Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Alexander Louis Tournier
born in Bebington (UK)
oral examination:

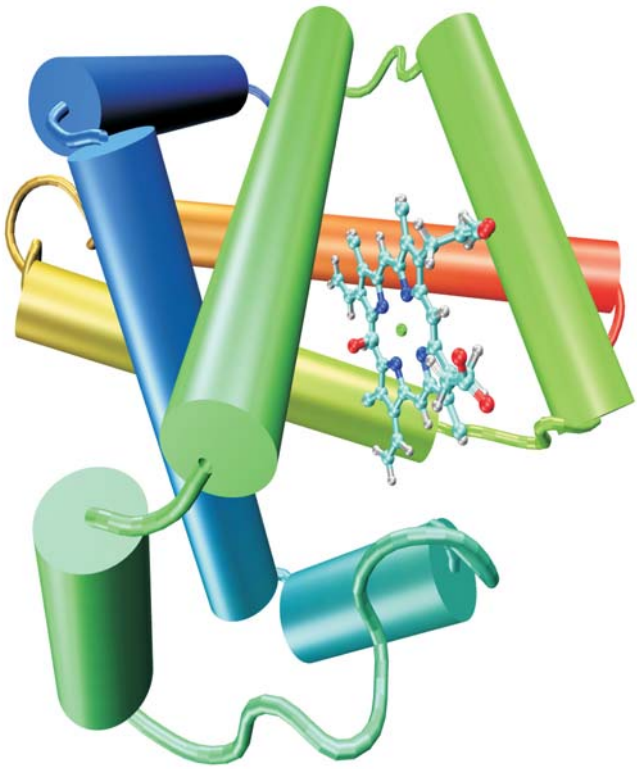
The Role of the Solvent in the Protein Dynamical Transition

Referees:

Prof. Dr. Jeremy C. Smith

Prof. Dr. Jörg Langowski

The Role of the Solvent in the Protein Dynamical Transition



Alexander Louis Tournier

ABSTRACT

Experimental and computer simulation studies have revealed the presence of a transition in the dynamics of hydrated proteins around 220 K. This transition has been compared with that of a glass phase transition. It manifests itself by a nonlinear behavior in the temperature dependence of the average atomic mean-square displacements and involves an increase of the amplitude of protein dynamics. This increase in flexibility has been correlated with the onset of protein activity. In this thesis, the mechanisms behind the protein dynamical transition are explored using molecular dynamics simulations and neutron scattering experiments.

The driving force behind the protein transition is investigated by performing simulations of myoglobin surrounded by a shell of water. A dual heatbath simulation method is used in which the protein and solvent are held at different temperatures, and sets of simulations are performed varying the temperature of the two components. The results show that the protein transition is driven by a dynamical transition in the hydration water that induces increased fluctuations primarily in side-chains in the external regions of the protein. The water dynamical transition involves activation of translational, but not rotational, diffusion and occurs even in simulations where the protein atoms are held fixed.

In order to determine the protein motions involved in the transition, longer molecular dynamics trajectories are decomposed using principal component analysis. The results indicate that the nonlinearity in mean-square displacement arises from only a very small number of principal components. These components, activated by the solvent:surface interaction, describe collective dynamics propagated through to the interior of the protein. The onset of the transition at $\sim$180 K is characterized by the appearance of a single double-well mode involving a global relative motion of two rigid-body groups of helices. As the temperature is raised a few more multiminimum and quasiharmonic principal components successively appear.

Finally, experimental results from neutron scattering on xylanase in solution at varying methanol concentrations reveal that the protein dynamics is strongly influenced by the dynamics of its surrounding solvent on short timescales. On longer timescales the results indicate the presence of a collaborative effect between the protein surface and the solvent which lowers the freezing temperature of the protein hydration layer. All together, the results indicate that the protein hydration shell plays a central role in the appearance of the transition in the temperature dependence of protein dynamics.

ZUSAMMENFASSUNG

Experimentelle Untersuchungen, wie auch Computersimulationen, zeigen markante Änderungen des dynamischen Verhaltens hydratisierter Proteine bei einer Temperatur von $\sim$ 200 K. Die experimentellen Beobachtungen zeigen charakteristische Gemeinsamkeiten mit dem Glasübergang komplexer Systeme und in Analogie spricht man vom Protein-Glasübergang. Kennzeichnend für diesen Übergang ist der nichtlineare Temperaturverlauf der mittleren quadratischen Auslenkung des Proteins. Bei Temperaturen über 200 K ist ein deutliches Ansteigen der Amplituden zu beobachten. In der vorliegenden Arbeit werden die Mechanismen dieses Übergangs mittels Molekulardynamik Simulationen und Neutronenstreuexperimenten untersucht.

Die Ursachen und Charakteristika des Protein-Glasübergangs werden anhand des Proteins Myoglobin untersucht. Mit Hilfe einer Doppeltenwärmebad-Simulation können Protein und die umgebende Wasserhülle auf unterschiedlichen Temperaturen gehalten werden. Dies ermöglicht, den Temperaturverlauf dynamischer Prozesse in Wasserhülle und Protein unabängig voneinander zu kontrollieren und gegenseitige Wechselwirkungen zu untersuchen. Die Ergebnisse der Simulationen zeigen, dass der Protein Glasübergang durch Änderungen im dynamischen Verhalten der Wassermoleküle verursacht wird. Dies betrifft vor allem die Translationsbewegung der Wassermoleküle, während Rotationen durchgehend normales Temperaturverhalten zeigen. Das Einsetzen der Translationsbewegungen führt zu erhöhten Fluktuation vor allem der Seitenketten und oberflächennahen Bereiche des Proteins.

Die Simulationen werden mit Hilfe einer Hauptkomponentenanalyse in charakteristische Bewegungsmoden zerlegt. Diese Analyse zeigt, dass die beobachtete Nichtliniarität der mittleren quadratischen Auslenkung von einer sehr geringen Anzahl an Moden verursacht wird. Diese Moden werden durch Wechselwirkungen zwischen Proteinoberfläche und Wasserhülle aktiviert und beschreiben kollektive Bewegungen, die sich bis ins innere des Proteins ausbreiten. Der beobachtete Übergang bei $\sim$ 180 K ist gekennzeichnet durch die qualitative Änderung einer einzigen Mode von einem quasi-harmonischen zu einem anharmonischen doppelminimum Verlauf. Mit steigender Temperatur zeigen weitere Moden diesen Übergang zu anharmonischem Profil.

Anhand von Lösungen des Proteins Xylanase in unterschiedlichen Methanolkonzentrationen, wurde die Abhängigkeit der Proteindynamik vom umgebenden Lösungsmittel mittels Neutronenstreuung untersucht. Auf kurzen Zeitskalen ($<$ 100 ps) kann eine deutliche Abhängigkeit der Proteindynamik vom umgebenden Lösungsmittel beobachtet werden. Langsamere Prozesse ($\sim$ 1 ns) deuten auf Wechselwirkungen zwischen Proteinoberfläche und Lösungsmittel hin, die das Gefrieren des Lösungsmittels in unmittelbarer Nähe der Proteinoberfläche verhindern.

Die Ergebnisse dieser Arbeit verdeutlichen die Bedeutung der Wasserhülle für ein Verständnis dynamischer Prozesse in biologischen Makromolekülen.

CONTENTS

ACKNOWLEDGEMENTS

Writing up this PhD thesis would not have been possible without the help of a number of people. First of all, my many thanks go to Professor Jeremy Smith head of the Computational Molecular Biophysics group and my PhD supervisor. Under his supervision I was able to explore my own ideas some of which actually turned out to give very interesting results.

Special thanks go to Andrea Vaiana and Torsten Becker for essential discussions and support in times of existential doubts and utter despair in the face of seemingly intractable problems. I also thank my dear roommates Laura Vulpescu, Durba Sengupta and Edda Kloppmann for great fun, support and many inspiring discussions. My thanks also go to the whole CMB group, it was very nice not only to work in but also to party with. I would also like to acknowledge the work of Jiancong Xu who wrote her C-practical under my supervision. Some of her results have been included in this thesis.

I thank the Deutsche Forschungsgemeinschaft (DFG) for funding part my research.

I also thank my brother, Jacques-Donald Tournier for helping me out with setting up the Latex formatting for this document. I also thank Dudu, Lars Meinhold and Torsten for proofreading and ideas in writing my thesis.

Very special thanks go to my parents, Bernard and Joan Tournier, who supported me all through my long studies.

PUBLICATIONS ARISING FROM THIS WORK

[1] TOURNIER, A. L., AND SMITH, J. C. Translational hydration water dynamics drives the protein glass transition. *Biophys. J.*, 2003, **85**(3).

[2] TOURNIER, A. L., XU, J., AND SMITH, J. C. Solvent caging of internal motions in myoglobin at low temperatures. *PhysChemComm*, 2003, **6**(2), 6–8.

[3] TOURNIER, A. L., HUANG, D., SCHWARZL, S. M., FISCHER, S., AND SMITH, J. C. Time-resolved computational protein biochemistry: solvent effects on interactions, conformational transitions and equilibrium fluctuations. *Faraday Discuss*, 2003, **122**, 243–51.

[4] TOURNIER, A. L., AND SMITH, J. C. Principal components of the protein dynamical transition. *Phys Rev Lett*, 2003, accepted for publication.

[5] BECKER, T., FISCHER, S., NOE, F., TOURNIER, A. L., ULLMANN, M. G., AND SMITH, J. C. Protein dynamics: Glass transition and mechanical function. *Adv. Sol. State Phys.*, 2003, **43**, 677–94.

[6] SMITH, J. C., COURNIA, Z., TALY, A., TOURNIER, A. L., MIHAILESCU, D., AND ULLMANN, M. G. Conformational transitions in proteins and membranes. In *NATO ASI series C: Mathematical and Physical Sciences*. 2003, in press.

CHAPTER 1

INTRODUCTION

CONTENTS

A BRIEF INTRODUCTION

Higher forms of life rely nearly entirely on the intricate palette of functions proteins can fulfill to perform the many task essential to cell survival. Proteins can be classified according to their biological activity as: enzymes or transport, storage, motile, structure, defense and regulatory proteins. Insight into how these proteins function leads to an understanding of the molecular basis of life. Hence, they have been the focus of intense research ever since they were first isolated at the end of the 19^{th} century.

In the last decades the picture of a fixed, rigid protein structure inherited from crystallographic studies has given way to a more flexible and dynamical view of the way proteins operate. Very specific motions are now seen to be essential to protein function. Although proteins have been the subject of intense research for decades in biology, chemistry and physics, understanding internal proteins dynamics remains a challenge.

An interesting physical property of proteins was first noticed in the 1970's when experiments showed that proteins undergo a change in their dynamical properties between 180 K to 220 K. The origin of this transition in dynamics has aroused much attention ever since and different theories have emerged to explain it. In particular, the striking analogy with the transition present in glass-forming materials has been the subject of much debate.

This dynamical transition has since then been studied through numerous experiments as well as through computer simulations which the extraordinary progress in computer power has made possible. Simulations of proteins offer the advantage of giving access to atomistic details and short timescales.

This dynamical transition was first observed in myoglobin, a globular oxygen transport protein. Such globular proteins in their natural state are surrounded by water in which they are able to fold and function correctly. The role of the solvation shell around proteins has, since then, been shown to play an essential role in the dynamical transition.

The study of the dynamical transition in hydrated proteins is the subject of the present thesis. In doing this, a lot of importance will be drawn to the essential role played by protein solvation. Another important aspect of the present study is the characterization of the protein dynamics involved in this dynamical transition. In order to achieve this, molecular dynamics techniques as well as neutron scattering experiments were employed.

The next section in this introductory chapter will present in more detail the experiments which revealed the dynamical transition in proteins. The molecular dynamics approach to the problem will then be presented, followed by an overview of the protein-glass analogy. The influence of the dynamical transition on protein function will then be discussed. This will be followed by a review of the role played by the solvent in proteins in general, and in the protein dynamical transition in particular. Finally the specific ques-

tions this thesis aims to answer will be presented along with an general outlook of this study.

1.1 EXPERIMENTAL EVIDENCE OF THE DYNAMICAL TRANSITION IN PROTEINS

Allosteric proteins such as hemoglobin show a conformational change upon ligand binding (eg: oxygen). The discovery of such allosteric proteins promoted the idea that conformational changes are an essential part of the way protein function.[1, 2] The study of protein dynamics thus reveals important aspects of the way proteins function.

1.1.1 FIRST DISCOVERY: MÖSSBAUER SPECTROSCOPY

Several techniques have appeared in order to investigate protein dynamics, one of the first such techniques is Mössbauer spectroscopy. Mössbauer obtained the 1961 Nobel prize in physics for his work on the nuclear effect that bears his name. This effect builds upon the fact that nuclei in atoms undergo a variety of energy level transitions which are associated with absorption and emission of gamma rays. Changes in the environment, electronic or magnetic, will have an influence on the energy levels which can be measured. The energy lines corresponding to the transition are very fine and a variety of methods use this effect to study the interaction of the nuclei with its environment. In proteins the very sharp nuclear transitions allied with the reasonable absorption cross-section of ^{57}Fe is used in investigating the dynamics of iron atoms. Measurements of the mean square displacement of the iron nuclei can be obtained by measuring the spread of the spectral line due to Doppler shift. Due to the energy resolution of the apparatus, the technique is sensitive to motions on timescales slower than $\sim$4 ps.

In 1971, using Mössbauer spectroscopy, Fritz Parak first reported a transition in the Mean-Square Deviation (MSD) of the iron site of myoglobin at around 200 K.[3] It took about 10 years for the study to be repeated and the effect investigated.[4–7] A sharp transition was seen in the MSD of the iron-containing proteins under investigation at temperatures ranging from 180 K to 220 K. A characteristic plot of the data obtained from Mössbauer can be seen in figure 1.1, the MSD is seen to increase linearly from 0 K up to $\sim$200 K at which point a break occurs and the MSD continues to rise with a much steeper slope. Such behavior is characteristic of the protein dynamical transition. Investigations of non iron-containing protein was made possible using neutron scattering, also the dynamics under scrutiny were not restricted to the neighborhood of the iron atoms.

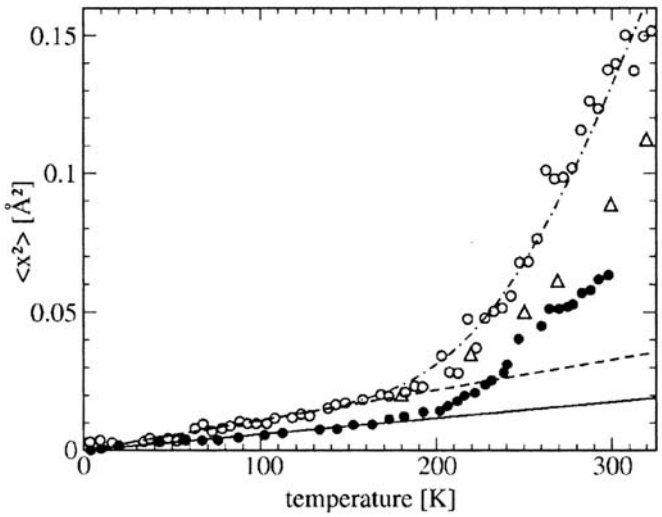


Figure 1.1: Mean square displacement from incoherent neutron scattering (Doster and Settles 1998[10]); ○ : time resolution 80 ps; △ : time resolution 12 ps. ● : Mean square displacements from Mössbauer absorption spectroscopy on Myoglobin crystals (from Parak 2003[11])

1.1.2 NEUTRON SCATTERING

In the 1980s, with the construction of the high flux reactor at the Institut Laue-Langevin in Grenoble, neutron scattering experiments were able to study the dynamics of proteins and were in an ideal position to investigate the temperature dependence of protein dynamics.[8, 9] Incoherent neutron scattering is a very versatile technique which enables many interesting experiments to be undertaken. Incoherent neutron scattering takes advantage of the fact that hydrogen has a very large incoherent scattering cross section compared to all other elements (by a factor 10 or more). Hydrogen being very abundant in biological systems the scattering originating from hydrogens dominates the signal. In this way selective deuteration enables the investigation of the part of the system left undeuterated. Incoherent neutron scattering uses thermal neutrons which are scattered by the atoms in the system thus carrying information about the dynamics of the atoms in the system. The neutron in such experiments being of thermal energies are particularly sensitive to thermal motions in the sample. This makes neutron scattering a particularly interesting technique to study protein dynamics. In particular this technique allows the calculation of the atomic mean-square fluctuation over a whole protein. A more in depth presentation of the theory behind incoherent neutron scattering is made in the methods chapter.

In 1989, Doster and Cusack were the first to report the dynamical transition in protein using incoherent neutron scattering.[12] The presence of the dynamical transition was reported for bacterioohodopsin at $\sim$230 K in 1993,[13] and in 1997.[14] It was also reported to be present in α-Amylase around 200 K in 1999.[15] The presence of a dynamical transition at $\sim$220 K was reported for the enzyme Xylanase in 2000.[16] Figure 1.1 shows the dynamical transition reported using neutron scattering experiments. A transition between two linear regimes is seen, below and above the transition temperature of $\sim$200 K. This transition is very similar to that seen using Mössbauer spectroscopy. The data from both

Mössbauer spectroscopy and neutron scattering show a linear temperature dependence below ∼200K followed by sharp transition at ∼200 K and a linear regime of much increased slope above ∼200 K.

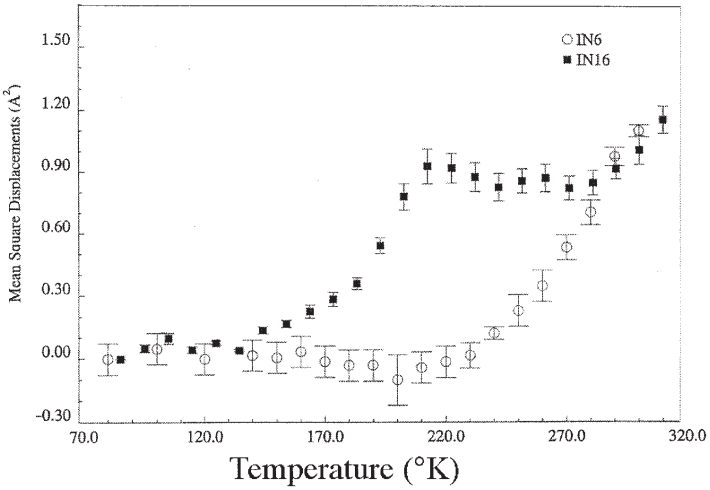


Figure 1.2: Mean square displacement from incoherent neutron scattering obtained on IN6 and IN16 detectors at the ILL in grenoble. IN6 is sensitive to motions on time-scales of ∼100ps whereas IN16 is sensitive to motions on time-scales up to ∼5ns.(Adapted from Daniel *et al* 1999[17])

An interesting effect has been reported using different instruments with different energy resolution and therefore able to perceive motions on different timescales.[17] The study compares the results from two neutron scattering spectrometers: IN6 and IN16 at the Institut Laue-Langevin reactor in Grenoble. Whereas IN6 probes motion on time scales faster than ∼100ps, those detected by IN16 extend to ∼5ns. As can be seen in figure 1.2, the IN6 data is very similar to that presented in Figure 1.1, the data from IN16, however, is very different and presents an early transition to increased dynamics at 160-170K and then plateaus between 180K to 280K to increase thereafter. The meaning of this timescale dependence is not completely elucidated and is the subject of research at the present time.

1.1.3 COMPLEMENTARY RESULTS FROM X-RAY DIFFRACTION AND FTIR

Dynamics of proteins were first studied using X-ray diffraction by Frauenfelder *et al* in 1979.[18] X-ray crystallographic refinement yields the Debye-Waller factors of individual atoms which in turn can be used to calculate the mean square fluctuation of individual atoms. Using this effect the presence of a dynamical transition in Ribonuclease A was reported at 180-200K in 1992.[19]

Proteins show a certain degree of flexibility around their average structure as revealed by X-ray crystallographic refinement.[18] Protein do not remain in a unique state of minimum free state but fluctuate around it through a large number of conformational substates which are nearly isoenergetic.[20, 21] These conformational substates are separated by free

energy barriers which have to be overcome during conformational changes. An illustration of such a free energy landscape is presented in figure 1.3.

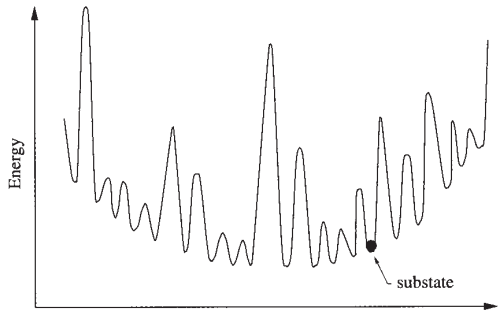


Figure 1.3: Simplified view of the multidimensional energy landscape of a protein. The protein fluctuates from substate to substate in the region around the global minimum of the energy (adapted from Frauenfelder *et al*[22]).

More recent studies performed in the group led by U. Nienhaus have combined X-ray crystallographic refinement with Fourier Transform Infra Red (FTIR) to study ligand rebinding kinetics in myoglobin.[23, 24] Looking at the IR bands associated with the bound CO, FTIR can follow the CO as it rebinds after photodissociation. X-ray gives information as to where the CO goes upon photodissociation. Using these techniques it was shown that below 180 K photodissociated ligands migrate to specific sites within an internal cavity of an essentially immobilized, frozen protein, from which they subsequently rebind by thermally activated barrier crossing. On photodissociation above 180 K, ligands can escape from the distal pocket, aided by protein fluctuations that transiently open exit channels.

1.2 MOLECULAR DYNAMICS SIMULATIONS

1.2.1 MOLECULAR DYNAMICS: A BRIEF HISTORY

With the advent of modern powerful computer came the idea of simulating interacting atoms from first principles. Molecular dynamics (MD) simulation consists in the integration of newtonian equations of motions for a system of interacting atoms, such as a protein, under a particular set of forces (van der Waals, electrostatic, etc ...). The outcome of the simulation consists in the trajectory of all the atoms during the time covered by the simulation. This trajectory can then be analyzed to reach a new understanding of the system based on the atomistic description of the system offered by MD. The trajectory also enables the calculation of observables which can then be compared to experimental values, such as neutron scattering spectra.

Adler and Wainwright in 1957[†] were the first to report the results of an MD simulation. They simulated a hard sphere system where particles only interact via instantaneous

[†]Reprints of early papers on MD and work in the area of computer simulation of liquids and solids published up to 1986 can be found in the book edited by Ciccotti *et al*[25]

collisions. In 1964, MD simulations enabled the study of liquid argon[†]. McCammon *et al* were the first to use MD to simulate a protein in 1977, using it for the study the protein BPTI (Bovine Pancreatic Trypsin Inhibitor).[26]

Figure 1.4: PC clusters are now commonly used to perform molecular dynamics simulations. Current software such as CHARMM scale well on 8 or even 16 processors in parallel, effectively dividing computation times by a factor of 6 or better.

1.2.2 VALIDATION OF MD: COMPARISON WITH EXPERIMENT

Molecular modelling relies extensively on the accuracy of the model or *force field* used to approximate the forces between the atoms in the system. Different force fields have been developed over the years, one such, which is widely used, and used in this work, is the CHARMM22 force field which describes forces acting between atoms in proteins.[27] Mathematical details of this force field will be presented in the chapter 2.

In order to validate the force field used in our modelling and simulations comparisons with experiment were made, in particular neutron scattering experiments which give a lot of information about the dynamics. To this effect quantities extracted from neutron scattering were compared to ones obtained from computations.

The simplest calculations are performed using normal modes. Normal modes describe the essential modes of motion of a protein and gives their respective frequencies. From such calculation one can obtain quantities such as neutron scattering factor and density of states which can then be compared to those obtained directly from experiment (for a review see Smith *et al* 1991).[28]

Neutron scattering factors are calculated more precisely from MD simulations. In this way it was shown that most of the scattering data could be accounted for solely from liquid-like rigid-body motion of the protein side-chains.[29] Other studies have shown the importance of correct sampling, *i.e.*, sufficiently long simulation times for the correct reproduction of experimental results.[30] Sufficient hydration of the protein (a hydration shell consisting of $\sim$350 water molecules in the case of myoglobin) has been shown to be necessary in order to reproduce experimental results for proteins in solution.[31–33] The

way the surroundings of the protein are simulated was shown to significantly influence the dynamics of the protein. The environment most appropriate for in-solution protein simulations was shown to be a hydrated crystal or powder environment with periodic boundary conditions.[34, 35]

The validity of certain approximations made in deriving properties such as mean-square-fluctuations from experimental data has been investigated using MD. By simulating the process by which experimental results are obtained it is possible to estimate the errors made in calculating them. Recent work has shown that errors of as much as 30% are made in deriving mean-square-fluctuations from experimental data when using the current approximations.[36]

1.2.3 THE DYNAMICAL TRANSITION FROM MD SIMULATIONS

The increase in computer power has made it possible to run whole sets of simulations at different temperatures in order investigate the temperature dependence of protein dynamics. In 1989, Wong *et al* were thus the first to reproduce the dynamical transition using MD.[37] They performed simulations of ferrocytochrome c and found the dynamical transition feature to be present at $\sim$220 K. Subsequent results performed using MD showed the presence of a dynamical transition in myoglobin at temperatures of $\sim$245 K and $\sim$210 K.[38, 39] The dynamical transition was also reported in superoxide dismutase simulations at $\sim$200 K.[34] Since then many simulations have been performed and have confirmed the presence of a dynamical transition in myoglobin at $\sim$220 K using MD.[33, 36, 40–43]

1.3 THE PROTEIN-GLASS ANALOGY

The experimental evidence presented in the previous sections suggests that hydrated proteins, as single molecules, may possess the complexity necessary to exhibit cooperative dynamics comparable with those of simpler glass-forming systems.[44] In glasses the transition occurs when the system does not have enough thermal free energy to overcome the energy barrier between local minima of the energy landscape and becomes suddenly locked in a local minima which is generally not the global minima: a glass is a 'liquid that has lost its ability to flow'[45] and appears as a 'structurally disordered solid'.[46] This behavior is now well understood in terms of Mode Coupling Theory (MCT).[47] In proteins a similar phenomenon is thought to occur in which proteins are locked in a harmonic local minima below the transition and are then able to explore surrounding local minimas above the transition temperature.[44–46, 48, 49] As the temperature goes down, the protein motions are confined to subregions of the energy landscape until finally they reach a state where they are trapped in a local minimum of the energy landscape.

Standard glass forming material are classified according to their sensitivity to temperature changes. So-called "strong" liquids have a built-in resistance to structural change and they show little reorganization despite wide temperature changes. So-called "fragile" liquids have glassy structures that teeter on the brink of collapse at their glass transition temperature, T_g, and which, with little provocation from thermal excitation, reorganize to structures that fluctuate over a wide variety of different particle orientations and coordinations states.[45, 48]

Mode Coupling Theory distinguishes between two types of motions in glass forming liquids: α and β relaxations. Secondary, β, relaxations are very fast processes due to local rearrangements involving low energy barriers. They lack the cooperative character of the primary, α, relaxations corresponding to slow diffuse processes.[45, 47, 48, 50]

The β processes follow a standard Arrhenius temperature dependence:

$$k(T) = A_0 \; exp(-E/k_B T) \tag{1.1}$$

where k_B is the Boltzmann constant and T the temperature. On the other hand, the α processes, such as diffusion, are strongly hindered below the transition temperature. Their relaxation function, $\phi_r(t)$ follows a stretched exponential, Kohlrausch-Williams-Watts (KWW) behavior with respect to time:[44, 51]

$$\phi_r(t) = e^{-(t/\tau)^\beta}, \qquad 0 < \beta < 1 \tag{1.2}$$

The temperature dependence of their viscosity (as well as diffusivity and relaxation times) can correctly be modelled by the Vogel-Fulcher-Tamman-Hesse (VFTH) equation:

$$\eta(T) = \eta_0 \exp[\frac{1}{K_f}\frac{T_0}{T - T_0}] \tag{1.3}$$

where K_f is a measure of the kinetic fragility of the material: more fragile liquids have larger K_f values.[46, 52]

A key characteristic of standard glass-forming systems, which is not seen in proteins, is their well defined glass transition temperature, T_g, as defined by a marked jump in heat capacity, C_p.[44] This common jump in heat capacity, C_p, is absent in proteins. This usually only happens in the case of very strong glass-forming liquids such as SiO_2, in which case the jump is much reduced. However, proteins have been shown to conform to mode coupling theory[53] usually a sign of a fragile glass-former. In proteins the jump in C_p seems to be much more spread over a range of temperatures. This in turn could be explained by considering that proteins possess not one but several α relaxation processes possibly activated at different temperatures which would make the transition much more

gradual than in standard glass forming materials.[46, 54]

1.4 THE DYNAMICAL TRANSITION *vs* PROTEIN FUNCTION

The transitions between conformational substates are generally considered to be essential to the function of proteins. According to this view, proteins would stop functioning below the dynamical transition temperature where the protein is locked in a single conformational substate. For this reason, the link between the presence of the dynamical transition in protein and their function has been the focus of attention.

In 1980, Parak *et al* showed that the chromatophores in *Rhodospirillum rubrum* membrane ceases to function below the dynamical transition at 180 K.[4] Subsequent work using X-ray crystallography showed that Ribonuclease A loses function below the dynamical transition at 220 K.[55] Such an effect has also been reported for crambin using X-ray scattering.[56] However, more recent studies indicate that the dynamical transition might not signify the loss of function below the transition temperature. Parallel measurements of activity and dynamics of glutamate dehydrogenase and xylanase showed that the function of these enzymes was not affected by the dynamical transition.[57, 58] Figure 1.5 shows data from Dunn *et al* 2000, a transition is clearly seen in the dynamics of xylanase ∼-60 K (213 K) whereas no such transition is seen in the activity measurements. The reason why some protein are sensitive to the dynamical transition while some seem not to be is not yet clear. A possible explanation has been put forward according to which some functional motions would be 'slaved' to the solvent dynamics and sensitive to the dynamical transition and some other insensitive to the solvent dynamics and thereby unaffected by the dynamical transition.[49] This aspect relating to the role of the solvent will now be explored in greater detail.

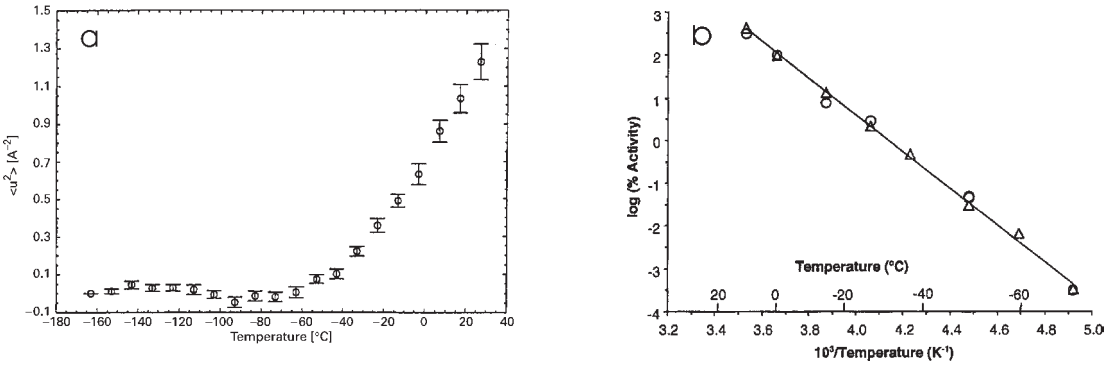


Figure 1.5: a) Effect of temperature on the activity of xylanase in 70% aqueous methanol; b) Effect of temperature on the dynamics of xylanase in fully deuterated 70% aqueous methanol, as measured by neutron scattering (from Dunn *et al* 2000[58]).

1.5 SOLVENT/COSOLVENT PROPERTIES AND EFFECTS ON PROTEIN STRUCTURE AND DYNAMICS

The presence of solvent is essential for proteins to fold and function properly. For this reason, the intricate interactions of proteins with their surrounding solvent has been and remains the focus of much attention. Apart from water, a number of cosolvents such as ions also play a role in protein folding and function.[59] In recent years a number of cosolvents, not usually present in nature, have found a wide range of uses in Biology: as cryosolvent, protein denaturant, protein fold stabilizer, protein function modulators, and more.[60] The interesting properties of these cosolvents will now be reviewed. Finally, in order to simulate protein hydration correctly, a number of water models have been proposed over the years. The present work focuses on the TIP3P model for which the CHARMM parameters have been optimized.[61]

1.5.1 HYDRATION EFFECTS ON PROTEIN STRUCTURE AND DYNAMICS

STRUCTURAL EFFECTS OF HYDRATION

A number of studies have shown the effect on protein structure and dynamics as hydration levels increase from the anhydrous state to solution (see Ref. 62 and references therein). Studies on lysozyme have shown four stages of hydration. In the first stage, from 0 to 0.07 h (g/g), proton redistribution occurs, mediated by the presence of mobile water molecules. This represents less then 55 waters molecules per protein molecule. In stage two (0.07 to 0.30 h, *i.e.*less than 250 water molecules per protein molecule) water binds to charged and polar sites. In stage three (0.30 to 0.50 h) the protein surface is gradually covered by a water monolayer. In stage four, above 0.50 h (more than 400 water molecules per protein molecules), the hydration shell builds further, the protein displays dynamical properties of a fully hydrated proteins.[33]

DYNAMICAL EFFECTS OF HYDRATION

Dry proteins are, at physiological temperatures, relatively rigid compared to hydrated proteins. The increase in flexibility due to hydration is generally thought to be important for function.[62] Neutron scattering experiments have shown that dry protein exhibits little motion at 300 K over that expected from a rigid body.[16, 63] Further Neutron scattering experiments on lysozyme, parvalbumine and bovine pancreatic trypsin inhibitor have shown protein dynamics to increase gradually upon hydration.[64–66] Using MD simula-

tions, the protein hydration shell has been shown to soften diffusive motions in proteins as one gradually moves away from the protein core.[67] In the absence of water protein sidechains form tight hydrogen bonds between them thus turning the protein into a rigid body. Upon hydration these hydrogen bonds are transferred to water molecules and the sidechains diffuse freely in the solvent.[60, 68, 69]

STRUCTURE OF WATER AT PROTEIN INTERFACES

The protein hydration layer has been studied using experimental techniques such as X-rays and neutron scattering[68, 70–76] as well as MD simulations.[42, 77–80] The protein first hydration shell has been shown to be more dense than bulk water by as much as 10%.[70, 78] This surface effect was studied in detail using MD simulations, and was found to be less important than previously thought due to the way densities are calculated close to surfaces. In this study the increase in density was then estimated at $\sim$5% around the protein.[80] The water orientation in this hydration layer was shown to be strongly correlated to the local topology of the protein surface and the electrostatic field present at that surface. Water diffusion was shown to be increased at the protein surface (up to 6 Å away), this effect being most important over hydrophobic patches and in protein cavities.[78] No correlation was found between diffusion properties and the neighboring atom or residue types.[74] Water has been found to form large-scale networks over charged and polar regions of the protein.[60, 76] However, these structures were shown to be uncorrelated with water residence times.[81] The presence of the protein has been shown to affect solvent properties as far as 15Å away from the protein surface.[74, 78, 79] Other water structures called clathrates have been found over hydrophobic areas of protein surface, in these structures water molecules form cage like structures such as the one shown in figure 1.6 around a methane molecule.

