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 Science

presented by

Diplom-Physiker:	Torsten Becker
born in:	Schongau
Oral examination:	17. November 2004

Protein Dynamics

Comparison of Incoherent Neutron Scattering and Molecular Dynamics Simulation

Referees: Prof. Dr. Jeremy C. Smith
PD Dr. Ilme Schlichting

The city really ain't no bigger than the friendly people that you meet.
Bill Withers: Lonely town, lonely street

Acknowledgments

This thesis would not have been possible without the help of many people. At this point i want to say thank you to all of them.

I want to especially thank my supervisor Jeremy C. Smith for letting me work on this interesting subject and for guiding the progress of this thesis with ideas and discussions.

Andrea Vaiana and Alexander Tournier shared with me the joy and pain of getting started in research. Thanks a lot for all the interesting discussions about science, politics and all the rest...

My special thanks to Erika Balog for working with me on the dimerization project and discovering the universal scaling law of errorbars.

I am grateful to Jennifer Hayward for providing me with trajectories to test my ideas about the dynamical transition.

If this thesis is written in a language approximating English it is due to the great effort Durba Sengupta, Lars Meinhold and Vandana Kurkal put into proof reading.

This thesis would lack its nicest graphics without the help of Sabine Radl. Thank you very much.

Finally, a special thank you to my parents who encouraged and supported me throughout my long education.

Zusammenfassung

Proteindynamik und deren Beziehung zur biologischen Funktion von Proteinen ist Gegenstand zahlreicher experimenteller und theoretischer Untersuchungen. In der vorgelegten Arbeit werden Aspekte der Proteindynamik im Pico- und Nanosekundenbereich untersucht. Experimentelle Untersuchungen haben eine Abhaengigkeit zwischen enzymatischer Aktivität und atomaren Fluktuationen auf diesen Zeitskalen aufgezeigt. Inkoherente Neutronenstreuung sowie Molekulardynamiksimulationen stellen geeignete Instrumente zur Untersuchung atomarer Dynamik auf diesen Zeitskalen zur Verfügung.

Ein interessantes Phänomen im Hinblick auf den Zusammenhang zwischen Flexibilität und Aktivität ist die *Dynamical Transition*, d.h. ein nicht-linearer Anstieg atomarer Fluktuationen bei einer charakteristischen Temperatur $T_0 \sim 200$K. Die explizite Einbeziehung des instrumentellen Auflösungsvermögens eines Neutronenstreuspektrometers in die theoretische Analyse der *Dynamical Transition* führt zu einer neuen Interpretation dieses Überganges. Mittels Molekulardynamiksimulationen wird die quantitative Übereinstimmung dieser alternativen Interpretation mit der Zeitskalen- wie auch der Temperaturabängigkeit der mittleren quadratischen Auslenkung eines Proteins in Lösung gezeigt. Darüberhinaus bietet diese Interpretation eine Erklärung der experimentell beobachteten Verschiebung der Übergangstemperatur, T_0, in Abhängigkeit der instrumentellen Auflösung.

Die Gauss'sche Näherung, grundlegend für die experimentelle Bestimmung der mittleren quadratischen Auslenkung, $\langle \Delta r^2 \rangle$, wird untersucht und Korrekturen hierzu diskutiert. Abweichungen von dieser Näherung werden für $Q \leq 6\text{Å}^{-2}$ auf die Heterogenität atomarer Bewegungen zurückgeführt.

Abschliessend wird eine Methode vorgeschlagen, um aus der inkoherenten inelastischen Streufunktion die Schwingungsdichte des Systems abzuleiten. Mittels dieser Methode werden Änderungen der Schwingungsdichte des Proteins Dihydrofolate Reduktase bei Bindung des Liganden Methotrexate bestimmt. Es wird gezeigt, dass die Änderungen im Spektrum dieser internen Freiheitsgrade einen signifikanten Beitrag zur freien Bindungsenergie dieses Systems beitragen.

