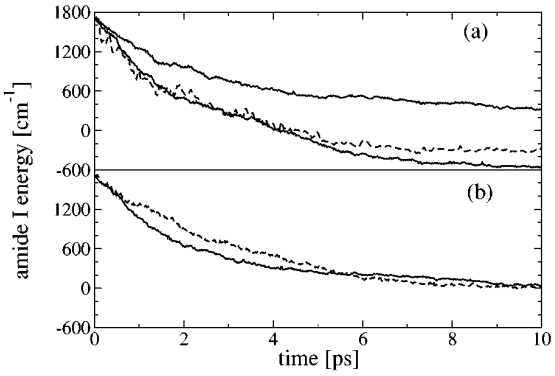

Figure 3: *Vibrational energy relaxation of the amide I mode of NMA in D_2O following $\nu 1 \rightarrow 0$ excitation as obtained for various quasiclassical initial conditions. For better comparison, the initially included zero-point energy is subtracted. (a) Comparison of action-angle initial conditions including zero point energy in all solute modes (dashed line), in the amide I mode only (lower solid line), and no solute mode (upper solid line). Note that in the first two cases the energy content of the amide I mode drops below zero for times ≥ 4 ps. (b) Including only 35% of the zero-point energy in all solute modes (dashed line) and in the amide I mode only (solid line), respectively, the amide I energy remains positive.*

(b) Photoinduced energy transfer. In another work, we investigate energy transport through an α-aminoisobutyric acid-based 3_{10}-helix (Aib) dissolved in chloroform in a combined experimental-theoretical approach[9]. The basic idea is that the vibrational energy is locally deposited at the N-terminus of the Aib by ultrafast internal conversion of a covalently attached, electronically excited, azobenzene moiety [Fig.4(A)]. Heat flow through the helix is detected with sub-picosecond time resolution by employing vibrational probes as local thermometers at various distances from the heat source. This peptide was chosen because it forms exceptionally a stable 3_{10}-helix even in a relatively short sequence of 8 amino acids. The apolar solvent chloroform was used to minimize potential loss of the heat flow into the surrounding solvent and moreover mimics the hydrophobic environment in the interior of a protein. We chose azobenzene molecule because it undergoes ultrafast internal conversion (cis and $trans$ isomerization) on a 200-fs timescale. To guide the interpretation of the experiments and obtain a microscopic picture of the molecular processes underlying energy transport in peptides, we have developed a simple computational strategy that allows us to extend well-established MD techniques to the description of photoinduced nonequilibrium dynamics[11, 12]. To model the laser-induced photoisomerization process, we use a minimal model for the corresponding potential-energy surfaces that diabatically connects the excited-state S_1 of the cis-isomer with the ground state S_0 of the $trans$-isomer [Fig.4(B)]. The photoexcitation of the system by an ultrafast laser pulse is mimicked by instantly switching from the ground-state N=N torsional potential to the excited-state potential. Following this nonequilibrium preparation at time $t = 0$, the system isomerizes along an excited-state N=N potential within 0.2 ps. After isomerization (i.e, for times ≥ 0.5 ps), the N=N torsional potential is switched back to its ground state form, and a constant-energy MD simulation is performed up to 100 ps. All simulations were performed with the GROMACS program suite[32], using the GROMOS96 force field 43a1[35]. Additional force field parameters for the azobenzene unit were derived from density functional theory. Following the nonequilibrium simulations, the time-dependent observables of interest are obtained via an ensemble average over 800 equilib-

rium conformations. We observe that after the photoexcitation at time $t = 0$, the kinetic energy is deposited into the azobenzene photoswitch within 0.1 ps. While the excitation of the photoswitch decays on a picosecond time scale, its energy is transferred to the Aib peptide (30%) and directly to the solvent (70%). The peptide energy rises within 0.3 ps and remains approximately constant up to 10 ps, before it decays with an 20-ps time constant. The solvent energy $E_{\mathrm{solv}}(t)$ rises with time constants 0.5 ps and 20 ps, where the shorter time scale reflects the initial momentum transfer of the isomerizing photoswitch to its surrounding solvent molecules. After the thermalization of the photoinduced energy, the 20-ps time scale of $E_{\mathrm{solv}}(t)$ accounts for the subsequent cooling of the peptide. We also observe that the photoexcitation does not lead to significant conformational changes of the peptide. To elucidate the energy transfer along the peptide chain, we displays the time evolution of the mean kinetic energy for most of the peptide groups. The kinetic energy of unit 1 (the linker unit to the photoswitch, Fig.3.2.1) reaches its peak at 0.3 ps, which nicely agrees with the experimental finding that unit 1 takes up substantial energy within the time resolution (0.2 ps). Unit 2 receives 50% of the energy from unit 1, and unit 3 receives 35% within 1 ps. According to the simulation, the terminal unit 9 still receives a small (20%) amount of the excess energy because of the photoexcitation, but this effect is not measurable in the experiment. However, we found that in terms of quantitative numbers, the MD simulation seems to overestimate the heat diffusion constant by a factor of five.

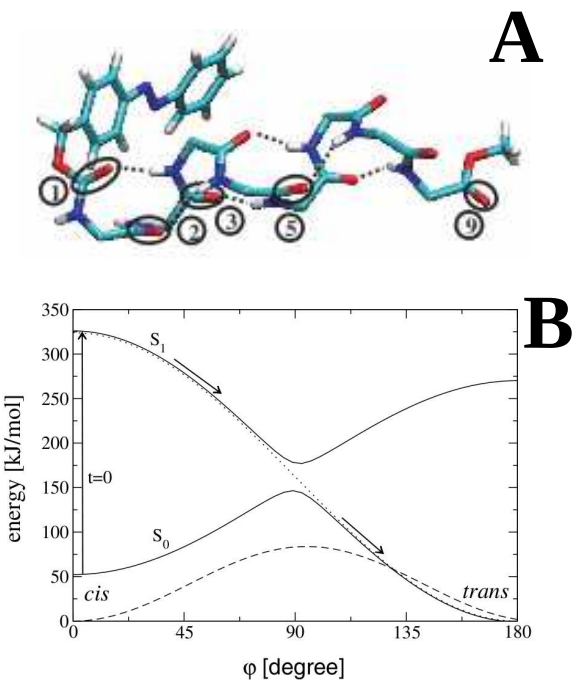

Figure 4: *(A) Structure of a stable Aib peptide where one end is attached to an azobenzene molecule which plays a role as a heater. (B) Scheme of the S_0 and the S_1 potential-energy curves of azobenzene as a function of the N=N $cis - trans$ isomerization coordinate ϕ. The solid lines represent the adiabatic potentials of a model which is designed to reproduce the experimental cis and $trans$ absorption bands and the ground-state $cis - trans$ energy barrier. The dashed line corresponds to the GROMOS force field potential of the N=N torsion, the dotted line shows a model of the cis-trans photoisomerization potential, which simply connects the S_1 cis state and the S_0 $trans$ state of azobenzene.*

To understand this discrepancy, we perform experiments with a lower energy excitation[10]. To this end, we carry out experiment and simulation with the direct excitation of a peptide C=O oscillator (amide I mode) of unit 1 with 0.2 eV (IR) photons. We should recall that the energy deposited through the excitation of the azobenzene chromophore is much higher, about 3 eV (UV) photons. The difference in the two excitation methods has not only a consequence for the amount of energy deposited, but also the form in which it is provided. The experiments show that heat transport through the peptide after excitation with low energy photons is at least 4 times faster than after UV excitation. On the other hand, the heat transport obtained by nonequilibrium MD simulations is largely insensitive to the kind of excitation. Thus, the calculations agree well with the experimental results for the low-frequency case. This finding suggests that the photoinduced energy gets trapped, if it is deposited in high amounts.

3.2.2 Photoinduced conformational rearrangements of peptides

In recent times, a number of experimental techniques have been developed which study biomolecular processes such as protein folding directly and in a time-resolved manner. For example, there have been various suggestions to include a molecular photoswitch such as azobenzene into peptides[36]. This guarantees that the light-induced structural changes of the chromophore upon photoisomerization around the central N=N double bond are directly transferred into the peptide chain. Femtosecond time-resolved pump-probe experiments with optical and infrared detection indicate that the main conformational changes of the peptide backbone are completed after only 20 ps, although the subsequent structural equilibration of the peptide continues for about 20 ns. These types of experi-

ments, especially in combination with 2D-IR probing, provide a new and promising way to study the folding and unfolding of peptides in unprecedented detail.

To guide the interpretation of the experiments and obtain a microscopic picture of the conformational dynamics of peptides, we carry out NEMD simulations of the experimentally well characterized octapeptide fragment H-Ala-Cys-Ala-Thr-Cys-Asp-Gly-Phe-OH which was connected head to tail via a azobenzene molecule[11, 12] [Fig.5]. Our replica-exchange MD (REMD) simulations of this bicyclic azobenzene peptide (bcAMPB) in its cis and $trans$ equilibrium states[37] are in direct agreement with nuclear magnetic resonance (NMR) results, which show that the $trans$-azopeptide is predominantly in a single conformational state, while there are many conformations of similar energy in the cis state of the peptide. To model the photoinduced conformational dynamics of the peptide we employ the simple model of the photoexcitation process described above (Section 2.2.1), which allows us to employ the standard GROMACS MD program package[32]. We used the GROMOS96 force field 43a1[35] to model the bcAMPB peptide. Additional force field parameters for the azobenzene unit were derived from density functional theory. For the initial structures, we selected 200 statistically independent conformations from the previous equilibrium REMD simulations, and the time-dependent observables of interest are obtained via an ensemble average over the initial distribution. Our NEMD simulation draws a detailed picture of the photoinduced dynamics of the photoswitchable peptide which occurs on at least four timescales, including (i) femtosecond $cis \rightarrow trans$ nonadiabatic photoisomerization of the azobenzene unit, (ii) picosecond-driven dynamics which include the rapidly redistribution of the photoinduced excess energy of the azobenzene chromophore to the vibrational modes of the peptide (1 ps) as well as to the surrounding solvent molecules (13 ps), and the ultrafast (0.2 and 14 ps) stretching of the peptide, (iii) fast conformational rearrangement (50 -100 ps) of most local (backbone dihedral angles) and global (radius of gyration, root mean squared deviation) reaction coordinates, and (iv) slow conformational equilibration (500 - 1000 ps) accounts for the subsequent conformational equilibration of the system. This process is diffusion-controlled, as the peptide needs solvent-driven fluctuations to escape from traps on the way to the final state. A quantitative comparison to experiment can only be obtained by a direct calculation of the measured transient spectra. Therefore, starting from semiclassical line shape theory, we have derived explicit expressions for transient one- and two dimensional IR spectra, which can be directly evaluated from nonequilibrium MD simulation trajectories[16]. Applying the method to the bcAMPB system, our results are in remarkable agreement with the experimental studies[38]. In summary, our simulation of transient IR spectra together with nonequilibrium photoinduced peptide dynamics simulation clearly suggest that transient IR spectroscopy on isotope-labeled biomolecules has a great potential to reveal conformational dynamics in unprecedented detail.

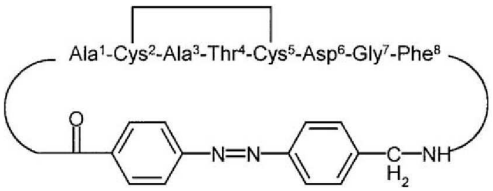

Figure 5: *Structure and amino acid labeling of the bicyclic azobenzene peptide bcAMPB*

3.2.3 Free energy landscapes of biomolecules

(a) Dihedral angle principal component analysis. The concept of describing biomolecular processes in terms of the molecule's energy landscape has promoted much of the recent progress in understanding the process of protein folding. However, since it is neither possible nor desirable to represent the free energy as a function of all $3N$-6 coordinates of a biomolecule, one needs to invoke a strategy to identify the most important conformational degrees of freedom of a simulation. To this end, the principal component analysis (PCA) method has been found useful[39]. The approach is based on the covariance matrix

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle. \tag{15}$$

where $x_1 \cdots x_{3N}$ are the mass-weighted Cartesian coordinates of the N-particle system and $\langle \cdots \rangle$ denotes the average over all sampled conformations. The eigenvectors and eigenvalues of σ yield the modes of collective motion and their amplitudes. It has been shown that a large part of the system's fluctuations can be described in terms of only a few principle components[39]

$$y_i = \sum_{k=1}^{3N} U_{ki} x_k \qquad i = 1, \cdots, 3N-1, \tag{16}$$

with U is the $(3N \times 3N)$ matrix where the ith column contains the ith eigenvector of the matrix σ [Eq.15]. Although the method has been widely used, it has two serious problems. First, the coordinates x_i in Eq.15 describe the internal motion, and must be separated from the overall motion. However, the completely elimination of overall rotation is impossible in the case of large amplitude motion such as protein folding/unfolding. As a result, the outcome of a PCA depends significantly on the reference structure chosen to remove the overall motion. As a solution to this problem, we develop the so-called dihedral angle PCA (dPCA), where we use the internal protein dihedral angles instead of Cartesian coordinates[17]. To avoid the periodicity of the dihedral angles, we transform each dihedral angle ϕ_i to two variables as $q_i = \cos(\phi_i)$, $q_{i+1} = \sin(\phi_i)$. Then, the PCA is performed for the new variables q_i.