1.5.2 EFFECTS OF COSOLVENTS ON PROTEIN STRUCTURE DYNAMICS AND FUNCTION

STRUCTURAL, DYNAMICAL AND FUNCTIONAL EFFECTS OF COSOLVENTS

In living organisms, ion concentrations and pH levels are closely monitored as they have a big impact on the folding and function of proteins (for example Calmodulin is very sensitive to Calcium concentration and will take a different fold in the presence of Calcium). Apart from these naturally present cosolvents, a number of different cosolvents are used in biology. For example urea is a protein denaturant and is thought to act by destroying the natural tetrahedral arrangements of water, thus removing the entropic effect necessary for the hydrophobic effect which stabilizes proteins.[59] Trifluoroethanol (TFE) also

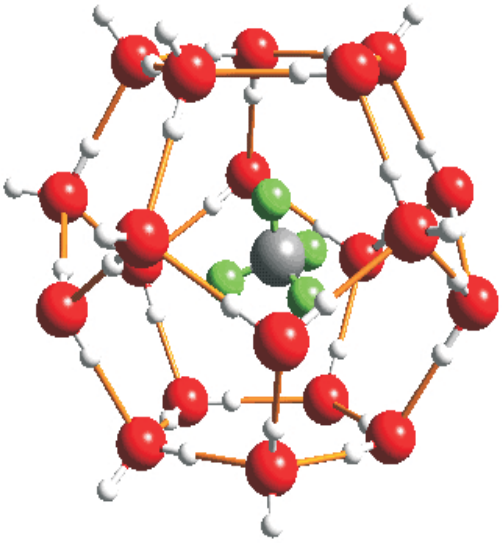


Figure 1.6: Typical clathrate water formation, in this case around a methane molecule

a protein denaturant has been found to unfold complex secondary structures such as beta strands and to stabilize and induce helical content. Dimethyl sulfoxide (DMSO) is a good hydrogen bond acceptor as well as having a more hydrophobic part is a good cryosolvent down to ∼235 K and does not seem to affect protein structure.[73] Organic solvents have also been used, such as ethanol which acts as a denaturant by dehydrating the protein surface, *i.e.* removing essential, stabilizing water molecules from the protein surface. It is also found to promote helical content.[72] Glycerol is also widely used as a cryosolvent.[63] Highly viscous cosolvents such as the sugar trehalose have been found to be very effective in preserving protein structure at low hydration and high temperatures.[82, 83] At high concentrations, the presence of organic solvents has also been found to modulate enzyme activity, altering substrate specificity and increasing thermostability. Catalysis of reaction not possible in water has been shown to be possible in organic solvents.[69] In the present work methanol was used as cryosolvent, the freezing point of 40% CD_3OD/ 60% D_2O being at ∼233 K. Methanol preserves protein structure at moderate concentration although at higher concentration it tends to denature protein structure, promoting helical content.[60]

STRUCTURE OF COSOLVENTS AT PROTEIN INTERFACES

Cosolvents have been show to interact with the protein surfaces in very specific ways.[60, 68, 72, 73, 75] Using neutron scattering techniques it has been possible to find the position of different cosolvent molecules on the protein surface. Ethanol was found to interact mostly with the hydrophobic parts of the protein,[72] whereas DMSO showed more H-bond interactions with the protein.[73] Different studies showed that different cosolvent molecules occupy similar positions on the protein surface.[72, 73, 75] Three categories of structural water molecules were found on protein surfaces: those that were present in all

structures, those that were present in some, and those too mobile to ever be resolved by diffraction measurements.[60, 84] The slightly disordered - category two - water molecules are thought to indicate potential ligand binding sites. In these areas the waters are not too tightly bound and not so loose that there is no entropic gain in removing them. This effect has been used to probe for possible binding sites on proteins surfaces.[60, 84]

PROPERTIES OF METHANOL/WATER MIXTURES

Water/cosolvent mixtures have been the focus of research using neutron scattering and small angle X-ray scattering. Water/methanol mixtures have been extensively studied at different relative concentrations.[85–88] These mixtures show excess properties: excess heat of mixing, excess partial molar volumes, excess heat capacity maximal at molar fraction of $X_m = 0.25$. These excess properties have been attributed to the inverse tendencies of water and methanol, water being dominated by hydrophilic behavior while the other is governed by hydrophobic interactions.[86] The different studies show that the hydrophobic groups of the methanol molecules tend to cluster together forming methanol clusters of 2 or more molecules. The hydrophilic group of the methanol molecules is left exposed to the solvent, forming hydrogen bonds with the water molecules and taking part in the water tetrahydral formations.[86, 88]

1.5.3 MODELLING WATER: THE TIP3P MODEL

In order to obtain accurate simulations in hydrated environment, the properties of the water must be adequately reproduced. In the present work the water molecule is modelled using to the TIP3P model.[89] Many different models exist for simulating water with different levels of accuracy and different computational costs. A general review of the water simulation field from simple empirical models to full ab-initio approaches can be found in Ref. 90.

In the TIP3P model the water molecule is simulated using three atoms: two hydrogens covalently bonded to a oxygen. The OH distance, HOH angle, van der Waals parameters and charges on the atom center of mass are parametrized so as to reproduce as closely as possible experimental results.[89] The TIP3P model reproduces satisfactorily the liquid phase properties of water such as average density and heat of vaporization.[91] The tetrahedrality of the model is considered somewhat too weak and the diffusion constant somewhat too high.[61, 91] The CHARMM22 force field was parametrized using the TIP3P model for water. The TIP3P model is therefore the preferred model for simulating water when using the CHARMM package.[61]

Other empirical water models exist such as the single point charge (SPC) model which

is identical to the TIP3P model but with different values for the parameters. The SPC model fits better the second peak in the O-O autocorrelation function (g_{OO}) of liquid water.[89] Other models such as the TIP4P model better the oxygen lone pair by moving the oxygen partial charge a parametrized distance along the monomer axis, away from the hydrogens. An extensions of the TIP4P model, the TIP5P model, places two negative charges symmetrically along the lone-pair direction. Another model extended from the SPCE model, the SPC/E model introduces polarizability into the model. These models reproduce more closely the thermodynamical properties of liquid water, however they also come at a computational cost and do not bring a significant improvement to the simulation of hydrated proteins.

1.6 ROLE OF SOLVENT IN THE DYNAMICAL TRANSITION

1.6.1 IMPORTANCE OF THE PRESENCE OF A HYDRATION SHELL FOR THE PROTEIN DYNAMICAL TRANSITION

MD simulations[36, 38] and neutron scattering experiments[92] have shown that isolated or dehydrated proteins also show dynamical transition behavior. However MD studies on dehydrated systems have placed the dynamical transition at widely differing temperature ranging from 110 K[36] to 240-250 K.[33, 38, 40] Neutron experiments showed that dry lysozyme shows no transition.[16, 63] This contrasts with the relative homogeneity of the transition temperature reported in the literature at 200-230 K for proteins in hydrated powders[12] and in cryosolvent solutions.[13, 14, 16, 19]

As early as 1989 the possible importance of the protein hydration shell in the protein dynamical transition was discussed.[12, 37] A number of experiments and simulations have indicated that when a protein is solvated the dynamical transition at $\sim$220 K is strongly coupled to the solvent. Neutron experiments found hydration to be essential to the presence of the dynamical transition feature in bacteriorhodopsin.[13, 93] MD simulations showed the dynamical transition feature in myoglobin to be much reduced in the absence of solvation.[33]

1.6.2 EFFECT OF HYDRATION ON THE PROTEIN DYNAMICAL TRANSITION

A number of studies have looked at the effect of solvation on the protein dynamical transition. Solution studies require the use of cryosolvents in order to prevent the freezing of the sample (see study by Réat *et al*[16]). Protein dynamics have been shown not to be

affected by the use of cryosolvents.[16, 94] Most experiments have been performed either on hydrated powders which do not suffer from the freezing problem or in glycerol:water mixtures. The effect of commonly used cryosolvents has been studied at different concentration levels and found not to influence the dynamical transition temperature as long as the glass transition temperature of the cryosolvent was below ∼230 K.[16]

Studies on pure glycerol have shown that glycerol stabilizes the protein and increases its dynamical transition temperature.[63] The dynamical transition of Lysozyme in pure glycerol was shown to be ∼240 K and to lower down to 210-200 K upon mixing with water.[92] It was also shown that in pure glycerol protein dynamics followed that of glycerol, the protein being slave to its glycerol environment. On the other hand, highly viscous solvents, such as trehalose, have been shown to completely suppress dynamical transition behavior.[82, 83, 95]

1.6.3 GLASS TRANSITION IN THE PROTEIN HYDRATION SHELL

The central role played by the solvent in the dynamical transition hints at a possible phase transition in the protein hydration shell which would induce the dynamical transition feature in the protein dynamics. Wong *et al* in their 1989 comparison of pure water and hydrated protein simulations were the first to indicate the presence of a dynamical transition in pure water simulations at ∼220 K.[37] A study using infrared spectroscopy on myoglobin has shown that the hydrogen bonding network in the solvent around a protein undergoes a transition ∼180 K, the H-bond network was seen to freeze-in protein motions below that temperature. Similar behavior was recently reported using neutron scattering.[96] Previous work performed using neutron scattering also showed that at low temperatures, protein motion were reduced in hydrated systems as compared to dry systems, thus demonstrating that the solvent had a caging effect on the protein motions.[97]

In 2000, Vitkup *et al* performed dual heatbath MD simulations in which the solvent and protein are held at different temperatures.[98] They were thus able to demonstrate that cold solvent strongly inhibits internal protein fluctuations while hot solvent slightly increases them. MD simulations have also shown that the water hydrogen bond network structure is rigid below ∼180 K while at the same temperature the water molecules are still able to rotate freely in their confined space.[41] This is a strong indication that the protein hydration layer undergoes a glass transition at that temperature.

By comparing rate coefficients of certain protein processes with those of solvent dielectric relaxation rate, recent work by Fenimore *et al* shows the presence of two types of processes: *slaved* and *nonslaved*.[49] Nonslaved processes are, according to their definition, all processes which are independent of the solvent dynamics, whereas slaved processes depend critically on solvent dynamics. Enzymatic activity such as bond formation would

be a typical nonslaved process whereas opening and closing of channels would be a slaved process.

OUTLOOK

The central questions this thesis work aims to address are:

- **Is the dynamical transition observed in hydrated protein controlled by the solvent or do the intrinsic anharmonicity of the protein dynamics also plays a role?**

- **What are the characteristics of the protein motions involved in the dynamical transition?**

- **What is the role played by cosolvents, in particular cryosolvents, in the protein dynamical transition.**

In order to investigate the role of the solvent, dual heatbath MD simulations were performed, in which the solvent and protein are held at different temperatures. Varying the temperature of one component (protein or solvent) while keeping the other temperature constant dissociates changes with temperature of features of the protein energy landscape from those inherent to the solvent. This enables the driving force behind the protein transition to be identified. MD dual-heatbath simulations of myoglobin were performed, the results show that the solvent effectively drives the dynamical transition in protein by caging protein motions below $\sim$220 K. This hypotheses was further strengthened by investigating the dynamical properties the solvation layer. These show clearly that the solvent undergoes a glass transition at the dynamical transition temperature $\sim$220 K.

The origin of the protein dynamical transition is hypothesized to be a transition in the protein dynamics from harmonic motions below the transition temperature to anharmonic, more diffusive motions above. In order to analyze the protein motions over the dynamical transition temperature range, Principal Component Analysis (PCA) was used. PCA is a powerful way of analyzing MD trajectories. Using PCA one is in a unique position to appreciate the anharmonicity of modes as a function of temperature, and thereby to appreciate whether a transition in harmonicity behavior appears $\sim$220 K. The results indicate that anharmonic motions occur well below the transition point, at temperatures as low as 120 K. They also indicate that the dynamical transition at 220 K does not correlate to any change in harmonicity of the protein motions. The results show that the low-frequency modes of motions contribute overwhelmingly to the total MSD. In themselves the non-harmonic motions are seen to be responsible for the transition feature seen in the MSD of

the protein. Further calculations of the damping experienced by the protein show that the solvent induces a transition in the damping experienced by the protein at $\sim$220 K.

Finally, experimental results from neutron scattering experiments performed at the ILL in Grenoble are presented. The dynamical transition in xylanase was measured in methanol/water mixtures of different concentration. The dynamical transition was seen at 0 C in pure water. The dynamical transition temperature was lowered to $\sim$240 K for methanol concentrations $>$ 10% on the 100 ps timescale and to $\sim$170 K for methanol concentrations $>$ 30% on the ns timescale. The results indicate that in the presence of the methanol the protein solvation layer freezes at much lower temperature than expected from bulk solvent properties of methanol/water mixtures.

The next chapters can be outlined in the following way:

- **Methods:** This chapter will present the general theory behind molecular dynamics simulation along with their analysis. It will also present the theory behind neutron scattering.
- **System Used in molecular dynamics Simulations:** This chapter will present the system used in the MD simulations.
- **Nosé-HooverDual Heatbath Simulations:** This chapter will present the theory behind dual heatbath simulations along with the initial results obtained using it.
- **Multiple Heatbath Simulations:** In this chapter the dual heatbath simulation method will be extended and the ensuing results presented. The properties of the solvent surrounding a protein are investigated.
- **Principal Component Analysis:** This chapter presents a PCA analysis of the protein motions over the dynamical transition temperature range. Different measures of protein anharmonicity are presented, enabling the characterization of protein motions in the dynamical transition.
- **Results from Neutron Scattering Experiments:** This chapter presents neutron scattering results obtained on protein solution of varying methanol concentrations over the temperature range of the dynamical transition.

REFERENCES

[1] PERUTZ, M. F. Stereochemical mechanism of oxygen transport by haemoglobin. *Proc R Soc Lond B Biol Sci*, 1980, **208**(1171), 135–62.

[2] GELIN, B. R., LEE, A. W., AND KARPLUS, M. Hemoglobin tertiary structural change on ligand binding. its role in the co-operative mechanism. *J Mol Biol*, 1983, **171**(4), 489–559.

[3] PARAK, F., AND FORMANEK, H. Untersuchung des schwingungsanteils und des kristallgitterfehleranteils des temperaturfaktors in myoglobin durch vergleich von mössbauer-absorptionsmessungen mit röntgenstrukturdaten. *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1971, **27**, 573–578.

[4] PARAK, F., FROLOV, E. N., KONONENKO, A. A., MOSSBAUER, R. L., GOLDANSKII, V. I., AND RUBIN, A. B. Evidence for a correlation between the photoinduced electron transfer and dynamic properties of the chromatophore membranes from rhodospirillum rubrum. *FEBS Lett.*, 1980, **117**(1), 368–72.

[5] PARAK, F., FROLOV, E. N., MOSSBAUER, R. L., AND GOLDANSKII, V. I. Dynamics of metmyoglobin crystals investigated by nuclear gamma resonance absorption. *J. Mol. Biol.*, 1981, **145**(4), 825–33.

[6] COHEN, S. G., BAUMINGER, E. R., NOWIK, I., OFER, S., AND YARIV, J. Dynamics of the iron-containing core in crystals of the iron-storage protein, ferritin, through mossbauer spectroscopy. *Phys. Rev. Lett.*, 1981, **46**(18), 1244–1248.

[7] KNAPP, E. W., FISCHER, S. F., AND PARAK, F. Protein dynamics from mossbauer spectra. the temperature dependence. *J. Am. Chem. Soc.*, 1982, **86**(26), 5042–5047.

[8] MIDDENDORF, H. D. Biophysical applications of quasi-elastic and inelastic neutron scattering. *Annu. Rev. Biophys. Bioeng.*, 1984, **13**, 425–51.

[9] CUSACK, S. Low frequency motion in proteins and its study by inelastic neutron scattering. *Comm. molec. cell. Biophys.*, 1986, **3**(4), 243–71.

[10] KLEINERT, T., DOSTER, W., LEYSER, H., PETRY, W., SCHWARZ, V., AND SETTLES, M. Solvent composition and viscosity effects on the kinetics of co binding to horse myoglobin. *Biochemistry*, 1998, **37**(2), 717–33.

[11] PARAK, F. Physical aspects of protein dynamics. *Rep. Prog. Phys.*, 2003, **66**, 103–129.

[12] DOSTER, W., CUSACK, S., AND PETRY, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature*, 1989, **337**(6209), 754–6.

[13] FERRAND, M., DIANOUX, A. J., PETRY, W., AND ZACCAI, G. Thermal motions and function of bacteriorhodopsin in purple membranes: effects of temperature and hydration studied by neutron scattering. *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**(20), 9668–72.

[14] FITTER, J., LECHNER, R. E., AND DENCHER, N. A. Picosecond molecular motions in bacteriorhodopsin from neutron scattering. *Biophys. J.*, 1997, **73**(4), 2126–37.

[15] FITTER, J. The temperature dependence of internal molecular motions in hydrated and dry alpha-amylase: the role of hydration water in the dynamical transition of proteins. *Biophys. J.*, 1999, **76**(2), 1034–42.

[16] REAT, V., DUNN, R., FERRAND, M., FINNEY, J. L., DANIEL, R. M., AND SMITH, J. C. Solvent dependence of dynamic transitions in protein solutions. *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**(18), 9961–6.

[17] DANIEL, R. M., FINNEY, J. L., REAT, V., DUNN, R., FERRAND, M., AND SMITH, J. C. Enzyme dynamics and activity: time-scale dependence of dynamical transitions in glutamate dehydrogenase solution. *Biophys. J.*, 1999, **77**(4), 2184–90.

[18] FRAUENFELDER, H., PETSKO, G. A., AND TSERNOGLOU, D. Temperature-dependent x-ray diffraction as a probe of protein structural dynamics. *Nature*, 1979, **280**(5723), 558–63.

[19] TILTON, R. F., J., DEWAN, J. C., AND PETSKO, G. A. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-a at nine different temperatures from 98 to 320 k. *Biochemistry*, 1992, **31**(9), 2469–81.

[20] FRAUENFELDER, H., PARAK, F., AND YOUNG, R. D. Conformational substates in proteins. *Annu. Rev. Biophys. Biophys. Chem.*, 1988, **17**, 451–79.

[21] FRAUENFELDER, H., SLIGAR, S. G., AND WOLYNES, P. G. The energy landscapes and motions of proteins. *Science*, 1991, **254**(5038), 1598–603.

[22] FRAUENFELDER, H., AND LEESON, D. T. The energy landscape in non-biological and biological molecules. *Nat Struct Biol*, 1998, **5**(9), 757–9.

[23] MCMAHON, B. H., MULLER, J. D., WRAIGHT, C. A., AND NIENHAUS, G. U. Electron transfer and protein dynamics in the photosynthetic reaction center. *Biophys J*, 1998, **74**(5), 2567–87.

[24] OSTERMANN, A., WASCHIPKY, R., PARAK, F. G., AND NIENHAUS, G. U. Ligand binding and conformational motions in myoglobin. *Nature*, 2000, **404**(6774), 205–8.

[25] CICCOTTI, G., FRENKEL, D., AND MCDONALDS, I. R. *Simulation of liquids and solids*. North-Holland, 1986.

[26] MCCAMMON, J. A., GELIN, B. R., AND KARPLUS, M. Dynamics of folded proteins. *Nature*, 1977, **267**(5612), 585–90.

[27] BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S., AND KARPLUS, M. Charmm: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Biol.*, 1983, **4**(2), 187–217.

[28] SMITH, J. C. Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. *Q. Rev. Biophys.*, 1991, **24**(3), 227–91.

[29] KNELLER, G. R., AND SMITH, J. C. Liquid-like side-chain dynamics in myoglobin. *J. Mol. Biol.*, 1994, **242**(3), 181–5.

[30] SOUAILLE, M., GUILLAUME, F., AND SMITH, J. C. Molecular dynamics simulation of n-nonadecane in urea inclusion compoud. i. comparison with quasielastic neutron scattering experiment. *J. Chem. Phys.*, 1996, **105**(4), 1516–1536.

[31] LEVITT, M., AND SHARON, R. Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**(20), 7557–61.

[32] STEINBACH, P. J., LONCHARICH, R. J., AND BROOKS, B. The effects of environment and hydration on protein dynamics: a simulation study of myoglobin. *Chem. Phys.*, 1991, **158**, 383–94.

[33] STEINBACH, P. J., AND BROOKS, B. R. Protein hydration elucidated by molecular dynamics simulation. *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**(19), 9135–9.

[34] MELCHIONNA, S., FALCONI, M., AND DESIDERI, A. Effect of temperature and hydration on protein fluctuations: Molecular dynamics simulation of cu, zn superoxide dismutase a six different temperatures. comparison with neutron scattering data. *J. Chem. Phys.*, 1998, **108**(14), 6033–41.

[35] TAREK, M., AND TOBIAS, D. J. The dynamics of protein hydration water: a quantitative comparison of molecular dynamics simulations and neutron-scattering experiments. *Biophys. J.*, 2000, **79**(6), 3244–57.

[36] HAYWARD, J. A., AND SMITH, J. C. Temperature dependence of protein dynamics: computer simulation analysis of neutron scattering properties. *Biophys. J.*, 2002, **82**(3), 1216–25.

[37] WONG, C. F., ZHENG, C., AND MCCAMMON, J. A. Glass transition in spc/e water and in a protein solution: a molecular dynamics simulation study. *Chem. Phys. Lett.*, 1989, **154**(2), 151–4.

[38] SMITH, J., KUCZERA, K., AND KARPLUS, M. Dynamics of myoglobin: comparison of simulation results with neutron scattering spectra. *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**(4), 1601–5.

[39] LONCHARICH, R. J., AND BROOKS, B. R. Temperature dependence of dynamics of hydrated myoglobin. comparison of force field calculations with neutron scattering data. *J. Mol. Biol.*, 1990, **215**(3), 439–55.

[40] STEINBACH, P. J., AND BROOKS, B. R. Hydrated myoglobin's anharmonic fluctuations are not primarily due to dihedral transitions. *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**(1), 55–9.

[41] TAREK, M., AND TOBIAS, D. J. Role of protein-water hydrogen bond dynamics in the protein dynamical transition. *Phys. Rev. Lett.*, 2002, **88**(13), 138101.

[42] BIZZARRI, A. R., AND CANNISTRARO, S. Molecular dynamics of water at the protein-solvent interface. *J. Phys. Chem. B*, 2002, **106**, 6617–6633.

[43] TOURNIER, A. L., AND SMITH, J. C. Translational hydration water dynamics drives the protein glass transition. *Biophys. J.*, 2003, **85**(3).

[44] GREEN, J. L., FAN, J., AND ANGELL, C. A. The protein-glass analogy: Some insights from homopeptide comparisons. *J. Phys. Chem.*, 1994, **98**(51), 13780–13790.

[45] ANGELL, C. A. Formation of glasses from liquids and biopolymers. *Science*, 1995, **267**, 1924–1935.

[46] IBEN, I. E., BRAUNSTEIN, D., DOSTER, W., FRAUENFELDER, H., HONG, M. K., JOHNSON, J. B., LUCK, S., ORMOS, P., SCHULTE, A., STEINBACH, P. J., XIE, A. H., AND YOUNG, R. D. Glassy behavior of a protein. *Phys. Rev. Lett.*, 1989, **62**(16), 1916–1919.

[47] SOKOLOV, A. P., KISLIUK, A., QUITMANN, D., KUDLIK, A., AND ROSSLER, E. The dynamics of strong and fragile glass formers: vibrational and relaxation contributions. *J. Non-Cryst. Solids*, 1994, **172-174**, 138–53.

[48] STILLINGER, F. H. A topographic view of supercooled liquids and glass formation. *Science*, 1995, **267**, 1935–1939.

[49] FENIMORE, P. W., FRAUENFELDER, H., MCMAHON, B. H., AND PARAK, F. G. Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc Natl Acad Sci U S A*, 2002, **99**(25), 16047–51.

[50] FRICK, B., AND RICHTER, D. The microscopic basis of the glass transition in polymers from neutron scattering studies. *Science*, 1995, **267**, 1939–1945.

[51] ABSEHER, R., SCHREIBER, H., AND STEINHAUSER, O. The influence of a protein on water dynamics in its vicinity investigated by molecular dynamics simulation. *Proteins*, 1996, **25**(3), 366–78.

[52] SASTRY, S. The relationship between fragility, configurational entropy and the potential energy landscape of glass-forming liquids. *Nature*, 2001, **409**(6817), 164–7.

[53] DOSTER, W., CUSACK, S., AND PETRY, W. Dynamic instability of liquidlike motions in a globular protein observed by inelastic neutron scattering. *Phys. Rev. Lett.*, 1990, **65**(8), 1080–1083.

[54] AUSTIN, R. H., BEESON, K. W., EISENSTEIN, L., FRAUENFELDER, H., AND GUNSALUS, I. C. Dynamics of ligand binding to myoglobin. *Biochemistry*, 1975, **14**(24), 5355–73.

[55] RASMUSSEN, B. F., STOCK, A. M., RINGE, D., AND PETSKO, G. A. Crystalline ribonuclease a loses function below the dynamical transition at 220 k. *Nature*, 1992, **357**(6377), 423–4.

[56] TEETER, M. M., YAMANO, A., STEC, B., AND MOHANTY, U. On the nature of a glassy state of matter in a hydrated protein: Relation to protein function. *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**(20), 11242–7.

[57] DANIEL, R. M., SMITH, J. C., FERRAND, M., HERY, S., DUNN, R., AND FINNEY, J. L. Enzyme activity below the dynamical transition at 220 k. *Biophys. J.*, 1998, **75**(5), 2504–7.

[58] DUNN, R. V., REAT, V., FINNEY, J., FERRAND, M., SMITH, J. C., AND DANIEL, R. M. Enzyme activity and dynamics: xylanase activity in the absence of fast anharmonic dynamics. *Biochem. J.*, 2000, **346 Pt 2**, 355–8.

[59] FRANKS, F. Protein stability: the value of 'old literature'. *Biophys Chem*, 2002, **96**(2-3), 117–27.

[60] MATTOS, C., AND RINGE, D. Proteins in organic solvents. *Curr Opin Struct Biol*, 2001, **11**(6), 761–4.

[61] MACKERELL, J. A. D., BASHFORD, D., BELLOTT, M., R. L. DUNBRACK, J., EVANSECK, J. D., FIELD, M. J., FISCHER, S., GAO, J., GUO, H., HA, S., JOSEPH-MCCARTHY, D., KUCHNIR, L., KUCZERA, K., LAU, F. T. K., MATTOS, C., MICHNICK, S., NGO, T., NGUYEN, D. T., PRODHOM, B., W. E. REIHER, I., ROUX, B., SCHLENKRICH, M., SMITH, J. C., STOTE, R., STRAUB, J., WATANABE, M., WIORKIEWICZ-KUCZERA, J., YIN, D., AND KARPLUS, M. All-atom empirical potential for molecular modelling and dyamics studies of proteins. *J. Phys. Chem.*, 1998, **102**(18), 3586–3616.

[62] DANIEL, R. M., DUNN, R. V., FINNEY, J. L., AND SMITH, J. C. The role of dynamics in enzyme activity. *Annu Rev Biophys Biomol Struct*, 2003, **32**, 69–92.

[63] TSAI, A. M., NEUMANN, D. A., AND BELL, L. N. Molecular dynamics of solid-state lysozyme as affected by glycerol and water: a neutron scattering study. *Biophys J*, 2000, **79**(5), 2728–32.

[64] SMITH, J., CUSACK, S., POOLE, P., AND FINNEY, J. Direct measurement of hydration-related dynamic changes in lysozyme using inelastic neutron scattering spectroscopy. *J Biomol Struct Dyn*, 1987, **4**(4), 583–8.

[65] ZANOTTI, J. M., BELLISSENT-FUNEL, M. C., AND PARELLO, J. Hydration-coupled dynamics in proteins studied by neutron scattering and nmr: the case of the typical ef-hand calcium-binding parvalbumin. *Biophys J*, 1999, **76**(5), 2390–411.

[66] PEREZ, J., ZANOTTI, J. M., AND DURAND, D. Evolution of the internal dynamics of two globular proteins from dry powder to solution. *Biophys J*, 1999, **77**(1), 454–69.

[67] DELLERUE, S., PETRESCU, A. J., SMITH, J. C., AND BELLISSENT-FUNEL, M. C. Radially softening diffusive motions in a globular protein. *Biophys. J.*, 2001, **81**(3), 1666–76.

[68] RINGE, D. What makes a binding site a binding site? *Curr Opin Struct Biol*, 1995, **5**(6), 825–9.

[69] COLOMBO, G., OTTOLINA, G., AND CARREA, G. Modelling of enzyme properties in organic solvents. *Monatsh. Chem.*, 2000, **131**, 527–547.

[70] SVERGUN, D. I., RICHARD, S., KOCH, M. H., SAYERS, Z., KUPRIN, S., AND ZACCAI, G. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc Natl Acad Sci U S A*, 1998, **95**(5), 2267–72.

[71] FINNEY, J. From ices to proteins: the organisation of water at interfaces. In *NATO Science Series, Series A: Life Sciences*, vol. 305 (Hydration Processes in Biology). IOS press, 1999, pp. 107–114.

[72] LEHMANN, M. S., MASON, S. A., AND MCINTYRE, G. J. Study of ethanol-lysozyme interactions using neutron diffraction. *Biochemistry*, 1985, **24**(21), 5862–9.

[73] LEHMANN, M. S., AND STANSFIELD, R. F. Binding of dimethyl sulfoxide to lysozyme in crystals, studied with neutron diffraction. *Biochemistry*, 1989, **28**(17), 7028–33.

[74] KOMEIJI, Y., UEBAYASI, M., SOMEYA, J., AND YAMATO, I. A molecular dynamics study of solvent behavior around a protein. *Proteins*, 1993, **16**(3), 268–77.

[75] BOUQUIERE, J. P., FINNEY, J., LEHMANN, M. S., AND MOGENS, S. Interaction of the tetramethylammonium ion with the lysozyme molecule, studied using neutron diffraction. *J. Chem. Soc., Faraday Trans.*, 1993, **89**(15), 2701–5.

[76] NAKASAKO, M. Large-scale networks of hydration water molecules around bovine beta-trypsin revealed by cryogenic x-ray crystal structure analysis. *J Mol Biol*, 1999, **289**(3), 547–64.

[77] BROOKS, C. L., R., AND KARPLUS, M. Solvent effects on protein motion and protein effects on solvent motion. dynamics of the active site region of lysozyme. *J Mol Biol*, 1989, **208**(1), 159–81.

[78] LOUNNAS, V., PETTITT, B. M., AND PHILLIPS, G. N., J. A global model of the protein-solvent interface. *Biophys J*, 1994, **66**(3 Pt 1), 601–14.

[79] MAKAROV, V. A., PETTITT, B. M., AND FEIG, M. Solvation and hydration of proteins and nucleic acids: A theoretical view of simulation and experiment. *Acc. Chem. Res.*, 2002, **35**, 376–384.

[80] MERZEL, F., AND SMITH, J. C. Is the first hydration shell of lysozyme of higher density than bulk water? *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**(8), 5378–83.

[81] FINNEY, J. L. Overview lecture. hydration processes in biological and macromolecular systems. *Faraday Discuss*, 1996(103), 1–18.

[82] CORDONE, L., FERRAND, M., VITRANO, E., AND ZACCAI, G. Harmonic behavior of trehalose-coated carbon-monoxy-myoglobin at high temperature. *Biophys. J.*, 1999, **76**(2), 1043–7.

[83] WALSER, R., MARK, A. E., AND VAN GUNSTEREN, W. F. On the temperature and pressure dependence of a range of properties of a type of water model commonly used in high-temperature protein unfolding simulations. *Biophys. J.*, 2000, **78**(6), 2752–60.

[84] RINGE, D., AND PETSKO, G. A. Mapping protein dynamics by x-ray diffraction. *Prog. Biophys. Mol. Biol.*, 1985, **45**(3), 197–235.

[85] SOPER, A. K., AND FINNEY, J. L. Hydration of methanol in aqueous solution. *Phys. Rev. Lett.*, 1993, **71**(26), 4346–4349.

[86] PALINKAS, G., AND HEINZINGER, K. Simulation of liquid mixtures. In *NATO ASI Series, Series C: Mathematical and Physical Sciences*, vol. 435(Hydrogen Bond Networks). Kluwer Academic Publisher, 1994, pp. 281–304.

[87] NOSKOV, S., KISELEV, M., KOLKER, A., AND RODE, B. Structure of methanol-methanol associates in dilute methanol-water mixtures from molecular dynamics simulation. *J. Mol. Liqu.*, 2001, **91**, 157–165.

[88] DIXIT, S., CRAIN, J., POON, W. C., FINNEY, J. L., AND SOPER, A. K. Molecular segregation observed in a concentrated alcohol-water solution. *Nature*, 2002, **416**(6883), 829–32.

[89] JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W., AND KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 1983, **79**(2), 926–934.

[90] KEUTSCH, F. N., CRUZAN, J. D., AND SAYKALLY, R. J. The water trimer. *Chem Rev*, 2003, **103**(7), 2533–77.

[91] JORGENSEN, W. L., AND JENSON, C. Temperature dependence of tip3p, spc, and tip4p water from npt monte carlo simulations: seeking temperatures of maximum density. *J. Comput. Chem.*, 1998, **19**(10), 1179–1186.

[92] PACIARONI, A., CINELLI, S., AND ONORI, G. Effect of the environment on the protein dynamical transition: a neutron scattering study. *Biophys. J.*, 2002, **83**(2), 1157–64.

[93] LEHNERT, U., REAT, V., WEIK, M., ZACCAI, G., AND PFISTER, C. Thermal motions in bacteriorhodopsin at different hydration levels studied by neutron scattering: correlation with kinetics and light- induced conformational changes. *Biophys J*, 1998, **75**(4), 1945–52.

[94] REAT, V., FINNEY, J. L., STEER, A., ROBERTS, M. A., SMITH, J., DUNN, R., PETERSON, M., AND DANIEL, R. Cryosolvents useful for protein and enzyme studies below -100 degrees c. *J Biochem Biophys Methods*, 2000, **42**(3), 97–103.

[95] HAGEN, S. J., HOFRICHTER, J., AND EATON, W. A. Protein reaction kinetics in a room-temperature glass. *Science*, 1995, **269**(5226), 959–62.

[96] BIZZARRI, A. R., PACIARONI, A., AND CANNISTRARO, S. Glasslike dynamical behavior of the plastocyanin hydration water. *Phys. Rev. E. Stat. Phys.*, 2000, **62**(3 Pt B), 3991–9.

[97] DIEHL, M., DOSTER, W., PETRY, W., AND SCHOBER, H. Water-coupled low-frequency modes of myoglobin and lysozyme observed by inelastic neutron scattering. *Biophys. J.*, 1997, **73**(5), 2726–32.

[98] VITKUP, D., RINGE, D., PETSKO, G. A., AND KARPLUS, M. Solvent mobility and the protein 'glass' transition. *Nat. Struct. Biol.*, 2000, **7**(1), 34–8.

CHAPTER 2

METHODS

CONTENTS

INTRODUCTION

This chapter presents the theoretical foundations of the present work. Molecular dynamics simulation techniques are presented followed by a general description of the way MD trajectories can be analyzed. Principal component analysis is presented and a brief introduction to neutron scattering is given.

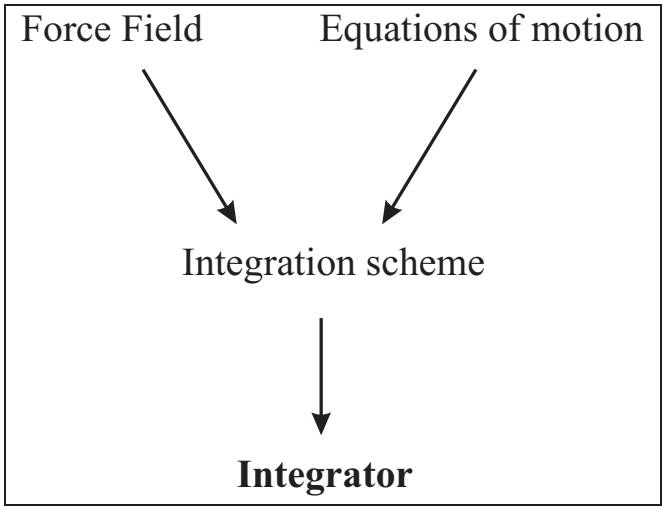


Figure 2.1: General scheme showing how a force-field combines with equations of motion in an integration scheme to achieve an integrator ready to perform molecular dynamics simulations.

2.1 MOLECULAR DYNAMICS SIMULATIONS

This section covers the theoretical background to Molecular Dynamics pertinent to the present work. For an in depth introduction to molecular dynamics the reader is referred to the book by A. Leach.[1]

2.1.1 GENERAL PRINCIPLES OF MOLECULAR DYNAMICS SIMULATIONS

Different elements are required to come together in order to perform molecular dynamics simulations. These different elements are schematized in Figure 2.1. First of all, a description of the interaction between the atoms in the system is required. The best available description of the system is at the quantum level, however performing simulations of a quantum system are at the present time unfeasible for systems larger than a few tens of atoms. Luckily interatomic interactions can be well approximated by empirical functions, these come together to form a description of the forces acting on the atoms called the force field. The force field used in the present work will be described in the next section.

The laws according to which the atoms are going to move also need to be clarified. The simplest set of equations of motion is Newton's equations of motion. More evolved sets of equations can take care of such things as temperature control of the system and pressure control for confined systems. These sets of equations will be discussed shortly.

Finally, the equations of motion need to be integrated, *i.e.*, recast into discreet

timesteps which a computer can calculate. For perfect precision one would need infinitely small timesteps which would also mean infinitely long simulation times. So in order to perform meaningful simulations in a reasonable amount of time, one has to find an integration scheme which gives a good approximation of the correct trajectory taken by the system. In practice, the effect of this approximation is negligible since the interesting equilibrium properties of the system are not affected by the exact trajectory of each atom in the system. Several integration schemes exist such as the Verlet algorithm, the velocity Verlet algorithm, and a number of others. These integrators will be presented along with a more recent integration scheme: the Liouville Operator approach.

2.1.2 FORCE FIELD DESCRIPTION

The force field description of the interatomic forces is split into two categories: the bonded terms and the non-bonded terms. The bonded terms regroup simple covalent binding as well as the more complex hybridization and π-orbital effects, these are the bonds, angles and dihedrals terms. These terms are schematically drawn in figure 2.2. The non-bonded terms describe the van der Waals forces and the electrostatic interactions between the atoms. The different terms will now be presented in more detail.

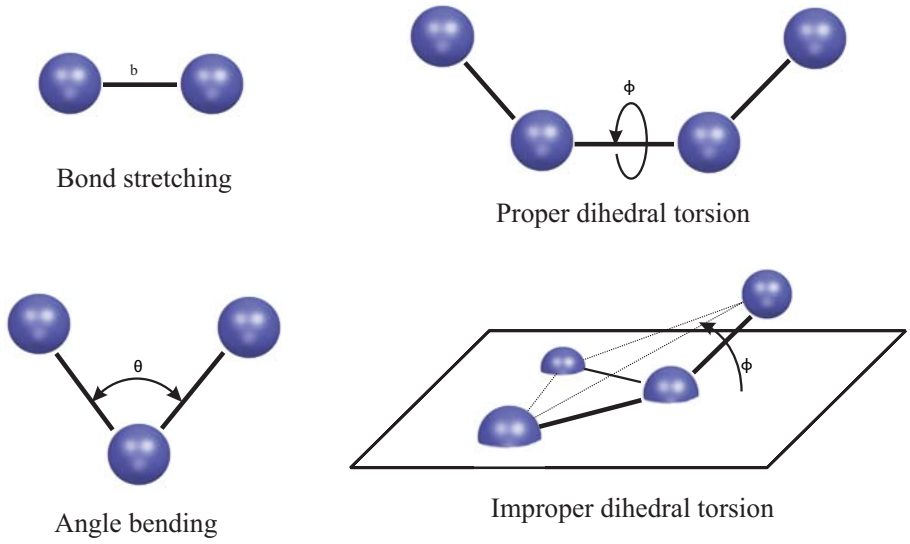


Figure 2.2: Schematic representation of the bonded interaction terms contributing to the force field: bond stretching, angle bending, proper and improper dihedrals.