Summary

Protein dynamics and its relation to protein function is the subject of various studies using both, theoretical and experimental techniques. In this thesis, several aspects of protein dynamics on short timescales are addressed. Motions in the pico- to nanosecond timescale have been experimentally shown to be intimately related to enzyme activity. Incoherent neutron scattering and molecular dynamics simulation are well suited and widely used to study motions on the above timescales.

A prominent phenomenon in the context of this observed flexibility-activity relationship is the dynamical transition, *i.e.* a non-linear increase in atomic fluctuations at a characteristic transition temperature of $T_0 \sim 200$K. By explicitly incorporating finite resolution of neutron spectrometers in the theoretical analysis of neutron scattering experiments, a novel interpretation of the dynamical transition arises. This alternative 'frequency window' interpretation is shown to reproduce the timescale and temperature dependence of mean-square displacements calculated from MD simulations of a protein in solution. The frequency window interpretation, furthermore, offers an explanation of the experimentally observed shift of T_0 with instrumental resolution. Implications of the new interpretation for the relation between the dynamical transition and enzyme activity are discussed.

Molecular dynamics simulations are further used to test the Gaussian approximation implicit in experimental data analysis. Deviations from Gaussian scattering in the calculated spectra for $Q^2 \leq 6\text{Å}^2$ are shown to be dominated by the distribution of $\langle \Delta r^2 \rangle$.

Finally, a method to derive the vibrational density of states on an absolute scale from low-temperature inelastic incoherent neutron scattering is suggested. The change in the vibrational density of states of the protein dihydrofolate reductase on binding the ligand methotrexate is determined. The vibrations of the complex soften significantly relative to the unbound protein. The resulting free energy change, which is directly determined by the density of states change, is found to contribute significantly to the binding equilibrium.

Publication list

Balog, E., Becker, T., Oettl, M., Lechner, R., Daniel, R., Finney, J. & Smith, J. (2004). Direct determination of vibrational density of states change on ligand binding to a protein. *Phys. Rev. Lett.*, **93**, 028103.

Becker, T. & Smith, J.C. (2003). Energy resolution and dynamical heterogeneity effects on elastic incoherent neutron scattering from molecular systems. *Phys. Rev. E*, **67**, 021904.

Becker, T., Fischer, S., Noe, F., Tournier, A., Ullmann, M. & Smith, J. (2003). Protein dynamics: Glass transition and mechanical function. In B. Kramer, ed., *Advances in Solid State Physics*, vol. 43, 677–694, Springer.

Becker, T., Hayward, J., Daniel, R., Finney, J. & Smith, J. (2004). Neutron frequency windows and the protein dynamical transition. *Bioph. J.*, **87**, 1–9.

Hayward, J., Becker, T. & Smith, J. (2002). The glass transition in proteins. In Krause, E. & Jäger, W., eds., *High Performance Computing in Science and Engineering'02*, 503–511, Springer.

Meinhold, L., Lammers, S., Becker, T. & Smith, J.C. (2004). Convergence properties of x-ray scattering calculated from protein crystal molecular dynamics simulations. *Physica B*, **350**, 127–131.

Contents

0 About the thesis **1**

1 Protein dynamics **3**

1.1 Structural diversity of proteins 3

1.2 Flexibility-activity relationship 4

1.3 The energy landscape of proteins 5

1.4 The dynamical transition in proteins 9

1.5 Protein association 13

2 Neutron scattering from proteins **15**

2.1 Theory of neutron scattering 15

2.2 Response function 19

2.3 Coherent and incoherent scattering 20

2.4 Elastic, quasielastic and inelastic scattering 22

2.5 Separation of motions 24

2.6 The Gaussian approximation 25

3 Computer simulations of proteins **27**

3.1 Molecular dynamics forcefield 28

3.2 Time evolution 32

3.3 Analysis of molecular dynamics simulations 35

4 Protein dynamics and neutron scattering **39**

4.1 Internal protein dynamics 40

4.2 Energy landscapes 41

4.3 Stretched exponential relaxation 43
4.4 Single exponential relaxation 50
4.5 Conclusions 56

5 The dynamical transition in proteins 59
5.1 Equilibrium and frequency window scenario 60
5.2 Dynamical transition in MD Simulations 64
5.3 Dynamical transition in experiment 68
5.4 Transition-function relationship 70
5.5 Conclusions 72