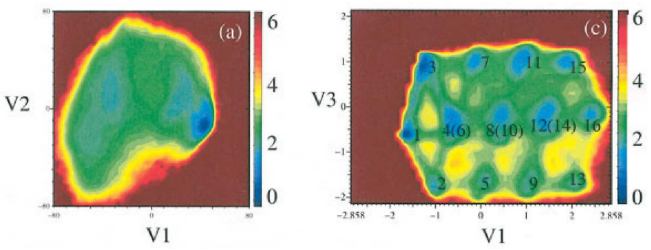

Figure 6: *Free energy landscape of the penta-alanine peptide plotted as a function of the first two principal components obtained from Cartesian coordinates (left) and transformed dihedral angles (right) PCA. The unit of free energy is kcal/mol.*

Fig.6 shows the free energy landscape of the penta-alanine peptide obtained using the Cartesian coordinates (left) and dPCA (right). The free energy surface obtained from the Cartesian PCA exhibits a smooth appearance and a simple single-minimum funnel-like shape, and these features are an artifact of the mixing of internal and overall motion. The dPCA shows that the true free energy landscape of penta-alanine is quite rugged, that is, it contains numerous minima of comparable probabilities and life times. The method is further developed and applied to visualize the free energy landscapes of peptides[18, 19], RNA[25] and to analyze the complexity of peptide folding[20]. The dPCA is now the method of choice in studies of the conformational analysis of peptides, proteins and implemented in the GROMACS simulation package[32].

(b) Nonlinear principal component analysis.

The second problem associated with the conventional PCA method described above is that the transformation from the original coordinates to the principal components [Eqs.15,16] is a linear transformation. And, in this respect it is not necessarily the most appropriate methodology for handling problems which exhibit nonlinear behavior. For complex biological systems, it is likely that nonlinear characteristics exist in some dynamical processes, like protein folding and unfolding. If that is the case then there are two obvious drawbacks in using PCA: (i) one still needs many principal components to construct a low dimensional space to represent meaningfully molecular conformations; (ii) some states in low dimensional space maybe mixed up by several conformations. To solve this drawbacks, I have developed a nonlinear PCA (NLPCA) method using the artificial neural network (ANN) theory[21]. By analogy to Eq. 16 one seeks a mapping in the form

$$\mathbf{y} = G(\mathbf{x}), \tag{17}$$

where G is an nonlinear vector function composed of m nonlinear functions $G = \{G_1, \cdots, G_m\}$. In terms of the components, this transformation is rewritten as follows

$$y_i = G_i(\mathbf{x}) \qquad i = 1 \cdots m. \tag{18}$$

The nonlinear functions G_i can be obtained by fitting Eq.18 to the following nonlinear function

$$v_k = \sum_{j=1}^{N_2} w_{jk2}\sigma\Big(\sum_{i=1}^{N_1} w_{ij1} + \theta_j\Big) \tag{19}$$

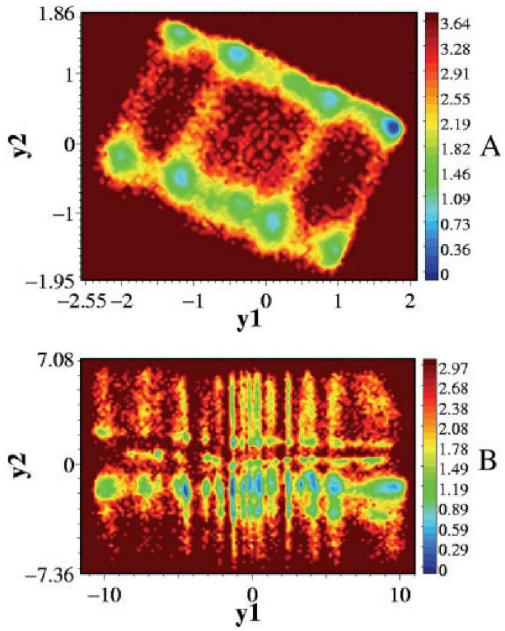

Figure 7: *Free energy landscape of hexaalanine (in kcal/mol) in water at 300 K as a function of the first two principal components y_1 and y_2. Results are obtained from PCA (A) and NLPCA (B).*

where $\sigma(x)$ is any continuous and monotonically increasing function with $\sigma(x) \to 1$ as $x \to \infty$ and $\sigma(x) \to 0$ as $x \to -\infty$, and the optimal values of the parameters w_{ijk}, θ, N_1, and N_2 are obtained from the training and learning modes of an ANN. As an example, Fig.7 shows the free energy landscape as a function of the first two principal components obtained from linear and nonlinear PCA of the 200 ns MD simulation trajectoty of the hexaalanine peptide in water at 300 K. As seen, the PCA yields a relatively well-structured free energy landscape, exhibiting 10 prominent local minima which cover nearly 60% of the configurations sampled in the simulation. The NLPCA, on the other hand, gives a somewhat much more complicated picture of the conformational distribution. In total, we can identify 31 main conformational states. In both cases, the folded state (α-helical conformations) are found. However, in the unfolded regions, the NLPCA reveals many different states which are hidden in the PCA map. This indicates that the NLPCA method appears to give an appropriate description of molecular conformations in the low-dimensional space. With this method, one needs fewer principal components to represent the full dimension of the original data but more accuracy to reveal the molecules conformational states. Thus, the NLPCA technique may be valuable to reveal a detailed picture of protein folding pathways.

(c) Principal-independent component analysis method. The primary aim of the linear or nonlinear PCA methods described above is to reduce the high dimensional space of a biomolecule to a low dimensional space, which allow one to capture the important conformations. If we limit ourself to a subspace which has only two dimensions then the analysis of conformations is simple because the states are clearly visualized. However, the state identification task becomes much more difficult if the subspace is formed by more than two principal components. This is mainly due to the fact that the states in this space can not be visualized. With this in mind, I propose a method which allows one to capture the distribution of molecular conformations in the form of one dimensional (1D) distribution plots[22]. The basic idea of our method is that we first reduce the dimension of the original multidimensional conformational space by performing the linear or nonlinear PCA to obtain the most important principal components (PCs). Subsequently, the so-called independent component analysis (ICA) method [40] is employed to obtain the independent components (ICs) from those PCs. Basically, the ICA allows one to decompose an input data set into components so that each component is as statistically independent from the others as possible. The task is now to identify the states in the subspace spanned by the ICs. Let $s_1, \cdots, s_m$ be those m ICs. Since the ICs are independent, the joint probability distribution $p(s_1, \cdots, s_m)$ can be factorized as $p(s_1, \cdots, s_m) = p_1(s_1) \cdots p_m(s_m)$. Thus, the number of states in the subspace spanned by the ICs is equal to the number of all possible combinations of the states obtained from each IC s_i ($i = 1 \cdots m$). Of course, the conformational states distributed along each IC are easily determined by locating the peaks in the distribution function of that IC. Thus the identification of conformations in the subspace is then straightforward.

3.3 Current research

My current research basically continues the directions which have been established during the postdoctoral study. The first research topic focuses on the aggregation of the amyloid- proteins. These systems are studied predominantly by means of computer simulations, through which we aim to realize our primary goals: first, to determine the structure of small oligomers - the most toxic species, and second to understand the aggregation mechanism. The insight thus gained allows the design of new anti-Alzheimer's disease drugs. The second topic concerns the nonequilibrium simulations of the laser-induced dissociation of various systems including amyloid fibrils, nanostructures, viruses and cellulose. Not only do these systems exhibit fascinating properties stimulating our scientific curiosity, but they are also strongly gaining importance in nanotechnological and bioengineering applications. Finally, despite the steady increase in available computer power, many of these problems hover on the verge of what is feasible. Therefore, in order to obtain scientifically worthwhile results within an acceptable time frame, it is essential to employ state-of-the-art techniques. We take an active interest in the development of new methodologies, both simulation techniques and advanced approaches to data analysis. Below I present some main achievements in details.

3.3.1 Protein aggregation

Amyloid fibrils are the production of an aggregation process, i.e., it represents an accumulation and clumping together, of mis-folded proteins. In our brains, fibrils are the cause of several significant neurodegenerative diseases,

such as Parkinson disease, Alzheimer's disease (AD)[41, 42]. There is increasing in vivo evidence that the amyloid-β (Aβ) protein and Aβ oligomers are the proximate neurotoxic agents in AD. This protein usually presents in the two forms 1 - 40 ($A\beta_{1-40}$) and 1 - 42 ($A\beta_{1-42}$) amino acids sequences. Currently, there are no effective treatments for this dreaded disease. As a consequence, extensive studies have been devoted to understand the fundamental aspects of the aggregation mechanism as well as to develop drugs and alternative approaches for treating AD[42, 43]. Below, I summary the main achievements of my research in protein aggregation.

(a) Mechanism of oligomer growth. To monitor the early events that direct assembly of amyloidogenic peptides we probe the dynamics of formation $(A\beta_{16-22})_{n-1} + A\beta_{16-22} \rightleftharpoons (A\beta_{16-22})_n$ of (A$\beta_{16-22})_n$ oligomers by adding a monomer to a preformed (A$\beta_{16-22})_{n-1}$ ($n = 4 - 6$) oligomer in which the peptides are arranged in an antiparallel β-sheet conformation[24].

To this end, we first carry out multiple long all-atom MD simulations in water using the GROMOS43a1 force field[35] for the trimer $(A\beta_{16-22})_3$. We find that the in-register antiparallel structure is most populated. We then place at randomly an unstructured $A\beta_{16-22}$ monomer at the vicinity of the preformed trimer, and carry out again multiple MD simulations for the new tetramer. Then, an unstructured monomer is added to the most populated tetramer and multiple MD simulations are carried out for the pentamer. This process is repeated with hexamer system. Fig.8 shows the time evolution of the β-strand content of the peptides of the reaction $(A\beta_{16-22})_4 + A\beta_{16-22} \rightleftharpoons (A\beta_{16-22})_5$. It shows clearly that the oligomer grows by a two-stage dock-lock mechanism. The largest conformational change in the added disordered monomer occurs during the rapid (≈ 50 ns) first dock stage in which the β-strand content of the monomer increases substantially from a low initial value. In the second slow-lock phase, the monomer rearranges to form in register antiparallel structures. Surprisingly, the mobile structured oligomers undergo large conformational changes in order to accommodate the added monomer. The time needed to incorporate the monomer into the fluid-like oligomer grows even when $n = 6$, which suggests that the critical nucleus size must exceed six. Stable antiparallel structure formation exceeds hundreds of nanoseconds even though frequent interpeptide collisions occur at elevated monomer concentrations used in the simulations. We suggest that the dock-lock mechanism should be a generic mechanism for growth of oligomers of amyloidogenic peptides.