BOND STRETCHING

The bond stretching term describes the forces acting between two covalently bonded atoms. The potential is assumed to be approximately harmonic:

$$V_b = k_b(b - b_0)^2 \tag{2.1}$$

where b is the distance between the two atoms. Two parameters characterize each bonded interaction: b_0 the average distance between them and a force constant k_b.

ANGLE BENDING

The angle bending terms describes the force originating from the deformation of the valence angles between three covalently bonded atoms. The angle bending term is described using a harmonic potential:

$$V_\theta = k_\theta(\theta - \theta_0)^2 \tag{2.2}$$

where θ is the angle between three atoms. There again two parameters characterize each angle in the system: the reference angle θ_0 and a force constant k_θ.

TORSIONAL TERMS

The torsional terms are weaker than the bond stretching and angle bending terms. They describe the barriers to rotation existing between four bonded atoms. There are two type of torsional terms: proper and improper dihedrals. Proper torsional potentials are described by a cosine function:

$$V_\phi = k_\phi[1 + \cos(n\phi - \delta)], \qquad n = 1, 2, 3, 4, 6 \tag{2.3}$$

where ϕ is the angle between the planes formed by the first and the last three of the four atoms. Three parameters characterize this interaction: δ sets the minimum energy angle, k_ϕ is a force constant and n is the periodicity.

The improper dihedral term is designed both to maintain chirality about a tetrahedral heavy atom and to maintain planarity about certain atoms. The potential is described by a harmonic function:

$$V_\omega = k_\omega(\omega - \omega_0)^2 \tag{2.4}$$

where ω is the angle between the plane formed by the central atom and two peripheral atoms and the plane formed by the peripheral atoms (see figure 2.2).

VAN DER WAALS INTERATIONS

Van der Waals interactions and electrostatic interactions are non-bonded interactions, *i.e.*, they act between atoms which are not covalently bonded together.

The van der Waals force acts on atoms in close proximity. It is strongly repulsive at short range and weakly attractive at medium range. The interaction is described by a Lennard-Jones potential:

$$V_{VdW} = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{2.5}$$

where r is the distance between two atoms. It is parameterized by σ: the collision parameter (the separation for which the energy is zero) and ϵ the depth of the potential well. The Lennard-Jones potential is represented in figure 2.3.

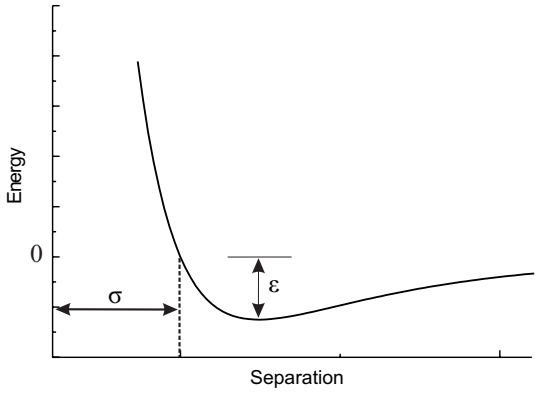


Figure 2.3: The Lennard-Jones potential. The collision parameter, σ, is shown along with the well depth, ϵ.

ELECTROSTATIC INTERACTIONS

Finally, the long distance electrostatic interaction between two atoms is described by Coulomb's law:

$$V_{Elec} = \frac{q_1 q_2}{4\pi\epsilon_0 r_{12}} \tag{2.6}$$

where q_1 and q_2 are the charges of both atoms and r_{12} the distance between them. ϵ_0 is the electric susceptibility of vacuum.

So finally, the equation for the potential energy describing the force field can be written:

$$\begin{aligned} V &= \sum_{bonds} k_B(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 \\ &+ \sum_{\substack{proper\\dihedrals}} k_\phi[1+\cos(n\phi-\delta)] + \sum_{\substack{improper\\dihedrals}} k_\omega(\omega-\omega_0)^2 \\ &+ \sum_{\substack{i,j\\i<j}} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\substack{i,j\\i<j}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \end{aligned} \tag{2.7}$$

2.1.3 EQUATIONS OF MOTION

NEWTONIAN EQUATIONS OF MOTION

The simplest set of equations of motion are the Newtonian equations:

$$\begin{aligned} \dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i(\mathbf{r}_1, ..., \mathbf{r}_N, t) \end{aligned} \tag{2.8}$$

where $\mathbf{r}_i$ and $\mathbf{P}_i$ are the position and momentum of atom i at a time t. $\mathbf{F}_i$ is the force acting on atom i: $\mathbf{F}_i = \nabla_i V$

Newtonian equations are the physically exact equations describing the motion of atoms in the system. However, they have the disadvantage for molecular dynamics simulations that the temperature of the system has to be periodically reset so as to mimic biological conditions. To avoid this problem different sets of equations have been developed which take care of the temperature as well as the pressure control of the system.[2]

NOSÉ-HOOVER CONSTANT TEMPERATURE ALGORITHM

The equations of motion for constant temperature or Nosé-Hoover thermostat equations are the following:[3]

$$\begin{aligned} \dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i - \frac{p_\eta}{Q}\mathbf{p}_i \\ \dot{\eta} &= \frac{p_\eta}{Q} \\ \dot{p_\eta} &= \sum_i \frac{\mathbf{p}_i^2}{m_i} - dNkT \end{aligned} \tag{2.9}$$

where η and p_η are the thermostat position and momentum and T is the temperature at which the system is to be regulated. N being the number of atoms in the system, k Boltzmann's constant and d the number of spatial dimensions. The parameter Q, given by $Q = dNkT\tau^2$, determines the time scale of the thermostat motion via the time scale parameter τ, which should be chosen corresponding to a characteristic time scale of the system, *e.g.*, a vibrational period.

The Nosé-Hoover scheme has the advantageous feature that it approximates the canonical distribution of temperature present in physical temperatures, *i.e.* the temperature of the system is not fixed at a given temperature but oscillates about it, as expected for small systems. An extension of this algorithm, the ***Nosé-Hoover-Chain algorithm***, has the advantage over the simple Nosé-Hoover algorithm that it reproduces the exact canonical distribution of temperature. This algorithm was used in the present work and will be presented in section 4.1. It also has the advantage of being more stable as will be shown in Chapter 4.

CONSTANT TEMPERATURE AND PRESSURE

Simultaneous regulation of temperature and pressure can also be taken care of through the equations of motions: the isothermal-isobaric equations of motion:[2]

$$\begin{aligned}
\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} + \frac{p_\epsilon}{W}\mathbf{r}_i \\
\dot{\mathbf{p}}_i &= \mathbf{F}_i - \left(1 + \frac{1}{N}\right)\frac{P_\epsilon}{W}\mathbf{p}_i - \frac{P_\eta}{Q}\mathbf{p}_i \\
\dot{V} &= \frac{dVp_\epsilon}{W} \\
\dot{p_\epsilon} &= dV(P_{\text{int}} - P_{\text{ext}}) + \frac{1}{N}\sum_i \frac{\mathbf{p}_i^2}{m_i} - \frac{p_\eta}{Q}p_\epsilon \\
\dot{\eta} &= \frac{P_\eta}{Q} \\
\dot{p_\eta} &= \sum_i \frac{\mathbf{p}_i^2}{m_i} + \frac{p_\epsilon^2}{W} - (dN+1)kT
\end{aligned} \tag{2.10}$$

where p_ϵ is a momentum conjugate to the logarithm of the volume, W is its associated mass parameter, $\epsilon = \ln(V/v_0)$, P_{ext} is the externally applied pressure, and P_{int} is the instantaneous internal pressure of the system given by:

$$P_{\text{int}} = \frac{1}{\mathrm{d}V}\left[\sum_i \frac{\mathbf{p}_i^2}{m_i} + \sum_i \mathbf{r}_i\mathbf{F}_i - (\mathrm{d}V)\frac{\partial U}{\partial V}\right] \tag{2.11}$$

Thus, the variable p_ϵ acts as a 'barostat' which drives the system to the steady state $\langle P_{\text{int}} \rangle = P_{\text{ext}}$. In this way, both temperature and pressure are regulated so as to reproduce exact canonical distributions.

Having gained an overview of the force field and of the equations of motion let us now turn to the methods by which these equations can be integrated over time.

2.1.4 METHODS FOR INTEGRATING THE EQUATIONS OF MOTION

SIMPLE INTEGRATION ALGORITHMS

Numeric integration of equations of motion is done step by step using Finite Difference methods. These methods are explicit and use the information available at time t to predict the system's coordinates and velocities at a time $t + \Delta t$, where Δt is a short time interval.

These integration schemes are based on a Taylor expansion of the position at time $t + \Delta t$, represented by the $n + 1$ subscript:

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \Delta t\, \mathbf{v}_n + \frac{\Delta t^2}{2}\, \mathbf{a}_n + ... \tag{2.12}$$

where $\mathbf{v}_n$ is the first derivative of the position $\mathbf{r}_n$, $\mathbf{a}_n$ is the second derivative of the position etc. For Newtonian equations this simply yields:

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \Delta t\, \mathbf{v}_n + \Delta t^2\, \frac{\mathbf{F}_n}{2m} + ... \tag{2.13}$$

The different simple integration algorithms vary in the way they implement this basic expansion.

The most basic and most common integration algorithm is the ***Verlet Integrator***. This integrator is based on two Taylor expansions, one forward and one backward:

$$\begin{aligned} \mathbf{r}_{n+1} &= \mathbf{r}_n + \Delta t\, \mathbf{v}_n + \Delta t^2\, \frac{\mathbf{F}_n}{2m} + ... \\ \mathbf{r}_{n-1} &= \mathbf{r}_n - \Delta t\, \mathbf{v}_n + \Delta t^2\, \frac{\mathbf{F}_n}{2m} - ... \end{aligned} \tag{2.14}$$

These two expansions are then added to give the basic Verlet integration formalism:

$$\mathbf{r}_{n+1} = 2\mathbf{r}_n - \mathbf{r}_{n-1} + \Delta t^2\, \frac{\mathbf{F}_n}{m} + \mathcal{O}(\Delta t^4)... \tag{2.15}$$

The simple Verlet approach has the advantage that it does not require the velocities, needs a single force calculation per cycle and is naturally reversible in time. It generates, however, rather large errors.

The ***Leap Frog Integrator*** is a variation of the Verlet algorithm designed to improve the velocity evaluations. Its name comes from the fact that the velocities are evaluated at the mid-point of the position evaluation and vice versa. The algorithm is as follows:

$$\begin{aligned} \mathbf{v}_{n+1/2} &= \mathbf{v}_{n-1/2} + \Delta t \, \frac{\mathbf{F}_n}{m} \\ \mathbf{r}_{n+1} &= \mathbf{r}_n + \Delta t \, \mathbf{v}_{n+1/2} \end{aligned} \tag{2.16}$$

This scheme has the advantage of providing a direct handle on the velocities which can be useful for temperature regulation. It has less error that the simple Verlet scheme.

A further improvement is the ***Velocity Verlet algorithm***:

$$\begin{aligned} \mathbf{r}_{n+1} &= \mathbf{r}_n + \Delta t \, \mathbf{v}(t) + \Delta t^2 \, \frac{\mathbf{F}_n}{2m} \\ \mathbf{v}_{n+1} &= \mathbf{v}_n + \frac{\Delta t}{2m} \, [\mathbf{F}_n + \mathbf{F}_{n+1}] \end{aligned} \tag{2.17}$$

This has the best evaluation of velocities and is the most widely used algorithm. It works very satisfactorily for Hamiltonian systems of equations such as Newtons equations. However for non-Hamiltonian systems such as temperature thermostating and temperature-pressure regulation the scheme has to be revised: the Velocity step is taken and then corrected for the non-Hamiltonian[†] contribution in an implicit iterative procedure. To illustrate this procedure, consider Milne's simple method for first-order equations of the form:

$$\dot{\mathbf{r}}_n = f\left(\mathbf{r}_n, t\right) \tag{2.18}$$

then the next step is given by:

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \int_{t_n}^{t_{n+1}} f\left(\mathbf{r}, t\right) dt \tag{2.19}$$

Expanding $f\left(\mathbf{r}, t\right)$ to third order in time, an expression for the next step in time can then be written:

$$\mathbf{r}_{n+1} = \mathbf{r}_{n-1} + {}^{\Delta t}/_{3}(\dot{\mathbf{r}}_{n-1} + 4\dot{\mathbf{r}}_n + \dot{\mathbf{r}}_{n+1}) + \mathcal{O}(\Delta t^5) \tag{2.20}$$

Since $\dot{\mathbf{r}}_{n+1} = f\left(\mathbf{r}_{n+1}\right)$, the above expression is a *corrector* equation is an implicit equation for $\mathbf{r}_{n+1}$. If Δt is sufficiently small and a first approximation for $\mathbf{r}_{n+1}$ can be found, the equation is solved through successive iterations. To provide the first approximation for $\mathbf{r}_{n+1}$, an explicit *predictor* formula is needed for which a simple step can be used such

as the ones presented previously.

The drawbacks of the schemes presented above are that the time reversal symmetry of the equations is not preserved and that they do not preserve the invariant phase space measure.[2] An elegant solution to this problem is the Liouville operator approach.

LIOUVILLE OPERATOR APPROACH

In this section a new approach to the integration problem is presented. This approach leads to time reversible methods that preserve the phase space metric. It has also the advantage of being directly useable on many systems. For an extended discussion of this topic the reader is referred to Tuckermann *et al* 2000 and references therein.[2]

The equations of motion of a Hamiltonian or non-Hamiltonian system can be recast in the general form:

$$\dot{\mathbf{x}} = iL\mathbf{x} \tag{2.22}$$

where **x** is the phase space vector and iL is the ***Liouville operator***. For a Hamiltonian system, the Liouville operator is given by:

$$iL \equiv \sum_{i=1}^{N} \left[\frac{\mathbf{p}_i}{m_i} \frac{\partial}{\partial \mathbf{r}_i} + \mathbf{F}_i \frac{\partial}{\partial \mathbf{p}_i} \right] \tag{2.23}$$

Equation 2.22 has the formal solution:

$$\mathbf{x}(t) = e^{iLt}\mathbf{x}(0) \tag{2.24}$$

The unitary operator, $\exp(iLt)$, is the ***classical propagator***. Its solution can not be determined analytically except for very simple cases. However an approximation can be introduced which enables the construction of numerical integrators. Consider that the Liouville operator can be written as a sum: $iL = iL_1 + iL_2$. The classical propagator can then be rewritten using the *Trotter Theorem*:

$$e^{iLt} = e^{(iL_1+iL_2)t} = \lim_{P\to\infty} \left[e^{iL_2 t/2P}\, e^{iL_1 t/P}\, e^{iL_2 t/2P} \right]^P \tag{2.25}$$

[†]Non-Hamiltonian systems are dynamical systems which cannot be expressed in the form:

$$\begin{aligned} \dot{\mathbf{r}}_i &= \frac{\partial H}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i &= -\frac{\partial H}{\partial \mathbf{r}_i} = -\frac{\partial U}{\partial \mathbf{r}_i} = \mathbf{F}_i(\,\mathbf{r}_1, ..., \mathbf{r}_N) \end{aligned} \tag{2.21}$$

where U is the potential energy of the system.

Defining ${}^{t}/_{P} = \Delta t$ for finite P, the following approximation can be made:

$$e^{iL\Delta t} \approx e^{{}^{iL_2\Delta t}/_2}\, e^{iL_1\Delta t}\, e^{{}^{iL_2\Delta t}/_2} + \mathcal{O}(\Delta t^3) \tag{2.26}$$

For long times:

$$e^{iLP\Delta t} \approx \prod_{k=1}^{P} e^{{}^{iL_2\Delta t}/_2}\, e^{iL_1\Delta t}\, e^{{}^{iL_2\Delta t}/_2} + \mathcal{O}(t\Delta t^2) \tag{2.27}$$

which yields a numerical integration procedure accurate to the second order in the time step.

For the Hamiltonian system of equation 2.23 the Liouville operator can be split into:

$$\begin{aligned} iL_1 &= \sum_{i=1}^{N} \frac{\mathbf{p}_i}{m_i} \frac{\partial}{\partial \mathbf{r}_i} \\ iL_2 &= \sum_{i=1}^{N} \mathbf{F}_i \frac{\partial}{\partial \mathbf{p}_i} \end{aligned} \tag{2.28}$$

The operator $\exp\left({}^{iL_2\Delta t}/_2\right)$ then becomes a translation operator on the momenta: $\mathbf{p}_i \rightarrow \mathbf{p}_i + {}^{\Delta t}/_2\mathbf{F}_i(\mathbf{r})$. Similarly, $e^{iL_1\Delta t}$ becomes a translation operator on the position: $\mathbf{r}_i \rightarrow \mathbf{r}_i + \Delta t\mathbf{p}_i/_{m_i}$. Combining these two actions allows the action of equation 2.26 to be expressed as the action of $\exp\left({}^{iL_2\Delta t}/_2\right)$ followed by the action of $e^{iL_1\Delta t}$ finally followed by that of $\exp\left({}^{iL_2\Delta t}/_2\right)$. For the simple Hamiltonian case, this procedure actually yields the Velocity Verlet algorithm presented in equation 2.17.

This procedure of translating each operator into an update step, which then can be directly turned into instructions for a computer, is called the ***direct translation technique***.[4] Although seemingly trivial in the case of Hamiltonian systems it becomes very powerful when facing non-Hamiltonian systems such as constant temperature or constant temperature/pressure algorithms. An important property of this technique is that it preserves the invariant phase space measure. This property is known as the *simplectic property*. The significance of this property is that the error is bounded, *i.e.*, there will be no secular growth in the energy conservation error, provided the time step is not too large.

Another advantage of the present approach is that it can easily be extended to treat systems with multiple time scale motions. The Liouville operator for such a system can be split up into two parts: L_f containing either fast motions or motions which are unstable and require careful integration and L_s containing a relatively slow varying contribution to the motion. Considering n substeps for the motions in L_f, L_f can be rewritten:

$$e^{iL_f\Delta t} = \left(e^{iL_f{}^{\Delta t}/_n}\right)^n \tag{2.29}$$

which combined with equation 2.26 gives a complete expression for the propagation operator which can be written in two ways:

$$e^{iL\Delta t} = e^{iL_s \Delta t/2} \left(e^{iL_f \Delta t/n}\right)^n e^{iL_s \Delta t/2} \quad (2.30)$$

$$e^{iL\Delta t} = \left(e^{iL_f \Delta t/2n}\right)^n e^{iL_s \Delta t} \left(e^{iL_f \Delta t/2n}\right)^n \quad (2.31)$$

In order to increase the accuracy of the algorithm higher order approximations of the evolution operator can be used.[5, 6] Instead of the Trotter approximation the following approximation can be used:

$$e^{iL\Delta t} = \prod_{i=1}^{n} e^{iL\frac{w_i \Delta t}{n}} \quad (2.32)$$

where the w_i are a precise set of numbers depending on the order of the approximation n. The values of w_i for $n = 3$ are $w_1 = w_3 = \frac{1}{2-2^{1/3}}$, $w_2 = 1 - 2w_1$. And for $n = 5$ the w_i values are given by: $w_1 = w_2 = w_4 = w_5 = \frac{1}{4-4^{1/3}}$, $w_3 = 1 - 4w_1$. Using these approximations the error goes as $\mathcal{O}(\Delta t/n^5)$.

This Liouville operator approach was used to integrate the equations of motion of the Nosé-Hoover-Chain algorithm for the multiple heatbath approach, as will be presented in chapter 4.

2.2 ANALYSIS OF MOLECULAR DYNAMICS TRAJECTORIES

2.2.1 MEAN-SQUARE FLUCTUATIONS

There are many different possible ways of analyzing MD trajectories. Different quantities can be calculated directly from the trajectories of the atoms in the system. Experimental observables can be calculated and then compared to the values obtained experimentally. Alternatively, it is also possible to calculate quantities inaccessible to experiments which enable new ways of understanding the system under scrutiny. Experimental methods will call on a whole range of observables specific to the technique: particular rates, scattering patterns, dynamical properties...

One such quantity commonly used is the atomic Mean-Square Deviation (MSD). It can be compared to results obtained using neutron scattering and give general information about the amount of motion present in the system. The MSD is defined as:

$$MSD = \frac{1}{N} \sum_{i=1}^{N} \langle \left(\mathbf{r}_i - \mathbf{r}_i^0\right)^2 \rangle \quad (2.33)$$

where $\mathbf{r}_i$ is the position of atom i, $\mathbf{r}_i^0$ the starting position of atom i, N the number of atoms and an average is made over a time interval Δt (for neutron scattering one might choose $\Delta t = 200\ ps$ as a meaningful value). Closely related to the MSD is the mass weighted MSD which will be used in this thesis when performing principal component analysis. The mass weighted MSD is defined as:

$$MSD = \frac{1}{N}\sum_{i=1}^{N} m_i \left(\mathbf{r}_i - \mathbf{r}_i^0\right)^2\rangle \tag{2.34}$$

where m_i is the mass of atom i.

Another closely related quantity is the Mean-Square Fluctuations (MSF). The MSF is defined as:

$$MSF = \frac{1}{N}\sum_{i=1}^{N}\langle(\mathbf{r}_i - \mathbf{r}_i^m)^2\rangle \tag{2.35}$$

where $\mathbf{r}_i^m$ is the mean position of atom i and the average is again taken over a time interval Δt. In practice MSD and MSF are identical since when calculating the MSD several starting values $\mathbf{r}_i^0$ will be taken in order to obtain better statistics, this, in effect, amounts to using the average position $\mathbf{r}_i^m$, *i.e.*, calculating the MSF. In this work the terms MSD and MSF will therefore be used interchangeably.

2.2.2 COMPARISON WITH EXPERIMENTAL DATA FROM NEUTRON SCATTERING

In order to compare MD trajectories with neutron scattering experiments more elaborate calculations can be performed. From the atomic trajectories it is possible to calculate the scattering intensity of the sample were the neutron experiment to be performed. This calculation uses information about the way the individual atoms scatter neutrons and averages it out over the time period of the simulation and the different atoms present in the system (although hydrogen atoms dominate the signal). Section 2.3 will present the equations which describe neutron scattering.

In calculating neutron scattering intensities one is in a unique position to validate the model used to perform the simulation. Neutron scattering yields critical information about the frequencies of the motion present in the system which gives a good measure of the accuracy of the model.

2.2.3 PRINCIPAL COMPONENT ANALYSIS AND NORMAL MODES ANALYSIS OF THE DYNAMICAL TRANSITION

Another powerful way of analyzing MD trajectories is Principal Component Analysis. PCA is a convenient method for representing the conformational space explored in an MD trajectory and has no experimental counterpart. PCA determines the essential motions present in the simulation: the principal component modes. The set of principal components is defined as the solution to the eigenvalue problem of the second-moment matrix, $\mathbf{A}$, of the mass-weighted internal atomic displacements. The diagonalization of $\mathbf{A}$ yields the eigenvectors $\mathbf{W}_k$, i.e., the principal components and their associated eigenvalues,ζ_k. The mathematical details are given in chapter 6

Normal modes are closely related to principal component modes, and are calculated by diagonalizing the Hessian matrix: the second derivative matrix of the potential energy function. Normal Mode Analysis is widely used to investigate domain motions in protein.[7] It has been used to look at low-frequency motions in proteins.[8] It has also been used in X-ray refinement[9–11] and NMR refinement[12] and in combination with PCA to investigate protein motions.[13–17]

Normal modes analysis relies on the assumption that the protein energy landscape can be considered harmonic, this assumption has been shown to be good enough for many purposes. However, it is known that hydration damps the low frequency modes, thereby introducing diffusive motions. In this regime the harmonic assumption does not hold. Principal Component Analysis does not suffer from this limitations. PCA has been used to investigated diffusive motions in proteins,[18–20] and interdomain motions.[21, 22] PCA is also useful in studying the conformational space explored by a protein[23] and the shape of this conformational space.[24] Current research is being performed to analyze the way energy propagates between protein dynamical modes of motion.[25–28]

2.3 NEUTRON SCATTERING EXPERIMENTS

2.3.1 NEUTRON SCATTERING THEORY BACKGROUND

Neutron scattering is a very valuable experimental method for studying protein dynamics. Thermal neutrons have the interesting property that they are especially sensitive to slow, thermal motions. Another advantage of neutron scattering is that neutron, being electrostatically neutral, only interact with the nuclei of the system and electrostatics need not be take into consideration. Neutrons are especially sensitive to hydrogen nuclei (a single proton) for which they have a very large scattering cross section, 10 times greater than for

any other nuclei. Using selective deuteration thus allows to focus on the interesting parts of the system, for example the protein.

Neutrons can be used in either of two ways: for spectroscopic measurements, using elastic and inelastic scattering to give dynamical information. They can also be used in crystallographic measurements, using elastic scattering on protein crystals to get structural information. In the present work, the experiment performed aimed at gaining insight into the dynamics of the system. Therefore the crystallographic methods will not be covered here.

DERIVATION OF THE SCATTERING EQUATIONS

This section presents a simple way of deriving the neutron scattering equations. For a more elaborate and in depth discussion, the reader is referred to the standard textbooks by M. Bée[29] and W.Marshall & S.W. Lovesey.[30]

Let us start from the probability per unit time of a change in the total system from the initial state $|k_n\rangle|k_s\rangle$ to the final state $|k'_n\rangle|k'_s\rangle$:

$$W_{k_n k'_n k_s k'_s} = \frac{2\pi}{\hbar} \left|\langle k_s|\langle k_n|H_c|k'_n\rangle k'_s\rangle\right|^2 \delta(\Delta E_s - \Delta E_n) \tag{2.36}$$

where the subscripts s and n stands for the sample and the neutron respectively. The probability of change in the state of the neutron is then given by:

$$W_{k_n k'_n} = \frac{1}{Z} \sum_{k_s, k'_s} W_{k_n k'_n k_s k'_s} e^{-\beta E'_s} \tag{2.37}$$

The dependence on the initial and final state of the neutron can be removed by using the operator:

$$\overline{H}_c = \langle k_n|H_c|k'_n\rangle \tag{2.38}$$

which acts on the states of the sample only. Equation 2.36 can then be rewritten:

$$W_{k_n k'_n} = \frac{2\pi}{\hbar^2 Z} \sum_{k_s k'_s} e^{-\beta E'_s} \left|\langle k_s|\overline{H}_c|k'_s\rangle\right|^2 \delta(\omega_s - \omega) \tag{2.39}$$

where

$$\hbar\omega = E'_n - E_n$$

and

$$\hbar\omega_s = E'_s - E_s$$

The Kröniker delta can be written as an integral:

$$\delta(\omega_a - \omega) \rightarrow \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(\omega_a - \omega)t} \, \mathrm{d}t \tag{2.40}$$

In the Heisenberg representation the Hamiltonian operator can be expressed as:

$$\overline{H}_c(t) = e^{i\omega_s t} \, \overline{H}_c(0) \, e^{i\omega'_a t} \tag{2.41}$$

So that part of equation 2.39 can be rewritten:

$$\begin{aligned} \sum_{k_s k'_s} e^{-\beta E'_s} |\langle k_s | \overline{H}_c | k'_s \rangle|^2 &= \sum_{k_a} e^{-\beta E'_s} \langle k_s | \overline{H}_c(0) \overline{H}_c(t) | k_a \rangle \\ &= \langle \overline{H}_c(0) \overline{H}_c(t) \rangle \end{aligned} \tag{2.42}$$

which is simply the self correlation function of the Hamiltonian operator. The probability $W_{k_n k'_n}$ of equation 2.39 is then:

$$W_{k_n k'_n} = \frac{1}{\hbar^2} \int_{-\infty}^{\infty} \langle \overline{H}_c(0) \overline{H}_c(t) \rangle e^{i\omega t} \, \mathrm{d}t \tag{2.43}$$

At this stage, a simple 'collision' energy function is introduced for the interaction potential in the Hamiltonian:

$$H_c = \frac{2\pi\hbar^2}{m} \sum_i b_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{2.44}$$

where the summation is over the different atoms types i, at positions $\mathbf{r}_i$, each with scattering lengths b_i, $\mathbf{r}$ being the position of the neutron. The scattering lengths, b_i depend on the spin states of the nuclei i and also takes into account the different nuclear isotopes.

For a given nuclear species the scattering lengths and cross section are defined as[†]:

Scattering lengths	Scattering cross sections
$b_{coh} = \langle b_s \rangle$	$\sigma_{coh} = 4\pi \langle b_s \rangle^2$
$b_{inc} = \left(\langle b_s^2 \rangle - \langle b_s \rangle^2 \right)^{1/2}$	$\sigma_{inc} = 4\pi \left(\langle b_s^2 \rangle - \langle b_s \rangle^2 \right)$

where the subscript s spans the possible spin states and isotopes of a given atomic species.

[†]Assuming a current of I_0 neutrons per second and per square centimetre incident on the sample, I_s and I_a, the number of scattering and absorption event occurring each second is defined in terms of the scattering and absorption cross sections σ_s and σ_a by:

$$I_s = I_0 \sigma_s$$

$$I_a = I_0 \sigma_a$$

σ_s and σ_a have dimension of surfaces. Their usual unit is the barn: 1 barn = 10^{-24} cm^2, and at low energies are roughly proportional to the neutron incident wavelength.

For a change in the neutron state $k_n \rightarrow k'_n$, the Hamiltonian is then written:

$$\begin{aligned} \overline{H}_c &= \langle k_n | H_c | k'_n \rangle \\ &= \frac{2\pi\hbar^2}{m} \sum_i b_i \exp(i\mathbf{q} \cdot \mathbf{r}_i) \end{aligned} \tag{2.45}$$

where the neutron-wavevector transfer has been introduced:

$$\mathbf{q} = \mathbf{k}'_n - \mathbf{k}_n$$

Therefore:

$$\langle \overline{H}_c(0) \overline{H}_c(t) \rangle = \left(\frac{2\pi\hbar^2}{m} \right)^2 \sum_{i,j} \langle b_i b_j \exp(i\mathbf{q} \cdot \mathbf{r}_i(t)) \ \exp(-i\mathbf{q} \cdot \mathbf{r}_j(0)) \rangle \tag{2.46}$$

Finally the probability that a neutron leaves the sample in a solid angle $\mathrm{d}\mathbf{\Omega}$ and with an energy exchange $\mathrm{d}\omega$ can be written:

$$\frac{\partial^2\sigma}{\partial\mathbf{\Omega}\partial\omega} = \frac{k'_n}{k_n} \frac{1}{2\pi} \sum_{i,j} b_i b_j \int_{-\infty}^{\infty} \langle \exp(i\mathbf{q} \cdot \mathbf{r}_i(t)) \exp(-i\mathbf{q} \cdot \mathbf{r}_j(0)) \rangle \ e^{-i\omega t} \, \mathrm{d}t \tag{2.47}$$

2.3.2 SCATTERING FROM HYDROGENOUS COMPOUNDS

The intensity expressed in equation 2.47 can be split into two parts as:

$$\frac{\partial^2\sigma}{\partial\mathbf{\Omega}\partial\omega} = \left(\frac{\partial^2\sigma}{\partial\mathbf{\Omega}\partial\omega} \right)_{coh} + \left(\frac{\partial^2\sigma}{\partial\mathbf{\Omega}\partial\omega} \right)_{inc} \tag{2.48}$$

the two parts on the r.h.s corresponding to the coherent and incoherent scattering respectively. Equation 2.47 describes the scattering from a sample with different types of atoms. Actually we are only concerned with scattering from hydrogenous compounds where the most common isotope is hydrogen. Hydrogen has a scattering cross section $\sigma = 81.66$ barns more than 10 times greater than any other scattering cross section among organic compounds. The incoherent scattering cross section of hydrogen being $\sigma_{inc} = 79.9$ barns compared to $\sigma_{coh} = 1.75$ barns for the coherent scattering cross section, such that the overwhelming majority comes from hydrogen incoherent scattering. With only one isotope the equations become much more tractable:

$$\frac{\partial^2\sigma}{\partial\mathbf{\Omega}\partial\omega} = \frac{1}{4\pi N} \frac{k'_n}{k_n} \left[\sigma_{coh} S_{coh}(\mathbf{q}, \omega) + \sigma_{inc} S_{inc}(\mathbf{q}, \omega) \right] \tag{2.49}$$

The scattering functions (or scattering laws), $S_{coh}(\mathbf{q},\omega)$, and, $S_{inc}(\mathbf{q},\omega)$, are time-Fourier transforms of the intermediate scattering functions:

$$S_{coh}(\mathbf{q},\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} I_{coh}(\mathbf{q},\omega)e^{i\omega t}\,\mathrm{d}t \tag{2.50}$$

and

$$S_{inc}(\mathbf{q},\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} I_{inc}(\mathbf{q},\omega)e^{i\omega t}\,\mathrm{d}t \tag{2.51}$$

with the intermediate functions themselves being inverse space-Fourier transforms of a pair-correlation and self-correlation function respectively:

$$I_{coh}(\mathbf{q},t) = \frac{1}{N}\sum_i\sum_j \langle \exp(i\mathbf{q}\cdot\mathbf{r}_i(t))\exp(-i\mathbf{q}\cdot\mathbf{r}_j(0))\rangle \tag{2.52}$$

$$G_{pair}(\mathbf{q},t) = \frac{1}{(2\pi)^3}\int I_{coh}(\mathbf{q},t)\exp(-i\mathbf{q}\cdot\mathbf{r})\mathrm{d}\mathbf{q} \tag{2.53}$$

and

$$I_{inc}(\mathbf{q},t) = \frac{1}{N}\sum_i \langle \exp(i\mathbf{q}\cdot\mathbf{r}_i(t))\exp(-i\mathbf{q}\cdot\mathbf{r}_i(0))\rangle \tag{2.54}$$

$$G_{self}(\mathbf{q},t) = \frac{1}{(2\pi)^3}\int I_{inc}(\mathbf{q},t)\exp(-i\mathbf{q}\cdot\mathbf{r})\mathrm{d}\mathbf{q} \tag{2.55}$$

where:

$$G_{pair}(\mathbf{r},t) = \frac{1}{N}\sum_i\sum_j \int \langle \delta(\mathbf{r}-\mathbf{r}'+\mathbf{r}_i(0))\delta(\mathbf{r}'-\mathbf{r}_j(t))\rangle \mathrm{d}\mathbf{r}' \tag{2.56}$$

and

$$G_{self}(\mathbf{r},t) = \frac{1}{N}\sum_i \int \langle \delta(\mathbf{r}-\mathbf{r}'+\mathbf{r}_i(0))\delta(\mathbf{r}'-\mathbf{r}_i(t))\rangle \mathrm{d}\mathbf{r}' \tag{2.57}$$

The entire scheme can be summarized the following way:

$$S_{coh}(\mathbf{q},\omega) \overset{\text{time FT}}{\Longleftarrow} I_{coh}(\mathbf{q},t) \overset{\text{space FT}}{\Longrightarrow} G_{pair}(\mathbf{r},t) \tag{2.58}$$

$$S_{inc}(\mathbf{q},\omega) \overset{\text{time FT}}{\Longleftarrow} I_{inc}(\mathbf{q},t) \overset{\text{space FT}}{\Longrightarrow} G_{self}(\mathbf{r},t) \tag{2.59}$$

2.3.3 GAUSSIAN APPROXIMATION: OBTAINING MEAN-SQUARE FLUCTUATIONS

Neutron scattering data yields useful information about the dynamics of the system. The mean-square fluctuation, $\langle u^2\rangle$ of the protein atoms can be extracted from neutron scattering data through the so-called gaussian approximation which will now be presented. For

a more complete discussion of the way the MSF is calculated and of the approximation being made, the reader is referred to the study by J. Hayward *et al* 2002[31] and references therein.

The mean-square displacement of an atom i, $\langle u_i^2 \rangle$ is defined as:

$$\langle u_i^2 \rangle = \langle (\mathbf{r}_i - \mathbf{r}_i^m)^2 \rangle \tag{2.60}$$

where $\mathbf{r}_i$ is the position of atom i and $\mathbf{r}_i^m$ its mean position, and the average is taken over time.

To obtain a value for the MSF, $\langle u^2 \rangle$ of the whole protein one starts from the incoherent intermediate scattering function $I_{inc}(\mathbf{q}, t)$ of equation 2.54 and make use of the cumulant expansion:

$$I_{inc}(\mathbf{q}, t) = \frac{1}{N} \sum_i b_i^2 \langle \exp(i\mathbf{q} \cdot \mathbf{r}_i(t)) \exp(-i\mathbf{q} \cdot \mathbf{r}_i(0)) \rangle \tag{2.61}$$

$$= \frac{1}{N} \sum_i b_i^2 \exp\left(-\frac{1}{2} \langle [\mathbf{q} \cdot (\mathbf{r}_i(t) - \mathbf{r}_i(0))]^2 \rangle \pm ... \right) \tag{2.62}$$

Neglecting higher order terms, and integrating over all $\mathbf{q}$ directions, $I_{inc}(q, t)$, can be rewritten:

$$I_{inc}(q, t) = \frac{1}{N} \sum_i b_i^2 \exp\left(-\frac{q^2}{6} \langle (\mathbf{r}_i(t) - \mathbf{r}_i(0))^2 \rangle \right) \tag{2.63}$$

For harmonic motions and restricted motion such as those found in proteins, the long time limit is taken and the incoherent scattering function, S_{inc}, can be expressed as:

$$S_{inc}(q, \omega = 0) = \frac{1}{N} \sum_i b_i^2 \exp\left(-\frac{q^2}{6} \langle u_i^2 \rangle \right) \tag{2.64}$$

Assuming all atoms appearing in the summation of equation **??** to be identical hydrogen atoms undergoing similar harmonic motions, the gaussians in equation **??** can be approximated by a single gaussian. $S_{inc}(q, \omega = 0)$ can be then rewritten:

$$S_{inc}(q, \omega = 0) = b^2 \exp\left(-\frac{q^2}{6} \langle u^2 \rangle \right) \tag{2.65}$$

such that the protein MSF, $\langle u^2 \rangle$ can be obtained from neutron scattering data using the following equation:

$$\ln S_{inc}(q, \omega = 0) = -\frac{q^2}{6} \langle u^2 \rangle + C \tag{2.66}$$

where C is a constant independent of q.

The above calculations of the mean-square fluctuations rely on a number of non trivial

assumptions:

- **Low q values**: The approximations made in order to obtain equation 2.66 are only valid for small q values. However owing to the sparsity of the data, one often has to resort to using values of q^2 up to ~ 1.4 $\mathring{A}^2$ which is the limit of validity of the approximation.

- **Long time limit**: in going from equation 2.63 to equation 2.64 the long time limit is taken. In practice the instruments have a finite energy resolution which mean they are sensitive to only a certain time range ($\sim$200 ps for IN6, $\sim$5 ns for IN16)[32] so that the values obtained are instrument dependent as was seen in the introduction, section:1.1.2.

- **Similarity of motion**: all contributions are considered to be coming from hydrogen atoms, and these hydrogens are considered to have identical motions. This assumed similarity of motion has been put under question.[31]

To summarize, the procedure to obtain mean square fluctuations using equation 2.66 is subject to many limitations. However, it does give us a useful measure of the dynamics in the protein system. The different approximations made have been shown to result in $\langle u^2 \rangle$ being under-estimate by as much as 30%.[31]

Another of obtaining is the mean-square fluctuations of the protein, $\langle u^2 \rangle$ is directly from the integrated peak intensity. This approach has the advantage that it does not rely on fitting low q region of the scattering the data this being often a difficult task owing to the noise in the data. This simpler approach uses the statistically stronger integrated peak intensities to calculate the $\langle u^2 \rangle$. These are calculated in the following way:

$$S_{inc}(q) = b^2 \exp\left(-\frac{q^2}{6}\langle u^2 \rangle\right) \tag{2.67}$$

Integration over the low q region is performed via a simple summation over the $S_{inc}(q_i)$'s:

$$\sum_i S_{inc}(q_i) = \sum_i b^2 \exp\left(-\frac{q_i^2}{6}\langle u^2 \rangle\right) \tag{2.68}$$

$$= C - \sum_i b^2 \, \frac{q_i^2}{6} \, \langle u^2 \rangle + \tag{2.69}$$

such that:

$$\langle u^2 \rangle = C - \frac{6}{b^2 \sum_i q_i^2} \sum_i S_{inc}(q_i^2) \tag{2.70}$$

Setting $\langle u^2 \rangle = 0$ at low temperature, $\langle u^2 \rangle$ is obtained directly from the integrated elastic peak intensities.

$$\langle u^2 \rangle = 1 - \frac{6}{b^2 \sum_i q_i^2} \sum_i S_{inc}(q_i^2) \tag{2.71}$$

This procedure was used to determined the $\langle u^2 \rangle$ from the neutron experiments in chapter 7.

REFERENCES

[1] LEACH, A. R. *Molecular Modelling Principles and Applications*. Prentice Hall, 2001.

[2] TUCKERMAN, M. E., AND MARTYNA, G. J. Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem.*, 2000, **194**, 159–178.

[3] HOOVER, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 1985, **31**(3), 1695–1697.

[4] MARTYNA, G. J., TUCKERMAN, M. E., TOBIAS, D. J., AND KLEIN, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.*, 1996, **87**(5), 1117–1157.

[5] YOSHIDA, H. Construction of higher order symplectic integrators. *Phys. Lett. A*, 1990, **150**(5,6,7), 262–68.

[6] SUZUKI, M. General theory of fractal path integrals with applications to many-body theories and statistical physics. *J. Math. Phys.*, 1991, **32**(2), 400–7.

[7] BROOKS, B., AND KARPLUS, M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc Natl Acad Sci U S A*, 1985, **82**(15), 4995–9.

[8] SMITH, J. C., CUSACK, S., TIDOR, B., AND KARPLUS, M. Inelastic neutron scattering analysis of low-frequency motions in proteins: Harmonic and damped harmonic models of bovine pancreatic trypsin inhibitor. *J. Chem. Phys.*, 1990, **93**(5), 2974–91.

[9] KIDERA, A., AND GO, N. Refinement of protein dynamic structure: normal mode refinement. *Proc Natl Acad Sci U S A*, 1990, **87**(10), 3718–22.

[10] KIDERA, A., AND GO, N. Normal mode refinement: crystallographic refinement of protein dynamic structure. i. theory and test by simulated diffraction data. *J Mol Biol*, 1992, **225**(2), 457–75.

[11] KIDERA, A., INAKA, K., MATSUSHIMA, M., AND GO, N. Normal mode refinement: crystallographic refinement of protein dynamic structure. ii. application to human lysozyme. *J Mol Biol*, 1992, **225**(2), 477–86.

[12] ABSEHER, R., HORSTINK, L., HILBERS, C. W., AND NILGES, M. Essential spaces defined by nmr structure ensembles and molecular dynamics simulation show significant overlap. *Proteins*, 1998, **31**(4), 370–82.

[13] ICHIYE, T., AND KARPLUS, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins*, 1991, **11**(3), 205–17.

[14] KITAO, A., HIRATA, F., AND GO, N. The effects of solvent on the conformational and the collective motions of protein: normal mode analysis ad molecular dynamics simulations of melittin in water and in vaccum. *Chem. Phys.*, 1991, **158**, 447–472.

[15] BROOKS, B. R., JANEZIC, D., AND KARPLUS, M. Harmonic analysis of large systems. i methodology. *J. Comput. Chem.*, 1995, **16**(12), 1522–1542.

[16] JANEZIC, D., AND BROOKS, B. Harmonic analysis of large systems. ii. comparison of different protein models. *J. Comput. Chem.*, 1995, **16**(12), 1543–53.

[17] JANEZIC, D., AND VENABLE, R. Harmonic analysis of large systems. iii. comparison with molecular dynamics. *J. Comput. Chem.*, 1995, **16**(12), 1554–66.

[18] GARCIA, A. E. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, 1992, **68**(17), 2696–2699.

[19] HAYWARD, S., KITAO, A., HIRATA, F., AND GO, N. Effect of solvent on collective motions in globular protein. *J Mol Biol*, 1993, **234**(4), 1207–17.

[20] AMADEI, A., LINSSEN, A. B., AND BERENDSEN, H. J. Essential dynamics of proteins. *Proteins*, 1993, **17**(4), 412–25.

[21] GARCIA, A. E., AND HARMAN, J. G. Simulations of crp:(camp)2 in noncrystalline environments show a subunit transition from the open to the closed conformation. *Protein Sci*, 1996, **5**(1), 62–71.

[22] HAYWARD, S., KITAO, A., AND BERENDSEN, H. J. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins*, 1997, **27**(3), 425–37.

[23] CAVES, L. S., EVANSECK, J. D., AND KARPLUS, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci*, 1998, **7**(3), 649–66.

[24] KITAO, A., HAYWARD, S., AND GO, N. Energy landscape of a native protein: jumping-among-minima model. *Proteins*, 1998, **33**(4), 496–517.

[25] BAHAR, I., ATILGAN, A. R., DEMIREL, M. C., AND ERMAN, B. Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.*, 1998, **80**(12), 2733–2736.

[26] MORITSUGU, K., MIYASHITA, O., AND KIDERA, A. Vibrational energy transfer in a protein molecule. *Phys. Rev. Lett.*, 2000, **85**(18), 3970–3.

[27] MORITSUGU, K., MIYASHITA, O., AND KIDERA, A. Temperature dependence of vibrational energy transfer in a protein molecule. *J. Phys. Chem. B*, 2003, **107**, 3309–17.

[28] YU, X., AND LEITNER, D. M. Vibrational energy transfer and heat conduction in a protein. *J. Phys. Chem. B*, 2003, **107**, 1698–1707.

[29] BÉE, M. *Quasielastic Neutron Scattering: Principles and Applications in Solid State Chemistry, Biology and Material Science*. Adam Hilger, 1988.

[30] MARSHALL, W., AND LOVESEY, S. W. *Theory of Thermal Neutron Scattering*. Clarendon Press, 1971.

[31] HAYWARD, J. A., AND SMITH, J. C. Temperature dependence of protein dynamics: computer simulation analysis of neutron scattering properties. *Biophys. J.*, 2002, **82**(3), 1216–25.

[32] DANIEL, R. M., FINNEY, J. L., REAT, V., DUNN, R., FERRAND, M., AND SMITH, J. C. Enzyme dynamics and activity: time-scale dependence of dynamical transitions in glutamate dehydrogenase solution. *Biophys. J.*, 1999, **77**(4), 2184–90.

CHAPTER 3

SYSTEM USED IN MOLECULAR DYNAMICS SIMULATIONS

CONTENTS

INTRODUCTION

This chapter describes the protein myoglobin which was used for the Nosé-Hoover, Nosé-Hoover-Chain simulations, as well as for performing the principal component analysis. The way the system was setup and relaxed is described.

3.1 MYOGLOBIN

The simulations were performed on the protein Myoglobin. Myoglobin is a monomeric heme protein found mainly in muscle tissue where it serves as an intracellular storage site for oxygen. During periods of oxygen deprivation oxymyoglobin releases its bound oxygen which is then used for metabolic purposes. The tertiary structure of myoglobin is that of a typical water soluble globular protein. Its secondary structure is unusual in that it contains a very high proportion (75%) of α-helical secondary structure. A myoglobin polypeptide is comprised of 8 separate right handed a-helices, designated A through H, that are connected by short non helical regions. Each myoglobin molecule contains one heme group inserted into a hydrophobic cleft in the protein. Each heme residue contains

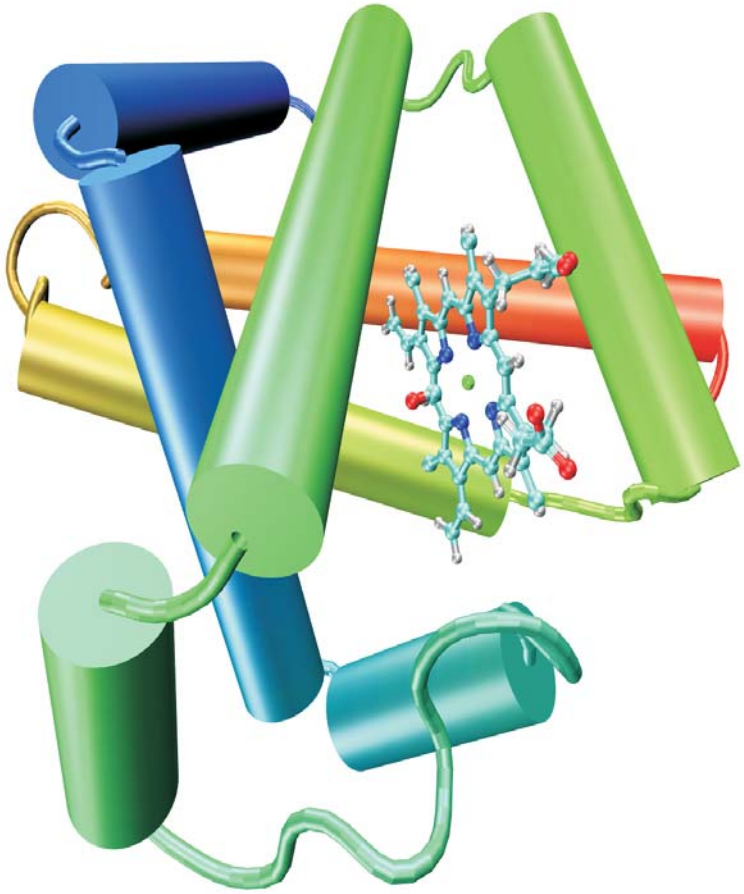


Figure 3.1: The protein myoglobin in a 'cartoon' representation showing the 8 helices, the heme group and the carbon-monoxide bound to it.

one central coordinately bound iron atom that is normally in the Fe2+, or ferrous, oxidation state. The oxygen carried by hemeproteins is bound directly to the ferrous iron atom of the heme prosthetic group. Oxidation of the iron to the Fe3+, ferric, oxidation state renders the molecule incapable of normal oxygen binding. Carbon monoxide also binds to heme iron atoms in a manner similar to that of oxygen, but the binding of carbon monoxide to the heme group is much stronger than that of oxygen. The preferential binding of carbon monoxide to the heme iron is largely responsible for the asphyxiation that results from carbon monoxide poisoning.

3.2 SIMULATION SYSTEM AND PARAMETERS

[†]The myoglobin structure used is actually that of carbonmonoxide-myoglobin or CO-Mb. The presence of carbon monoxide in the heme pocket is not thought to influence the system's dynamics to any degree.

The model consists of one myoglobin molecule† surrounded by a shell of water molecules, constructed by placing the protein in a box of water and retaining those 492 waters closest to the protein.† The model system mimics the hydrated powder sample used in neutron scattering experiments in Doster *et al* 1989.[2] The myoglobin coordinates were taken from the Protein Data Bank[3] (from the RCSB site: www.rscb.org) structure 1A6G,[4] solved at 1.15 $\mathring{A}$ resolution using X-ray crystallography.

The CHARMM package version 27b2 was used to perform the simulations.[5] The TIP3P potential function was used to represent the water molecules.[6] The all-atom parameter set 22 was used throughout the simulations.[7] A shift function with a 12 $\mathring{A}$ cutoff was used to truncate the electrostatic interactions and a switch function was used to truncate the van der Waals contributions over 10-12 $\mathring{A}$‡. A relative dielectric constant of 1 was used. A time step of 1 fs was used.

3.3 SYSTEM RELAXATION

In order to relax the system the following calculations were sequentially performed:

- 200 steps of minimization using Steepest Descent (SD) with harmonic constraints on the protein of 10 kcal mol$^{-1}\mathring{A}^{-2}$.
- 200 steps of minimization (SD) with harmonic constraints on the protein of 1 kcal mol$^{-1}\mathring{A}^{-2}$.
- 5 ps heating phase up to 180 K by increments of 5 K every 50 steps, fixing the protein atoms.
- 5 ps equilibration at 180 K with harmonic constraints on the protein of 1 kcal mol$^{-1}\mathring{A}^{-2}$.
- 5 ps equilibration at 180 K without constraints.

†The model system constructed and the potential function used were that of Vitkup *et al* 2000, as much as was made possible from the detail published and personal communications.[1]

‡Van der Waals and electrostatic interactions exist between every pair of atoms in the system, the number of interaction to be calculated thus goes as N^2, N being the number of atoms in the system. However, this might include pairs of atoms so far away from each other that their interaction is negligible. To save computation time, a **pair list** is determined of pairs of atoms in sufficiently close proximity that their interaction is significant (the maximum distance is usually of the order of $15\mathring{A}$). The interaction energies are then only calculated for the pairs of atoms present in the list, thereby gaining considerable computation time. The list is periodically updated during the simulation so as to make sure all atom-pairs whose interactions should be taken into account are indeed present in the list.

Using such a list in effect introduces a cutoff in the electrostatics. The abruptness of this cutoff can lead to spurious results when a charged atoms is at the limit of the cutoff. To avoid this, two schemes can be used: the **shift** cutoff in which the electrostatic potential is shifted so as to equal to zero at the cutoff distance. The second is the **switch** scheme, it introduces a switching region in which the electrostatic potential is replaced by a switch function (sigmoidal function) which links the non interacting region with the interacting region in a continuous way.

- 200 steps of minimization using Steepest Descent (SD).

- 100 steps of minimization Adopted Basis Newton-Raphson (ABNR).[5]

After relaxation the system was considered ready for simulations. The relaxed system was then used as starting point for the MD simulations presented in the next chapters.

REFERENCES

[1] VITKUP, D., RINGE, D., PETSKO, G. A., AND KARPLUS, M. Solvent mobility and the protein 'glass' transition. *Nat. Struct. Biol.*, 2000, **7**(1), 34–8.

[2] DOSTER, W., CUSACK, S., AND PETRY, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature*, 1989, **337**(6209), 754–6.

[3] BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N., AND BOURNE, P. E. The protein data bank. *Nucleic Acids Res*, 2000, **28**(1), 235–42.

[4] VOJTECHOVSKY, J., CHU, K., BERENDZEN, J., SWEET, R. M., AND SCHLICHTING, I. Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophys. J.*, 1999, **77**(4), 2153–74.

[5] BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S., AND KARPLUS, M. Charmm: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Biol.*, 1983, **4**(2), 187–217.

[6] JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W., AND KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 1983, **79**(2), 926–934.

[7] MACKERELL, J. A. D., BASHFORD, D., BELLOTT, M., R. L. DUNBRACK, J., EVANSECK, J. D., FIELD, M. J., FISCHER, S., GAO, J., GUO, H., HA, S., JOSEPH-MCCARTHY, D., KUCHNIR, L., KUCZERA, K., LAU, F. T. K., MATTOS, C., MICHNICK, S., NGO, T., NGUYEN, D. T., PRODHOM, B., W. E. REIHER, I., ROUX, B., SCHLENKRICH, M., SMITH, J. C., STOTE, R., STRAUB, J., WATANABE, M., WIORKIEWICZ-KUCZERA, J., YIN, D., AND KARPLUS, M. All-atom empirical potential for molecular modelling and dyamics studies of proteins. *J. Phys. Chem.*, 1998, **102**(18), 3586–3616.

CHAPTER 4

NOSÉ-HOOVER DUAL HEATBATH SIMULATIONS

CONTENTS

INTRODUCTION

In this chapter the simulations performed by Vitkup *et al* 2000[1] are reproduced. In attempting to do so problems occured in the simulations, these problems took the form of unphysical temperature oscillations in the system. In order to overcome this problem a different implementation of the algorithm was made. The results thus obtained are compared to the previous results. Some previous conclusions made from the results of Vitkup *et al* 2000[1] are shown to be incorrect.

4.1 PRINCIPLE OF THE MULTIPLE HEATBATH METHOD

The idea of the multiple heatbath approach is in essence very simple: one part of the simulation system, the protein say, is set to a temperature T_p while another part, the solvent

say, is set at another temperature T_s. Relatively modest modifications to the temperature regulation algorithms allow this type of simulations to be performed. Of course such a scheme is unphysical and no pretence as to its physicality is made.

The motivation behind such a scheme is to be able to investigate the motions of one part of the system, here the protein, and show the influence the other part, here the solvent, has on it. The dynamics of the two parts usually remain very much interconnected and it is very difficult to determine the origin of certain measured properties. The multiple heatbath scheme offers the possibility of dissociating the contribution to the dynamics of the system coming from both parts of the system. Figure 4.1 illustrates the way the Nosé-Hoover algorithm functions.

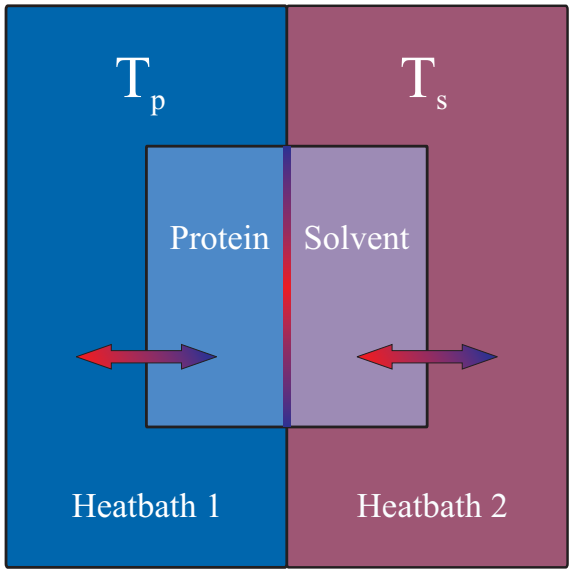


Figure 4.1: Schematic representation of the way the Nosé-Hoover algorithm functions. Two parts of the system (in this case the protein and its solvent) are each connected to a different heatbath so as to be regulated at their different temperatures: T_p and T_s.

4.2 SIMULATION PROTOCOL

The multiple heatbath approach was first used by Vitkup *et al* in 2000.[1] Following their study, a coupling constant of 200 kcal mol^{-1}ps^{-2} was used for the Nosé-Hoover algorithm, together with a tolerance of 10^{-10} kcal mol^{-1} and a maximum of 10 cycles for convergence of the predictor-corrector method.

Following the procedure used in the previous study, temperatures of 180 K and 300 K were used in different combinations: Protein Cold / Solvent Cold (PC/SC), Protein Cold / Solvent Hot (PC/SH), Protein Hot / Solvent Cold (PH/SC) and Protein Hot / Solvent

Hot (PH/SH). Two simulation protocols were used (Methods 1 and 2). The simulation protocols are divided into three sequential stages:

1. System relaxation

2. System preparation

3. Nosé-Hoover simulation

The system relaxation and Nosé-Hoover simulation stages were the same for both methods. The two methods differ in the system preparation stage of the simulation protocol. The system relaxation was performed as described in the previous chapter.

METHOD 1

Method 1 resembles as closely as was practically possible the protocol used in Ref. 1. Personal communication established that in the work in Ref. 1 a 3 ps to 5 ps heating phase to 300 K in 5 K increments was used. This was therefore adopted in Method 1.

METHOD 2

Method 2 avoids the instantaneous temperature reduction of the protein from 300 K to 180 K present in the protein cold simulations at the beginning of the third (i.e., Nosé-Hoover simulation) stage of Method 1. The system was heated during 3 ps to the temperature the protein was to have during the Nosé-Hoover run, *i.e.*, 180 K for PC/SC, PC/SH and 300 K for PH/SC and PH/SH. The system was subsequently equilibrated at that temperature during 50 ps.

The Nosé-Hoover simulation stage was the same for all simulations and follows the protocol used in Ref. 1. The Nosé-Hoover thermostat was turned on, setting the protein and the solvent at their respective chosen temperatures. In all methods 50 ps of equilibration were performed followed by 100 ps for the production run.

The simulation model and force field used in the present work (Methods 1 and 2) and in Ref. 1 are the same. Consequently, any differences in the results depend on the method of system preparation. The question therefore arises as to which preparation method is most suitable. Methods that equilibrate the system at the temperature at which the protein will be fixed during the Nosé-Hoover run (e.g. Method 2, here) may be the least perturbative. These methods avoid the instantaneous reduction of the protein temperature from 300 K to 180 K that was present in Ref. 1 and in the closely analogous Method 1 simulation here.

The mean-square fluctuations of the protein, $\langle u^2 \rangle$, were calculated from the 100 ps production runs and compared with each other. Errors in $\langle u^2 \rangle$ were estimated by calculating the standard error for the $\langle u^2 \rangle$ in 10 bins along the 100 ps of the production runs.

4.3 INSTABILITIES IN NOSÉ-HOOVER ALGORITHM: THE TODA DEMON

4.3.1 INITIAL RESULTS

When first trying out the Nosé-Hoover algorithm present in the CHARMM package,[2] great difficulty was encountered in getting the system to behave in a stable enough manner, fit for subsequent analysis. Figure 4.2 shows typical runs obtained when running the algorithm making best use of the available information. The PC/SC run is clearly unstable and would often simply break down due to overflow problems in the CHARMM internal variables. Other runs such as PC/SH and PH/SH although less prone to instability still showed intermittent signs of undesirable instabilities.

4.3.2 THE TODA DEMON OSCILLATIONS

On the way towards finding a solution to these instabilities it was discovered that the Nosé-Hoover algorithm is known to be subject to instability.[3] These instabilities known as Toda demon appear in the form of unphysical oscillations in the temperature of the system (see for example figure 4.2-PC/SC).

An initial analysis of the trajectory through Fourier Transform showed the frequency of the oscillations to be well defined, thus the presence of the Toda demon problem was considered a definite possibility.

The Toda oscillations appear in the following way (from Holain 1995[3]). The rate of change of the internal energy of the atoms in the system can be related to the heat flow into the system from the thermal bath in the following way (cf equation 2.9):

$$\begin{aligned} E &= \sum_i \frac{1}{2} m_i \mathbf{v}_i^2 + V(\mathbf{r}) \\ \dot{E} &= \sum_i m_i \mathbf{v}_i \cdot \dot{\mathbf{v}}_i + \sum_i \frac{\partial V}{\partial \mathbf{r}_i} \cdot \mathbf{v}_i \\ &= \sum_i \mathbf{v}_i \cdot (\mathbf{F}_i - m_i \frac{p_\eta}{Q} \mathbf{v}_i - \sum_i \mathbf{F}_i \cdot \mathbf{v}_i \\ &= -\frac{p_\eta}{Q} \sum_i m_i \mathbf{v}_i^2 \\ &= -\frac{p_\eta}{Q}\, dNkT \end{aligned} \tag{4.1}$$

Now, suppose that the average be taken over the high-frequency oscillations in the system; then, the internal energy can be decomposed into the zero-temperature potential energy, V_0, and the thermal part, given by the heat capacity times the temperature:

$$\begin{aligned} \overline{E}_0 &= V_0 + C_v \overline{T} \\ \dot{\overline{E}}_0 &= C_v \dot{\overline{T}} \end{aligned} \tag{4.2}$$

where the bar over the symbols indicates coarse grained averaging over roughly on vibrational period. Now considering that the temperature is not stable at T_0 but can oscillate around it, and using the fact that in most condensed-phase systems, $C_v \approx dNk$, equations 4.1 and 4.2 can be combined as:

$$C_v \dot{\overline{T}} = -\frac{p_\eta}{Q}\, dNkT \Longrightarrow \frac{\dot{T}}{T} = \frac{p_\eta}{Q} \tag{4.3}$$

At this stage a collective variable, q for the system, is introduced:

$$\begin{aligned} q &= -\ln \frac{T}{T_0} \Longrightarrow \frac{T}{T_0} = e^{-q} \\ \dot{r} &= -\frac{\dot{T}}{T} = \frac{p_\eta}{Q} \\ \ddot{r} &= \frac{1}{Q} \dot{p}_\eta \\ &= \frac{dNkT_0}{Q} \left[\frac{T}{T_0} - 1 \right] \\ &= \frac{dNkT_0}{Q} (e^{-r} - 1) \end{aligned}$$

such that, to first order approximation:

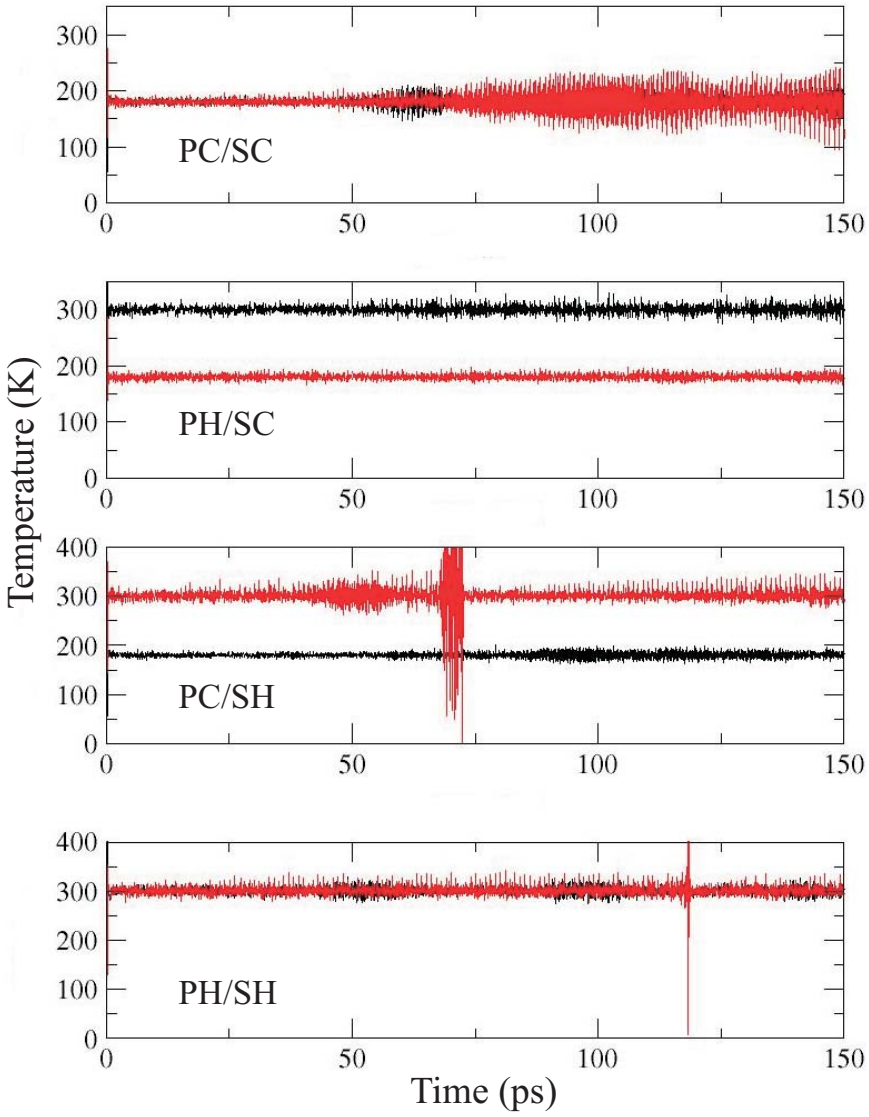


Figure 4.2: Typical temperature trajectories during production runs obtained using the Nosé-Hoover algorithm, as found in the previous implementation. The protein temperature is represented in black, the solvent temperature in red.

$$\ddot{q} = -\omega^2 q \tag{4.4}$$

where the fundamental harmonic frequency of oscillation of this pseudoparticle is given by:

$$\omega = \sqrt{\frac{dNkT_0}{Q}} \tag{4.5}$$

The best way to prevent this type of unwanted oscillation is to choose Q in a way such that the fundamental frequency of oscillation ω matches a fundamental frequency of the system. For condensed phase systems such as the present one, a suitable frequency is $\sim$2 ps.

These oscillations are due to temperature oscillations building up between the system and its heatbath. A more robust way out of this problem is to use a heatbath chain, in this setup an extra heatbath is added which regulates the first heatbath thus preventing any oscillations from building up between the system and the first heatbath. This scheme will be introduced in greater detail in the next chapter along with the results obtained using it.

4.4 NEW IMPLEMENTATION OF THE ALGORITHM

Having discovered this possible solution to the Toda problem, it was decided to implement the Nosé-Hoover-Chain algorithm in the CHARMM package. A detailed description of this algorithm is given in chapter 5.

Having implemented the Nosé-Hoover-Chain the presence of some residual unwanted oscillations in the temperature was still noticed. The presence of these oscillations was found to be dependent on the timestep used. In an attempt to remedy this, multiple timesteps and Yoshida-Suzuki steps where implemented thereby increasing the accuracy of the algorithm (cf. section 2.1.4). However some unwanted oscillation and instabilities remaind nevertheless. After a wild and epic hunt through the code, a few mistakes were found in the previous implementation, in particular the keyword 'NTRFRQ' which is commonly used to remove any drift or rotation of the system about its axis was found to also perform a full reset of the NH/NHC internal variables thereby creating unwanted oscillations. This was fixed and the multiple timesteps and Yoshida-Suzuki steps were found to be very effective in stabilizing the system. The temperature trajectories where then suitably stable, as can be seen in figure 4.3

4.5 REPRODUCTION OF PREVIOUS RESULTS

In order to compare with the previous results obtained using the Nosé-Hoover algorithm, the Nosé-Hoover-Chain was reduced to chains with only one heatbath thereby effectively reducing it to the Nosé-Hoover algorithm.

The protein $\langle u^2 \rangle$ results for Methods 1 and 2, described in the previous chapter, are presented in Table 4.1. Comparison of the $\langle u^2 \rangle$ obtained for PC/SC and PH/SC with those obtained for the hot solvent simulations indicate that the cold solvent has a distinct caging effect on the protein atoms. This effect is significantly stronger using Method 2, as

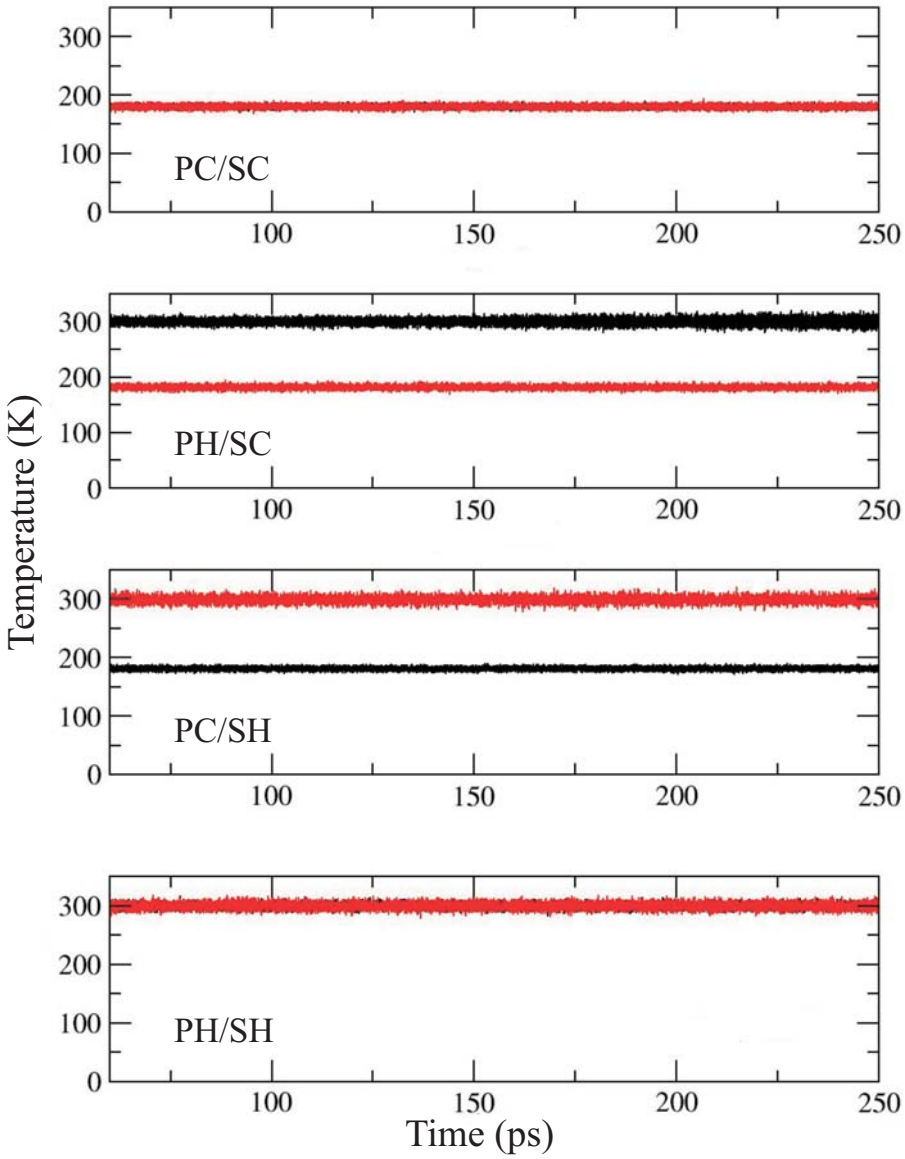


Figure 4.3: Typical temperature trajectories during production runs obtained using the Nosé-Hoover newly implemented algorithm. The protein temperature is represented in black, the solvent temperature in red.

can be deduced from the fact that the ratio of the backbone (and, in brackets, side-chain) atom fluctuations of PH/SC to those of PC/SC is 1.49 (1.5) in Method 1 and 1.28 (1.18) in Method 2, and that the ratio of the backbone atom fluctuations of PH/SH to those of PC/SC is 2.68 (3.0) in Method 1 and 3.28 (3.45) in Method 2. Comparing PC/SH with PC/SC shows that in both Methods 1 and 2 an effect of hot solvent is to double the protein $\langle u^2 \rangle$. This remarkable result agrees with that found in Ref. 1.

Of particular interest is the question as to what fraction of the PH/SH fluctuations can be obtained by heating only the solvent (i.e. in PC/SH). The data obtained using both

methods (Method 1 and Method 2) shows the $\langle u^2 \rangle$ from PC/SH to lie between PC/SC and the PH/SH. One measure of the effect of the solvent on the high temperature protein fluctuations is given by:

$$S = \frac{\langle u^2 \rangle_{PC/SH} - \langle u^2 \rangle_{PC/SC}}{\langle u^2 \rangle_{PH/SH} - \langle u^2 \rangle_{PC/SC}} \tag{4.6}$$

If solvent were to drive the high temperature protein fluctuations S should approach 1.0. For the backbone (and, in brackets, side-chain) atoms S is 0.55 (0.50) for Method 1 and 0.37 (0.37) for Method 2. These compare with higher values of 0.64 (0.65) obtained in Ref. 1. For comparison, table 4.2 presents the results obtained using the Nosé-Hoover-

Table 4.1: Comparison of the $\langle u^2 \rangle$ results obtained using Methods 1 and 2

	PC/SC		**PC/SH**		**PH/SC**		**PH/SH**	
	Meth 1	*Meth 2*	*Meth 1*	*Meth 2*	*Meth 1*	*Meth 2*	*Meth 1*	*Meth 2*
$\langle u^2 \rangle$ backbone	0.067	0.07	0.13	0.13	0.10	0.09	0.18	0.23
atoms ($\AA^2$)	± 0.003	± 0.007	± 0.01	± 0.012	± 0.005	± 0.007	± 0.01	± 0.03
$\langle u^2 \rangle$ non-H	0.10	0.11	0.20	0.21	0.15	0.13	0.30	0.38
atoms ($\AA^2$)	± 0.005	± 0.009	± 0.019	± 0.018	± 0.009	± 0.012	± 0.03	± 0.05

Chain dual heatbath algorithm (results taken from the Nosé-Hoover-Chain simulation presented in the next chapter). The results are very similar to the ones obtained using methods 1 & 2. The corresponding S value is 0.35 (0.31), which is even lower than the values obtained for methods 1 & 2, thereby showing again that the increase in protein fluctuations is not dominated by the solvent.