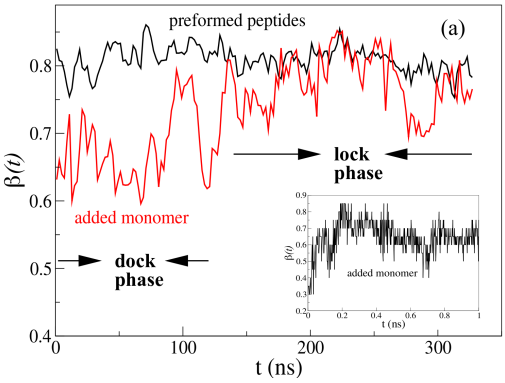

Figure 8: *Probes of lockdock mechanisms. Average (over the four trajectories) value of the dependence of the β-strand content for the preformed pentamer and the added monomeras a function of time. The value of $\beta(t)$ for the structured oligomer remains high and fluctuates around its initial value of $\beta \approx 0.8$. The β-strand content of the added monomer is considerably less than that of the structured pentamer for $t < 120$ ns that roughly corresponds to the dock phase. For $t > 120$ ns, which represents the lock phase, the β-strand content coincides with the value in the ordered state. (Inset) The changes in the β-strand of the monomer for $t \leq 1$ ns. It is clear that the maximum change in the extent of β-strand conformation occurs immediately upon docking*

(b) Structure and dynamics of Aβ oligomers. As mentioned, oligomers, but not mature fibrils, are the most toxic components in AD. Therefore, it is important to understand the structure and dynamics of oligomers. For example, any effective drug design strategy targeting early formed oligomers requires detailed knowledge on the structure of Aβ oligomers in aqueous solution. Currently, most all-atom MD simulations were performed on small oligomers, and there are only a few all-atom simulations of large oligomers with sizes ranging from 8- to 18-mers and time scales of 50-500 ns[44, 45, 46, 47]. This is because, at least partly, the all-atom simulations are too expensive to capture the time scales associated with fibril formation. Thus, we are still lacking detailed information on the equilibrium structure and dynamics of large oligomers. With this in mind, we carry out an extensive all-atom REMD simulation of 16-mer $A\beta_{37-42}$ peptide in explicit water for 460 ns/replica [47] using CHARMM27 force field. This peptide was chosen because of its counter intuitive feature. That is, the amyloid fibril structure reveals antiparallel β-sheets and parallel β-strands within individual sheets, although the peptide

has two opposite charges at the extremities [48]. The extensive conformational sampling allows us to address important questions: what is the equilibrium population distribution of oligomers and β-sheets? what is the fraction of mixed parallel/antiparallel β-sheets? what is the critical nucleus size? what is the mechanism of forming ordered aggregates? Our simulation shows that the peptide assembly follows the "condensation-polymerization" mechanism observed experimentally[49]: In the first stage, polymerization driven by hydrophobic collapse take places resulting in various oligomer sizes with little and small β-strand formed. Then, within the amorphous aggregates, two small sheets extend by essentially monomer addition and further conversion of peptides from random coil to β-strand can either lead to longer β-sheetsor new β-sheet layer.

While this is the longest simulation of such large amyloid oligomer starting from randomized and disordered chains, this approach is too expensive if one wants to study other large oligomers. Thus, our strategy is to develop further a coarse-grained lattice model[50], where a residue is represented by a bead located on a corner of a simple cubic lattice, and its side chain is described by an unit vector. The interaction between residues is modeled by the coarse-grained off-lattice OPEP force field[51]. Our bottom-up approach starts with the determination of the best lattice force field parameters for the Aβ_{16-22} dimer by fitting its equilibrium parallel and anti-parallel β-sheet populations to all-atom simulation results. Surprisingly, the calibrated force field is transferable to the trimer of Aβ_{16-22} and the dimer and trimer of Aβ_{37-42}. The dominant structure of the Aβ_{16-22} decamer matches the microcrystal structure. In contrast, the Aβ_{37-42} decamer is largely disordered with mixed by parallel and antiparallel chains, suggesting that the nucleus size is > 10 peptides. Our refined force field coupled to this on-lattice model should provide useful insights into the critical nucleation number associated with neurodegenerative diseases[52].

In another work, we carry out long all-atom REMD simulation of the dimeric system formed by the full-length $A\beta_{1-40}$ peptide for 400 ns/replica using the CHARMM22* force field[53]. The dimer is chosen because it is the smallest toxic specie in AD. We find transient configurations with an unstructured N-terminus and multiple β-hairpins spanning residues 17-21 and 30-36, but the antiparallel and perpendicular peptide orientation is preferred over the parallel organization. Short-lived conformational states also consist of all α topologies, and one compact peptide with beta-sheet structure stabilized by a rather extended peptide with α-helical content. Overall, this study provides, for the first time, insights into the equilibrium structure of $A\beta_{1-40}$ dimer in explicit aqueous solution, opening a new avenue for a comprehensive understanding of the impact of pathogenic and protective mutations in early-stage Alzheimer disease at a molecular level.

Finally, we study the interaction between the well-studied inhibitor 1,4-naphthoquinon-2-yl-L-tryptophan(NQTrp, Fig.9) with the $A\beta_{1-28}$ monomer and dimer[54, 55]. This compound has been reported to inhibit aggregation, to rescue cells from Aβ toxicity, and showed complete phenotypic recovery in an in vivo AD model[56]. The Aβ_{1-28} segment, rather than the full-length Aβ_{1-40} or Aβ_{1-42} peptides was chosen for the following reasons: Aβ_{1-28} forms fibrils with in register parallel β-sheets, as in the Aβ_{1-40} or Aβ_{1-42} fibrils[57], and the γ-secretase generates mostly peptides from Aβ_{1-36} to Aβ_{1-43}, but Aβ_{1-30} and Aβ_{1-26} peptides were reported by mass spectrometry analyses of human AD brains[58]. Electrospray ionization mass spectrometry showed that low molecular weight compounds bind similarly to Aβ_{1-42}, Aβ_{1-40} and Aβ_{1-28} peptides[59]. Finally, as Aβ_{1-28} is much less prone to aggregation than the Aβ_{1-40} or Aβ_{1-42} peptides, our theoretical results can be more easily tested experimentally. The simulations show that while the secondary structure contents and the intrinsic disorder of the Aβ_{1-28} monomer with and without NQTrp are very similar, and the lifetimes of the salt-bridges remain constant, the population of β-hairpin is reduced by a

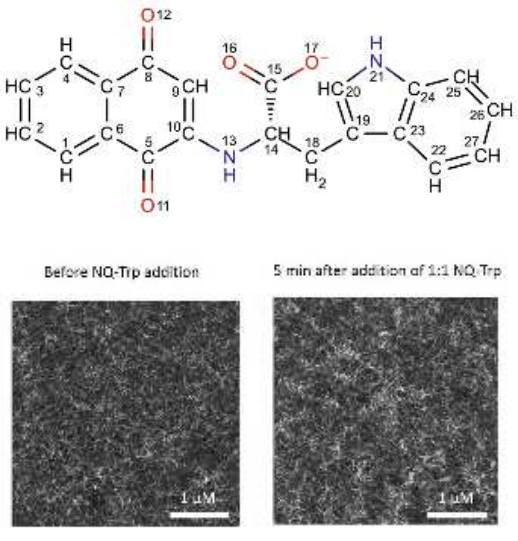

Figure 9: *(Above) The chemical structure of NQTrp. (Below) AFM images of pre-aggregated Aβ_{1-28} fibrils after 70 h before (left) and after (right) the addition of 1 equivalent of NQTrp (500 mm aggregated peptide diluted to 200 mm for imaging).*

factor of 1.5 and the population of α-helix in the region 17-24 is increased by a factor of two upon NQTrp binding. These two factors, which impact the free energy barrier for nucleation, provide a first explanation for the reported reduced Aβ_{1-40} and Aβ_{1-42} aggregation kinetics in the presence of NQTrp. However, the population of the free Aβ_{1-28} monomer (does not bind to NQTrp) is rather high (20-25%) at the concentration of 17.5 mM, and this shows that the affinity of NQTrp is low and hence its inhibitory activity is not very strong[54]. Furthermore, the interaction of NQTrp with the dimer of Aβ_{1-28} is very diffuse and not very specific, as most residues display probability contacts between 1 and 3% with the NQTrp molecules. These findings are confirmed by our experimental results obtained by fluorescence spectroscopy, FTIR spectroscopy, NMR spectroscopy as well as by cell viability studies[55]. For example, as seen from Fig.9, which displays the AFM images acquired immediately before and after the addition of NQTrp, there is no significant change, as in both cases images with high concentrations of fibrils were observed, even when fibrils were incubated for 6 h with NQTrp. Taken together, our study suggests that the reported anti-AD activity of NQTrp-type molecules in in vivo models has to involve another mechanism. This study has revealed the potential pitfalls in the development of aggregation inhibitors for amyloidogenic peptides, which are of general interest for all the molecules studied in the context of inhibiting the formation of toxic aggregates.

(c) Effects of mutation on the oligomer structure. Because of the disordered structure of the N-terminal in the fibril structure, the role of this region in aggregation, toxicity and pathogenesis was considered for a long time to be negligible, thus remained largely unexploited. Therefore, most of mutation studies have focused on the C-terminal of the Aβ_{1-40} or the Aβ_{1-42} peptides. However, several recent studies indicate the possible important of the N-terminal in the Aβ structure, aggregation and toxicity. For examples, Stefansson and co-workers first discovered the single-point Ala2→Thr2 (A2T) mutation effects by comparing the complete genome sequences of 1795 Icelanders with their medical histories[60]. The variant is rare, but it has a huge impact on those fortunate enough to inherit even a single copy of it. About 0.5% of Icelanders have a protective gene that prevents mental deterioration in old age. Compared with their countrymen who lack the mutation, Icelanders who carry this mutation are more than five times more likely to reach 85 without being diagnosed with AD. In another work, the Italian team leaded by Tagliavini has shown that the single-point mutation Ala2→Val2 (A2V) causes dementia in homozygous carriers, whereas heterozygous carriers were found to be protected[61].

To explain these experimental findings, we carry out extensive all-atom REMD simulations (400 ns/replica) for the wild-type (WT) and A2V mutation of the Aβ_{1-28} monomer, and also for the WT-WT, WT-A2V, WT-A2T dimers of the Aβ_{1-40} peptide. We show that upon A2V mutation, the population of β-hairpins of Aβ_{1-28} monomer is increased significantly (by a factor of 4), its intrinsic disorder is reduced by a factor of 2, and the free energy landscape is completely different as compared to the wild type counterpart[62]. Our results also show that for the homozygous A2V-A2V dimer, the A2V mutation enhances the intermolecular interaction between two N-terminals, thus reduces the intramolecular interaction between N-terminal with CHC and C-terminal of each monomer. This means that all three regions N-terminal, CHC, and C-terminal have strong tendency to participate into the aggregation as compared to WT-WT dimer, leading to the acceleration of the aggregation of A2V chains. In contrast, for the heterozygous WT-A2V dimer, the A2V mutation enhances the intramolecular N-terminal-CHC interaction, leading to the double-hairpin structure topology. This means that the N-terminal and CHC have weaker tendency to form intermolecular contacts as compared to WT-WT dimer, leading to the slowdown of aggregation of the wild-type and mutated A2V chains[63].

In contrast to A2V and A2T mutations, the other single-point mutations His6→Arg6 (H6R), Asp7→Asn7 (D7N) and Asp7→His7 (D7H) are reported to alter the monomer misfolding and produce oligomers that are toxic compared with the wild-type. To explain this, we carry out MD simulations for both monomers and dimers of Aβ_{1-40} and Aβ_{1-42} WT and mutant counterparts[64, 65, 66]. The long simulations with total time of 3-4 μs allows us to understand for the first time the microscopic effects of the mutations. For example, for both Aβ_{1-40} and Aβ_{1-42} systems, we did not find an increase of β-strand content upon D7N mutation. Rather, we found that the enhanced formation rate of Aβ_{1-40} fibrils comes essentially from the formation of the loop Asp23-Lys28 in the monomer. In contrast, the enhanced rate of Aβ_{1-42} fibrils does not result from the formation of the loop Asp23-Lys28 in monomer, but may result from the formation of the same loop in dimer, the increased turn propensity of the residues 28-30 in monomer, and the higher population of β -hairpin conformations involving the residues 19-20 with residues 31-32 in the dimer.

3.3.2 Nonequilibrium MD simulation of laser-induced dissociation of nanostructures

Recently, new mid-infrared free-electron laser (FEL) having specific oscillation characteristics of a picosecond pulse structure, a tunable wavelength within infrared frequencies and a high photon density has been developed and applied to the amyloid field. By tuning the laser frequency to the amide I bands of the amyloid, experiments were able to refold amyloid-like fibrils into native form, and to dissociate fibrils of short peptides[67, 68, 69, 70] . These findings are very interesting, however, the microscopics of the dissociation process is largely unknown. To this end, I have developed a comprehensive laser-induced NEMD simulation method, where a time-dependent electric field

$$E(t) = E_0 \exp[-\frac{(t-t_0)^2}{2\sigma^2}] \cos[2\pi c\omega(t-t_0)], \tag{20}$$

is applied to mimic a laser pulse. Here, E_0 represents amplitude of the electric field, σ is the pulse width, t is the time after the pulse maximum t_0, c is the speed of light and ω is the frequency. By scanning ω, we identify the frequency ω where the resonance between the vibrational modes of the system and laser frequency takes place, and this results in quick dissociation of the system. The method has been applied to study the dissociation of various systems including amyloid fibrils[71], peptide nanotube[72], virus[73] and cellulose.

As an example, Fig.10 shows the dissociation of the Poliovirus after 300 ps laser-induced excitation with the laser frequency $\omega = 1675$ cm^{-1} and intensity E_0=2 Vnm^{-1}. We find that the process is determined by a balance between the formation and dissociation of the protein shell, reflecting the highly plasticity of the virus. Furthermore, our method should provide a feasible approach to simulate viruses, which is otherwise too expensive for conventional equilibrium all-atom simulations of such very large systems. Our work shows a proof of concept which may open a new, efficient way to cleave or to recycle virus-based material, provide an extremely valuable tool to elucidating mechanical aspects of viruses, and may well play an important role in future fighting against virus-related diseases. The use of laser to dissociate amyloid fibrils may open up new venues to investigate the complex phenomena associated with amyloidogenesis. Finally, for the peptide nanostructure our method should provide a motivation for future experimental developments with the final aim is to open a new and efficient way to cleave or to recycle peptide-inspired. The dissolution of cellulose using laser may provide an alternative approach to the current widely used ionic liquids which might be toxic, thus providing a green method for producing biofuels.

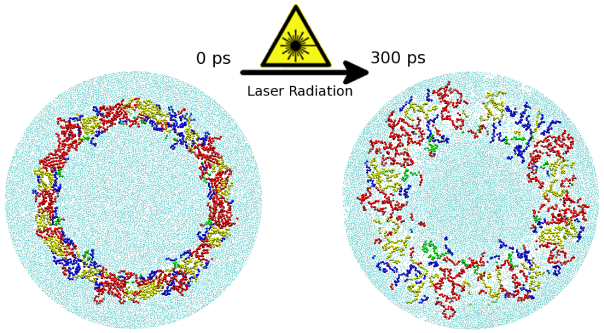

Figure 10: *The initial (left) and after 300 ps (right) structures of the poliovirus following a NEMD laser-induced dissociation simulation. For clarity, a cross section is shown.*

3.3.3 Enhanced sampling in molecular dynamics simulations

In general, a biomolecule is represented by a multidimensional rough free energy landscape with a huge number of local minima separated by high barriers. Therefore, sampling of conformational space becomes an extremely hard computational task. My idea is to develop simple but efficient methods to enhance the sampling.

a) Speed up convergence in the Wang-Landau simulation. In the first study, I propose a simple technique to speed up the convergence of the Wang-Landau method[74]. Briefly, given some temperature T, the canonical probability distribution is given by

$$P(E) = g(E) \exp(-\beta E), \tag{21}$$

where $\beta = 1/kT$, k is Boltzmanns constant, and $g(E)$ is the density of states (DOS) as a function of energy E. In the Wang-Landau method the probability $P_{\rm acc}$ of accepting a move from an energy level E_1 to a new trial level E_2 is given by

$$P_{\rm acc}(E_1 \to E_2) = \min\Big(1, \frac{g(E_1)}{g(E_2)}\Big) \tag{22}$$

where $g(E)$ denotes the current estimate DOS for the system. As the simulation progresses, DOS is updated and the process is repeated until the histogram of energies $H(E)$ becomes flat, i.e, the system samples all possible energy states. This method has been applied in various contexts, ranging from random spin models and quantum systems to simple models for proteins. Despite the advantages, it can still be difficult to obtain a reliable DOS with the Wang-Landau method. With this in mind, my idea is to apply a general and controllable method for smoothing a potential energy surface (PES)[75]

$$E_q(x) = \frac{q}{\beta(q-1)} \ln(1 + \beta(q-1)E(x)), \tag{23}$$

where $E(x)$ is the original PES of the system, and the smoothness of PES is controlled the parameter $q \geq 1$. On this smoothed surface $E_q(x)$, one begins a standard Wang-Landau calculation to provide a crude, but well-sampled initial estimate of DOS. In a second step, one restores the true surface $E(x)$ and continues to improve the estimation of the density of states. Applied the method to a simple Go-type model of 20-residue peptide, the new and original Wang-Landau simulations required on average $5.9 \pm 4 \times 10^6$ and $16.6 \pm 2 \times 10^6$ trial moves, respectively, thus resulting in roughly a threefold speedup. Because of the simple generalized effective potential for the seeding calculations, the method is general as well as very easy to implement and a wide range of applications is foreseen.

(b) Reduce the number of replicas in the replica-exchange simulation. Currently, the REMD method is widely used to enhance conformational sampling in biomolecular simulations. However, it has been shown that the number of replicas M increase quickly $M \sim \sqrt{f}$, where f is the number of degrees of freedom of the system. As a consequence, for large systems such as proteins in explicit solvent, this technique requires many processors and is clearly inappropriate. To this end, I propose a variant of the standard REMD method, employing the concept of solvent viscosity dependence of dynamics of biomolecules[76]. The physics behind this approach is straightforward: intuitively, the lower the viscosity, the faster a protein chain will diffuse, and the faster it will sample the conformational space. Quantitatively, the Kramers theory suggested that the rate of protein folding is inversely proportional to solvent viscosity, therefore, decreasing the viscosity will increase the folding rate. The basic idea of the method is to run the system of interest at a normal temperature and normal solvent viscosity, while other replicas run not only at different temperatures but also at artificially low solvent viscosities. Because both temperature and viscosity are used to enhance sampling, the method requires fewer replicas than the conventional REMD method. In a simulation the solvent viscosity can be reduced by a factor of λ if mass of solvent molecules is scaled by a factor of λ^2. Of course, the change of mass is probably the easiest task for any biomolecular simulation users. That makes the method is as simple as the conventional REMD method but more efficient. I show that (i) given the same number of replicas, the new method always converges to equilibrium faster than the conventional REMD simulations, (ii) the new method achieves similar sampling quality as the conventional REMD method but could reduce the number of replicas by a factor of 1.5 - 2 at least for the test systems in this work (16-residue β-hairpin peptide in explicit water). For very large systems, the efficiency of the new method will become even more pronounced.

(c) Simulated tempering with on-the-fly weight determination. Another widely used enhanced sampling method is the simulated tempering (ST) which does not have limitation to large systems like REMD. However, it does require the determination of a priori unknown weight parameters to ensure a uniform random walk in temperature space and this is non-trivial and very tedious for complex systems. In the ST simulation, temperature itself becomes a dynamical variable which could take discrete values T_m ($T_1 < T_2 < \cdots < T_M$). Probability distribution of a state at temperature T_m and potential energy E is given by the following generalized canonical distribution:

$$W_{\mathrm{ST}}(E, \beta_m) = \exp(\beta_m E + f_m), \tag{24}$$

where $\beta_m = 1/kT$. If the weight parameters f_m are chosen as $f_m = \ln(\int dE n(E) \exp(\beta_m E))$ ($n(E)$ is the density of states) then it follows immediately that the distribution of temperature is flat, i.e., a free random walk in temperature space is realized, which in turn induces a random walk in potential energy space and allows the system to escape from local energy minima. Here, I suggest a simple method to determine these weights on-the-fly based on the recursion formulae

$$f_{n+1} = f_n + (\beta_{n+1} - \beta_n)(\bar{E}_{n+1} + \bar{E}_n)/2, \tag{25}$$

where $\bar{E}_n$ is the average potential energy at temperature T_n at time t. With this scheme, simulated tempering requires neither prior trial simulations nor complicated update schemes. The advantage of our method over REMD

simulations has been demonstrated with the study of the folding of the 20-residue alanine peptide and the aggregation of a trimer formed by the Alzheimers peptide fragment Aβ_{16-22} using the implicit solvent coarse grained OPEP force field[77].

The method was then applied to the folding or aggregation of seven proteins with the CHARMM, OPLS, and AMBER protein, and the SPC and TIP3P water force fields[78]. The sampling with ST is found to be more efficient than with REMD for a much lower CPU cost. For example, starting from unfolded or extended conformations, the WW domain and the Trp-cage peptide fold to their NMR structures with a backbone RMSD of 2.0 and 1 Å. Remarkably, the ST simulation explores transient non-native topologies for Trp-cage that have been rarely discussed by other simulations. Taken together, these results open the door to the study of the configurations of single proteins, protein aggregates, and any molecular systems at atomic details in explicit solvent using a single normal CPU. They also demonstrate that our ST scheme can be used with any force field ranging from quantum mechanics to coarse-grain and atomistic.

(d) Multiple force fields in a single simulated tempering simulation. Two most frequent questions of any molecular dynamics simulations are: how robust are the results with respect to the force field used and has configurational sampling converged? Clearly, the results should be validated using various force fields such as OPLS/ AA, AMBER and CHARMM, and enhanced sampling methods such as replica exchange (RE) or simulated tempering (ST) must be employed. However, repeating RE or ST simulation with different force fields is very time-consuming and rarely done. To this end, I propose a simple and automatic method which allows in one single simulation to satisfy three goals: (i) efficient configurational sampling, (ii) force field dependence of the results, and (iii) temperature-dependence of the results modeled by each force field. Let's consider two independent states m and n associated with the configurational spaces x_m and x_n and the potential energy functions U and V, respectively. The probability distribution of the configuration x_m at temperature T_m and potential energy U is given by the following generalized canonical distribution:

$$P(x_m, \beta_m) = e^{-\beta_m U(x_m)+u_m}, \tag{26}$$

where $\beta = 1/k_{\rm B}T$ ($k_{\rm B}$ is the Boltzmann constant) and u_m is the weight parameter. The temperature transition from state m to n, while fixing configuration x_m, is given by the following Metropolis-like probability[79, 80]:

$$W_{\rm ST}(T_m, T_n) = \min[1, e^{-\beta_n V(x_m)+\beta_m U(x_m)+(v_n-u_m)}], \tag{27}$$

where $V(x_m)$ is the potential energy of x_m and v_n is the weight parameter of state n. As shown in Ref.[81], the acceptance probability, Eq.27, can be reformulated in terms of the instantaneous energy fluctuations $V(x_m) - \langle V(x_m)\rangle$ and $U(x_m) - \langle U(x_m)\rangle$. These fluctuations can result in significant temperature transition probability, even if U and V are different, and this is the essential physics that makes our method work. Applications to a 1D system and the all-atom chignolin peptide have shown that this new and efficient method should be very useful to study the folding of any biomolecules in explicit solvent.

3.3.4 General method to determine structure of amyloid fibrils

As mentioned in Section 2.3.1 it is important to understand the structure oligomers. Many reaction coordinates such as the radius of gyration, the β-sheet size, the oligomer size distribution, the number of contacts, the order parameter P_2, the connectivity length are often used to characterize the structures, but each variable captures only one feature of the self-assembly. The characterization and classification of the conformations remain difficult for systems characterized by degenerate states, that is, identical oligomer conformations differing only by the permutations of the molecules. To avoid the degeneracy problem, I propose a new method allowing the characterization of the structures of amyloid oligomers having arbitrary shapes and position-orientation arrangements[82]. The method takes into account both the intermolecular and intramolecular degrees of freedom, and treats correctly the degenerate states. Briefly, the intermolecular structures of an oligomer are described in terms of the combination of double-molecule states. These states describe the position-orientation arrangements of a double-chain in the presence of the other double-chains, and can be obtained through the PCA method of the inverse distances between side-chain's centers of mass of two chains of the double-molecule trajectory. The intramolecular structures of an oligomer are described in terms of the combination of single-molecule states. These states describe the

structures of a single-chain in the presence of the other chains, and can be obtained through the dPCA method of the single-molecule trajectory. The overall structure of the oligomer is given in terms of the product basis of the intermolecular and intramolecular structures. We have shown that our method can be applied to any oligomer size for two reasons. First, the identification of the single-molecule and double-molecule states involves the PCA of single- and double-molecule trajectories, and this is straightforward. Second, while the number of combinations of the single-molecule or double-molecule states increases with the number of chains, especially for highly flexible molecules, the calculation of the combinations is simple and fast, and the identification of the overall structures is a simple matter. In addition, many states can be marginally populated and safely discarded. This method should be a general method to obtain the overall structure of oligomers from simulation trajectory in unprecedented details.

3.3.5 Calculation of the configurational entropy of large biomolecules

Configurational entropy, which measures the number of available configurations that are occupied by a molecule in 3D space [83, 84], plays an important role in many chemical and biological processes[84] and structure-based drug design[85]. Currently, there are several methods to evaluate configurational entropy, including the thermodynamic integration, hypothetical scanning technique and the factorization of the multidimensional probability density function into lower-order probability density functions. Although these methods have proven to be quite accurate, their main drawbacks are that they tend to be computationally demanding, in particular for large systems such as proteins in explicit solvent, reducing therefore their possible applications. Our primary aim is to develop simple, fast yet accurate methods to estimate configurational entropy of large systems. Let us consider a molecule containing N atoms with Cartesian coordinates $\mathbf{x} = \{\mathbf{x}_i\}$, where $\mathbf{x}_i = (x_{3i-2}, x_{3i-1}, x_{3i})^T$ denotes the position of the ith atom. The *configurational entropy* of the molecule, first introduced by Kushick and Karplus [86], is

$$S = -k_B \int \mathrm{d}\mathbf{x} \rho(\mathbf{x}) \ln \rho(\mathbf{x}), \tag{28}$$

k_B is Boltzmann's constant and $\rho(\mathbf{x})$ is the canonical probability density distribution function of the system.