Table 4.2: Comparison of the $\langle u^2 \rangle$ results obtained using the Nosé-Hoover-Chainalgorithm

	PC/SC	**PC/SH**	**PH/SC**	**PH/SH**
$\langle u^2 \rangle$ backbone	0.077	0.13	0.10	0.23
atoms ($\AA^2$)	± 0.007	± 0.013	± 0.01	± 0.02
$\langle u^2 \rangle$ non-H	0.11	0.20	0.16	0.40
atoms ($\AA^2$)	± 0.010	± 0.024	± 0.019	± 0.03

CONCLUSION

Nosé-Hoover dual heatbath molecular dynamics simulations were performed to investigate solvent effects on fast (picosecond timescale) internal protein dynamics. Two main results are seen

(i) Low temperature solvent cages the protein fluctuations.

(ii) Heating the solvent while keeping the protein cold drives the protein fluctuations to values intermediate between those in the fully cold and fully hot systems.

The present Nosé-Hoover results thus confirm, in accord with Refs 4 and 1, that solvent strongly influences the dynamical transition in proteins. However, there is no clear evidence from the present work that, in the hydrated myoglobin system, the high-temperature protein fluctuations are dominated by the solvent - rather, both the protein and the solvent contribute.

The results presented in this chapter have been published in *Faraday Discussions*.[5]

REFERENCES

[1] VITKUP, D., RINGE, D., PETSKO, G. A., AND KARPLUS, M. Solvent mobility and the protein 'glass' transition. *Nat. Struct. Biol.*, 2000, **7**(1), 34–8.

[2] BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S., AND KARPLUS, M. Charmm: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Biol.*, 1983, **4**(2), 187–217.

[3] HOLIAN, B. L., VOTER, A. F., AND RAVELO, R. Thermostatted molecular dynamics: How to avoid the toda demon hidden in nose-hoover dynamics. *Phys. Rev. E. Stat. Phys.*, 1995, **52**(3), 2338–2347.

[4] REAT, V., DUNN, R., FERRAND, M., FINNEY, J. L., DANIEL, R. M., AND SMITH, J. C. Solvent dependence of dynamic transitions in protein solutions. *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**(18), 9961–6.

[5] TOURNIER, A. L., HUANG, D., SCHWARZL, S. M., FISCHER, S., AND SMITH, J. C. Time-resolved computational protein biochemistry: solvent effects on interactions, conformational transitions and equilibrium fluctuations. *Faraday Discuss*, 2003, **122**, 243–51.

CHAPTER 5

MULTIPLE HEATBATH SIMULATIONS

CONTENTS

INTRODUCTION

This chapter presents the results obtained from dual heatbath simulations performed using the Nosé-Hoover-Chain thermostat. The Nosé-Hoover-Chain dual heatbath method is described along with the simulation protocol used. The protein or the solvent were held at 80 K, 180 K and 300 K while the temperature of the other component was varied from 80 K to 300 K. The effect of solvent on the protein dynamics is analyzed in detail through comparison of the protein dynamics in the different setups. The diffusion properties of the hydration layer are examined and compared with the dynamical properties of the protein. The solvent is shown to undergo a glass transition at $\sim$220 K which induces the protein dynamical transition.

5.1 PRINCIPLE OF THE NOSÉ-HOOVER-CHAIN METHOD

The equations of motion for the Nosé-Hoover-chain temperature thermostating algorithm are the following:[1]

$$\begin{aligned}
\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} \\
\dot{\mathbf{p}}_i &= \mathbf{F}_i - \frac{p_{\eta_1}}{Q_1}\mathbf{p}_i \\
\dot{\eta}_k &= \frac{p_{\eta_k}}{Q_k} \qquad\qquad k = 1, ...M \\
\dot{p}_{\eta_k} &= G_k - \frac{p_{\eta_{k+1}}}{Q_{\eta_{k+1}}} p_{\eta_k} \qquad k = 1, ...M-1 \\
\dot{p}_{\eta_M} &= G_M
\end{aligned} \tag{5.1}$$

with the thermostat forces G_k given by:

$$\begin{aligned}
G_1 &= \sum_i \frac{\mathbf{p}_i^2}{m_i} - dNkT \\
G_k &= \frac{p_{\eta_{k-1}}^2}{Q_{\eta_{k-1}}} - kT \qquad k = 2, ..., M
\end{aligned} \tag{5.2}$$

η and p_η are the thermostat position and momentum and T is the temperature at which the given part of the system is to be regulated. N is the number of atoms in the system, k Boltzmann's constant and d the number of spatial dimensions. The parameter Q_k, given by $Q_k = dNkT\tau_k^2$, determines the time scale of the thermostat motion via the time scale parameter τ_k, which should be chosen corresponding to a characteristic time scale of the system. M is the number of thermostat in the chain, in the case where $M = 1$, the equations reduce to the Nosé-Hoover equations of equation 2.9. However, for these equations of motion to reproduce the exact canonical distribution of temperature requires $M \geq 2$. For this reason and for reasons of computer efficiency M was set to its minimal value of $M = 2$ in the present work.

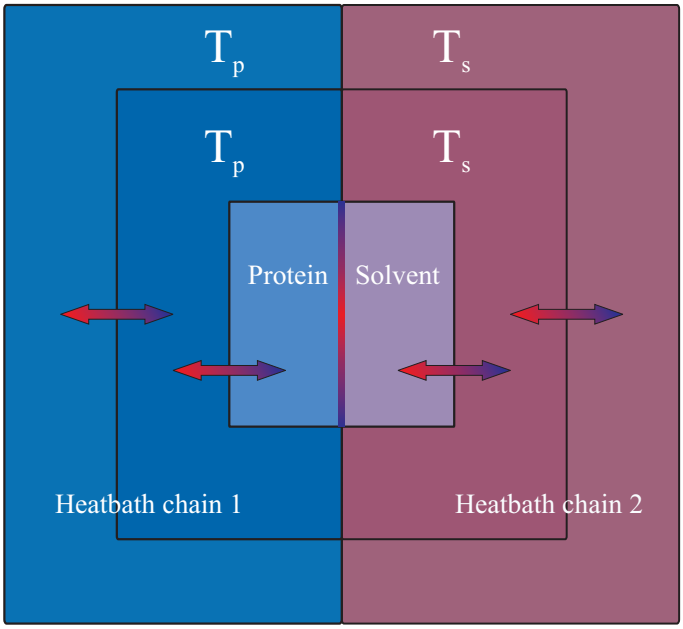


Figure 5.1: Schematic representation of the way the Nosé-Hoover-Chain algorithm functions. Two parts of the system are each connected to a different heatbath chain so as to be regulated at their different temperatures: T_p and T_s.

For the multiple heatbath scheme, each part of the system is regulated by one such set of equations and the total set of equations becomes:

$$\begin{aligned}
\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} \\
\dot{\mathbf{p}}_i &= \mathbf{F}_i - \frac{p_{\eta_1}}{Q_1}\mathbf{p}_i \\
\dot{\eta}_k &= \frac{p_{\eta_k}}{Q_k} \qquad\qquad k = 1, 2 \\
\dot{p}_{\eta_1} &= \sum_i \frac{\mathbf{p}_i^2}{m_i} - dNkT - \frac{p_{\eta_2}}{Q_{\eta_2}} p_{\eta_1} \\
\dot{p}_{\eta_2} &= \frac{p_{\eta_1}^2}{Q_{\eta_1}} - kT
\end{aligned} \tag{5.3}$$

One such a set of equations is set for each part of the system: protein part, solvent part and possibly more parts. Each part can thus be set at a particular temperature $T_{protein}, T_{solvent}$.... The Nosé-Hoover-Chain method functions in that a first heatbath, p_{η_1} regulates the systems while it is itself regulated by p_{η_2}. Figure 5.1 shows a schematic representation of how the Nosé-Hoover-Chain algorithm operates.

The Nosé-Hoover-Chain algorithm with the extensions for multiple heatbath was implemented into the CHARMM package. Extra features improving the stability of the algorithm such as multiple timesteps and Yoshida-Suzuki steps were also implemented. Details of the implementation are presented in Appendix II.

5.2 SIMULATION PROTOCOL

All Nosé-Hoover-Chain simulations were performed with the same protocol. The system was heated to 180 K over 5 ps and in a further phase of heating the protein and solvent were brought to their desired temperatures in 5 K increments every 200 fs and the system was equilibrated for 20 ps. The subsequent production phase was 200 ps long for each simulation and the data from this phase was used for the analysis.

Eight sets of simulations were performed. In six of these the protein or the solvent temperature was held at 80, 180 or 300 K while varying the temperature of the other component from 80 K to 300 K in 20 K steps below 140 K and 10 K steps above 140K. In a seventh set of simulations the protein atoms were fixed and the solvent temperature varied from 80 to 300 K. Finally, a 'control' set of simulations was also performed with the solvent and protein held at the same temperature. Each simulation required 8 hours on 4 processors (800MHz) in a Linux cluster. The 100 simulations performed required 3200 hours of CPU.

NHC has the advantages over the original Nosé-Hoover algorithm[2] that exact canonical behavior is reproduced and the simulations are not prone to unphysical temperature oscillations.[3] The characteristic time for the thermostat motion adopted was $\tau = 0.2$ ps, a value commonly used for condensed phase molecular systems. With the above method the variation of the protein surface temperature was found not to exceed 10 degrees in all simulations except those in which the solvent was held at 300K. For the solvent 300K simulations it was found at low temperatures that there was a temperature gradient leading to significant heating of the protein surface, the most extreme case being a surface heating of 40 degrees in the protein 80 K/solvent 300 K simulation. Although this surface heating was found to not significantly alter the average fluctuation properties examined here, an additional set of solvent 300 K simulations was performed with improved surface temperature properties (see section 5.3.2).

5.3 MULTIPLE HEATBATH RESULTS

A control set of simulations was first performed where protein and solvent were set at the same temperature in each simulation. The results are presented in figure 5.2, a clear transi-

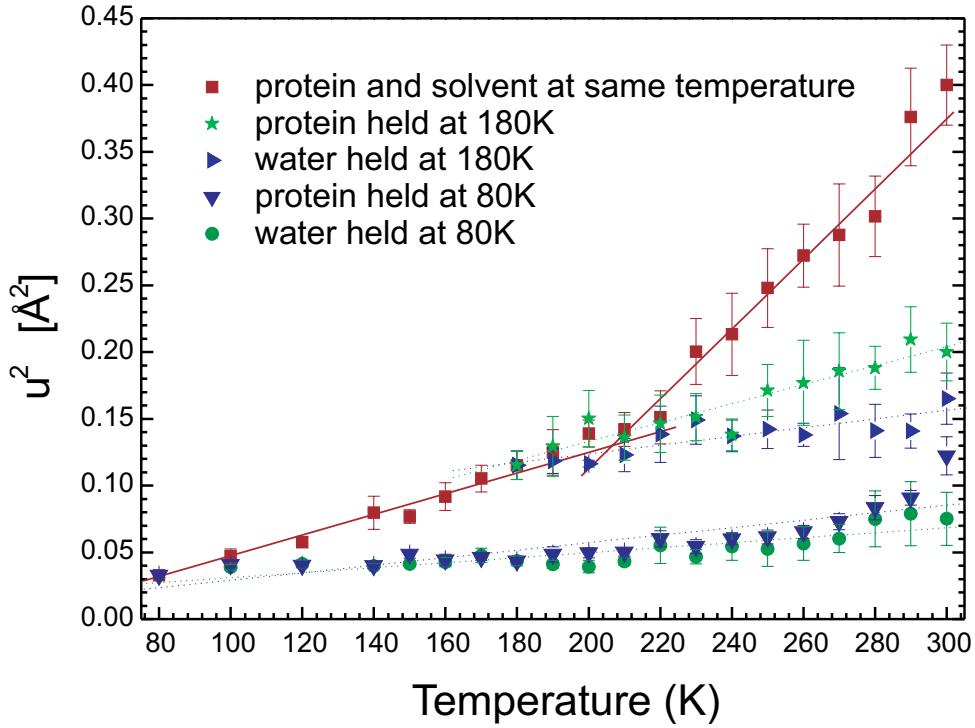


Figure 5.2: Mean-square fluctuations, $\langle u^2 \rangle$, of the protein non-hydrogen atoms for different sets of simulations. ■, control set with protein and solvent at same temperature, ▼, protein held at 80 K; ●, solvent held at 80 K; ★, protein held at 180 K; ►, solvent held at 180 K.

tion in mean-square deviation is visible at ∼220 K. The dynamical transition in myoglobin is well reproduced which then allows for comparison with other sets of simulations.

No instabilities were present in the simulations and the maximum root-mean-square deviation of backbone heavy-atoms with respect to the crystallographic structure was 1.15 $\AA$, indicating that the protein structure remained stable.

5.3.1 SOLVENT CAGING OF PROTEIN DYNAMICS AT LOW TEMPERATURES

In figure 5.2 are presented the protein fluctuations calculated from the control set of simulations together with those obtained by fixing the temperature of one component at a temperature below the dynamical transition while varying the temperature of the other. Fixing the solvent temperature at 80 K or 180 K suppresses the dynamical transition, the protein $\langle u^2 \rangle$ increasing linearly with temperature up to 300 K. Therefore, low temperature solvent is seen to cage the protein dynamics, as has been previously seen.[4]

Figure 5.2 also shows that holding the protein temperature constant at 80 K or 180 K and varying the solvent temperature also abolishes the dynamical transition behavior in

the protein. This contrasts with the previous study by Vitkup *et al* 2000[4] in which it was concluded that when the protein is held at 180 K but the solvent at 300 K then the protein fluctuations are almost identical to those at 300 K.[4] In summary, then, figure 5.2 demonstrates that holding either component at a low temperature suppresses the protein dynamical transition.

5.3.2 ABSENCE OF DYNAMICAL TRANSITION FEATURE IN THE PROTEIN ENERGY LANDSCAPE

Figure 5.3 shows the effect of holding the solvent above the transition temperature, at 300 K, while varying the temperature of the protein. Here due to increased instability in the temperature regulation a third heatbath chain was introduced. Three shells were used: the solvent, the protein surface (protein atoms less than $2.5\mathring{A}$ from any water atom) and the rest of the protein. This led to a temperature stability of the surface residues of 2 degrees. Figure 5.3 presents the results from both sets of simulations.

The fluctuations in the improved set of simulations are reduced at low temperatures compared to the previous set of simulations. Both sets show clearly that holding the solvent temperature at 300 K leads to increased protein fluctuations at most temperatures relative to the other simulation sets. However, neither set show a clear deviation from linearity, *i.e.*, no dynamical transition behavior is present when the solvent is hot.

5.3.3 PRESENCE OF A DYNAMICAL TRANSITION WHEN THE PROTEIN IS HELD HOT

Figure 5.4 shows the results when holding the protein at 300 K and changing the solvent temperature. Fixing the protein temperature at 300 K and varying the solvent temperature recovers the dynamical transition behavior although at a slightly lower temperature than in the control set at $\sim$200 K.

Holding the solvent at 300 K removes the influence of the solvent on the dynamical changes present in the protein as its temperature is changed. If the dynamical transition feature is a intrinsic property of the protein energy landscape, *i.e.*, a property of proteins as *biopolymers,* one would expect it to persist when the solvent is held at a constant temperature of 300 K. However, when the protein is held at 300 K, variations with temperature in the sampled solvent landscape trigger the protein transition. These results indicate the central role played by the solvent in the dynamical transition feature seen in hydrated proteins.

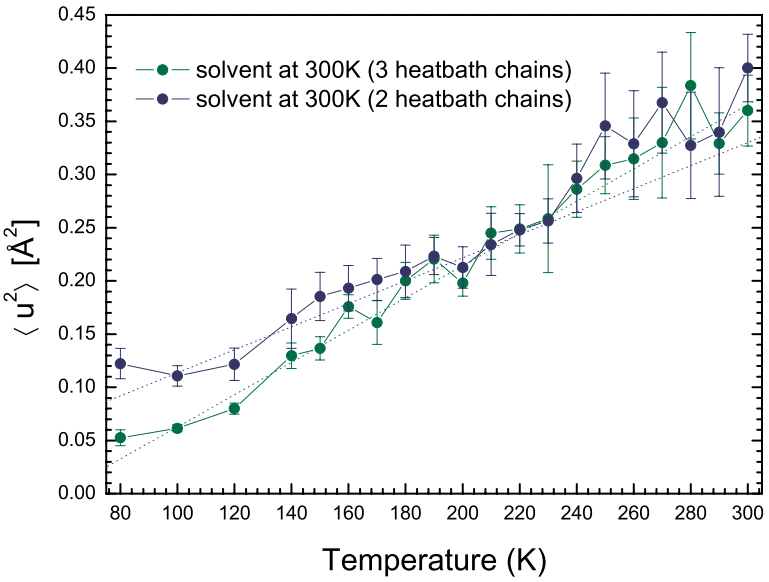


Figure 5.3: Mean-square fluctuations, $\langle u^2 \rangle$, of the protein non-hydrogen atoms with solvent held at 300 K. Two sets of simulation: in blue, the preliminary set with 2 heatbath chains and in green, the improved set with 3 heatbath chains.

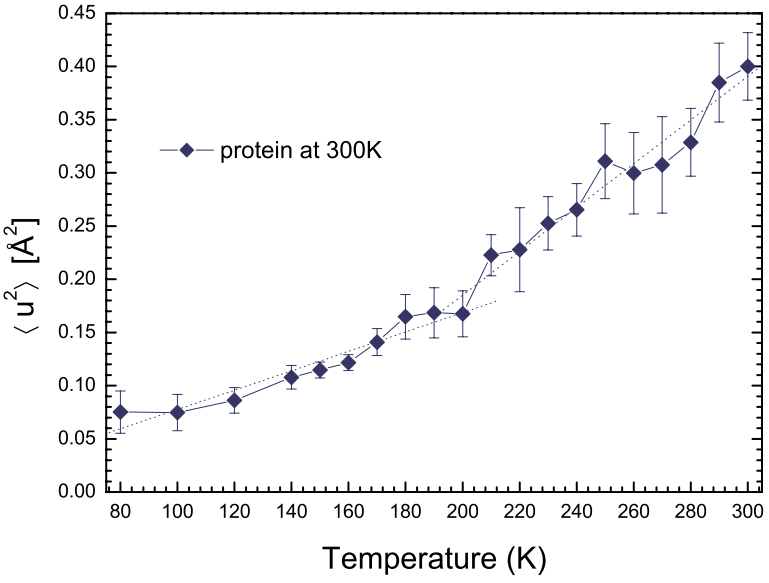


Figure 5.4: Mean-square fluctuations, $\langle u^2 \rangle$, of the protein non-hydrogen atoms with the protein held at 300 K.

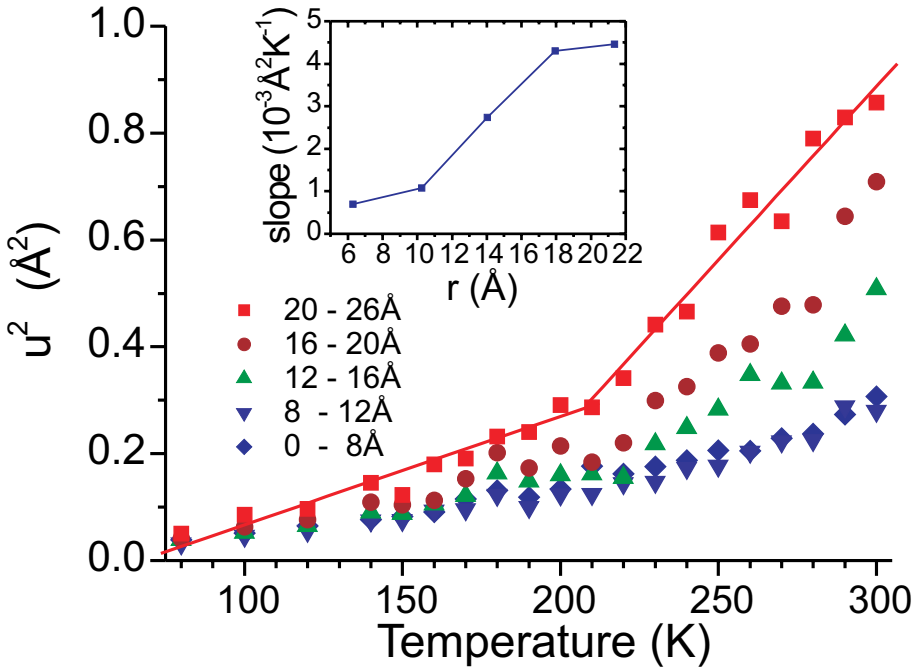


Figure 5.5: Mean-square fluctuations, $\langle u^2 \rangle$, of the protein side-chain heavy atoms for 5 different shells, each 4 $\AA$ thick (except for the inner shell (8 $\AA$) and the outer shell (6 $\AA$)). The inset shows the difference in slope of line fitted below and above 220 K as a function of distance from the protein centre of mass.

5.4 PROTEIN PARTS AFFECTED BY THE DYNAMICAL TRANSITION

The above findings leads to the question of which part of the protein is activated during the dynamical transition. Additional work performed by Jiancong Xu investigated the different parts of the protein to find out which ones are affected by the dynamical transition. The dynamical transition feature of the following different classes of atoms from the control set of simulations was analyzed:

- Hydrophobic *vs* hydrophilic.
- Secondary structure elements, *i.e.* the 8 helices.
- Side-chains *vs* backbone atoms.
- 5 concentric shells centered around the protein center of mass.

Hydrophillic/hydrophobic parts of the protein did not show any difference in the way they underwent the dynamical transition. Similarly no difference was found between the different helical secondary structure elements. The side-chains and the back bone showed

the same dynamical transition feature to be present at ∼220 K, however the fluctuations of the side-chains atoms were found to be ∼2 times greater than those of the backbone atoms.

Distance to the protein core proved to be a crucial factor it the dynamical transition feature. In figure 5.5 are shown the side-chain fluctuations in the control simulations as a function of distance from the protein center of mass. The dynamical transition is seen to be most pronounced in the outer parts of the protein, *i.e.*, those close to the solvent shell - above the transition the outer shells exhibit both stronger fluctuations and a larger change in gradient (inset to figure 5.5) than the inner atoms.

5.5 SOLVENT TRANSLATIONAL DIFFUSION AND ROTATIONAL AUTOCORRELATION TIME

The previous results raise the question of which properties of the solvent are responsible for the protein dynamical transition. To examine this the translational diffusion constant and dipole rotational correlation times were calculated for the water molecules.

Figure 5.6 presents the solvent translational diffusion, D_{trans}, as a function of temperature[†]. A qualitative transition is present in the temperature dependence of the water translational diffusion constant at the dynamical transition. If the diffusion behaves as an activated process, *i.e.*, follows an Arrhenius behavior: $D_{trans} \propto e^{-\alpha/T}$, then straight line behavior in the inset figure 5.6 would be expected. What is seen is two regimes of linear behavior, below and above the transition, involving a lowering of the effective activation energy for water translational diffusion above the transition. This is true even when the protein atoms are fixed, showing that this water transition is inherent to the solvent shell and is independent of the protein dynamics. The linear D_{trans} versus T scale on figure 5.6 makes clear that above the transition the translational diffusion increases rapidly with temperature.

Figure 5.7 compares the excess water translational diffusion constant, *i.e.*, that over and above the effective diffusion constant for harmonic motion with the excess mean-square fluctuation of the protein (again, the excess over the harmonic part). The two quantities vary nearly identically with temperature. Thus, water translational diffusion is

[†] D_{trans} is calculated in the following way:

$$D_{trans} = \lim_{\Delta t \to \infty} \left\langle |\mathbf{r}(\Delta t + t_0) - \mathbf{r}(t_0)|^2 / 6\Delta t \right\rangle_{t_0} \tag{5.4}$$

where $\mathbf{r}(t)$ is the position of a water molecule oxygen atom at time interval Δt after an initial time t_0. For practical reasons Δt was set to 20 ps. D_{trans} was calculated as the mean over 10 time intervals each 20 ps long. The errors were estimated using the standard deviation over the 10 intervals.

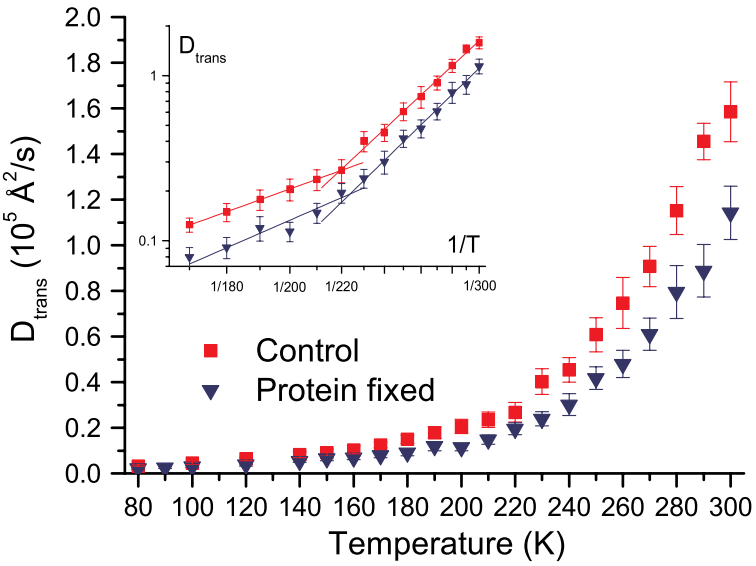


Figure 5.6: Translational diffusion constant, D_{trans}, for different sets of simulations. ■, protein and solvent at the same temperature (control); ▼, protein fixed.*Inset*:the same data plotted as $\log D_{trans}$ versus $1/T$. Straight line fits below and above 220 K are also shown.

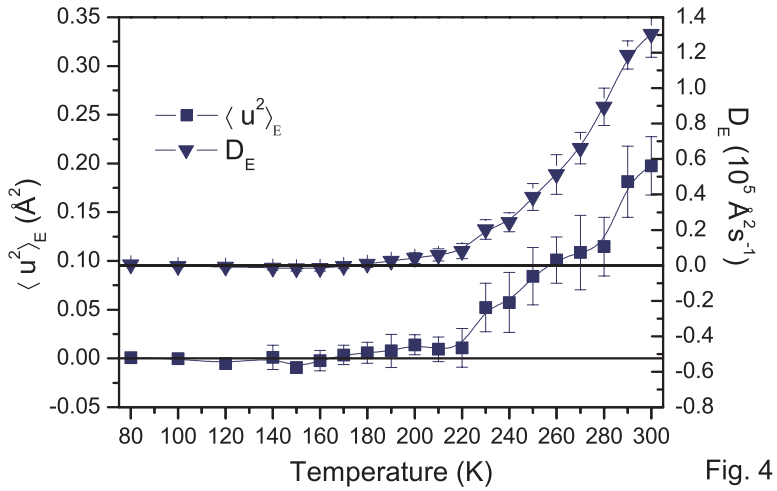


Figure 5.7: Excess mean-square fluctuations, $\langle u^2 \rangle_E$ and excess water translational diffusion constant, D_E versus temperature, $\langle u^2 \rangle_E$ is defined as $\langle u^2 \rangle - \langle u^2 \rangle_H$ where $\langle u^2 \rangle_H$ is the linear part of $\langle u^2 \rangle$ obtained by fitting to the data below 220 K. D_E is calculated from $D_{trans} - D_H$ where D_H is the linear part of D_{trans} obtained by fitting to the data below 220 K. All data calculated from the control simulations (protein and solvent at the same temperature).

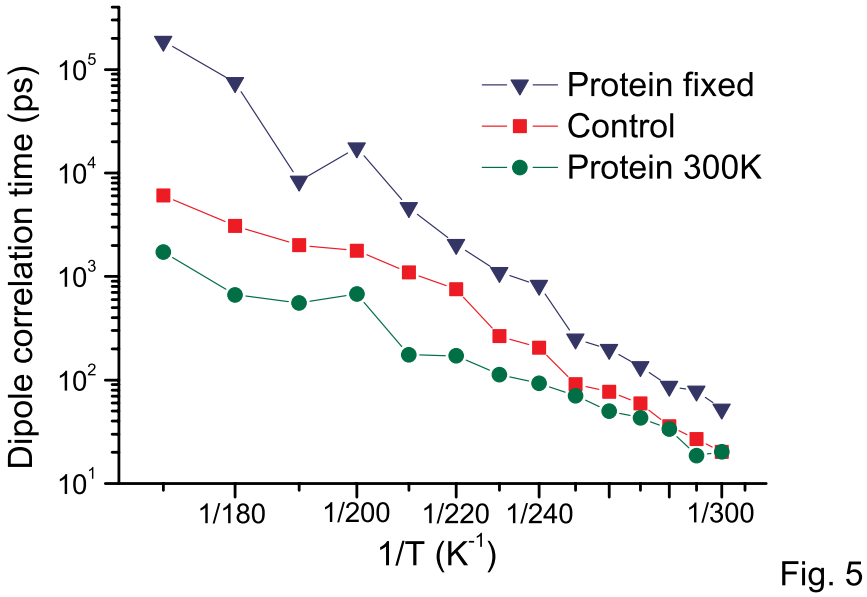


Figure 5.8: Dipole rotational autocorrelation time on a logarithmic scale versus $^1/_T$. ■, protein and solvent at the same temperature (control); ▼, protein fixed; ●, protein held at 300 K.

seen to drive the protein dynamical transition.

Figure 5.8 presents the water dipole rotational autocorrelation time for the control set of simulations as well as for the sets of simulation with the protein atoms held fixed and the protein temperature held at 300 K[‡]. Figure 5.8 shows that the water dipole rotational autocorrelation time undergoes no qualitative change through the dynamical transition - this was also found when holding the protein fixed and holding the protein temperature at 300 K. Thus the water rotational reorientation is completely decoupled from the transition in the translational diffusion seen in this simulation. In the control simulations a small change in slope is seen suggesting a change in the rotational properties. However, this change is much smaller and less sharp than that seen for translational diffusion. This decoupling of translational motion from reorientational properties is indicative of the presence of a glass transition in the solvent.[5]

[‡]The dipole rotational autocorrelation time was calculated by fitting to the correlation function $\langle\cos(\theta(t))\rangle$ where $\cos(\theta(t))$ denotes the scalar product of the corresponding dipole vectors of unit length separated by a time t. The correlation time were calculated by fitting the correlation curves between 1.5 and 20ps with a stretched exponential function.

CONCLUSION

The multiple heatbath technique is seen to be a very valuable simulation technique capable of discerning between dynamics inherent to the protein and dynamics induced by the solvent. Using this technique the cold solvent, *i.e.*, below the dynamical transition temperature, was shown to cage protein dynamics, thereby preventing the dynamical transition feature from appearing. Holding the protein cold was also seen to abolish dynamical transition behavior. On the other hand myoglobin was shown not undergo any dynamical transition in hot solvent demonstrating that the dynamical transition feature is not inherent to the protein energy landscape. The hot protein does undergo a dynamical transition when the temperature of the solvent is changed, showing the central role the solvent plays in the protein dynamical transition.

The dynamical transition was shown to affect only the outer parts of the protein. The surrounding solvent was shown to undergo a transition in its translational diffusion properties which strongly parallels the transition seen in the protein dynamics. The transition in the solvent diffusion is decoupled from the rotational properties, indicating that the solvent undergoes a glass transition at the protein dynamical transition. In conclusion, the solvent transition from a glass to a liquid $\sim$220 K is seen to be responsible for the presence of the dynamical transition measured in protein dynamics at that temperature.

The results presented in this chapter have been published in *Biophysical Journal*.[6] Preliminary results from this chapter were published in the online journal *PhysChemComm*.[7]

REFERENCES

[1] TUCKERMAN, M. E., AND MARTYNA, G. J. Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem.*, 2000, **194**, 159–178.

[2] HOOVER, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 1985, **31**(3), 1695–1697.

[3] HOLIAN, B. L., VOTER, A. F., AND RAVELO, R. Thermostatted molecular dynamics: How to avoid the toda demon hidden in nose-hoover dynamics. *Phys. Rev. E. Stat. Phys.*, 1995, **52**(3), 2338–2347.

[4] VITKUP, D., RINGE, D., PETSKO, G. A., AND KARPLUS, M. Solvent mobility and the protein 'glass' transition. *Nat. Struct. Biol.*, 2000, **7**(1), 34–8.

[5] ANGELL, C. A. Formation of glasses from liquids and biopolymers. *Science*, 1995, **267**, 1924–1935.

[6] TOURNIER, A. L., AND SMITH, J. C. Translational hydration water dynamics drives the protein glass transition. *Biophys. J.*, 2003, **85**(3).

[7] TOURNIER, A. L., XU, J., AND SMITH, J. C. Solvent caging of internal motions in myoglobin at low temperatures. *PhysChemComm*, 2003, **6**(2), 6–8.

CHAPTER 6

PRINCIPAL COMPONENT ANALYSIS

CONTENTS

INTRODUCTION

This chapter presents the results obtained using principal component analysis on 1 ns molecular dynamics trajectories of Mb-CO simulated over a range of temperatures from 80 K to 300 K. The free energy profiles along the principal modes were first calculated and the anharmonicity of these modes then measured using the *Anharmonicity factor* and the *Gaussian fit* scores. The protein mean-square fluctuations were analyzed in terms of the contributions from individual principal modes and their anharmonicity score. Finally the role played by solvent in the protein dynamical transition was further investigated

by looking at the dynamical transition behavior along individual principal modes. The damping exerted by the solvent along the principal modes was also analyzed, confirming the central role of the solvent in the protein dynamical transition.

6.1 PRINCIPAL COMPONENT ANALYSIS

6.1.1 PRINCIPAL COMPONENTS

PCA is a powerful method for analyzing the conformational space explored in an MD trajectory.[1–5] Principal mode analysis has found uses in X-ray refinement,[6, 7] NMR refinement[8] as well as many different types of simulation analysis.[9–12] PCA determines the essential motions present in the simulation: the principal component modes. The set of principal components is the solution to the eigenvalue problem of the second-moment matrix, **A**, of the mass-weighted internal atomic displacements. The elements of the mass-weighted second moment matrix, A_{ij}, are given by:

$$A_{ij} = \sqrt{m_i m_j}\langle(\mathbf{r}_i(t) - \mathbf{r}_i^m)\left(\mathbf{r}_j(t) - \mathbf{r}_j^m\right)\rangle \tag{6.1}$$

where $\mathbf{r}_i(t)$ is the position of atom i and $\mathbf{r}_i^m$ its mean position and the average is taken over the different time frames of the trajectory. The diagonalization of **A** yields the eigenvectors, $\mathbf{w}_k$, *i.e.*, the principal components and their associated eigenvalues, ζ_k. ζ_k, which is referred to as the variance along mode $\mathbf{w}_k$, is the mass-weighted mean-square fluctuations along mode $\mathbf{w}_k$.

The mass-weighted fluctuations, ζ_k, can be used to calculate the averaged mass-weighted atomic fluctuation in the protein in the following way:

$$\zeta_w^{prot} \equiv \frac{1}{N}\sum_{i=1}^{N}\langle m_i\left(\mathbf{r}_i(t) - \mathbf{r}_i^m\right)^2\rangle \tag{6.2}$$

$$= \frac{1}{N}\sum_{i=1}^{3N}\zeta_k \tag{6.3}$$

where m_i is the mass of atom i and N is the number of atoms in the protein and the average is taken over the different time frames of the trajectory. For comparison the standard, non mass-weighted atomic fluctuation or mean square fluctuation is given by:

$$\zeta_{st}^{prot} = \overline{\langle\left(\mathbf{r}_i(t) - \mathbf{r}_i^m\right)^2\rangle} \tag{6.4}$$

where the mean $\bar{\cdot}$ is taken over all the atoms in the protein.

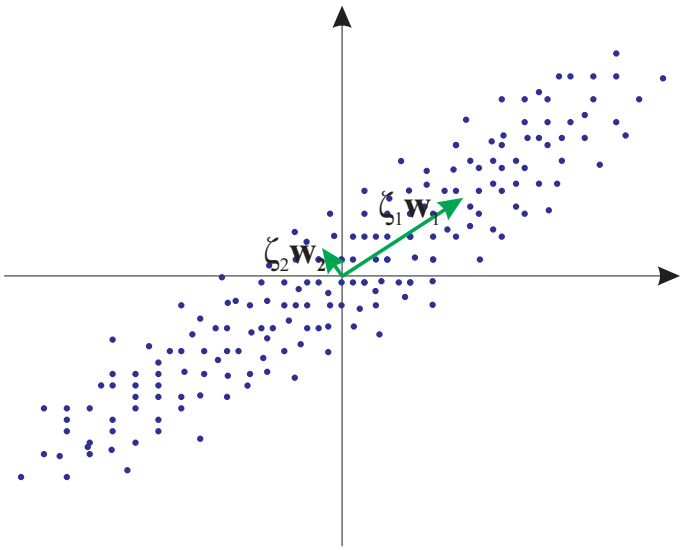


Figure 6.1: Schematic representation of the way Principal Component Analysis determines the most important directions of motion within a data set. Here, the most important axis of motion is along $\mathbf{w}_1$ which therefore has a correspondingly large value for ζ_1. $\mathbf{w}_1$ and $\mathbf{w}_2$ are orthonormal vectors.