(a) Calculation of entropy using variable separation approach. The configurational entropy S, [Eq. 28], can be, in principle, calculated by performing numerically the multidimensional integrations of the right-hand-side of Eq. 28. In practice, this is difficult due to high dimensionality of the system. The basis idea of our method is as follows. Physically, the coordinates $\mathbf{x}$ in Eq.28 are strongly correlated as they are imposed by covalent bonds, non-bonded repulsions, and other atomic forces. Mathematically, since those coordinates are just the integration variables, a one-to-one linear transformation $\mathbf{q} = \mathbf{A}\mathbf{s}$ can be performed, resulting in the new representation of the entropy as follows:

$$S = -k_B \int \mathrm{d}\mathbf{s} \rho(\mathbf{s}) \ln \rho(\mathbf{s}) + k_B \ln |A|. \tag{29}$$

where $|A|$ is the absolute value of the determinant of the Jacobian square matrix $A_{ij} = \partial q_i / \partial s_i$. The key feature of Eq.29 is the dependence of the entropy on the new variables $\mathbf{s} = s_1, \cdots, s_n$ which can be obtained via a transformation such that they are quasi-independent, i.e., $\rho(s_1, \cdots, s_n) \approx \rho_1(s_1) \cdots \rho_n(s_n)$, and entropy can be approximated as

$$S_{\mathrm{ica}} = -k_B \sum_i^n \int ds_i \rho_i(s_i) \ln \rho_i(s_i) + k_B \ln |A|. \tag{30}$$

The problem now is finding a transformation such that the transformed variables s_i are quasi-independent. To this end, I employ the independent component analysis (ICA) mentioned in the Section 2.2.3 which allows for performing such a transformation [28,29]. This approach takes into account the whole dependence of coordinates, and is very fast fast even for large systems. Applying the method to the 16-residue β-hairpin peptide shows a dramatic improvement in the entropy, S_{ica}, which is much closer to the exact entropy, S, than other methods such as the quasi-harmonic approach. The method is still very fast even for large proteins. Details of the method is described in Ref.[87].

(b) Calculation of entropy using configurational temperature approach. The well-known quasiharmonic (QH) method approximates the system probability density distribution function by a multivariate Gaussian density

distribution function

$$\rho(\mathbf{x}) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}-\langle\mathbf{x}\rangle)^T\sigma^{-1}(\mathbf{x}-\langle\mathbf{x}\rangle)\right]}{(2\pi)^{n/2}|\sigma|^{1/2}}. \tag{31}$$

This function describes the distribution of the fluctuations of configurations around an averaged configuration $\langle\mathbf{x}\rangle$. Physically, this implies that the configurational fluctuation of the system is governed by an quadratic effective potential of the form

$$U_{\text{eff}}(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\langle\mathbf{x}\rangle)^T\mathbf{F}(\mathbf{x}-\langle\mathbf{x}\rangle), \tag{32}$$

With this approximation, the integration in Eq. 28 can be calculated analytically, resulting in the configurational entropy in the QH approximation of the form [86]

$$S_{\text{qua}} = \frac{1}{2}nk_B + \frac{1}{2}k_B\ln\left[(2\pi)^n|\sigma|\right], \tag{33}$$

where $|\sigma|$ is the determinant of the covariance matrix $\sigma_{ij} = \langle(x_i - \langle x_i\rangle)(x_j - \langle x_j\rangle)\rangle$. It has been shown that for macromolecules, the QH method yields entropy which is usually higher than the true entropy. This suggests that the system described by the effective potential, Eq.32, is too soft. Thus, my idea is to increase the effective force constants of the effective potential. To this end, I introduce a constraint such that the effective configurational potential[88]

$$T_c = \frac{\langle\sum_i|\frac{\partial U_{\text{eff}}}{\partial x_i}|^2\rangle}{k_B\langle\sum_i\frac{\partial^2 U_{\text{eff}}}{\partial x_i^2}\rangle}, \tag{34}$$

must be equal to the system temperature T: $T_c \equiv T$. This can be done by scaling all the diagonal elements of the eigenvalue matrix D, that diagonalizes the covariance matrix σ, by a factor α. This results in the new QH configurational entropy

$$\begin{aligned} S_{\text{new}} &= \frac{1}{2}nk_B + \frac{1}{2}k_B\ln\left[(2\pi)^n|\sigma|\right] - \frac{1}{2}nk_B\ln[\alpha], \\ &= S_{\text{qua}} - \frac{1}{2}nk_B\ln[\alpha]. \end{aligned} \tag{35}$$

As seen from this equation, the last term plays a role as a correction term. The new entropy S_{new} is identical to S_{qua} if α is equal to 1. This is the case where the system potential is purely harmonic. Otherwise, S_{new} is always smaller than S_{qua}. The calculation of the scaling factor α from a MD simulation trajectory is straightforward. This makes the estimation of the new entropy from Eq. 35 as simple as the conventional QH method, and the study of large macromolecules in implicit and explicit solvents possible. Details of the method is described in Ref.[89].

3.4 Scientific production

(a) Book chapters

1. *Nonequilibrium molecular dynamics simulation of photoinduced energy flow in peptides: theory meets experiment.*
Nguyen, Phuong H. and Hamm, P and Stock, G
in: Proteins: Energy, Heat and Signal Flow, eds. D. Leitner and J. Straub, CRC Press, 151 (2009)

2. *Energy transport in peptide helices around the glass transition.*
E. H. G. Backus and **Nguyen, Phuong H.** and Botan, V and Pfister, R and Moretto, A and Crisma, M and Toniolo, C and Stock, G and Hamm, P
in: ULTRAFAST PHENOMENA XVI Book Series: Springer Series in Chemical Physics Volume: 92 Pages: 532-534 (2009)

3. *Exploring the energy landscape of small peptides and proteins by molecular dynamics simulations.*
Stock, G and Jain, A and Riccardi, L and **Nguyen, Phuong H.**
in: Protein and Peptide Folding, Misfolding, and Non-Folding, ed. R. Schweitzer-Stenner, Wiley (2011)

(b) Papers

65. *Communication: Dissolution of cellulose with picosecond Infrared laser: a nonequilibrium all-atom molecular*

dynamics simulation study
Domin, Dominik and Viet, Man and Derreumaux, Philippe and Nguyen-Thi, Van-Oanh and **Nguyen, Phuong H.**
PHYSICAL CHEMISTRY CHEMICAL PHYSICS (submitted 2016)

64. *Oligomerization mechanism of Alzheimer Aβ40 peptides: the high content of intra-peptide β-sheets in A2V and A2T heterozygous dimers retards amyloid fibril formation.*
Nguyen, Phuong H. and Sterpone, Fabio and Pouplana, Ram on and Derreumaux, Philippe and Campanera, Josep M.
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY (submitted 2016)

63. *Communication: Nonequilibrium all-atom molecular dynamics simulation of the ultrasound induced bubble vibration and application to dissociate amyloid fibrils.*
Viet, Man and Derreumaux, Philippe and **Nguyen, Phuong H.**
PHYSICAL CHEMISTRY CHEMICAL PHYSICS (submitted 2016)

62. *Self-assembly of Aβ_{41} is rather similar to Aβ_{40} than to Aβ_{42}: in silico and in vitro study.*
Nguyen, Hoang Linh. and Tran, Minh Thu and Phan, Minh Truong and Man, Viet and **Nguyen, Phuong H.** and Ly, Anh Tu and Chen, Yi-Cheng and Li, Mai Suan.
JOURNAL OF PHYSICAL CHEMISTRY B (submitted 2016).

61. *Lattice model for amyloid peptides: OPEP force field parametrization and applications to the nucleus size of Alzheimer's peptides.*
Thanh Thuy, Tran and **Nguyen, Phuong H.** and Derreumaux, Philippe.
JOURNAL OF CHEMICAL PHYSICS (accepted 2016)

60. *Impact of A2V mutation on the heterozygous and homozygous Aβ_{1-40} dimer structures from atomistic simulations.*
Nguyen, Phuong H. and Sterpone, Fabio and Campanera, Jose Maria and Nasica-Labouze, Jessica and Derreumaux, Philippe.
ACS CHEMICAL NEUROSCIENCE (10.1021/acschemneuro.6b00053, 2016).

59. *Communication: Picosecond Infrared laser-induced all-atom nonequilibrium molecular dynamics simulation of dissociation of viruses*.
Viet, Man and Nguyen-Thi, Van-Oanh and Derreumaux, Philippe and Li, Mai Suan and Roland, Christopher and Sagui, Celeste and **Nguyen, Phuong H.**
PHYSICAL CHEMISTRY CHEMICAL PHYSICS (10.1039/C5CP07711G, 2016).

58. *Communication: Picosecond melting of peptide nanotubes using an infrared laser: a nonequilibrium simulation study*.
Viet, Man and Phan, Truong and Derreumaux, Philippe and Li, Mai Suan and Roland, Christopher and Sagui, Celeste and **Nguyen, Phuong H.**
PHYSICAL CHEMISTRY CHEMICAL PHYSICS 17, 27275 (2015)

57. *Communication: Multiple atomistic force fields in a single enhanced sampling simulation*.
Viet, Man and Derreumaux, Philippe and **Nguyen, Phuong H.**
JOURNAL OF CHEMICAL PHYSICS 143, 021101 (2015)

56. *Structures of the Alzheimer's wild-type Aβ_{1-40} dimer from atomistic simulations*.
Tarus, Bogdan and Tran, Thanh and Nassica-Labouze, Jessica and Sterpone, Fabio and **Nguyen, Phuong H.** and Derreumaux, Philippe.
JOURNAL OF PHYSICAL CHEMISTRY B 119, 10478 (2015)

55. *Folding atomistic proteins in explicit solvent using simulated tempering*.
Zhang, Tong and **Nguyen, Phuong H.** and Nassica-Labouze, Jessica and Mu, Yuguang and Derreumaux, Philippe.
JOURNAL OF PHYSICAL CHEMISTRY B 119, 6941 (2015)

54. *Picosecond dissociation of amyloid fibrils with infrared laser: a nonequilibrium simulation study*.
Viet, Man and Derreumaux, Philippe and Li, Mai Suan and Roland, Christopher and Sagui, Celeste and **Nguyen, Phuong H.**
JOURNAL OF CHEMICAL PHYSICS 143, 155101 (2015)