The trajectory, $q_k(t)$, of a principal mode k is given by the projection of the MD trajectory onto the k^{th} principal component vector $\mathbf{w}_k$ [†]:

$$q_k(t) = (\mathbf{r}(t) - \mathbf{r}^m) \cdot \mathbf{w}_k \tag{6.5}$$

where here $\mathbf{r}(t)$ is the protein position vector in the 3N conformational space at time t, and $\mathbf{r}^m$ the mean vector.

FREE ENERGY LANDSCAPE

Let $P_k(q)$ be the probability of finding the protein at position q along mode k. The effective free energy along the i^{th} mode, is then defined as:

$$\mu_k(q) = -k_B T \ln P_k(q) \tag{6.6}$$

where k_B is Boltzmann's constant and T is the temperature. If the motions are assumed to be harmonic, using equipartition of thermal energy, the effective angular frequency, ω_k,

[†]Another, more rigorous, expression for $q_k(t)$ takes account of the mass weighting:

$$q_k(t) = \sum_{i=1}^{N} \sqrt{m_i}\,(\mathbf{r}_i(t) - \mathbf{r}_i^m) \cdot \mathbf{w}_{k,i}$$

where $\mathbf{r}_i$ is the position of atom i, $\mathbf{r}_i^m$ the mean position of atom i, $\mathbf{w}_{k,i}$ the coordinate vector corresponding to the motion of atom i along the principal component vector $\mathbf{w}_k$.

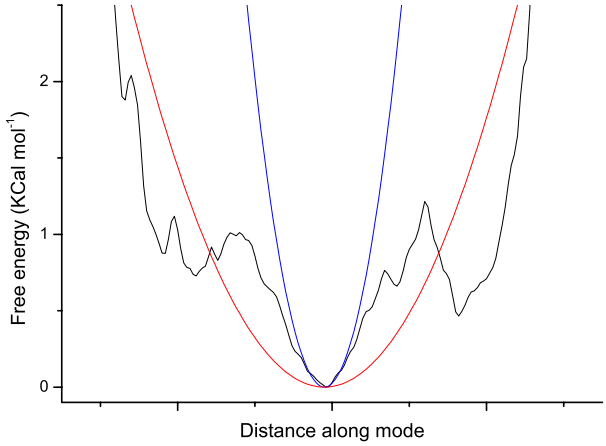


Figure 6.2: Typical free energy landscape obtained from equation 6.6. In red and blue are represented the gaussian approximations obtained through principal component analysis and normal modes respectively.

associated with the motion along a mode is defined as follows:

$$\omega_k^2 = \frac{k_B T}{\zeta_k} \tag{6.7}$$

6.1.2 MODE CLASSIFICATION

Normal modes were used for comparison with the Principal Modes. The normal modes were calculated in the standard way from the minimized conformation of the model system.[13] Normal modes are obtained through diagonalization of the Hessian matrix $\mathcal{H}$, the second order derivative of the potential energy function:

$$\mathcal{H}_{mn} = \frac{\partial^2 V(\mathbf{x})}{\partial x_m \partial x_n} \tag{6.8}$$

where $V(\mathbf{x})$ is the potential energy at position $\mathbf{x}$ and the x_n are the coordinates of the system in conformational space. Normal modes, like principal component modes, give a representation of the main modes of motion present in the system. However, normal modes considers only one minima of the potential energy function. From this, considering the protein motions to be harmonic, it extrapolates the protein conformational space. PCA on the other hand uses a MD trajectory to explore conformational space. The deviation of PC modes from normal modes gives information about the harmonicity of the modes.

ANHARMONICITY FACTOR

The anharmonicity factor, ρ_k, is a measure the deviation of a mode k from harmonicity.[5] To determine this a coefficient projecting the normal mode vector space onto that of the principal components is calculated:

$$c_{kl} = \left(\mathbf{w}_k \cdot \mathbf{w}_l^{NM}\right)^2 \tag{6.9}$$

The harmonic variance along the i^{th} mode due to harmonic motions, ζ_k^H, is then given by:

$$\zeta_k^H = \sum_l c_{kl}\zeta_l^{NM} \tag{6.10}$$

The anharmonicity factor is then defined as:

$$\rho_k = \sqrt{\frac{\zeta_k}{\zeta_k^H}} \tag{6.11}$$

ρ is a number greater than 1. Modes with ρ close to 1 are considered harmonic. Figure 6.2 exemplifies how different the mean-square fluctuations obtained through normal modes and principal component analysis can be.

THE GAUSSIAN FIT

When the energy surface is harmonic, the probability distribution, P_k, along mode k, follows a Gaussian distribution with variance ζ_k:

$$P_k(q) = (2\pi\zeta_k)^{-1/2} e^{-q_k^2/2\zeta_k} \tag{6.12}$$

Deviation from a gaussian distribution is an indication of anharmonicity in the effective potential. The standard error, σ, of a Gaussian fit to the probability distribution therefore gives an alternative measure of the degree of harmonicity of this mode. Figure 6.3 illustrates how big the error between the free harmonic fit (gaussian probability distribution) and the actual probability distribution can be.

The error was defined as $\sigma^2 \equiv 1000 \cdot \overline{(P(q) - G(q))^2}$, taken over 3 standard deviations of the probability distribution, where $G(q)$ is the Gaussian fit. The threshold for this measure is arbitrary as denotes the presence of the factor 1000 in the definition of σ. Upon visual inspection, the definition of σ given above was found to appropriately describe the harmonicity of the modes.

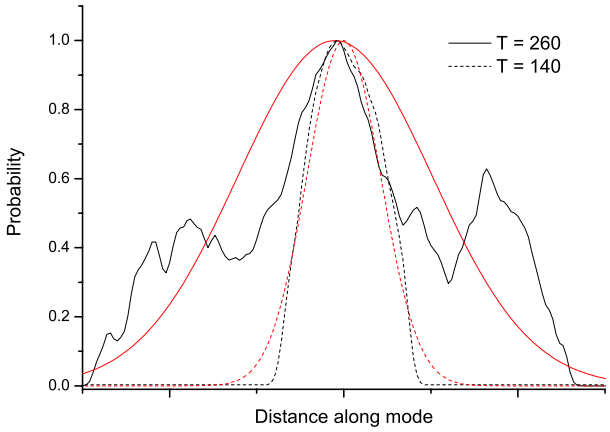


Figure 6.3: Probability distribution along principal mode number 1 at 260 K and 140 K with their corresponding gaussian fits.

MODE CLASSIFICATION

(i) ***Harmonic*** modes are those which exhibit approximately Gaussian probability distributions (defined here as $\sigma < 1$) and for which the anharmonicity is low $(\rho < 2)$.

(ii) ***Quasiharmonic*** modes exhibit approximately Gaussian probability distributions $(\sigma < 1)$ but possess considerable anharmonicity $(\rho > 2)$.

(iii) ***Multiminimum*** modes cannot be well fitted by a Gaussian $(\sigma > 1)$ and are also anharmonic $(\rho > 2)$

6.1.3 DAMPING COEFFICIENTS ALONG PRINCIPAL COMPONENT MODES

Frictional damping gives a measure of the effect of solvent on the motions of the protein. For this, the individual principal component modes can be considered as independent one-dimensional oscillators described by the Langevin equation:[11, 14]

$$\frac{\mathrm{d}v_k}{\mathrm{d}t} + \Gamma v_k + \omega_0^2 q_k = A(t) \tag{6.13}$$

where q_k is the coordinate along the principal component mode k, v_k its corresponding velocity, Γ the damping coefficient, ω_0 the frequency of the undamped oscillator, and $A(t)$ a random force term. From this the velocity time-correlation-function, C_{v_k} for mode k is

given by:

$$C_{v_k}(t) = \frac{\langle v_k(0)v_k(t)\rangle}{\langle v_k(0)v_k(0)\rangle} \quad (6.14)$$

$$= \frac{-\Gamma+\omega}{2\omega}e^{(-\Gamma+\omega)t/2} + \frac{\Gamma+\omega}{2\omega}e^{(-\Gamma-\omega)t/2} \quad (6.15)$$

where:

$$\omega = \sqrt{\Gamma^2 - 4\omega_0^2} \quad (6.16)$$

If $\Gamma > 2\omega_0$, then the motion is overdamped, and if $\Gamma < 2\omega_0$, the motion is underdamped.

The damping coefficient for each mode can thus be calculated using the following equation:

$$\Gamma = -\frac{\mathrm{d}}{\mathrm{d}t}C_{v_i}(t)\Big|_{t=0} \quad (6.17)$$

6.2 PRINCIPAL COMPONENT ANALYSIS PROTOCOL

In the simulations the system underwent an initial relaxation stage as described in chapter 3, followed by a heating phase where the protein and solvent were brought to their desired temperatures in 5 K increments every 200 fs. The system was equilibrated for 20 ps with velocity rescaling every 100 steps and for a further 40 ps without velocity rescaling. The subsequent production phase was 1 ns long for each simulation and the data from this phase was used for the analysis. The model system was simulated at 80 K, 140 K, 180 K, 200 K, 210 K, 220 K, 230 K, 240 K, 260 K and 300 K. The Nose-Hoover-Chains algorithm[15, 16] was used to thermostat the system. The simulations were performed on a Linux cluster (4 processors). Each simulation required 8 hours representing a total of 640 hours of CPU time. The analysis took 14 hours per run on a single processor machine, *i.e.*, a further 140 hours of CPU.

Previous PCA studies have demonstrated the importance of having sufficient simulation time so as to correctly sample the conformational space of the protein.[10, 17] A simulation time of the order of nanoseconds is recommended. The present 1 ns production are not sufficient to explore the protein energy landscape in detail, they are, however, long enough to give a precise indication of the protein motions over the experimental timescales of $\sim$200 ps.

In order to perform the damping coefficient calculations, runs of 100 ps were made, saving velocities every 5 fs so as to obtain precise velocity autocorrelation functions. The fit of autocorrelation data was made over only the first 45 fs of the autocorrelation function. Calculating damping coefficient over such short timescales means that only the damping due to the elasticity of the immediate environment is taken into account. These

damping coefficients cannot be compared to classical friction coefficients, they reflect the degree of energy loss due to dissipative collisional encounters in the system.

6.3 FREE ENERGY LANDSCAPES

The effective free energy landscapes of low frequency principal components was calculated using equation 6.6 for the first few modes at the different temperatures. In what follows the mode numbering ascends with the effective frequency, ω_k , with the lowest mode labelled number 1. Figure 6.4 and 6.5 show the free energy profiles of modes 1, 2 and 5.

For Mode 1 the profile is approximately quadratic below 180 K. The onset of the dynamical transition is characterized by the appearance at 180 K in Mode 1 of double-well behavior with a free-energy barrier of $\sim k_B T$. Above $\sim$240 K the Mode 1 profile is highly rugged. Mode 2 is approximately harmonic for T$\sim$200 K, above which multiple minima again appear. Mode 5 is an example of a mode that is harmonic at low temperatures (here T$<$200 K) and becomes quasiharmonic above.

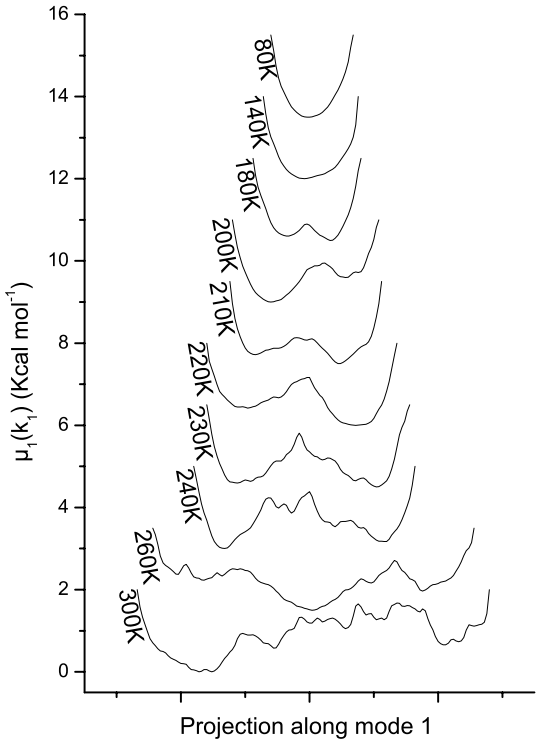


Figure 6.4: Free energy profile of the lowest principal component mode

Graphical examination of Modes 1-5 over the temperature range 180-300 K shows that these principal components all involve collective dynamics distributed over most of

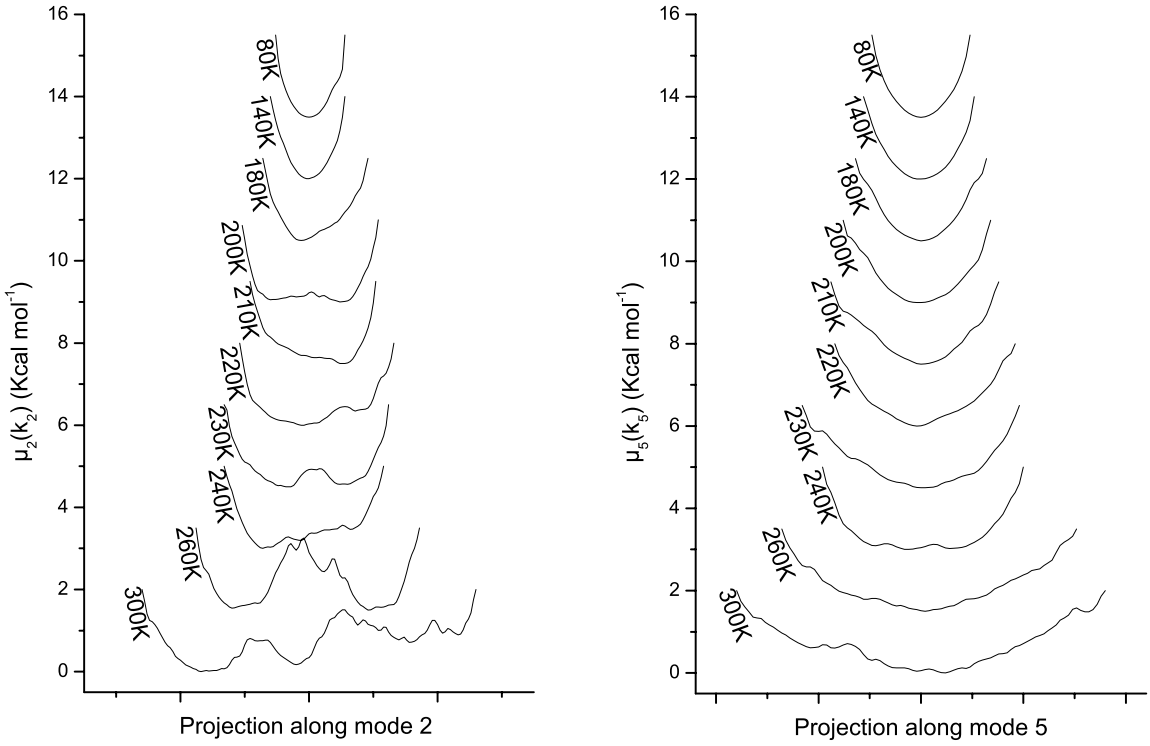


Figure 6.5: Free energy profile of the 2^{nd}, and 5^{th}-lowest principal component modes

the myoglobin molecule. In Mode 1 the transition between the free energy minima at 200 K comprises the relative motion of two large rigid-body blocks of ∼900 atoms. One of these blocks contains helices A, H and F and the other B,C,D and G. The relative motion involves helices G and H rotating in opposite directions around the axis of the stationary helix E. Figures 6.6(a) and (b) illustrate the myoglobin structures in the two free energy minima of Mode 1. Rigid-body helix translations and rotations are apparent. Due to rotations around the helix axes the side-chain displacements are generally larger than those of the main-chain atoms, as illustrated by Asp126 on Helix H in Figure 6.6(a).

6.4 MEASURES OF THE HARMONICITY

6.4.1 ANHARMONICITY FACTOR

The anharmonicity factor, ρ_k, is a measure of the deviation of mode k from harmonicity. Figure 6.7 shows ρ_k calculated for the first 30 modes at the different temperatures investigated. A value higher than 2 indicates a deviation from purely harmonic behavior, such modes are considered non-harmonic. There is no abrupt, qualitative change in the harmonicity of the principal components at the onset of the transition ∼180-210 K.

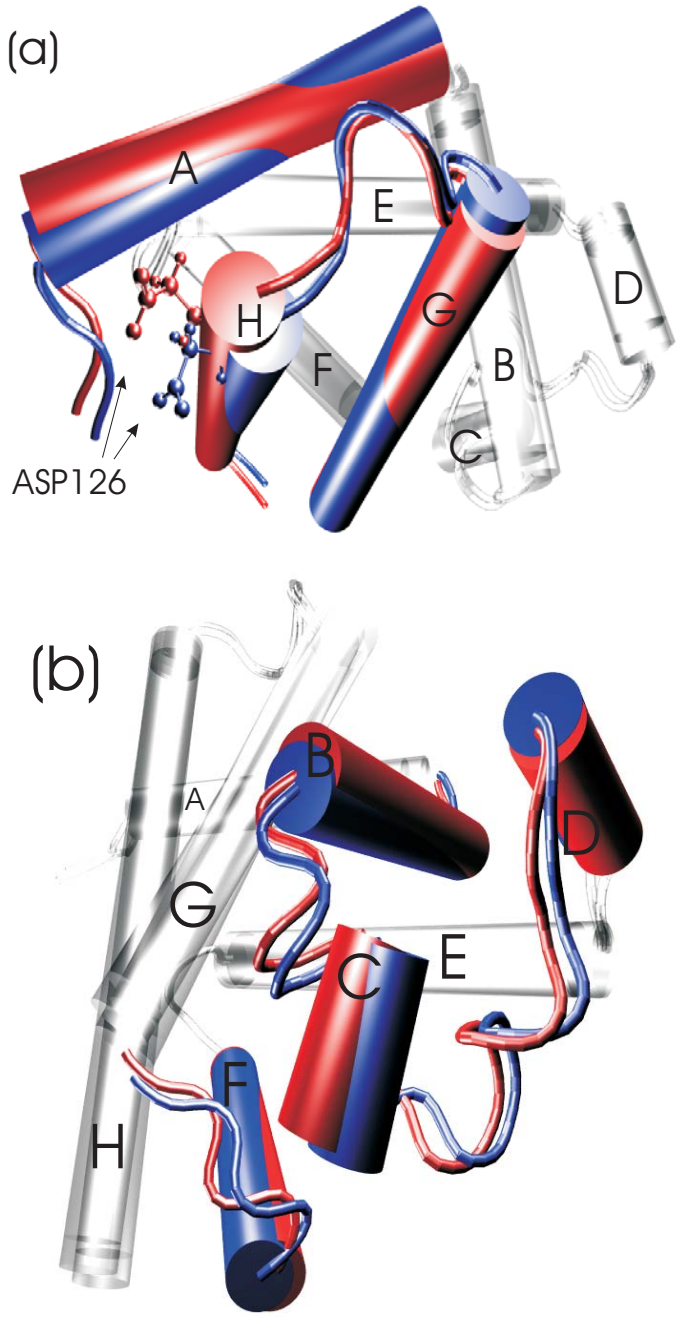


Figure 6.6: Motion of myoglobin along the first principal mode calculated at 200 K. The structure of myoglobin are shown in two minima (red and blue) of the mode. The helices are represented as cylinders. For ease of viewing the amplitude of the motion is multiplied by a factor of 6. Part (a) also shows the sidechain of ASP126.

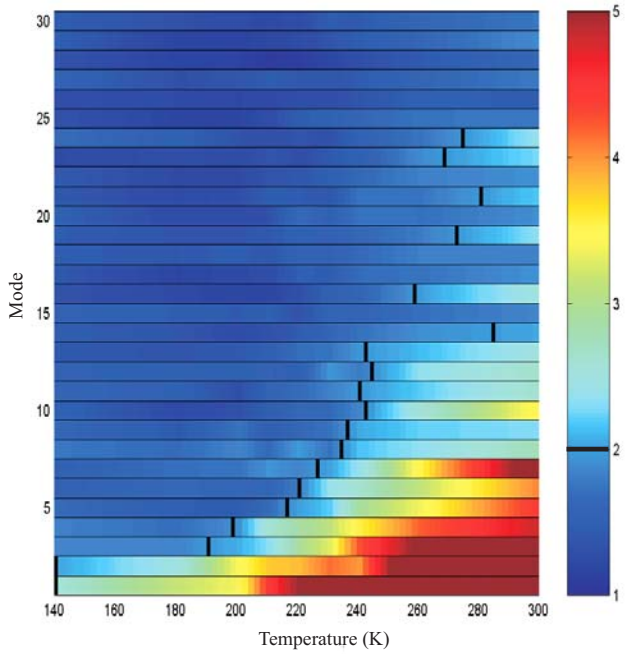


Figure 6.7: Anharmonicity factor, ρ for the first 30 principal component modes. For each principal component the temperature at which $\rho = 2$ is reached is indicated with a black bar.

Rather, the temperature dependence of the anharmonicity is highly mode dependent. As a guide to the eye, for each mode the temperature at which $\rho = 2$ is shown with a bar. The lowest two modes are anharmonic at temperatures as low as 80 K but no other principal components reach $\rho = 2$ below 190 K. At higher temperatures a few other low modes successively become anharmonic. However, at 300 K still only 20 of the 6536 components have reached $\rho = 2$.

6.4.2 GAUSSIAN FIT

The standard error, σ_k, of a Gaussian fit to the probability distribution along a mode, $P_k(q_k)$, is zero for harmonic modes and low for modes with close-to-quadratic effective potentials, *quasiharmonic* modes. σ_k is higher for multiminimum modes with non-harmonic free energy profiles. Figure 6.8 shows σ_k for the lowest 15 modes. The temperature at which $\sigma = 1$ is indicated with a bar. Here again very few principal components become non-Gaussian in the temperature range examined. Significant deviation is seen in Modes 1 and 2 at low temperatures - these two modes reach $\sigma_k = 1$ at 140 K and 170 K, respectively. Only 5 further modes reach $\sigma_k = 1$ by $T = 300$ K, all other modes being close to Gaussian at all temperatures examined.

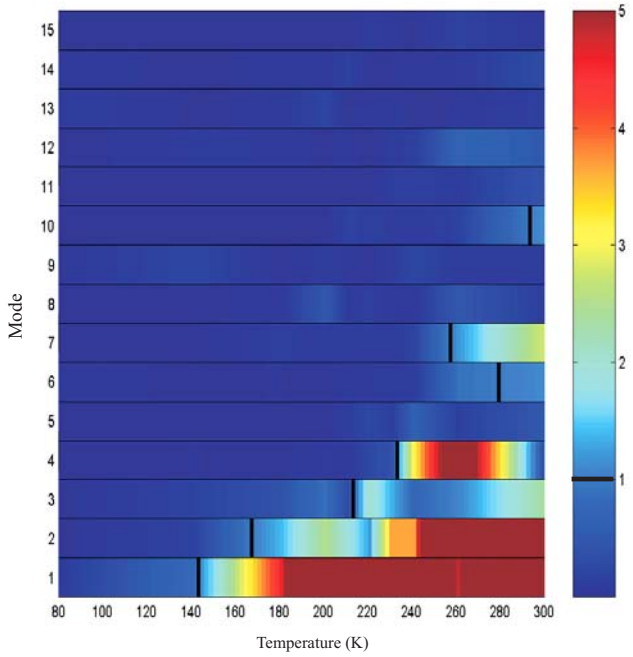


Figure 6.8: Standard deviation, σ of the Gaussian fit to the probability distribution of the first 15 principal component modes. For each principal component mode the temperature at which $\sigma = 1$ is reached, is indicated.

6.5 DECOMPOSITION OF THE TOTAL PROTEIN MSF

6.5.1 INDIVIDUAL MODE CONTRIBUTIONS TO THE TOTAL MSF

The protein mean-square displacement, $\langle u^2 \rangle$ can be decomposed in terms of the contributions made by the principal component modes. This was done in two ways: according to their effective frequency and according to their harmonic classification.

Figure 6.9 shows their contribution to the total $\langle u^2 \rangle$ according to their effective frequency. All contributions show a transition around 220 K, relatively constant below 220 K and increasing steadily above. The first mode has the single greatest contribution to the total $\langle u^2 \rangle$. The first mode accounts for 3.7% of the $\langle u^2 \rangle$ at 80 K increasing up to 14.5% of the total MSF at 300 K. The first 5 modes account for 11% of $\langle u^2 \rangle$ at 80 K increasing up to 40% at 300 K. The first 50 modes (only 0.6% of the total number of modes) account for 44% of $\langle u^2 \rangle$ at 80 K increasing up to 74% of at 300 K. In contrast the contribution from the remaining 99.4% of the modes is 41.2% at 80 K, dropping down to only 17.8% at 300 K. Thus 0.6% of the modes (50 modes) control most of the protein $\langle u^2 \rangle$ temperature dependence.

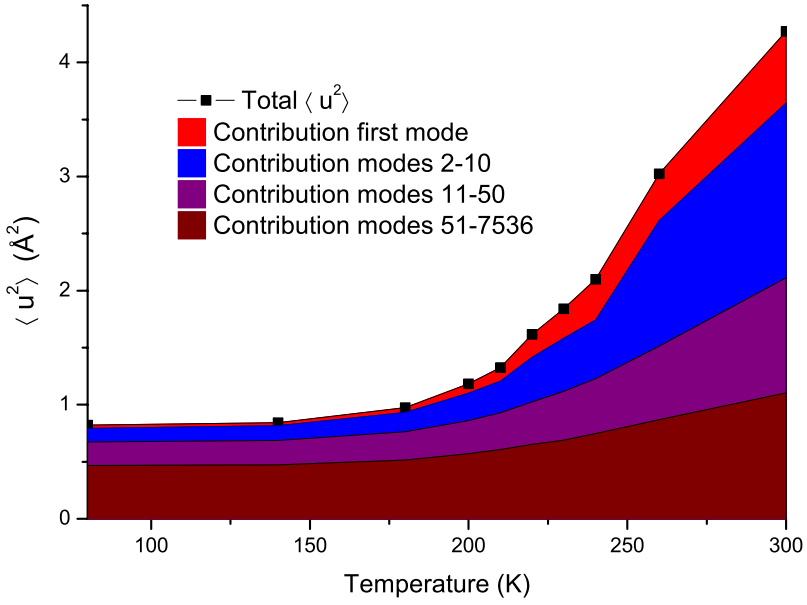


Figure 6.9: Decomposition of the total protein $\langle u^2 \rangle$ in terms of the principal component modes. The contribution of the 1^{st} mode, first 10 modes, and of the first 50 modes are indicated.

6.5.2 HARMONIC, QUASI-HARMONIC AND MULTIMINIMA MODES CONTRIBUTIONS

The contributions made to the mean-square displacement, $\langle u^2 \rangle$ by the three types of principal components (Harmonic, Quasi-harmonic, or Multiminima) are shown in Figure 6.10. The increase in $\langle u^2 \rangle$, signalling the incipient phase of the transition ~180-210 K, is seen to arise from a very small number of principal components. By 210 K, only four PCA modes are not harmonic - three of these are multiminimum and one is quasi-harmonic. At 210 K, 75% of the increase over the linear $\langle u^2 \rangle$ (the "excess" $\langle u^2 \rangle$) is due to multiminimum dynamics. As the temperature increases more multiminimum and quasiharmonic components appear. However, at 300 K still only 20 modes deviate from harmonic behavior, *i.e.*, only 0.3% of the total number of modes in the protein. 70% of the excess $\langle u^2 \rangle$ at 300 K arises from 7 multiminimum modes, with the remaining 30% originating from 13 quasiharmonic principal components. Figure 6.10 also shows that the normal mode analysis and the *harmonic* principal modes concur in that their contribution to $\langle u^2 \rangle$ is ~1.6 $\mathring{A}$ at 300 K.

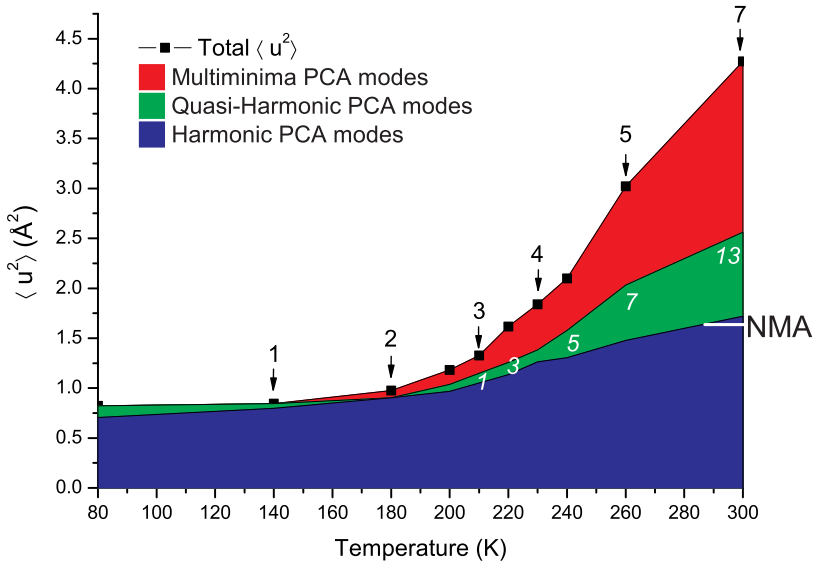


Figure 6.10: Decomposition of the mass-weighted protein, $\langle u^2 \rangle$ into contributions arising from the harmonic, quasi-harmonic and multiminimum classes of principal component. Vertical arrows indicate number of multiminimum modes and italics the number of quasiharmonic modes at certain temperatures. The $\langle u^2 \rangle$ calculated at 300 K from the normal mode analysis using the same model system and potential function is indicated.

6.6 DYNAMICAL TRANSITION IN THE MSF ALONG INDIVIDUAL MODES

The way each mode undergoes the dynamical transition, *i.e.*, the transition in $\langle u^2 \rangle$ along each mode was also investigated. Figure 6.11 shows the individual normalized contributions $\langle u^2 \rangle_N$ of the first 4000 modes to the total $\langle u^2 \rangle$[†]. The normalization procedure makes it possible to follow the dynamical transition in all the modes although the amplitude of the $\langle u^2 \rangle$ along the different modes varies greatly in amplitude. A transition is seen to occur in all modes around ∼220 K independently of the mode numbers. This result seems surprising at first since the high frequency modes are harmonic in nature and therefore should show a linear increase in MSF. Higher frequency modes describe simple dynamics such as hydrogens bond stretching motions for example. However PCA does

[†]The normalized mean-square fluctuation, $\langle u^2 \rangle_N$, is defined as:

$$\langle u^2 \rangle_N(T) = \frac{\langle u^2 \rangle(T) - \langle u^2 \rangle_{80K}}{\langle u^2 \rangle_{300K}}$$

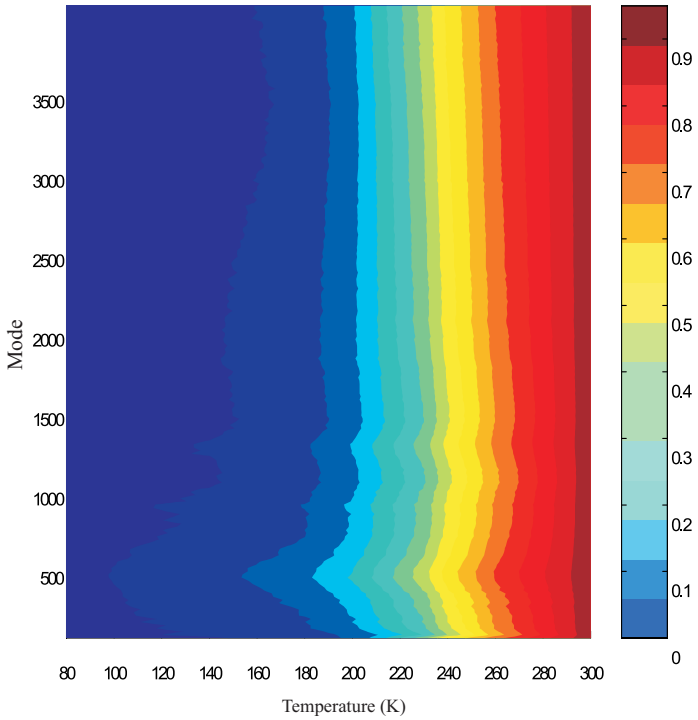


Figure 6.11: Normalized mean-square deviation, $\langle u^2 \rangle_N$ along the lowest 4000 principal modes.The $\langle u^2 \rangle$ values at 80 K are set to 0 and the $\langle u^2 \rangle$ is normalized to 300 K so as to show the transition in each mode.

not see the dynamics of the hydrogens in isolation but as attached to the rest of the protein which undergoes the dynamical transition. This induces a dynamical transition feature in the $\langle u^2 \rangle$ of the hydrogen modes of motion since the rest of the protein to which it is attached shows increased motion due to the dynamical transition. For the low frequency modes one might expect the mode to have different transition temperature reflecting their marked differences in ρ and σ scores. This however is seen not to be the case and indicates that the dynamical transition feature is decoupled from the harmonicity of the modes.

6.7 CHANGE IN DAMPING COEFFICIENTS

In the previous chapters it was shown that the surrounding hydration layer plays a crucial role in the dynamical transition feature. To further investigate the effect of solvent on the protein dynamics, the damping experienced by the protein in the presence and absence of solvent was calculated.

Figure 6.12 shows the damping coefficient calculated at each temperature for the solvated and dry protein. As would be expected, the presence of solvent is seen to increase

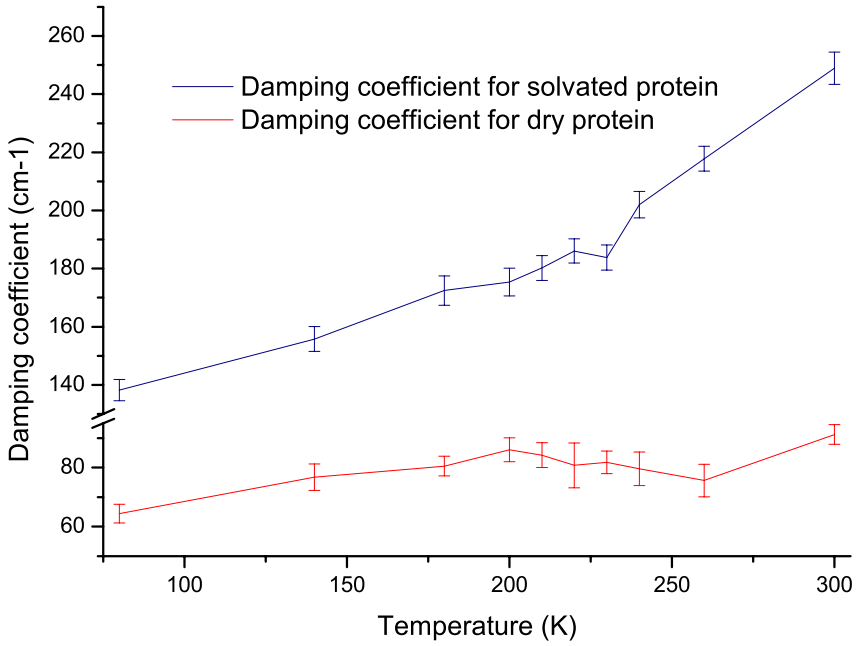


Figure 6.12: Damping coefficients for dry protein and hydrated myoglobin over the range 80 K to 300 K. The damping coefficient was defined as the average of the damping along the first 50 principal component modes. For details of calculations see section 6.1

the damping experienced by the protein. A transition around 220 K appears in the damping coefficient of the hydrated protein whereas no such transition is to be seen in the data for the dry protein. The fact that the damping experienced by the hydrated protein actually increases with temperature is explained by the fact that in ice (at low temperatures) the experienced damping is less (motions being more harmonic) than that experienced at higher temperatures were the solvent is more viscous. In effect the damping experienced by the protein due to the solvent undergoes a transition $\sim$220 K thus showing in a different way how essential the solvent is to the dynamical transition feature.

CONCLUSION

In this chapter the harmonicity of the motions involved in the dynamical transition were investigated using principal component analysis. The different modes of motion are seen to become anharmonic at different temperatures in a manner uncorrelated with the dynamical transition feature. The free energy profile along the first few modes is seen to be harmonic or quasiharmonic at low temperature. With increasing temperature anhar-

monicities appear and modes with multiple minimas appear. The protein motion along the lowest modes involves simple motions such as rotations and translations of different part of the protein. The whole protein motion being dominated by as few as 20 modes (0.3% of the total number of modes). The increase in the protein $\langle u^2 \rangle$ at the dynamical transition temperature correlates with the anharmonic (quasiharmonic + multimima) contributions to the $\langle u^2 \rangle$. The presence of the solvent is shown to induce a transition in the damping experienced by the protein at the dynamical transition temperature $\sim$220 K which confirms the central role played by the solvent in the protein dynamical transition.

REFERENCES

[1] KARPLUS, M., AND KUSHICK, J. N. Methos for estimating the configurational entropy of macromolecules. *Macromolecules*, 1981, **14**, 325–332.

[2] AMADEI, A., LINSSEN, A. B., AND BERENDSEN, H. J. Essential dynamics of proteins. *Proteins*, 1993, **17**(4), 412–25.