53. *Combined experimental and simulation study suggests a revised mode of action of the anti-Alzheimer disease drug NQ-Trp*.
Berthoumieu, Olivia and **Nguyen, Phuong H.** and del Castillo-Frias Maria and Ferre, Sabrina and Tarus, Bogdan

and Nasica-Labouze , Jessica and Noel, Sabrina and Saurel, Olivier and Rampon, Claire and Doig, Andrew and Derreumaux, Philippe and Faller, Peter.
CHEMISTRY 21, 12657 (2015)
52. *Amyloid β-protein and Alzheimer's disease: when computer simulations complement experimental studies.*
J. Nasica-Labouze , **Nguyen, Phuong H.**, O. Berthoumieu, F. Sterpone, N-V. Buchete, S. Cote, A. De Simone, A. Doig, P. Faller, A. Garcia, A. Laio, Mai Suan Li, S. Melchionna, N. Mousseau, Y. Mu, A. Paravastu, S. Pasquali, D. Rosenman, B. Strodel, B. Tarus, J. Viles, T. Zhang, C. Wang, and P. Derreumaux.
CHEMICAL REVIEWS 115, 3518 (2015)
51. *Preformed template fluctuations promote fibril formation: insights from lattice and all-atom models.*
Kouza, Masksim and Truong, Co and **Nguyen, Phuong H.** and Kolinski, Andrzej and Li, Mai Suan.
JOURNAL OF CHEMICAL PHYSICS 142, 145104 (2015)
50. *Effect of the English familial disease mutation (H6R) on the monomers and dimers of $A\beta_{40}$ and $A\beta_{42}$.*
Viet, Man and **Nguyen, Phuong H.** and Derreumaux, Philippe and Li, Mai Suan.
ACS CHEMICAL NEUROSCIENCE 5, 646 (2014)
49. *Molecular Structure of the NQTrp inhibitor with the Alzheimer $A\beta_{1-28}$ monomer.*
Tarus, Bogdan and **Nguyen, Phuong H.** and Berthoumieu, Olivia and Faller, Peter and Doig, Andrew J. and Derreumaux, Philippe
Eur. J. MED. CHEM. 91, 43 (2015)
48. *Replica Exchange Molecular Dynamics Simulation for Understanding the Initial Process of Amyloid Peptide Aggregation.*
Nishikawa, Naohiro and **Nguyen, Phuong H.** and Derreumaux, Philippe and Okamoto, Yuko
MOLECULAR SIMULATION 41, 1041 (2015)
47. *Effect of Taiwan mutation (D7H) on Structures of Amyloid-β peptides: replica exchange molecular dynamics study.*
Truong, Phan Minh and Viet, Man and **Nguyen, Phuong H.** and Hu, Chin-Kun and Li, Mai Suan.
JOURNAL OF PHYSICAL CHEMISTRY B 118, 8972 (2014)
46. *The OPEP coarse-grained Protein Model: From Single Molecules, Amyloid Formation, Role of Macromolecular Crowding and Hydrodynamics to RNA/DNA Complexs.*
Sterpone, Fabio and Melchionna, Simone and Tuffery, Pierre and Pasquali, and Mousseau, Normand and Cragnolini, Tristan and Chebaro, Yassmine and Saint-Pierre, Jean-Francois, and Kalimeri Maria, and Barducci, Alessandro and Laurin, Yohan and Tek, Alex and Baaden, Marc and **Nguyen, Phuong H.** and Derreumaux, Philippe.
CHEMICAL SOCIETY REVIEWS 43, 4871 (2014)
45. *Amyloid Oligomer Structure Characterization from Simulations: A General Method.*
Nguyen, Phuong H. and Li, Mai Suan and Derreumaux, Philippe
JOURNAL OF CHEMICAL PHYSICS 140, 094105 (2014)
44. *The familial Alzheimer A2V mutation reduces the intrinsic disorder and completely changes the free energy landscape of the $A\beta_{1-28}$ monomer.*
Nguyen, Phuong H. and Tarus, B. and Derreumaux, Philippe
JOURNAL OF PHYSICAL CHEMISTRY B 118, 501 (2014)
43. *Understanding amyloid fibril nucleation and $A\beta$ oligomer/drug interactions from computer simulations.*
Nguyen, Phuong H. and Derreumaux, Philippe
ACCOUNTS OF CHEMICAL RESEARCH 47(2):603-11 (2014)
42. *Effect of the Tottori familial disease mutation (D7N) on the monomers and dimers $A\beta_{40}$ and $A\beta_{42}$.*
Viet, Man and **Nguyen, Phuong H.** and Ngo, Son and Li, Mai and Derreumaux, Philippe
ACS CHEMICAL NEUROSCIENCE 4, 1446 (2013)
41. *Important of the ion-pair interactions in the OPEP coarse-grained force field: parametrization and validation.*
Sterpone, Fabio and **Nguyen, Phuong H.** and Kalimeri, Maria and Derreumaux, Philippe
JOURNAL OF CHEMICAL THEORY AND COMPUTATION 9, 4574 (2013)
40. *Conformational ensemble and polymorphism of the all-atom Alzheimer's $A\beta_{37-42}$ amyloid peptide oligomers.*
Nguyen, Phuong H. and Derreumaux, Philippe
JOURNAL OF PHYSICAL CHEMISTRY B 117, 5831 (2013)
39. *Communication: Simulated tempering with fast on-the-fly weight determination.*

Nguyen, Phuong H and Okamoto, Yuko and Derreumaux, Philippe
JOURNAL OF CHEMICAL PHYSICS 138, 061102 (2013)
38. *Construction of the Free Energy Landscape of Peptide Aggregation from Molecular Dynamics Simulations*.
Riccardi, Laura and **Nguyen, Phuong H.** and Stock, Gerhard
JOURNAL OF CHEMICAL THEORY AND COMPUTATION 8, 1471 (2012)
37. *Structures of $A\beta_{1742}$ Trimers in Isolation and with Five Small-Molecule Drugs Using a Hierarchical Computational Procedure*.
Chebaro, Yassmine and Jiang, Ping and Zang, Tong and Mu, Yuguang and **Nguyen, Phuong H.** and Mousseau, Normand and Derreumaux, Philippe
JOURNAL OF PHYSICAL CHEMISTRY B (2012)
36. *Configurational entropy: an improvement of the quasiharmonic approximation using configurational temperature*.
Nguyen, Phuong H. and Derreumaux, Philippe
PHYSICAL CHEMISTRY CHEMICAL PHYSICS 14, 877 (2012)
35. *Simulation of transient infrared spectra of a photoswitchable peptide*.
Kobus, Maja and Lieder, Martin and **Nguyen, Phuong H.** and Stock, Gerhard
JOURNAL OF CHEMICAL PHYSICS, 135, 225102 (2011)
34. *Real Time Observation of Ultrafast Peptide Conformational Dynamics: Molecular Dynamics Simulation vs Infrared Experiment*.
Nguyen, Phuong H. and Staudt, Heike and Wachtveitl, Josef and Stock, Gerhard
JOURNAL OF PHYSICAL CHEMISTRY B 115, 13084 (2011)
33. *Coherent vibrational energy transfer along a peptide helix*.
Kobus, Maja and **Nguyen, Phuong H.** and Stock, Gerhard
JOURNAL OF CHEMICAL PHYSICS 134, 124518 (2011)
32. *Effects of all-atom force fields on amyloid oligomerization: replica exchange molecular dynamics simulations of the $A\beta_{16-22}$ dimer and trimer*.
Nguyen, Phuong H. and Li, Mai Suan and Derreumaux, Philippe
PHYSICAL CHEMISTRY CHEMICAL PHYSICS 13, 9778 (2011)
31. *Infrared signatures of the peptide dynamical transition: A molecular dynamics simulation study*.
Kobus, Maja and **Nguyen, Phuong H.** and Stock, Gerhard
JOURNAL OF CHEMICAL PHYSICS 133, 034512 (2010)
30. *Replica exchange simulation method using temperature and solvent viscosity*.
Nguyen, Phuong H.
JOURNAL OF CHEMICAL PHYSICS 132, 144109 (2010)
29. *Nonequilibrium molecular dynamics simulation of the energy transport through a peptide helix.*
Nguyen, Phuong H. and Park, Sang-Min and Stock, Gerhard
JOURNAL OF CHEMICAL PHYSICS 132, 025102 (2010)
28. *Free-Energy Landscape of RNA Hairpins Constructed via Dihedral Angle Principal Component Analysis.*
Riccardi, Laura and **Nguyen, Phuong H.** and Stock, Gerhard
JOURNAL OF PHYSICAL CHEMISTRY B 113, 16660 (2010)
27. *Molecular dynamics simulation of cooling: Heat transfer from a photoexcited peptide to the solvent.*
Park, Sang-Min and **Nguyen, Phuong H.** and Stock, Gerhard
JOURNAL OF CHEMICAL PHYSICS 131, 184503 (2009)
26. *Energy Flow and Long-Range Correlations in Guanine-Binding Riboswitch: A Nonequilibrium Molecular Dynamics Study.*
Nguyen, Phuong H. and Derreumaux, Philippe and Stock, Gerhard
JOURNAL OF PHYSICAL CHEMISTRY B 113, 9340 (2009)
25. *Estimating configurational entropy of complex molecules: A novel variable transformation approach.*
Nguyen, Phuong H.
CHEMICAL PHYSICS LETTERS 468, 90 (2009)
24. *Structural Flexibility of a Helical Peptide Regulates Vibrational Energy Transport Properties.*
Backus, Ellen H. G. and **Nguyen, Phuong H.** and Botan, Virgiliu and Moretto, Alessandro and Crisma, Marco and

Toniolo, Claudio and Zerbe, Oliver and Stock, Gerhard and Hamm, Peter
JOURNAL OF PHYSICAL CHEMISTRY B 112, 15487 (2009)
23. *Energy transport in peptide helices: A comparison between high- and low-energy excitations.*
Backus, Ellen H. G. and **Nguyen, Phuong H.** and Botan, Virgiliu and Pfister, Rolf and Moretto, Alessandro and Crisma, Marco and Toniolo, Claudio and Stock, Gerhard and Hamm, Peter
JOURNAL OF PHYSICAL CHEMISTRY B 112, 9091 (2008)
22. *Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis.*
Altis, Alexandros and Otten, Moritz and **Nguyen, Phuong H.** and Hegger, Rainer and Stock, Gerhard
JOURNAL OF CHEMICAL PHYSICS 128, 245102 (2008)
21. *Nonadiabatic vibrational dynamics and spectroscopy of peptides: A quantum-classical description.*
Kobus, Maja and Gorbunov, Roman D. and **Nguyen, Phuong H.** and Stock, Gerhard
CHEMICAL PHYSICS 347, 208, 2008
20. *Energy transport in peptide helices.*
Botan, Virgiliu and Backus, Ellen H. G. and Pfister, Rolf and Moretto, Alessandro and Crisma, Marco and Toniolo, Claudio and **Nguyen, Phuong H.** and Stock, Gerhard and Hamm, Peter
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 104, 12749 (2007)
19. *Dihedral angle principal component analysis of molecular dynamics simulations.*
Altis, Alexandros and **Nguyen, Phuong H.** and Hegger, Rainer and Stock, Gerhard
JOURNAL OF CHEMICAL PHYSICS 126, 244111 (2007)
18. *Conformational states and folding pathways of peptides revealed by principal-independent component analyses.*
Nguyen, Phuong H.
PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS 67,579 (2007)
17. *Structure and dynamics of the homologous series of alanine peptides: A joint molecular dynamics/NMR study.*
Graf, Juergen and **Nguyen, Phuong H.** and Stock, Gerhard and Schwalbe, Harald
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY 129, 1179 (2007)
16. *Quantum-classical description of the amide I vibrational spectrum of trialanine.*
Gorbunov, Roman D. and **Nguyen, Phuong H.** and Kobus, Maja and Stock, Gerhard
JOURNAL OF CHEMICAL PHYSICS 126, 054509 (2007)
15. *How complex is the dynamics of peptide folding?*
Hegger, Rainer and Altis, Alexandros and **Nguyen, Phuong H.** and Stock, Gerhard
PHYSICAL REVIEW LETTERS 98, 028102 (2007)
14. *Monomer adds to preformed structured oligomers of Aβ-peptides by a two-stage dock-lock mechanism.*
Nguyen, Phuong H. and Li, Mai Suan and Stock, Gerhard and Straub, John E. and Thirumalai, D.
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 104, 111 (2007)
13. *Complexity of free energy landscapes of peptides revealed by nonlinear principal component analysis.*
Nguyen, Phuong H.
PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS 65, 898 (2006)
12. *Photoinduced conformational dynamics of a photoswitchable peptide: A nonequilibrium molecular dynamics simulation study.*
Nguyen, Phuong H. and Gorbunov, Roman D. and Stock, Gerhard
BIOPHYSICAL JOURNAL 91, 1224 (2006)
11. *Improved Wang-Landau sampling through the use of smoothed potential-energy surfaces.*
Nguyen, Phuong H. and Mittag, E and Torda, AE and Stock, Gerhard
JOURNAL OF CHEMICAL PHYSICS 124, 154107 (2006)
10. *Nonequilibrium molecular dynamics simulation of a photoswitchable peptide.*
Nguyen, Phuong H. and Stock, Gerhard
CHEMICAL PHYSICS 323, 36 (2006)
9. *Free energy landscape and folding mechanism of a β-hairpin in explicit water: A replica exchange molecular dynamics study.*

Nguyen, Phuong H. and Stock, G and Mittag, E and Hu, CK and Li, MS
PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS 61, 795 (2005)
8. *Structure and energy landscape of a photoswitchable peptide: A replica exchange molecular dynamics study.*
Nguyen, Phuong H. and Mu, YG and Stock, G
PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS 60, 485 (2005)
7. *Reply to the Comment on "Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis.*
Mu, YG and **Nguyen, Phuong H.** and Stock, G
PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS 64, 798 (2006)
6. *Energy landscape of a small peptide revealed by dihedral angle principal component analysis.*
Mu, YG and **Nguyen, Phuong H.** and Stock, G
PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS 58, 45 (2005)
5. *Nonequilibrium molecular-dynamics study of the vibrational energy relaxation of peptides in water.*
Nguyen, Phuong H. and Stock, G
JOURNAL OF CHEMICAL PHYSICS 119, 11350 (2003)
4. *Local structure in nematic and isotropic liquid crystals.*
Nguyen, Phuong H. and Schmid, F
JOURNAL OF CHEMICAL PHYSICS 119, 1214 (2003)
3. *The direct correlation function in nematic liquid crystals from computer simulation.*
Nguyen, Phuong H. and Germano, G and Schmid, F
COMPUTER PHYSICS COMMUNICATIONS 147, 350 (2002)
2. *Spatial order in liquid crystals: Computer simulations of systems of ellipsoids.*
Schmid, F and **Nguyen, Phuong H.**
MORPHOLOGY OF CONDENSED MATTER: PHYSICS AND GEOMETRY OF SPATIALLY COMPLEX SYSTEMS, Lecture Notes in Physics 600, 172 (2002)
1.*Elastic constants from direct correlation functions in nematic liquid crystals: A computer simulation study.*
Nguyen, Phuong H. and Germano, G and Schmid, F
JOURNAL OF CHEMICAL PHYSICS 115, 7227 (2001)

(c) Selected oral communications

1. 2002-2008: 8 talks at the Huefeld workshop "Computer simulation and theory of macromolecules" organized every year, Germany.
2. 2007: Seminar at the Centre for Self Organising Molecular Systems, University of Leeds, UK.
3. 2007: Seminar at the School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore
4. 2011: Seminar at the School of Biological Sciences, Nanyang Technological University, Singapore
5. 2012: Speaker at the CECAM workshop "Anchoring simulations to experiments: challenges for understanding and treating Alzheimer's disease", Paris, France.
6. 2013: Seminar at the Department of Physics, Freiburg University, Germany
7. 2013: Speaker at the "Journees Modelisation", Paris, France.
8. 2014: Invited speaker at the "International Conference on Computational Science and Technology", Warsaw, Poland.