[3] LEVY, R. M., SRINIVASAN, A. R., OLSON, W. K., AND MCCAMMON, J. A. Quasi-harmonic method for studying very low frequency modes in proteins. *Biopolymers*, 1984, **23**(6), 1099–112.

[4] BROOKS, B. R., JANEZIC, D., AND KARPLUS, M. Harmonic analysis of large systems. i methodology. *J. Comput. Chem.*, 1995, **16**(12), 1522–1542.

[5] KITAO, A., HAYWARD, S., AND GO, N. Energy landscape of a native protein: jumping-among-minima model. *Proteins*, 1998, **33**(4), 496–517.

[6] KIDERA, A., AND GO, N. Refinement of protein dynamic structure: normal mode refinement. *Proc Natl Acad Sci U S A*, 1990, **87**(10), 3718–22.

[7] MIZUGUCHI, K., KIDERA, A., AND GO, N. Collective motions in proteins investigated by x-ray diffuse scattering. *Proteins*, 1994, **18**(1), 34–48.

[8] ABSEHER, R., HORSTINK, L., HILBERS, C. W., AND NILGES, M. Essential spaces defined by nmr structure ensembles and molecular dynamics simulation show significant overlap. *Proteins*, 1998, **31**(4), 370–82.

[9] HAYWARD, S., KITAO, A., HIRATA, F., AND GO, N. Effect of solvent on collective motions in globular protein. *J Mol Biol*, 1993, **234**(4), 1207–17.

[10] DE GROOT, B. L., VAN AALTEN, D. M., AMADEI, A., AND BERENDSEN, H. J. The consistency of large concerted motions in proteins in molecular dynamics simulations. *Biophys J*, 1996, **71**(4), 1707–13.

[11] HAYWARD, S., KITAO, A., AND BERENDSEN, H. J. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins*, 1997, **27**(3), 425–37.

[12] CAVES, L. S., EVANSECK, J. D., AND KARPLUS, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci*, 1998, **7**(3), 649–66.

[13] BROOKS, B., AND KARPLUS, M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc Natl Acad Sci U S A*, 1985, **82**(15), 4995–9.

[14] SMITH, J., KUCZERA, K., AND KARPLUS, M. Dynamics of myoglobin: comparison of simulation results with neutron scattering spectra. *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**(4), 1601–5.

[15] MARTYNA, G. J., TUCKERMAN, M. E., TOBIAS, D. J., AND KLEIN, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.*, 1996, **87**(5), 1117–1157.

[16] TUCKERMAN, M. E., AND MARTYNA, G. J. Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem.*, 2000, **194**, 159–178.

[17] HESS, B. Convergence of sampling in protein simulations. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2002, **65**(3 Pt 1), 031910.

CHAPTER 7

RESULTS FROM NEUTRON SCATTERING EXPERIMENTS

CONTENTS

INTRODUCTION

This chapter presents the results obtained from the neutron scattering experiments performed on the IN6 spectrometer at the Institut-Laue-Langevin in Grenoble. Neutron scattering was measured on solutions of Xylanase with 0%, 3.5%, 7%, 15%, 20%, 25% and 40% methanol:water volume ratios (see methods section 7.1 for details). Previous results obtained by V. Réat on the IN16 spectrometer are also presented. In that experiment the neutron scattering measurement was performed on Xylanase solutions with 0%, 15%, 40% and 70% methanol/water concentration (v/v). The integrated elastic peak scattering intensities and protein mean-square fluctuations were determined and compared (cf: theory in section 2.3). The results are also compared with the properties of methanol/water solution properties.

7.1 NEUTRON SCATTERING EXPERIMENTAL PROTOCOL

7.1.1 EXPERIMENTAL APPARATUS

Neutrons can be obtained either from a spallation source or from a nuclear reactor. A spallation source consists of particles coming out of an accelerator hitting a heavy metal target (Tantalum for example) thereby extracting neutrons from the heavy nuclei. Another method for obtaining neutrons is from a nuclear reactor which emits neutron as a byproduct of the fission reaction. The present experiments where performed at the nuclear reactor facilities of the Institut Laue Langevin (ILL) in Grenoble.

In the case of the IN6 spectrometer, the neutrons are slowed down by passing through Hydrogen atoms (water for example). As can be seen in figure 7.1 a series of graphite mono-chromater are used to select the neutrons of the correct energy. These are then passed through a beryllium filter to make sure only the right (low energy) neutron are selected, finally a Fermi chopper perform the time-focusing so that the elastically scattered neutrons all arrive at the same time at the spectrometer. IN6 is a time-of-flight instrument, *i.e.* the excess / deficit in the time taken by a neutron to travel the distance from the sample to the detector bank is used to calculate the energy of the scattered neutron.

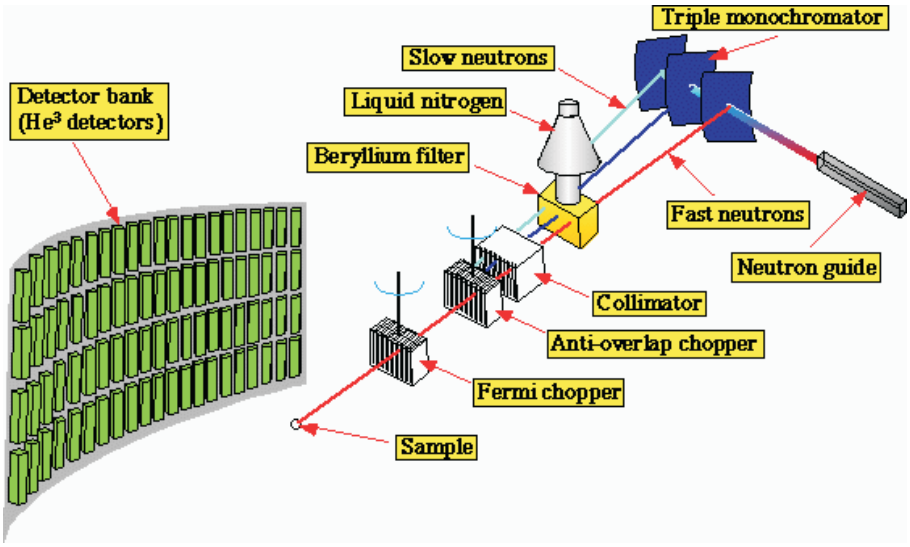


Figure 7.1: The IN6 detector at the Laue-Langevin Institut in Grenoble.

Réat *et al* previously performed experiments using the same system on the IN16 instrument. The IN16 instrument is a back-scattering spectrometer with a better energy resolution than IN6, and is therefore capable of resolving motions on the nanosecond timescale whereas IN6 is limited to timescales of $\sim$100 ps. A schematic representation of the IN16 spectrometer is presented in figure 7.2. In this configuration a moving monochromator reflects highly monochromated neutrons in such a way that the energy of the incoming neutrons on the sample will fluctuate with time due

to doppler shifting. These neutrons are then scattered by the sample and reach the analyzer bank. The analyzers (Si (111) crystals) will only reflect those neutron with a very well defined energy, so that only the neutrons that have lost or gained the required amount of energy in the sample will be backscattered onto the detectors. Knowledge of the movement of the monochromator then enables to determine the scattering intensity around the elastic peak with high precision.

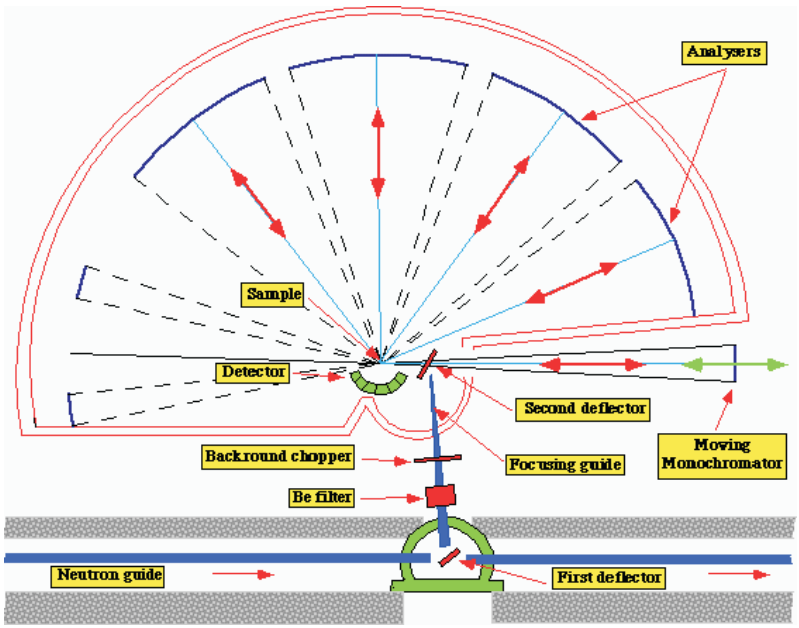


Figure 7.2: The IN16 detector at the Laue-Langevin Institut in Grenoble.

7.1.2 EXPERIMENTAL SYSTEM: XYLANASE

The present experiment were performed on a thermophilic xylanase enzyme. Xylanase is found in many microorganisms, it is involved in the hydrolysis of xylan polymers. Xylans are highly branched polysaccharides usually found in tight association with other biopolymers. Xylans belong to the major constituents of plant cell walls. As the most abundant hemicellulose, they account for more than 30% of the dry weight of terrestrial plants.

Xylanase is composed of a simple subunit with a central β-barrel motif (with 8 beta sheets) surrounded by 15 stabilizing α-helices. Figure 7.3 gives representation of xylanase in a 'cartoon' representation.

The stability of structure of xylanase has been studied at various temperatures and methanol concentrations. The structure of xylanase has been shown to be stable over the range of temperatures (100 K to 300 K) and methanol concentrations (0% to 70%) used in this experiment over the time of the experiment ($\sim$8 hours per sample).[1]

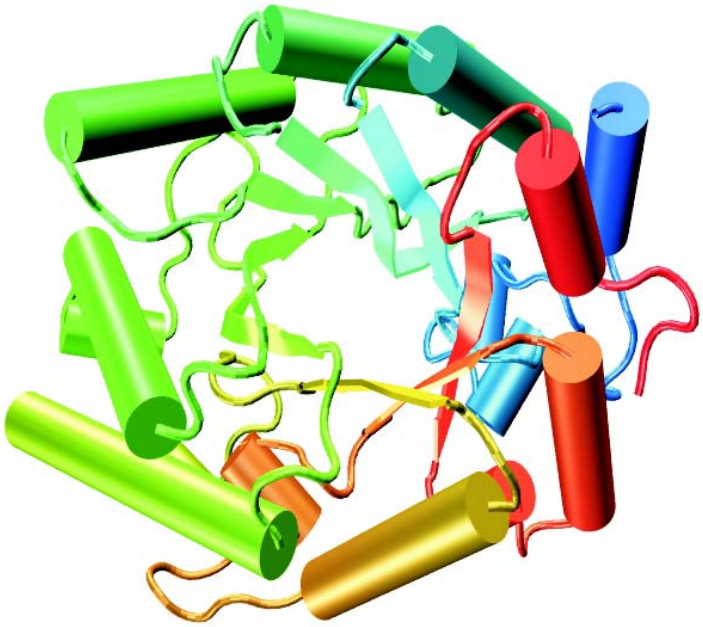


Figure 7.3: The protein xylanase in a 'cartoon' representation.

7.1.3 EXPERIMENTAL PROTOCOL

Dynamic neutron scattering measurements were performed on the IN6 time-of-flight spectrometer at the Insitut-Laue-Lagevin in Grenoble. The incident wavelength used was 5.12 Å. The energy resolution of the apparatus was $50\ \mu eV$. All data were collected with the sample holder oriented 135 relative to the incident beam. The samples were contained in aluminium flat-plat cells of 0.5 mm thickness.

Neutron scattering measurements were performed on the enzyme xylanase at different CD_3OD/D_2O concentrations. Neutron scattering intensities were recorded on the following samples (with their measured transmissions in parentheses and the data gathering time used for each temperature):

1. 70 mg/ml Xylanase in 3.5% CD_3OD/ 96.5% D_2O (0.91, 21min)
2. 72 mg/ml Xylanase in 7% CD_3OD/ 93% D_2O (0.88, 20min)
3. 68 mg/ml Xylanase in 15% CD_3OD/ 85% D_2O (0.91, 13min)
4. 71 mg/ml Xylanase in 20% CD_3OD/ 80% D_2O (0.89, 24min)
5. 76 mg/ml Xylanase in 25% CD_3OD/ 75% D_2O (0.92, 22min)
6. 66 mg/ml Xylanase in 40% CD_3OD/ 60% D_2O (0.90, 23min)
7. 63 mg/ml in D_2O (0.91, 15min)

Data was gathered over the temperature range from 100 K to 300 K for each methanol concentration. The samples were first cooled to ∼100 K and then slowly reheated back to 300 K over a period ∼6 hours, the data being gathered during the whole reheating process.

Data had previously been gathered on the IN16 spectrometer by V. Réat *et al*. Neutron scattering measurements were performed on the enzyme Xylanase at different CD_3OD/D_2O concentrations. The following protein samples were run:

1. Xylanase in 15% CD_3OD/ 85% D_2O
2. Xylanase in 27% CD_3OD/ 73% D_2O
3. Xylanase in 40% CD_3OD/ 60% D_2O
4. Xylanase in 70% CD_3OD/ 30% D_2O

The data gathering on IN16 used a similar 'temperature ramping' data collection technique as in IN6.

7.1.4 DATA ANALYSIS

The scattering function $S(q, \omega)$ was extracted from the IN6 experimental data using the INX data reduction software. Detectors were grouped 4 at a time (*Angle grouping*). The detectors were calibrated by normalizing with respect to a standard vanadium sample. INX also performed a number of corrections to take account of: sample holder scattering, sample size and geometry, detector efficiency, relative efficiency of the different detectors, and also removed the background intensity.

The incoherent scattering originating from hydrogen scattering dominates the scattering signal. Because hydrogens are evenly distributed in a protein, the protein scattering gives a global view of the protein motions. In order to maximize the contribution from the protein motions, the present experiments were performed using fully deuterated solvents and hydrogen/deuterium exchanged proteins (twice dissolved in D_2O and freeze-dried). This deuteration procedure serves to ensure that the solvent is fully deuterated (and not partially hydrogenated by exchange of hydrogen from the enzyme) and that no change in the deuteration of the enzyme or solvent occurs during the experiment. The consequent partial deuteration of the enzyme is limited to the exchangeable hydrogens, which means that the hydrogens left are essentially all in the protein.

Having obtained the scattering function, $S(q, \omega)$, the elastic peak intensity, $S(q, \omega = 0)$, was obtained by fitting the data around the peak with a gaussian. For this 5 energy channels were used: from -0.035 meV to 0.035 meV, which covers most of the elastic intensity peak. $S(q, \omega = 0)$ was then normalized to 115 K to remove coherent contributions (a linear interpolation was used to get data at 115 K).

The integrated elastic peak, $S_{INT}(T)$, intensities were determined by summing $S(q, \omega = 0, T)$ from $q = 0.35$ Å to $q = 1.05$ Å. $S_{INT}(T)$ provides a qualitative guide of dynamic transition behavior, with good counting statistical accuracy. The improved statistics

are obtained by integrating over a range of q chosen to lie in the dynamic scattering region ($0.35 < q < 1.05$ Å^{-1}).

The $\langle u^2 \rangle$ were then determined directly from the integrated elastic peak intensities using the approximation outlined in section 2.3.3:

$$\langle u^2 \rangle = 1 - \frac{6}{b^2 \sum_i q_i^2} \sum_i S_{inc}(q_i^2) \tag{7.1}$$

In the present work the $\langle u^2 \rangle$ thus determined is equal to $\langle u^2 \rangle_T - \langle u^2 \rangle_{115K}$, where $\langle u^2 \rangle_{115K}$ is the absolute mean-square displacement at 115 K.

The data analysis for IN6 followed as much as possible the methods used for the IN16 data analysis.

7.2 RESULTS AND DISCUSSION

7.2.1 INTEGRATED ELASTIC PEAK INTENSITIES

The integrated elastic peak intensities, $S_{INT}(T)$, for the different samples on IN6 are presented in figure 7.4. The pure D_2O data shows a slow decrease with temperature down to 275 K, with no marked transition but a gradual increase in curvature toward 275 K. Between 275 K and 280 K there is a sharp transition occurring at the solvent melting point (at 277 K in D_2O). This strongly nonlinear increase of $\langle u^2 \rangle$ with T is consistent with the activation of anharmonic dynamics on the $\sim$100 ps timescale of the experiment.

The general form of $S_{INT}(T)$ for xylanase in the different methanol/water solutions is similar to that of xylanase in pure D_2O. The decrease with temperature from 120 K to $\sim$240 K is larger than in pure D_2O, indicating that the presence of methanol - even in small concentrations - allows more motion than does pure D_2O. The onset of the transition moves to lower temperatures with increasing methanol concentrations: 275K at 0% CD_3OD; 270 K at 3.5% CD_3OD; 267 K at 7% CD_3OD; 270 K at 15% CD_3OD; 260 K at 20% CD_3OD; 255 K at 25% CD_3OD; 241 K at 40% CD_3OD(cf: figure 7.11).

The transition also becomes smoother with increasing methanol concentration. The 0% CD_3OD plot presents a very sharp transition between two states: below 270K and above 280 K. Whereas this feature is still visible in the 3.5% CD_3OD data set, it disappears at higher concentrations. This is consistent with a gradual release of the motion in the presence of methanol qualitatively different from the sharp melting of D_2O in the 0% CD_3OD case.

Figure 7.4 also shows that the maximum motion at high temperature reduces with increasing concentrations of methanol. This is consistent with the decrease in protein diffusion motions following the increase in viscosity in the solvent due to the presence of methanol.[2] Another possibility is that the solvent contributes largely to the elastic peak intensity at high methanol concentrations. The role of the background scattering due to presence of the solvent is discussed in

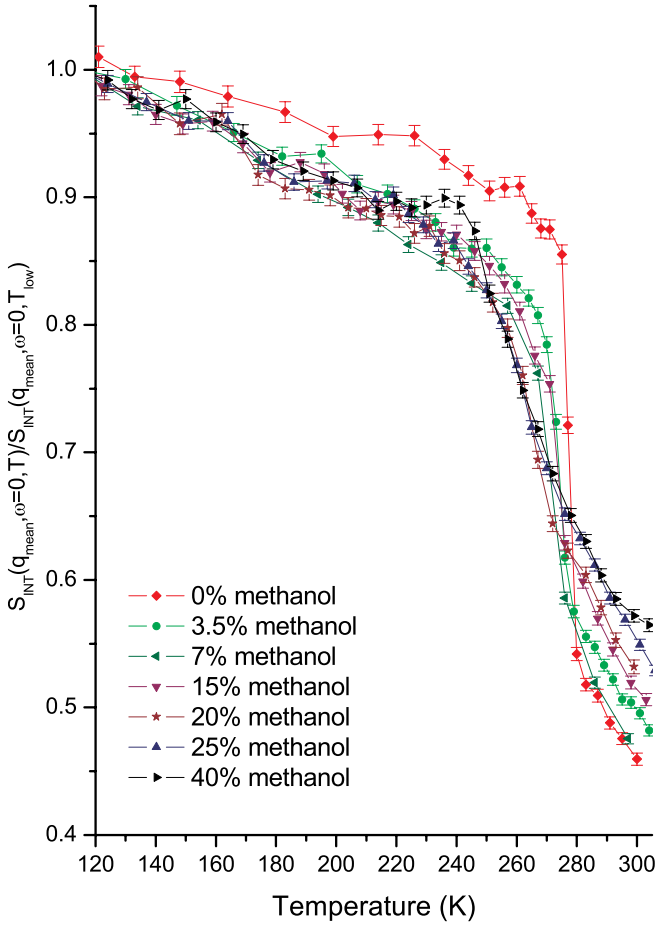


Figure 7.4: Normalized integrated elastic intensity from IN6. The q integration range is $0.35 < q < 1.05$ $\AA^{-1}$. The data for the following samples are presented: pure D_2O; 3.5% CD_3OD/ 96.5% D_2O; 7% CD_3OD/ 93% D_2O; 15% CD_3OD/ 85% D_2O; 20% CD_3OD/ 80% D_2O; 25% CD_3OD/ 75% D_2O; 40% CD_3OD/ 60% D_2O.

section 7.2.3.

The 40% methanol data shows an increase in $S_{INT}(T)$ around 230 K. This feature is not consistent with the data from samples with other methanol concentrations, some amount of contamination by small angle scattering may have occurred in the 40% CD_3OD sample around 230 K, creating an increase in $S_{INT}(T)$.

The corresponding data obtained from IN16 is shown in figure 7.5. The 0% methanol data shows a slow increase in dynamics until ∼260 K at which point a marked transition appears at 277 K coinciding with the melting of the D_2O solution. At 15% methanol concentration dynamics start increasing above ∼195 K and show a marked transition at 270 K. At the other concentrations the onset of the dynamics appears ∼175 K constituting a sharp change in dynamical behavior from the linear regime below 175 K to a steeper linear increase in dynamics above 175 K. The elastic peak intensity then reaches a plateau at lower and lower temperatures with increasing temperature

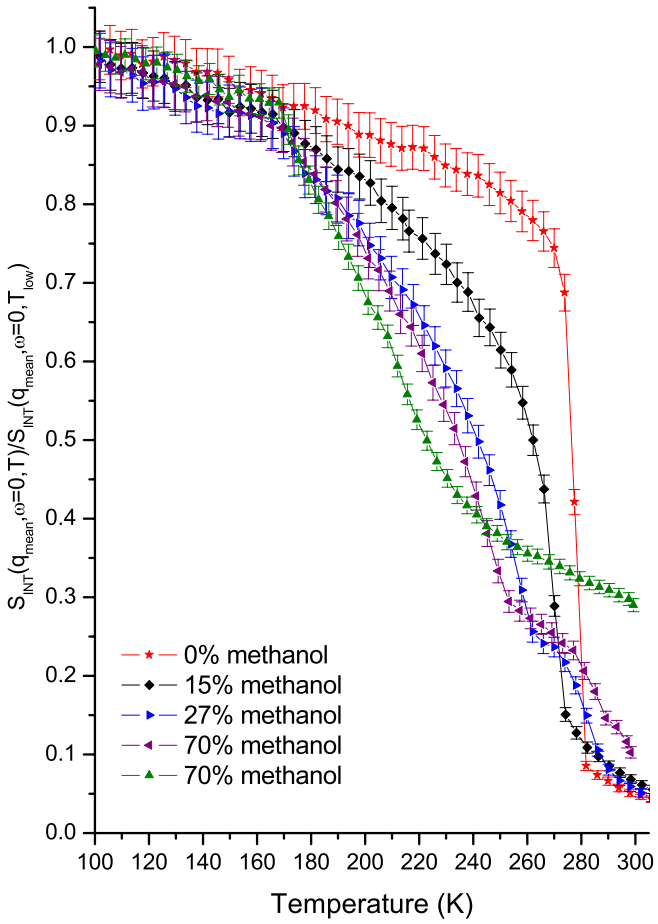


Figure 7.5: Normalized integrated elastic intensity from IN16. The q integration range is $0.43 < q < 1.06$ $\AA^{-1}$. The data for the following samples are presented: pure D_2O; 15% CD_3OD/ 85% D_2O; 27% CD_3OD/ 73% D_2O; 40% CD_3OD/ 60% D_2O; 70% CD_3OD/ 30% D_2O;

starting at temperatures as low as 240 K for 70% methanol. Also the amount of dynamics at high temperatures diminishes with increasing methanol concentration.

7.2.2 MEAN-SQUARE-FLUCTUATION MEASUREMENTS

The $\langle u^2 \rangle$s are calculated directly from the integrated elastic peak intensities as described in section 2.3.3. The $\langle u^2 \rangle$s obtained from IN6 and IN16 data are presented in figures 7.6 and 7.7 respectively. In the IN6 data, the 0% CD_3OD case shows a sharp transition in dynamics at 277 K between a low dynamics regime below ~275 K and a high dynamics regime above ~280 K. In the presence of methanol, the transition is not as sharp and shows a more gradual increase in dynamics. The 3.5% methanol case shows a pronounced increase in dynamics over the range from 270 to 275 K. The onset of the dynamical transition is seen at ~265 K in 7% CD_3OD, 270 K in 15% CD_3OD,

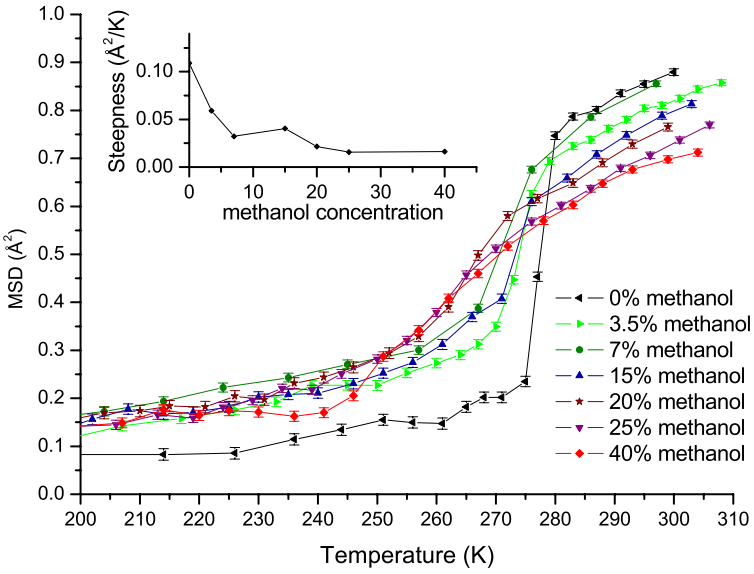


Figure 7.6: Mean-square displacement *vs* temperature for pure D_2O; 3.5% CD_3OD/ 96.5% D_2O; 7% CD_3OD/ 93% D_2O; 15% CD_3OD/ 85% D_2O; 20% CD_3OD/ 80% D_2O; 25% CD_3OD/ 75% D_2O; 40% CD_3OD/ 60% D_2O. The maximum slope of the $\langle u^2 \rangle$ data for the different methanol concentrations is presented as an insert.

260 K in 20% CD_3OD, 255 K in 25% CD_3OD and 240 K in 40% CD_3OD. In the 3.5%, 7%, 15% and 20% CD_3OD cases the dynamics tend to level off above ∼275 K (cf: figure 7.11). Overall the amount of motion above 275 K tends to decrease with increasing methanol concentrations. The 'sharpness' of the dynamical transition can be measured by looking at the maximum slope of the different curves as shown in the inserts of figures 7.4 and 7.5. The transition is seen to become much smoother for concentrations above ∼10% CD_3OD.

The IN16 data is presented in figure 7.5, there again a sharp dynamical transition in the 0% methanol data is present between the low dynamics, linear regime below 270 K and the high dynamics regime above 280 K. When methanol is present, the dynamics show the onset of the dynamical transition at ∼170 K independently of the methanol concentration in the sample. The transition becomes less steep with increasing methanol concentration as can be seen from the insert in figure 7.5. The transition levels off at different temperatures, depending on the methanol concentration: levelling of at ∼270 K for 15% methanol, 260 K for 27%, 250 K for 40% and 240 K for 70% methanol concentration.

Figures 7.6 and 7.7 also present the value of the maximum slope along the $\langle u^2 \rangle$ data for the different methanol concentration. They show the same qualitative increase in slope with decreasing methanol concentration.

Figures 7.8 and 7.9 presents the $\langle u^2 \rangle$ data for IN6 and IN16 in the form of contour plots, giving $\langle u^2 \rangle$ as a function of temperature and methanol concentration. The IN6 data presented in

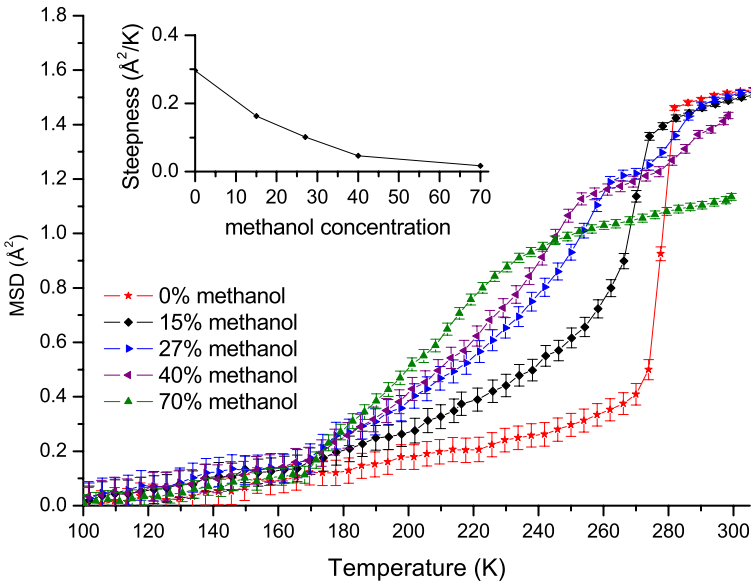


Figure 7.7: Mean-square displacement *vs* temperature for pure D_2O; 15% CD_3OD/ 85% D_2O; 27% CD_3OD/ 73% D_2O; 40% CD_3OD/ 60% D_2O; 70% CD_3OD/ 30% D_2O. The maximum slope of the $\langle u^2 \rangle$ data for the different methanol concentrations is presented as an insert.

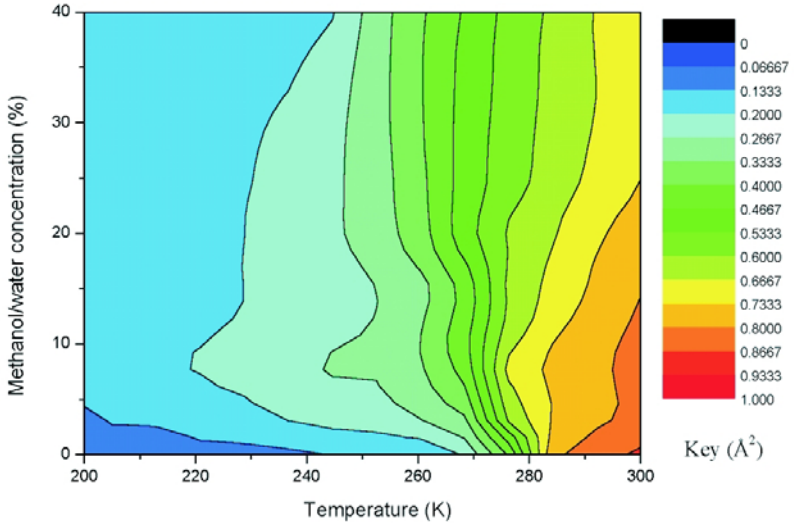


Figure 7.8: Contour plot presenting the protein $\langle u^2 \rangle$ as a function of temperature and methanol concentration (only the temperature range of interest is shown: 200 K to 300 K).

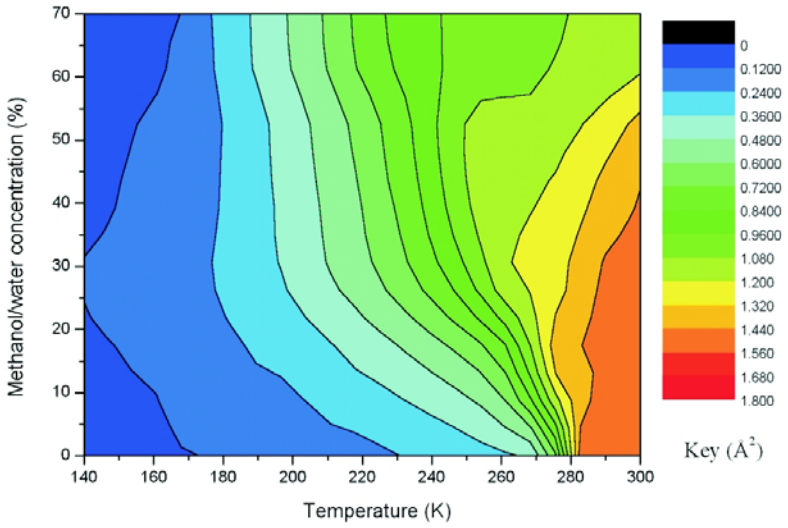


Figure 7.9: Contour plot presenting the protein $\langle u^2 \rangle$ as a function of temperature and methanol concentration (only the temperature range of interest is shown: 140 K to 300 K).

figure 7.8 shows that the transition temperature stabilizes between 230 K and 240 K for concentration greater than ∼10%. The effect of viscosity of the solution at higher methanol concentrations is clearly seen as a general broadening of the transition and decrease of the amount of motion for concentrations greater than 15% and 30% on IN6. The IN16 data shown in figure 7.9 shows a similar trend with a stabilization of the transition temperature between 180 and 200 K for methanol concentrations above ∼30%. Overall the dynamical transition becomes broader with increasing methanol concentration and the amount of motion at higher temperatures goes down with increasing methanol concentrations.

Comparison of the contour plots for IN6 and IN16 show similar trends: sharp transition at 277 K for 0% methanol, stabilization of the transition temperature above a certain methanol concentration and overall broadening of the transition with increasing methanol concentration.

The sharp transition at 277 K seen in both the IN6 and IN16 data for 0% methanol coincides with the freezing point of D_2O. This suggests that pure D_2O completely determines the motions available to the protein. Below 277 K the protein is caged in ice and cannot move hence the very low dynamics.

The stabilization of the onset of the transition in $\langle u^2 \rangle$ for methanol concentration >10% observed in IN6 and for >30% methanol IN16 data is an indication that above a relatively low methanol concentration the protein environment remains unfrozen and therefore independent of the methanol concentration. This independence relative to the methanol concentration was also observed in previous neutron scattering results published by Réat *et al* on Xylanase on IN6 in methanol and DMSO.[1] This effect will be discussed further in comparison with the freezing prop-

erties of methanol/water mixtures in section 7.2.4.

Another common feature between the IN6 and IN16 data: that the amount of motion present at high temperatures (>280 K) decreases with increasing methanol concentration, could be explained by the reduction in the amount of diffusion in the sample due to the increase in viscosity of the solution with increasing methanol concentration.[2] However this hypothesis would not explain why the broadening continues beyond 40% where the viscosity actually starts to go down with increasing viscosity as can be seen in figure 7.10. Another hypothesis is that the methanol molecule being bigger and having a larger scattering cross section than water molecules, contributes largely to the scattering at high methanol concentrations. The effect of this is that the elastic peak intensity has a stronger solvent contribution which does not vanish at high temperatures thereby lowering the $\langle u^2 \rangle$ observed.[3, 4] This effect is discussed in more detail in the next section.

There are two main differences between the IN6 data and the IN16 data: firstly the dynamical transition temperature is generally lowered in the IN16 data by ∼40 K, only the 0% data maintains its transition at ∼0 C. On explanation for this effect is found in the time scales of the dynamics observed in both apparatuses.[5] IN6 is sensitive to motions on the ∼100ps time scale whereas IN16 can observe motion on the ns time scale. IN16 is then able to observe motions on time scales much longer than IN6 such that motions which appear on the ns time scale of IN16 would only appear at higher temperatures in the ∼100 ps time scale of IN6.

Another important distinction between the IN6 and IN16 data is the fact that the $\langle u^2 \rangle$ stabilizes for methanol concentrations above ∼10 % on IN6 but only stabilizes above ∼30% methanol on IN16. A possible explanation for this effect is that IN16, having a better energy resolution, is able to resolve motions in the more viscous solutions with higher methanol concentrations.

7.2.3 SOLVENT BACKGROUND SCATTERING

Knowledge of the isotopic scattering cross sections and isotopic compositions of the samples indicates that, for the protein solutions, about 50% of the scattering is incoherent and 50% coherent. 70% of the incoherent scattering originates from the protein, because of the strong hydrogen contribution.[1] 30% originates from the solvent atoms and influences the observed scattering profile in a manner dependent on the solvent dynamics and the region of q, ω space examined. The isotopic coherent scattering cross sections and protein solution sample composition indicate that most of the coherent scattering originates from the solvent. The self-coherent contribution to this, which is strongest in the low q regime considered here (below $\sim 1\ \mathring{A}^{-1}$), is dynamic and identical in form to the incoherent scattering from the same atom (the amplitude is not identical because it is weighted by the coherent cross section). In addition, there is a contribution originating from cross-correlations (i.e., distance distributions). The intramolecular 'Bragg' part of this contribution lies at q values higher than those considered here. However, a structural 'small-angle' scattering contribution can also exist in the same q-range as the self-coherent and incoherent scattering. The small-angle scattering can be distinguished from the dynamic self-coherent scattering by its q-dependence. To remove the temperature independent small-angle scattering, the scattering

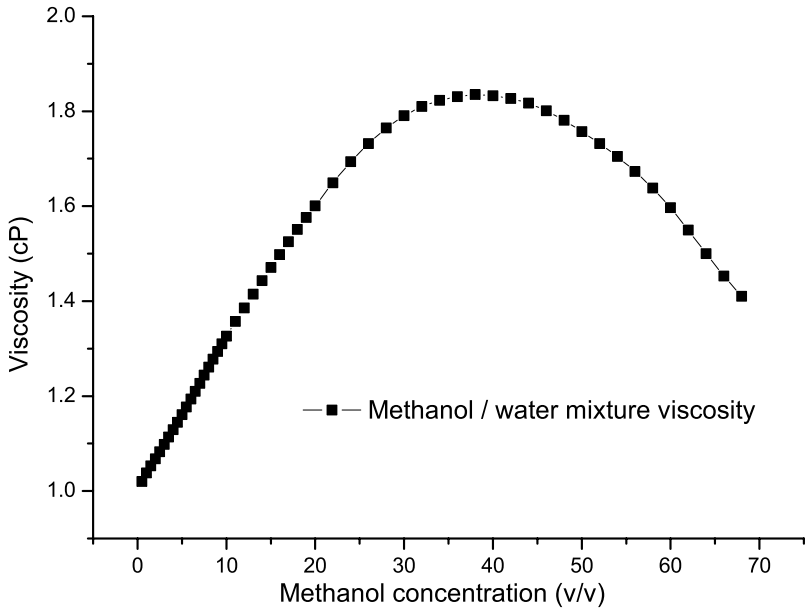


Figure 7.10: Viscosity of methanol/water mixtures as a function of methanol concentration (v/v). Data from the CRD Handbook of Chemistry and Physics.[2]

profiles were normalized with respect to the intensities at 115 K.

However the remaining 30% background scattering due to the solvent has been shown to contribute significantly to the amount of motion calculated from neutron scattering data.[3] The solvent is expected to contribute a constant background to the elastic peak intensity until the solution reaches its melting temperature, at this point the contribution from the solvent diminishes sharply as the motions of the solvent move out of the time windows of the instrument. This produces the saturation feature seen in the integrated peak intensities in IN6 and IN16 above $\sim$280 K. The protein contribution to the measured $\langle u^2 \rangle$ adds to the solvent contribution as a steady increase with temperature of the $\langle u^2 \rangle$ above the protein dynamical transition temperature.[3] In the present experiments it is difficult to dissociate the contributions coming from the protein from those coming from the solvent as the maximum slope in the transition is seen to coincide with the solvent melting point.

The general lowering of the transition with increasing methanol concentration could be due to the contribution of the solvent. Methanol is a bigger molecule than water and is less mobile. Scattering from methanol contributes 65% of the total solvent background scattering at 70% methanol concentration. This slow component to scattering would make a contribution to the elastic peak intensity which would not vanish at high temperatures, artificially inducing a lower value to the average $\langle u^2 \rangle$ at high temperatures.

7.2.4 COMPARISON WITH PROPERTIES OF METHANOL/WATER MIXTURES

Figure 7.11 presents the methanol/water freezing point as a function of temperature and methanol/water concentration (in volume per volume ratio) taken from the litterature.[2, 6] For comparison the point of maximum slope of $\langle u^2 \rangle$ from the IN6 and IN16 experimental results are presented on the same plot. The $\langle u^2 \rangle$ from both experiments follow the methanol/water freezing point closely. The IN6 results present a shift to higher temperature compared to the IN16 results due to the shorter time scales under investigation in IN6 as compared to IN16. On IN16 all results apart from the 40% methanol/water concentration are at higher temperatures than the methanol/water freezing point indicating the relatively short timescale sampled on IN16.

Figure 7.11 also presents the temperatures corresponding to the onset of the dynamical transition for the different methanol concentrations. A qualitatively different picture emerges for IN6 and IN16 results: whereas IN6 results follow the curve of methanol/water freezing point, the IN16 results do not. The IN16 data shows that the onset of the dynamical transition appears at temperatures much lower than the melting point of the solution. This means either the protein prevents glass formation in its vicinity or that the protein has preferential methanol solvation. The second hypothesis has been mostly invalidated by experiments which show that in the presence of cosolvents the concentration of these around a protein tended to be lower than in the bulk.[7] Cosolvents have been shown to solvate proteins by displacing some of the looser water molecule on protein surface.[7–9] This leads to the conclusion that in the methanol/water nixtures the protein disturbs the solvent around it in such a way that it freezes at much lower temperatures than the bulk. Another possible explanation would be that the solvent is frozen but that the experiments are observing only internal motion in the protein. However protein have been shown not to present any dynamical transition when trapped in a glassy environment.[10–13] This points towards the existence of a collaborative effect between the protein and methanol which produces a non-freezing hydration shell around the protein.

A corollary of this results is that - the bulk of the solvent being frozen - the neutron data obtained between the onset of the dynamical transition ($\sim$170 K in IN16) and the point of maximum slope (which corresponds to the solvent melting point) actually corresponds solely to the dynamics of the protein in the sample. The absence of this effect in the IN6 data would tend to indicate that the collaborative effect between the protein and methanol which prevents the hydration shell from freezing happens on long timescales and is therefore not picked up by IN6.

CONCLUSION

In this chapter the results obtained from neutron scattering experiments on xylanase in methanol:water solutions at different concentrations have been presented. The results obtained from the IN6 and IN16 spectrometers at the ILL in Grenoble indicate that pure solvent completely

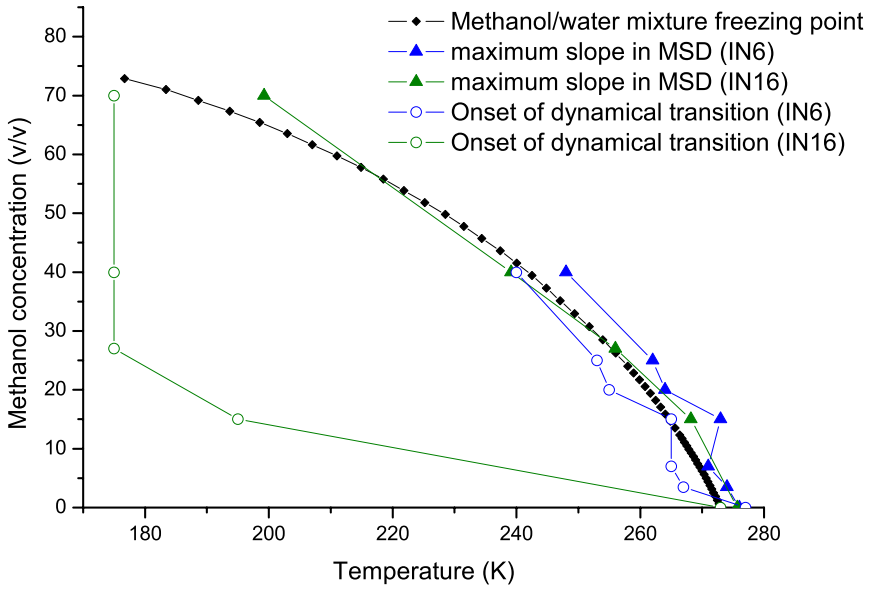


Figure 7.11: Methanol/water mixture freezing point as a function of methanol/water concentration (v/v). For comparison the plot presents the temperature of the onset of the transition as well as temperature of steepest slope of the MSD plots for IN6 and IN16.

dictates the motions of the protein, dynamics only being activated at 277 K corresponding to the solvent melting temperature. The presence of methanol at concentrations as low as 3.5% considerably loosens protein dynamics compared to pure solvent. The onset of the transition in the integrated peak intensity is seen to go down with increasing methanol concentrations: going from 270 K for pure water down to 240 K for 10% methanol on IN6 and down to 170 K for 30% methanol on IN16. The difference in temperature at which the transition stabilizes in IN6 and IN16 is shown to be due to the different timescales probed by the two instruments.

At high temperatures (>280 K) the amplitude of the motions lower with increasing methanol concentration, this effect seems linked to the increase in viscosity of the solution, but could also be due to the strong scattering background this viscous solution makes at high temperatures.

Comparison of the protein dynamics at different methanol concentrations with the bulk properties of methanol/water mixtures strongly suggests a collaborative effect between the protein surface and methanol molecules which induces a non-freezing solvation shell around the protein at temperatures where the bulk of the solvent is frozen.

REFERENCES

[1] REAT, V., DUNN, R., FERRAND, M., FINNEY, J. L., DANIEL, R. M., AND SMITH, J. C. Solvent dependence of dynamic transitions in protein solutions. *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**(18), 9961–6.

[2] CRC Handbook of Chemistry and Physics. CRC Press.

[3] HAYWARD, J. A., FINNEY, J. L., DANIEL, R. M., AND SMITH, J. C. Molecular dynamics decomposition of temperature-dependent elastic neutron scattering by a protein solution. *Biophys J*, 2003, **85**(2), 679–85.

[4] TEHEI, M., MADERN, D., PFISTER, C., AND ZACCAI, G. Fast dynamics of halophilic malate dehydrogenase and bsa measured by neutron scattering under various solvent conditions influencing protein stability. *Proc Natl Acad Sci U S A*, 2001, **98**(25), 14356–61.

[5] DANIEL, R. M., FINNEY, J. L., REAT, V., DUNN, R., FERRAND, M., AND SMITH, J. C. Enzyme dynamics and activity: time-scale dependence of dynamical transitions in glutamate dehydrogenase solution. *Biophys. J.*, 1999, **77**(4), 2184–90.

[6] LIN, J., AND BROWN, C. W. Near-infrared spectroscopic determination of compositions and properties of methanol-water mixtures. *Vib. Spectrosc.*, 1994, **7**, 117–123.

[7] LEHMANN, M. S., MASON, S. A., AND MCINTYRE, G. J. Study of ethanol-lysozyme interactions using neutron diffraction. *Biochemistry*, 1985, **24**(21), 5862–9.

[8] RINGE, D. What makes a binding site a binding site? *Curr Opin Struct Biol*, 1995, **5**(6), 825–9.

[9] MATTOS, C., AND RINGE, D. Proteins in organic solvents. *Curr Opin Struct Biol*, 2001, **11**(6), 761–4.

[10] HAGEN, S. J., HOFRICHTER, J., AND EATON, W. A. Protein reaction kinetics in a room-temperature glass. *Science*, 1995, **269**(5226), 959–62.

[11] CORDONE, L., FERRAND, M., VITRANO, E., AND ZACCAI, G. Harmonic behavior of trehalose-coated carbon-monoxy-myoglobin at high temperature. *Biophys. J.*, 1999, **76**(2), 1043–7.

[12] VITKUP, D., RINGE, D., PETSKO, G. A., AND KARPLUS, M. Solvent mobility and the protein 'glass' transition. *Nat. Struct. Biol.*, 2000, **7**(1), 34–8.

[13] TOURNIER, A. L., AND SMITH, J. C. Translational hydration water dynamics drives the protein glass transition. *Biophys. J.*, 2003, **85**(3).

CONCLUSIONS

Experimental and simulation studies have revealed the presence of a transition in the dynamics of proteins around 220 K. Previous studies have demonstrated that the dehydrated proteins also present a glasslike dynamical transition behavior. However, the present thesis deals with the more biologically relevant case of hydrated proteins. This thesis aimed at answering three main questions: does the solvent control this dynamical transition in proteins? What types of motions do proteins have at the dynamical transition? And, what is the role played by cosolvents in the protein dynamical transition? In order to find answers to these questions molecular dynamics simulations and neutron scattering experiments were performed.

The first question addressed was that of the role of the solvent in the protein dynamical transition. Previous MD work had shown that solvent fluctuations strongly influence internal protein dynamics.[1] In the present work, the dual heatbath simulation method was used in which the protein and its surrounding solvent are set at different temperatures during the simulations. The results thus obtained demonstrate that the dynamical transition in hydrated proteins is driven by changes with temperature in the solvent dynamics. Two main results are seen from these MD simulations: (i) low temperature solvent cages the protein fluctuations. (ii) Heating the solvent while keeping the protein cold drives the protein fluctuations to values intermediate between those in the fully cold and fully hot systems. Further, more detailed, simulations showed that holding the solvent at high or low temperatures abolishes the protein dynamical transition, while varying the solvent temperature with the protein held at 300 K was shown to recover the dynamical transition.

The dynamics specific to the surrounding water shell was subsequently analyzed. The hydration shell was found to undergo itself a glasslike dynamical transition at $\sim$220K, leading to the conclusion that the driving force for the protein transition is a qualitative change in the hydration water shell dynamics. This transition has been seen previously, in MD simulation[2] and in neutron scattering experiments.[3, 4] The solvent transition is found here to be independent of the protein dynamics in that it is present even in simulations in which the protein atoms are fixed. This solvent dynamical transition involves activation of translational solvent diffusion without qualitative change in the water reorientational dynamics. Below the transition the absence of water translational diffusion cages the protein atoms, preventing anharmonic protein dynamics. This is in accord with MD work on ribonuclease, which showed that complete structural relaxation of the protein at 300 K requires relaxation of the water hydrogen-bond network and that the short time-

scale water H-bond lifetime is not affected by the dynamical transition.[5] Decoupling of 'rattling' motions from more global translational diffusion is a characteristic of the glass transition,[6] indicating that the water transition is also glasslike. The solvent transition drives dynamical transition behavior primarily in the side-chain atoms of the external protein regions, *i.e.*, those closest to the solvent. Recent MD results have reported similar results showing that the amount of diffusive motions in proteins at room temperature gradually increases with increasing distance from the protein core.[7] Hydration water interacts principally with the surface atoms, which correspondingly have been shown to exhibit the largest change at the dynamical transition.[8]

The second aim of the present thesis was to investigate the characteristics of protein motion at the dynamical transition. Principal component analysis was used to analyze protein motions over the dynamical transition temperature range. The results provide a description of the subnanosecond protein dynamics activated at the transition. The present PCA results indicate that the solvent:surface interaction propagates to the interior of the protein via collective dynamics that can be described by a very small number of principal components. At 210 K only two quasiharmonic modes and two multiminimum modes are found to deviate substantially from harmonic behavior. The modes initiating the transition are global and distributed over the protein. The largest displacements arise from the activation of barrier-crossing, multiminimum components. In the mode with the largest contribution to the onset of the transition in the present system the barrier crossing involves blocks of supersecondary structural elements moving relative to each other. This rigid-body motion qualitatively resembles structural changes seen in proteins in different functional states.[9] However, it is unclear to what extent motions activated in the dynamical transition describe functional protein dynamics. Additional calculations of the effective damping experienced by the protein along principal modes indicates a solvent induced transition, thereby again illustrating the central role played by the solvent in the protein dynamical transition feature.

The third aim of this thesis was to elucidate the role played by cosolvents used in experiments observing the dynamical transition in hydrated proteins. Previous investigations have suggested that solvation condition influence the intrinsic anharmonicity of the protein potential energy surface.[3, 10–15] Previous results showed the dynamical transition feature to be largely independent of the cryosolvent used.[16] Here the protein dynamical transition feature was investigated at different methanol concentrations levels and on different timescales using neutron scattering experiments. The results indicate that dynamical transition temperature depends on methanol concentration. A sharp transition is seen at 277 K in pure water marking the phase transition of the solvent. The dynamical transition temperature is then seen to go down in temperature with increasing methanol concentration, stabilizing to $\sim$240 K for methanol concentrations greater than $\sim$10% on the 100 ps timescale and stabilizing at $\sim$170 K for methanol concentrations greater than 30% on the nanosecond timescale. Comparison with the bulk properties of methanol/water mixtures indicate that, on short timescales, the dynamical properties of the protein closely follow that of the solvent. On the nanosecond timescale however, the protein dynamics do not follow that of the solvent, rather they show increased dynamics at low temperature where the bulk of the solvent is frozen. These results would indicate that the presence of the protein disturbs the solvent around it. This effect

is thought to be due to a collaborative effect between the protein surface and methanol molecules in such a way that the hydration layer freezes at much lower temperatures than the bulk. The structure of the solvent shell surrounding the protein in the presence of cosolvents is known to be independent of the specific cosolvent used.[17–19] This may explain why the dynamical transition feature is independent of the actual cryosolvent used.[16]

The nonlinear increase in $\langle u^2 \rangle$ is much more gradual for concentration greater than 15%. This behavior is consistent with the presence of a distribution of energy barriers, with successively higher barriers being crossed.[16] This is contrasted by the abruptness of the transition in pure water associated with the D_2O melting point. The protein mean-square-fluctuations temperature at 300 K are seen to decrease with increasing methanol concentration. This is consistent with the existence of a damping effect exerted on protein dynamics by the presence of methanol in the solution.[20] However this does not explain all the data and it is thought that the solvent background scattering may play an important role in this effect.

In conclusion, the present work shows firstly that the observed dynamical transition in protein mean-square-displacements is inseparable from the effect of solvent. The water shell phase transition at $\sim$220 K dictates the behavior of the protein surface and thereby controls the whole of the protein dynamics. Secondly, the protein motions are seen to be of a simple nature over the dynamical transition temperature range, anharmonic motions being present well below the protein dynamical transition. Thirdly, on the 100 ps timescale protein dynamics follow closely that of the solvent in the presence of cosolvents such as methanol, confirming the central role of the solvation layer in the protein dynamical transition. Interactions between the cosolvent and the protein surface take place on the nanosecond timescale that allow the protein to undergo large motions at low temperatures where the bulk of the solvent is frozen.

The present thesis illustrates the complex interactions at work between proteins and their surrounding solvent. These interactions are crucial for the proper folding and function of proteins. In future the correct understanding of these interactions will benefit the efforts made to predict the correct fold of the numerous proteins sequences made available by the genome project. Knowledge of protein solvation will also contribute to making better *in silico* predictions of protein/protein interactions. It will thus assist in the current efforts to map out the complex network of reactions that regulate cellular activity. It is hoped that the work presented in this thesis will have achieved a significant - if small - step towards that goal.

REFERENCES

[1] VITKUP, D., RINGE, D., PETSKO, G. A., AND KARPLUS, M. Solvent mobility and the protein 'glass' transition. *Nat. Struct. Biol.*, 2000, **7**(1), 34–8.

[2] BIZZARRI, A. R., PACIARONI, A., AND CANNISTRARO, S. Glasslike dynamical behavior of the plastocyanin hydration water. *Phys. Rev. E. Stat. Phys.*, 2000, **62**(3 Pt B), 3991–9.

[3] FITTER, J. The temperature dependence of internal molecular motions in hydrated and dry alpha-amylase: the role of hydration water in the dynamical transition of proteins. *Biophys. J.*, 1999, **76**(2), 1034–42.

[4] PACIARONI, A., CINELLI, S., AND ONORI, G. Effect of the environment on the protein dynamical transition: a neutron scattering study. *Biophys. J.*, 2002, **83**(2), 1157–64.

[5] TAREK, M., AND TOBIAS, D. J. Role of protein-water hydrogen bond dynamics in the protein dynamical transition. *Phys. Rev. Lett.*, 2002, **88**(13), 138101.

[6] ANGELL, C. A. Formation of glasses from liquids and biopolymers. *Science*, 1995, **267**, 1924–1935.

[7] DELLERUE, S., PETRESCU, A. J., SMITH, J. C., AND BELLISSENT-FUNEL, M. C. Radially softening diffusive motions in a globular protein. *Biophys. J.*, 2001, **81**(3), 1666–76.

[8] TEETER, M. M., YAMANO, A., STEC, B., AND MOHANTY, U. On the nature of a glassy state of matter in a hydrated protein: Relation to protein function. *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**(20), 11242–7.

[9] GERSTEIN, M., ANDERSON, B. F., NORRIS, G. E., BAKER, E. N., LESK, A. M., AND CHOTHIA, C. Domain closure in lactoferrin. two hinges produce a see-saw motion between alternative close-packed interfaces. *J. Mol. Biol.*, 1993, **234**(2), 357–72.

[10] FERRAND, M., DIANOUX, A. J., PETRY, W., AND ZACCAI, G. Thermal motions and function of bacteriorhodopsin in purple membranes: effects of temperature and hydration studied by neutron scattering. *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**(20), 9668–72.

[11] LEHNERT, U., REAT, V., WEIK, M., ZACCAI, G., AND PFISTER, C. Thermal motions in bacteriorhodopsin at different hydration levels studied by neutron scattering: correlation with kinetics and light- induced conformational changes. *Biophys J*, 1998, **75**(4), 1945–52.

[12] SMITH, J., KUCZERA, K., AND KARPLUS, M. Dynamics of myoglobin: comparison of simulation results with neutron scattering spectra. *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**(4), 1601–5.

[13] ZANOTTI, J. M., BELLISSENT-FUNEL, M. C., AND PARELLO, J. Hydration-coupled dynamics in proteins studied by neutron scattering and nmr: the case of the typical ef-hand calcium-binding parvalbumin. *Biophys J*, 1999, **76**(5), 2390–411.

[14] SMITH, J. C. Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. *Q. Rev. Biophys.*, 1991, **24**(3), 227–91.

[15] DIEHL, M., DOSTER, W., PETRY, W., AND SCHOBER, H. Water-coupled low-frequency modes of myoglobin and lysozyme observed by inelastic neutron scattering. *Biophys. J.*, 1997, **73**(5), 2726–32.

[16] REAT, V., DUNN, R., FERRAND, M., FINNEY, J. L., DANIEL, R. M., AND SMITH, J. C. Solvent dependence of dynamic transitions in protein solutions. *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**(18), 9961–6.

[17] BOUQUIERE, J. P., FINNEY, J., LEHMANN, M. S., AND MOGENS, S. Interaction of the tetramethylammonium ion with the lysozyme molecule, studied using neutron diffraction. *J. Chem. Soc., Faraday Trans.*, 1993, **89**(15), 2701–5.

[18] RINGE, D. What makes a binding site a binding site? *Curr Opin Struct Biol*, 1995, **5**(6), 825–9.

[19] MATTOS, C., AND RINGE, D. Proteins in organic solvents. *Curr Opin Struct Biol*, 2001, **11**(6), 761–4.

[20] CRC Handbook of Chemistry and Physics. CRC Press.

FUTURE PERSPECTIVES

The present work has explored some of the many aspects of the intricate interactions proteins make with their surrounding solvation layer. Through the study of the protein dynamical transition important aspects of these interactions such as the protein dependence on its environment for the appearance of diffusive motions important for function, have been brought to light. The harmonic characteristics of protein motions have been explored using PCA and have brought new light on the diffusive nature of important functional modes. There remains, however, much to be investigated and understood. Further inquiries could be envisaged which would enlarge the present understanding of protein motion and function.

PROTEIN-CRYOSOLVENT INTERACTIONS

Although many aspects of the dynamical transition are presently understood, it is not yet clear how proteins interact with cryosolvents at low temperature. An MD study would be well suited to the investigation of such a question along with ever more elaborate neutron scattering experiments. Such a study would look at the structure and dynamics of the solvent/cosolvent at the protein interface. Factors such as concentration and temperature dependence should also be explored.

TIMESCALE PROBLEM

Recent publications have raised the issues of the timescale in the context of the protein dynamical transition.[1–3] The dynamics of both solvent and protein appear to be qualitatively different on different timescales, therefore a proper understanding of the mechanisms involved would be very useful. A study could use MD simulations to probe different timescales, the present computer power now makes it possible to probe efficiently timescales up to tens of nanoseconds in hydrated systems.

PROTEIN MODES OF MOTION

Another aspect of hydrated protein dynamics which requires further investigation is the characterization and prediction of protein modes of motion. A clear appreciation of protein motion is essential to the understanding of protein function. The techniques currently used include normal

modes analysis and principal component analysis. Normal mode analysis suffers from the assumption that modes are harmonic while many functional modes involve diffusive motions. Principal component analysis on the other hand suffers from the assumption that motions follow simple cartesian vectors thereby being quite inappropriate for describing rotational motions. Performing PCA on the protein internal coordinates instead of normal cartesian coordinate appear as an elegant solution to this problem. Although such analysis have been performed on simple chemical compounds, it has as yet not been performed on proteins. PCA requires long trajectories in order to obtain meaningful results and will therefore greatly benefit from the ongoing increase in computer power which is making long, nanoseconds, trajectories more and more accessible, thereby eliminating the sampling problem.[4]

SOLVENT MODELS

Finally, another interesting focus of research would concentrate on improving of the existing continuum solvent models so as to take into account the existing knowledge of the protein hydration layer. The presence of the first hydration layer has been shown to be the most important in reproducing solution dynamics. Therefore, hybrid solutions could be envisaged in which the more stable/bound water molecules would be simulated implicitly while the bulk of the solvent would be simulated using a continuum approximation. In this way computational power would be concentrated on simulating the protein atoms while at the same time accurately taking into account the effect of solvent. Thus longer simulation times could be reached.

These aspects only represent the 'next step' viewed from the perspective of the present thesis and are in no way exhaustive. There are undoubtedly many more aspects of protein hydration and motion that need to be and will be addressed in the future than the few listed above. Fully understanding proteins, their folding, dynamics and interactions with solvent, ligand and other proteins remains the final goal towards which this thesis work and many of the present efforts in Biology and Biophysics are dedicated.

REFERENCES

[1] DANIEL, R. M., FINNEY, J. L., REAT, V., DUNN, R., FERRAND, M., AND SMITH, J. C. Enzyme dynamics and activity: time-scale dependence of dynamical transitions in glutamate dehydrogenase solution. *Biophys. J.*, 1999, **77**(4), 2184–90.

[2] FENIMORE, P. W., FRAUENFELDER, H., MCMAHON, B. H., AND PARAK, F. G. Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc Natl Acad Sci U S A*, 2002, **99**(25), 16047–51.

[3] HAYWARD, J. A., FINNEY, J. L., DANIEL, R. M., AND SMITH, J. C. Molecular dynamics decomposition of temperature-dependent elastic neutron scattering by a protein solution. *Biophys J*, 2003, **85**(2), 679–85.

[4] HESS, B. Convergence of sampling in protein simulations. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2002, **65**(3 Pt 1), 031910.

APPENDIX I: PREVIOUS NOSÉ-HOOVER IMPLEMENTATION

This is the main loop of the algorithm as found when looking into the CHARMM code of the previous Nosé Hoover implementation[†]

$v'_\eta \leftarrow v_\eta + \frac{\Delta t/2}{Q} F_\eta$
$\eta \leftarrow \eta + \Delta t/2 \; v'_\eta$
$\mathbf{v}_i \leftarrow \mathbf{v}_i - \Delta t/2 \; v_\eta \mathbf{v}_i$

$\mathbf{v}_i \leftarrow \mathbf{v}_i + \frac{\Delta t}{2m_i}\mathbf{F}_i$
$\mathbf{r}_i \leftarrow \mathbf{r}_i + \Delta t \; \mathbf{v}_i$
get new forces, $\mathbf{F}_i$ (Energy call)
$\mathbf{v}_i \leftarrow \mathbf{v}_i + \frac{\Delta t}{2m_i}\mathbf{F}_i$

This is the standard Velocity Verlet algorithm

$\mathbf{v}_i \leftarrow \mathbf{v}_i - \Delta t/2 \; v_\eta \mathbf{v}_i$

Predictor

$E' \leftarrow E$
E update
$F'_\eta \leftarrow E - NKT$
$v_\eta \leftarrow v'_\eta + \frac{\Delta t/2}{Q} F'_\eta$
IF E' and E converged then continue
IF Ncycle not reached then loop.

Correction loop

E update
$F_\eta \leftarrow E - NKT$
$v_\eta \leftarrow v'_\eta + \frac{\Delta t/2}{Q} F_\eta$

[†]The following variables are used in the previous implementation, there meaning and there translation in a more understandable format is given here:
SNHF, SNHF1: thermostat force, F_η and F'_η
SNHV, SNHV1 : thermostat velocity, v_η and v'_η
SNH : thermostat position, η
EPTKX, EPTKE1 : Kinetic energies of the atoms, E and E'

To summarize this algorithm:

- Half a step is taken along v_η
- A Velocity Verlet step is taken
- Half a step is taken along v_η again
- Looking around v'_η, v_η is corrected in an iterative way so that F_η, E and $\mathbf{v}_i$ be consistent with the equation of motion at that instant

A few things are not quite clear with this algorithm:

- Why is the first step taken along v_η and not v'_η which has been updated at that point?
- Why is the guess taken along v_η and not along v'_η when at this point it would clearly make sense?
- The last three items of the algorithm are actually redundant.
- In the correction loop a step along an updated v_η is taken, a step along the same v_η is then taken in the first part of the next step. Which mean two half steps are taken along the same v_η.

Such an implementation works. It is however not a very precise nor very efficient implementation. It owes a lot to the correction step, without which it wouldn't work at all. Formally, it is not time symmetric.

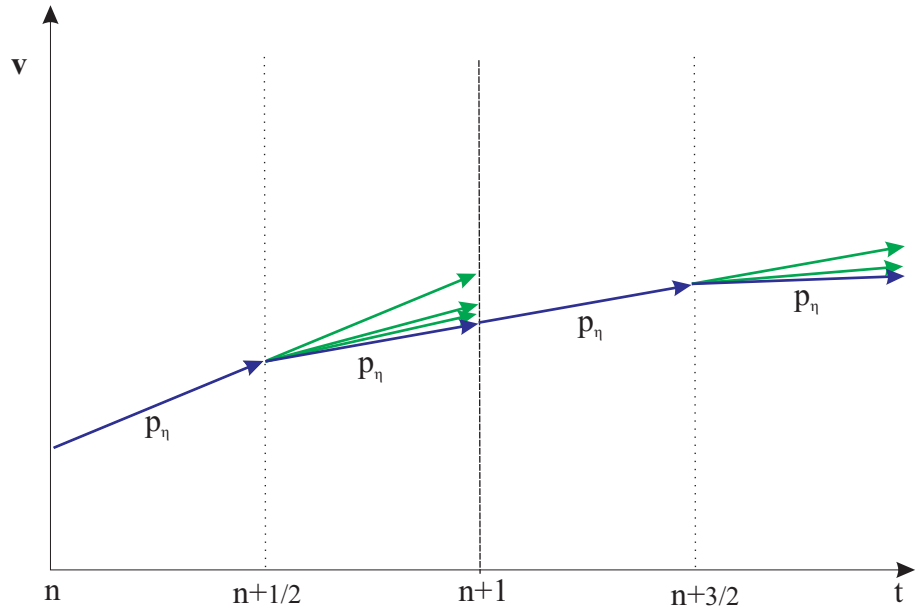


Figure 7.12: Schematic presentation of the time steps taken by the previous implementation of the Nosé-Hoover algorithm. The green lines indicate estimates which of p_η which are then refined in the correction loop.

APPENDIX II: NOSÉ-HOOVER-CHAIN IMPLEMENTATION

This appendix presents the implementation made of the Nosé-Hoover-Chain algorithm which the author added to the CHARMM package. It basically follows that of G. Martyna, M. Tuckerman, D. Tobias and M. Klein.[1] Their algorithm was adapted for multiple heatbaths, parallel processing, heating procedure, and constant heatflux procedure.

Following the ***direct translation technique*** described in chapter 2 the evolution operator can be written:

$$e^{iL\Delta t} = e^{iL_{NHC}\Delta t/2}\, e^{iL_1\Delta t/2}\, e^{iL_2\Delta t}\, e^{iL_1\Delta t/2}\, e^{iL_{NHC}\Delta t/2}$$

where iL_1 and iL_2 are the same as in equation 2.28. The Nosé-Hoover-Chain part (NHC) is then be broken down to use multiple time steps and Yoshida-Suzuki steps following equations 2.30 and 2.32. The Liouville operator can be written:

$$iL = iL_{Newton} + iL_{NHC}$$

with:

$$iL_{NHC} = -\sum_{i=1}^{N} v_{\eta_i}\mathbf{v}_i \cdot \nabla_{\mathbf{v}_i} + \sum_{k=1}^{2} v_{\eta_k}\frac{\partial}{\partial \eta_k} + \left(G_1 - v_{\eta_2}v_{\eta_1}\right)\frac{\partial}{\partial v_{\eta_1}} + G_2\frac{\partial}{\partial v_{\eta_2}}$$

and:

$$iL_{Newton} = \sum \mathbf{v}_i \cdot \nabla_{\mathbf{r}_i} + \sum_i \left(\frac{\mathbf{F}_i(\mathbf{r})}{m_i}\right) \cdot \nabla_{\mathbf{v}_i}$$

The NHC operator can then be split up into:

$$\begin{aligned}
iL_1^{NHC} &= -\sum_{i=1}^{N} v_{\eta_i}\mathbf{v}_i\nabla_{\mathbf{v}_i} \\
iL_2^{NHC} &= \sum_{k=1}^{2} v_{\eta_k}\frac{\partial}{\partial \eta_k} \\
iL_3^{NHC} &= \left(G_1 - v_{\eta_2}v_{\eta_1}\right)\frac{\partial}{\partial v_{\eta_1}} \\
iL_4^{NHC} &= G_2\frac{\partial}{\partial v_{\eta_2}}
\end{aligned}$$

and iL_3 is further split up as:

$$iL_{3.1}^{NHC} = G_1\frac{\partial}{\partial v_{\eta_1}}$$

$$iL_{3.2}^{NHC} \quad = \quad v_{\eta_2} v_{\eta_1} \frac{\partial}{\partial v_{\eta_1}}$$

iL_1 and iL_2 are completely independent from each other so that the full evolution operator can be expressed as:

$$\begin{aligned} e^{iL^{NHC}\Delta t} = \quad & e^{iL_4^{NHC}\Delta t/2}\, e^{iL_{3.2}^{NHC}\Delta t/4}\, e^{iL_{3.1}^{NHC}\Delta t/2}\, e^{iL_{3.2}^{NHC}\Delta t/4} \\ & e^{iL_1^{NHC}\Delta t/2}\, e^{iL_2^{NHC}\Delta t/2} \\ & e^{iL_{3.2}^{NHC}\Delta t/4}\, e^{iL_{3.1}^{NHC}\Delta t/2}\, e^{iL_{3.2}^{NHC}\Delta t/4}\, e^{iL_4^{NHC}\Delta t/2} \end{aligned}$$

such that:

$$e^{iL\Delta t} = e^{iL^{NHC}\Delta t/2}\, e^{iL^{Newton}\Delta t}\, e^{iL^{NHC}\Delta t/2}$$

i.e. half a step of NHC followed by a step of Newton and an another step of NHC. The NHC half step can be translated into a computer algorithm, the algorithm for each chain translates into:

$scale \leftarrow 1.0$
CALL GETKE
$G_2 \leftarrow (KE - kT)/Q_2$
$v_{\eta_2} \leftarrow v_{\eta_2} + G_2 \Delta t/4$
$G_1 \leftarrow (KE - NkT)/Q_1$
$v_{\eta_1} \leftarrow v_{\eta_1} \cdot exp(-v_{\eta_2} \Delta t/8)$
$v_{\eta_1} \leftarrow v_{\eta_1} + G_1 \Delta t/4$
$v_{\eta_1} \leftarrow v_{\eta_1} \cdot exp(-v_{\eta_2} \Delta t/8)$
$scale \leftarrow scale \cdot exp(-v_{\eta_1} \Delta t/2)$
$KE \leftarrow KE \cdot scale^2$
$\eta_1 \leftarrow \eta_1 + v_{\eta_1} \Delta t/2$
$\eta_2 \leftarrow \eta_2 + v_{\eta_2} \Delta t/2$
$G_1 \leftarrow (KE - NkT)/Q_1$
$v_{\eta_1} \leftarrow v_{\eta_1} \cdot exp(-v_{\eta_2} \Delta t/8)$
$v_{\eta_1} \leftarrow v_{\eta_1} + G_1 \Delta t/4$
$v_{\eta_1} \leftarrow v_{\eta_1} \cdot exp(-v_{\eta_2} \Delta t/8)$
$G_2 \leftarrow (KE - kT)/Q_2$
$v_{\eta_2} \leftarrow v_{\eta_2} + G_2 \Delta t/4$
$V \leftarrow V \cdot scale$

The subroutine NHCINT is also presented, it is used before and after each Velocity Verlet step in the algorithm[†]:

```
    SUBROUTINE NHCINT
C Variable used are:
C NOBL      : Number of parts in the system / heatbath chains
C EPTKX     : Kinetic Energie of the different part of the system
C NHX1,NHX2 : Position of the two heatbaths in the chain
C NHV1,NHV2 : Velocity of the two heatbaths in the chain
C NHG1, NHG2: Forces on the two heatbaths in the chain
C RTMPR     : Reference temperature
```

[†]CHARMM specific commands and non essential parts have been removed.

```
C SQM1, SQM2 : Thermal mass of the thermostat
C NHCAPR     : Number of Yoshida-Suzuki step = 1, 3 or 5
C NHCW       : Yoshida-Suzuki parameters
C NHCMTS     : Number of time step per cycle: Multiple time steps

C Update Kinetic Energies
    CALL NHKE
    DO I=1,NOBL
        DO J=1,NHCMTS
            DO K=1,NHCAPR
                DT = SA2X*NHCW(K)/NHCMTS
                NHSCAL(I) = 1.
                NHV2(I) = NHV2(I) + DT/2*NHG2(I)
                AA = exp(-DT/4*NHV2(I))
                NHG1(I) = (EPTKX(I)- NDGN(I)*KBOLTZ*PTMPR(I))/SQM1(I)
                NHV1(I) = NHV1(I)*AA*AA + DT/2*NHG1(I)
                NHSCAL(I) = NHSCAL(I) * exp(-DT*NHV1(I))
                EPTKX(I) = EPTKX(I)*NHSCAL(I) *NHSCAL(I)
                NHX1(I)=NHX1(I)+DT*NHV1(I)
                NHX2(I)=NHX2(I)+DT*NHV2(I)
                NHG1(I) = (EPTKX(I)- NDGN(I)*KBOLTZ*PTMPR(I))/SQM1(I)
                AA = exp(-DT/4*NHV2(I))
                NHV1(I) = NHV1(I)*AA*AA+ DT/2*NHG1(I)
                NHG2(I)=(SQM1(I)*NHV1(I)*NHV1(I)-KBOLTZ*PTMPR(I))/SQM2(I)
                NHV2(I) = NHV2(I) + DT/2*NHG2(I)
            ENDDO
        ENDDO
        DO J=ATFRST,ATLAST
            VX(J) = VX(J)*NHSCAL(I)
            VY(J) = VY(J)*NHSCAL(I)
            VZ(J) = VZ(J)*NHSCAL(I)
        ENDDO
    ENDDO
    RETURN
    END
```

REFERENCES

[1] MARTYNA, G. J., TUCKERMAN, M. E., TOBIAS, D. J., AND KLEIN, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.*, 1996, **87**(5), 1117–1157.