4 Future research projects

4.1 Large-scale nonequilibrium molecular dynamics simulations and experiments to study the focused ultrasound-induced opening of the blood-brain-barrier

In the near future my research will focus on the study of the ultrasound-induced blood-brain-barrier (BBB) opening for drug delivery to treat brain diseases. Because of the BBB, designing therapeutic agents for the brain is very challenging. Indeed, only about 5% of the 7000 known potential molecular drugs can cross the BBB to treat neurodegenerative diseases. For AD, only up to 0.1% of anti-Aβ antibodies administered peripherally can reach the brain, most remain in the bloodstream. Even small drugs can cross the BBB, but they are usually so dynamic

and do not bind specifically to metastable Aβ oligomers-the most toxic species, thus do not prevent the aggregation of the Aβ peptides. Therefore, although great progresses have been made in recent years toward understanding of these diseases, and numerous small- and large-molecule products have been developed for the treatment, no cures are currently available.

A variety of approaches have been developed to open the BBB for facilitation of drug delivery, however, none has achieved clinical applicability. Recently, increasing evidences suggest that focused ultrasound (FUS) in combination with microbubbles might be useful for delivery of drugs to the brain through transient opening of the BBB. As shown schematically in Fig.11, under the ultrasound, injected bubbles in the bloodstream vibrate in size. If the oscillation amplitude is small then the contraction and expansion of size is approximately symmetric. This phenomenon, called stable cavitation, produces rapid flows of liquid around the bubble and induces shear stress on nearby objects. In contrast, the large amplitude oscillation may lead to asymmetric contraction and expansion followed by a violent collapse called inertial cavitation. This generates a very high temperature, pressure and releases a shock wave that propagates at supersonic speed radially from the collapse site. In both cases, the bubble cavitation exerts mechanical forces on the BBB resulting in BBB opening. This technique offers a unique non-invasively and regionally avenue to deliver a wide range of drugs to the brain and promises to provide treatments for central nervous system disorders without unnecessary drug exposure. However, it is currently unknown the exact molecular mechanisms in opening the BBB, thus limiting the future development and application of this technology in clinical applications. So far, this technique is still in its infancy.

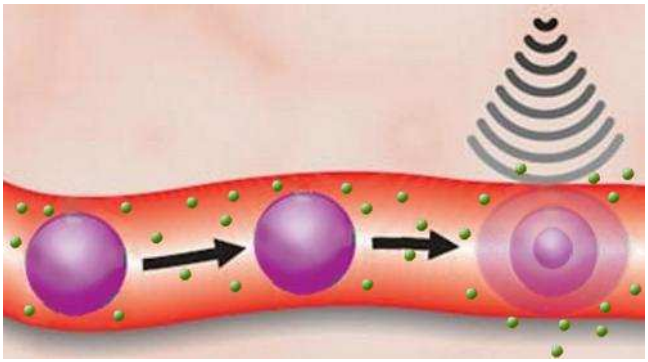

Figure 11: *Principle of the FUS technique: focused ultrasound is applied through the skull to the brain region of interest. Microbubbles (magnenta) are injected into the bloodstream. When the bubbles reach the ultrasound field, they vibrate. This changes the property of the blood vessel wall, making it more permeable. This temporary effect allows therapeutic molecules (green) to move from the blood to the brain.*

This project aims to combine systematically the molecular dynamics simulation and magnetic resonance imaging guided focused ultrasound (FUS) experiment to reveal at a molecular level the BBB opening mechanism. Because currently there are no molecular structures of the BBB, we first construct a BBB model, taking into account the experimental information of the claudin protein of the tight junction. Second, we develop, for the first time, a method for performing FUS-induced bubble cavitation molecular dynamics simulation. Having established the model and simulation method, we carry out simulations and experiments to understand (i) the dependence of the BBB opening amplitude on the ultrasound and bubble parameters, (ii) different cavitation effects on the BBB opening, (iii) kinetic of opening and closure of the BBB and (iv) possible routes of the BBB opening. In the second stage of the project, we employ simulation and experiment to study the delivery mechanisms of the drug and toxic substances as well as the effects of FUS on the amyloid fibril of the Alzheimer's disease mice brain. Overall, this project integrates different and complementary expertises from structural biology, biophysics, computer simulations to in vivo experiments on mice, aims at elucidating fundamental unsolved questions related to the FUS method for drug delivery, in general and for the Alzheimer's disease treatment, in particular. This may be important for translating and widespread adoption of this technology in clinical applications.

4.2 Resolving high-resolution 3D structure of functional amyloids using bioinformatics, computer simulation and experiment approaches

The second research project in the near future will be about the functional amyloids. Since the discovery of amyloid fibrils more than a century ago by Alois Alzheimer, amyloid concept is always associated with diseases, for example AD, type II diabetes, Parkinson and Creutzfeld-Jacob diseases. Thus, it is not surprised that extensive research has been driven by this "bad reputation" of amyloid for many decades with the final aim is to fight against these diseases. Within the past decade, a new concept-functional amyloid, has been emerg-

ing. This new idea states that many organisms can employ the biophysical properties of amyloid for their benefit. The research field of functional amyloid has just began with the first discovery of the Curli protein in 2002 and a breakthrough in 2003 with the discovery of the Pmel17 in humans. Compared to disease-amyloid, our knowledge of functional amyloid is much behind. To date, no high-resolution 3D models for both amyloid fibrils as well as partially formed oligomers are available. Solid state NMR, electron microscopy and homology modeling were only able to suggest that functional fibrils formed by the Curli proteins CsgA and CsgB do not contain in-register parallel -sheet structures but rather -helical structures as shown schematically in Fig.12. Whether this is also the common structure of other functional amyloids? If this is the case then the stability and dynamic of the assembly of functional amyloids should be affected by this architecture. The difference between in-register parallel -sheet structures and -helical structures is also reflected in their seeding propensity. Indeed, the seeding interface in a -helical structure should be 1 whereas in an in-register parallel -sheet structure it will be larger than 1. With this in mind, the primary aim of this project is to propose a novel computational approach to identify high-resolution 3D structures of functional amyloid proteins using the bioinformatics, computer simulation methods and results are accomplished by our experiments.

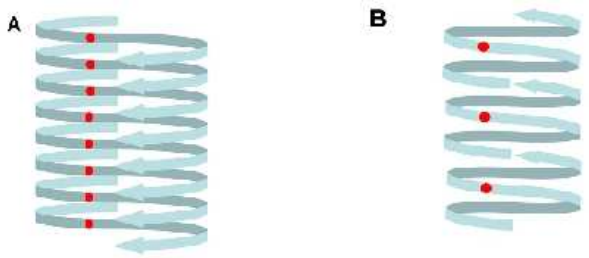

Figure 12: *A representative schematic of the in-register parallel β-sheet structure (A) and -helical structure (B)*

As mentioned, the determination of partially formed disease-oligomers are extremely difficult for both experiments and simulations. But why our method proposed in this project should be able to predict 3D structures of functional oligomers and fibrils? This is because we know how to explore valuable information encoded by nature in functional amyloid proteins: in contrast to disease-amyloid proteins, functional amyloid proteins are healthy molecules that are evolved by nature and evolutionally conserved across species. This suggests that nature probably have performed mutations on residues which should be close in space such that the stability and function of the protein are maintained. Therefore, by analyzing the contact signals between covarying residues in the sequence we might be informed about tertiary structures. The contact signals are used to drive molecular dynamics simulations to correct structures which will be subsequently accomplished by experiments.

5 Teaching activity and student advising

5.1 Teaching activity

1997: Lecturer in General Physics, Hue University of Sciences, Vietnam (45 hours)
2001: Teaching assistant in Statistical Thermodynamics, Bielefeld University, Germany (one semester)
2002-2008: Teaching assistant in Mathematics for Natural Sciences Students, Frankfurt University, Germany (every semester)
2002: Teaching assistant in Statistical Thermodynamics and Kinetic, Frankfurt University, Germany (one semester)
2003: Teaching assistant in Chemical Binding, Molecular Spectroscopy, Frankfurt University, Germany (one semester)
2007: Teaching assistant in Theoretical Chemistry, Frankfurt University, Germany (one semester)
2015: Lecturer in Advanced Methods in Molecular Dynamics Simulation, University Paris VII, France (10 hours)

5.2 Student advising

L3 supervision:
I supervised 3 students at the Institute of Physical and Theoretical Chemistry, Frankfurt University, Germany, March-June 2006, on the simulations of energy transfer along peptides.
M2 supervision:
Chu Huyen Trang, Master 2 Biology and Biotechnology, University of Science and Technology (USTH), Hanoi, Vietnam, March - November 2015 (with Dr. Van-Oanh Nguyen-Thi, Laboratory of Chemical Physics, University

Paris-Sud). I worked closely with the student on the effects of force fields on the laser-induced disssociation of amyloid fibrils.

PhD supervision:

1. Sang-Min Park, PhD student, Institute of Physical and Theoretical Chemistry, Frankfurt University, Germany, 2005 - 2008 (Official advisor: Professor Gerhard Stock). I worked directly with the student on the simulations of energy transfer along peptides. We had 2 papers together.
2. Jessica Koplin, PhD student, Institute of Physical and Theoretical Chemistry, Frankfurt University, Germany, 2002 - 2006 (Official advisor: Professor Gerhard Stock). I worked directly with the student on the simulations of RNA systems.
3. Thanh Thuy Tran, PhD student, Laboratoire de Biochimie Theorique, UPR 9080 CNRS, 2013-2016 (with Professor Philippe Derreumaux). I work directly with the student on the construction of the coarse-grained protein model for studying of the amyloid protein aggregation. We have had 2 papers together.

6 Administrative and community service activities

- Regular reviewer for the scientific journals including J. Am. Chem. Soc., J. Phys. Chem. B, J. Chem. Phys., Phys. Chem. Chem. Phys, Nanoscale, Chem. Phys, Lett., RSC Advances, Int. J. Mol. Sci., Biochemistry, Biophysics Report, Expert Opinion on Drug Discovery, ACS Chemical Neuroscience.
- Member de conseil de Laboratoire de Biochimie Theorique, UPR 9080 CNRS.

References

[1] P. G. de Gennes and J. Prost. *The Physics of Liquid Crystals*. Oxford University Press, Oxford, 1993.

[2] B. J. Berne and P. Pechukas. *J. Chem. Phys.*, 56:4213, 1975.

[3] J. P. Hansen and I. R. McDonald. *Theory of Simple Liquids*. Academic, London, 1986.

[4] P. H. Nguyen and F. Schmid. *J. Chem. Phys.*, 119:1214, 2003.

[5] D. Forster. *Hydrodynamic fluctuations, broken symmetry and correlation functions*. Frontiers in Physics, Benjamin, Reading, MA, 1975.

[6] C. Frank. *Discuss. Faraday Soc.*, 25:19, 1958.

[7] P. H. Nguyen, G. Germano, and F. Schmid. *J. Chem. Phys.*, 115:7227, 2001.

[8] P. H. Nguyen and G. Stock. *J. Chem. Phys.*, 119:11350, 2003.

[9] V. Botan, E. H. G. Backus, R. Pfister, et al. *Proc. Natl. Acad. Sci. USA*, 104:1274912754, 2007.

[10] E. H. G. Backus, P. H. Nguyen, V. Botan, et al. *J. Phys. Chem. B*, 112:9091, 2008.

[11] P. H. Nguyen and G. Stock. *Chem. Phys.*, 323:36, 2006.

[12] P. H. Nguyen, R. D. Gorbunov, and G. Stock. *Biophys. J*, 91:1224, 2006.

[13] P. H. Nguyen, H. Staudt, J. Wachtveitl, and G. Stock. *J. Phys. Chem. B*, 115:13084–13092, 2011.

[14] R. D. Gorbunov, P. H. Nguyen, M. Kobus, and G. Stock. *J. Chem. Phys.*, 126:054509, 2007.

[15] M. Kobus, R. D. Gorbunov, P. H. Nguyen, and G. Stock. *Chem. Phys.*, 347:208, 2008.

[16] M. Kobus, M. Lieder, P. H. Nguyen, and G. Stock. *J. Chem. Phys.*, 135:225102, 2011.

[17] Y. Mu, P. H. Nguyen, and G. Stock. *Proteins*, 58:45–52, 2005.

[18] A. Altis, P. H. Nguyen, R. Hegger, and G. Stock. *J. Chem. Phys.*, 126:244111, 2007.

[19] A. Altis, O. Moritz, P. H. Nguyen, R. Hegger, and G. Stoc. *J. Chem. Phys.*, 128:245102, 2008.

[20] R. Hegger, A. Altis, P. H. Nguyen, and G. Stock. *Phys. Rev. Lett.*, 98:028102, 2007.

[21] P. H. Nguyen. *Proteins*, 65:898, 2006.

[22] P. H. Nguyen. *Proteins*, 67:579, 2007.

[23] Juergen J. Graf, P. H. Nguyen, Stock G, and H. Schwalbe. *J. Am. Chem. Soc.*, 129:1179, 2007.

[24] P.H. Nguyen, M. S. Li, J. E. Staub, and D. Thirumalai. *Proc. Natl. Acad. Sci. USA*, 104:111–116, 2007.

[25] L. Riccardi, P. H. Nguyen, and G. Stock. *J. Phys. Chem.*, 113:16660, 2010.

[26] P. H. Nguyen, P. Derreumaux, and G. Stock. *J. Phys. Chem. B*, 113:9340, 2009.

[27] S. Mukamel and R. M. Hochstrasser. *Chem. Phys.*, 266:No. 2,3, 2001 (special issue).

[28] P. Hamm, M. Lim, and R. M. Hochstrasser. *J. Phys. Chem. B*, 102:6123, 1998.

[29] S. Woutersen, Y. Mu, G. Stock, and P. Hamm. *Proc. Natl. Acad. Sci. U.S.A.*, 98:11254, 2001.

[30] K. A. Peterson, C. W. Rella, J. R. Engholm, and H. A. Schwettman. *Phys. Chem. B*, 103:557, 1999.

[31] Q. Shi and E. Geva. *J. Chem. Phys.*, 118:7562, 2003.

[32] E. Lindahl, B. Hess, and D. van der Spoel. *J. Mol. Mod.*, 7:306–317, 2001.

[33] R. Schinke. *Photodissociation Dynamics*. Cambridge University Press, Cambridge, 1993.

[34] M. Buchner, B. M. Ladanyi, and R. M. Stratt. *J. Chem. Phys.*, 97:8522, 1992.

[35] W.F. van Gunsteren, S. R. Billeter, A. A. Eising, et al. *Biomolecular Simulation: The GROMOS96 Manual and User Guide.* Vdf Hochschulverlag AG an der ETH, Zurich, 1996.

[36] S. Spoerlein, H. Carstens, C. Renner, et al. *Proc. Natl. Acad. Sci. USA.*, 99:7998–8002, 2002.

[37] P. H. Nguyen, Y. Mu, and G. Stock. *Proteins*, 60:485, 2005.

[38] J. Bredenbeck, J. Helbing, A. Sieg, et al. *Proc. Natl. Acad. Sci. USA.*, 100:6452, 2003.

[39] A. E. Garcia. *Phys. Rev. Lett.*, 68:2696–2699, 1992.

[40] A. Hyvarinen, J. Karhznen, and E. Oja. *Independent component analysis*. Wiley, New York, 2001.

[41] F. Chiti and C. M. Dobson. *Annu. Rev. Biochem.*, 75:333–366, 2006.

[42] J. Nasica-Labouze, O. Berthoumieu, P. H. Nguyen, et al. *Chem. Rev.*, page DOI: 10.1021/cr500638n, 2015.

[43] P. H. Nguyen and P. Derreumaux. *Acc. Chem. Res.*, 47:603–611, 2014.

[44] D. Matthes, V. Gapsys, and B. L. de Groot. *J. Mol. Biol.*, 421:390–416, 2012.

[45] D. Matthes, V. Gapsys, V. Daebel, and B. L. de Groot. *PLoS ONE*, 6:e19129, 2011.

[46] F. Baftizadeh, X. Biarnes, F. Pietrucci, F. Affinito, and A. Laio. *J. Am. Chem. Soc.*, 134:3886, 2012.

[47] P. H. Nguyen and P. Derreumaux. *J. Phys. Chem B*, 117:5831–5840, 2013.

[48] M. R. Sawaya, S. Sambashivan, R. Nelson, et al. *Nature*, 447:453–457, 2005.

[49] T. R. Serio, A. G. Cashikar, A. S. Kowal, et al. *Science*, 25:1317–1321, 2000.

[50] S. Abeln, M. Vendruscolo, C. Dobson, and D. Frenkel. *PloS ONE*, 9:e85185, 2014.

[51] F. Sterpone, S. Melchionna, P. Tuffery, et al. *Chem. Rev. Soc.*, 43:4871, 2015.

[52] T. T. Trand, P. H. Nguyen, and P. Derreumaux. *J. Chem. Phys.*, 144:205103, 2016.

[53] B. Tarus, T. Tran, J. Nassica-Labouze, et al. *J. Phys. Chem B*, 119:10478–10487, 2015.

[54] B. Tarus, P. H. Nguyen, O. Berthoumieu, et al. *Eur. J. MED. CHEM.*, 91:43–50, 2015.

[55] O. Berthoumieu, P. H. Nguyen, M. del Castillo-Frias, et al. *Chemistry- a European journal*, DOI: 10.1002/chem.201500888, 2015.

[56] R. Scherzer-Attali, R. Pellarin, M. Convertino, et al. *PLOS One.*, 5:e11101, 2010.

[57] E. Mikros, D. Benaki, E. Humpfer, et al. *Angew. Chem. Int. Ed. Engl*, 40:3603e3605, 2001.

[58] B.D. Moore, P. Chakrabarty, Y. Levites, et al. *Alzheimers Res. Ther.*, 4:18, 2012.

[59] E. Martineau, J.M. de-Guzman, L. Rodionova, et al. *J. Am. Soc. Mass. Spectrom*, 21:1506e1514, 2010.

[60] T. Jonsson, J. K. Atwal, S. Steinberg, et al. *Nature*, 488:96–99, 2012.

[61] G. DiFede, M. Catania, M. Morbin, et al. *Science*, 323:1473–1477, 2009.

[62] P. H. Nguyen, B. Tarus, and P. Derreumaux. *J. Phys. Chem. B*, 118:501, 2014.

[63] P. H. Nguyen, B. Tarus, and P. Derreumaux. in preparation.

[64] M. H. Viet, P.H. Nguyen, P. Derreumaux, and M.S. Li. *ACS Chem. Neuroscience*, 5:646, 2014.

[65] M. H. Viet, P.H. Nguyen, N. S. Tung, M.S. Li, and P. P. Derreumaux. *ACS Chem. Neuroscience*, 4:1446, 2013.

[66] P. M. Truong, M. H. Viet, P.H. Nguyen, C. K. Hu, and M.S. Li. *J. Phys. Chem. B*, 118:8972, 2014.

[67] T. Kawasaki, J. Fujioka, T. Imai, and K. Tsukiyama. *The Protein Journal*, 31:710–716, 2012.

[68] T. Kawasaki, J. Fujioka, T. Imai, K. Torigoe, and K. Tsukiyama. *Lasers in Medical Science*, 29:1701–1707, 2014.

[69] T. Kawasaki, T. Imai, and K. Tsukiyama. *Journal of Analytical Sciences, Methods and Instrumentation*, 4:9–18, 2014.

[70] T. Kawasaki, T. Yaji, T. Imai, T. Ohta, and K. Tsukiyama. *American Journal of Analytical Chemistry*, 5:384–394, 2014.

[71] M. H. Viet, P. Derreumaux, M. S. Li, et al. *J. Chem. Phys.*, 143:155101, 2015.

[72] M. H. Viet, P. M. Truong, P. Derreumaux, et al. *Phys. Chem. Chem. Phys*, 17:27275, 2015.

[73] M. H. Viet, V. O. Nguyen, P. Derreumaux, et al. *Phys. Chem. Chem. Phys.*, 10.1039/C5CP07711G, 2016.

[74] F. Wang and D. P. Landau. *Phys. Rev. Lett.*, 86:2050, 2001.

[75] P. H. Nguyen, E. Mittag, A. E. Torda, and G. Stock. *J. Chem. Phys.*, 124:154107, 2006.

[76] P. H. Nguyen. *J. Chem. Phys.*, 132:144109, 2010.

[77] P. H. Nguyen, Y. Okamoto, and P. Derreumaux. *J. Chem. Phys.*, 138:061102, 2013.

[78] T. Zhang, P. H. Nguyen, J. Nassica-Labouze, Y. Mu, and P. Derreumaux. *J. Phys. Chem. B*, 119:6941, 2015.

[79] A. P. Lyubartsev, A. A. Martinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov. *J. Chem. Phys.*, 96:1776–1783, 1992.

[80] E. Marinari and G. Parisi. *Europhys. Lett.*, 19:451–458, 1992.

[81] M. H. Viet, P. Derreumaux, and P. H. Nguyen. *J. Chem. Phys.*, 143:021101, 2015.

[82] P. H. Nguyen, M. S. Li, and P. Derreumaux. *J. Chem. Phys.*, 140:094105, 2014.

[83] D. Chandler. *Introduction to modern statistical mechanics*. Oxford university press, New York, 1987.

[84] K. A. Dill. *Biochemistry*, 29:7133–7155, 1990.

[85] R.S. Spolar and M.T. Record Jr. *Science*, 263:777, 1994.

[86] M. Karplus and J. N. Kushick. *Macromolecules*, 14:325, 1981.

[87] P. H. Nguyen. *Chem. Phys. Lett.*, 468:90, 2009.

[88] H. H. Rugh. *Phys. Rev. Lett.*, 78:999, 1997.

[89] P. H. Nguyen and P. Derreumaux. *Phys. Chem. Chem. Phys.*, 14:877–886, 2012.

Acknowledgments

I am very grateful to all my supervisors for special supports that they had always gave to me along the years. I must in particular thanks Dr. Sandro Scandolo who introduced me to the field of computer simulation of condensed matter physics during my first difficult days in Italy. Moving to Germany, I received best supports from my supervisor Professor Friederike Schmid, and I am very grateful for her persistence and constructive feedback throughout my PhD, particular during those times where I felt defeated by my research. I must thank her for having trained me so deeply in the field of soft condensed matter simulations. I especially want to thank Professor Gerhard Stock for the opportunity he gave me to enter in a new field of research in biophysical chemistry and his continuous encouragement during my postdoctoral.

I especially want to thank all of my collaborators for their expertise, friendship and valuable ideas during my career. I especially mention Professor Mai Suan Li who introduced me to the field of protein aggregation, and open a long-term fruitful collaboration. We have shared many interesting discussion on the general topics of physical chemistry. A special thanks to Professor Peter Hamm for his collaboration on the picosecond Infrared experiments and Professor Harald Schwalbe for NMR experiments. I would like to express my thanks to Dr. Man Hoang Viet for working together as well as for discussion and sharing many interesting ideas.

I would like to express my gratitude to Professor Philippe Derreumaux for the opportunity to join the Laboratoire de Biochimie Théorique as CR1 and start fruitful and productive collaboration. I am really lucky to have had opportunity to work with him as a world-leader in the field of computer simulation of protein folding and protein aggregation.

Finally, I would like to thank all members of the jury committee for agreeing to judge this work. I would like to express my thanks to Professor Pedro de Oliveira for his administrative efficiency in the HDR application.