6 Gaussian approximation of the elastic scattering function 75
6.1 Q-dependence of $S(Q,0)$ 76
6.2 Non-Gaussian scattering in MD Simulations 79
6.3 Conclusions 86

7 Protein association - a neutron scattering study 87
7.1 Thermodynamics of association processes 88
7.2 Neutron scattering and vibrational density of states 89
7.3 Experimental data analysis 91
7.4 From $G(\omega)$ to $g(\omega)$. 92
7.5 Absolute vs relative mean-square displacements 94
7.6 Ligand binding of dihydrofolate reductase 99
7.7 Conclusions 103

8 Final Remarks 105
8.1 Protein dynamics 105
8.2 Association processes – ligand binding 107
8.3 Outlook 108

Appendix A i

References i

Chapter 0

About the thesis

Protein dynamics has been the subject of vast experimental and theoretical work in the last decades. Various experimental techniques, such as Mössbauer-spectroscopy or neutron scattering are well suited and have been applied to probe protein motions on the pico- to nanosecond timescale. Furthermore, advances in computational speed enable molecular dynamics simulations to give atomic detailed descriptions of dynamical processes on these fast timescales. The results of these investigations show that protein dynamics shares features commonly assigned to complex systems such as glasses. A phenomenon that attracted much attention is the *dynamical transition* of proteins, characterized by a non-linear increase in the measured mean-square displacement above a transition temperature $T_0 \sim 200k$. This transition was shown to correlate with the onset of (*e.g.* ribonuclease A) or qualitative changes in (*e.g.* myoglobin) the measurable activity of several enzymes (Frauenfelder *et al.*, 1999; Rasmussen *et al.*, 1992). Whether a coupling of fast, picosecond dynamical processes to slower, biological important motions of the enzyme is a general property of proteins is still under debate (Brunori *et al.*, 1999; Daniel *et al.*, 1998). Furthermore, it is important to know whether the dynamical transition is an inherent property of proteins or whether protein dynamics is *slaved* by the surrounding solvent. For a protein in solution, glutamate dehydrogenase GDH in methanol/water cryosolvents, the transition temperature has been shown to depend on the instrumental resolution (Daniel *et al.*, 1999).

In this thesis, a theoretical analysis of neutron scattering in the context of the dynamical transition is given. By explicitly incorporating finite resolution of neutron spectrometers, a new and alternative interpretation of the dynamical transition is presented and shown to be consistent with both, simulation and experiment.

Molecular dynamics simulations are further used to test the Gaussian approximation implicit in experimental data analysis. Deviations from Gaussian scattering in the calculated spectra for $Q^2 \leq 6\text{Å}^2$ are shown to be dominated by the distribution of $\langle \Delta r^2 \rangle$.

Finally, a method to derive the vibrational density of states on an absolute scale from low-temperature inelastic incoherent neutron scattering is suggested. The change in the vibrational density of states of the protein dihydrofolate reductase (DHFR) on binding the ligand methotrexate (MTX) is determined.

The outline of this thesis is as follows:
In **Chapters 1-3** a general introduction to the field of protein dynamics and an outline of the theory and methods used in this thesis are given. **Chapter 1** puts the field of protein dynamics into a broader context, introduces the general concepts underlying theoretical investigations and gives a short summary of the state of research on the dynamical transition. **Chapter 2** provides the basic theory of neutron scattering as needed for Chapters 4-6. The forcefield, algorithms and analysis used for molecular dynamics simulations of biological macromolecules are introduced in **Chapter 3**.

Chapter 4 presents an analysis of the temperature dependence of the incoherent intermediate scattering function calculated from molecular dynamics simulations. The scattering function is shown to allow for an interpretation where all the temperature dependence of slow relaxation processes resides in the corresponding relaxation frequency. This 'frequency window' model is discussed with respect to the opposite 'equilibrium' interpretation where changes of the scattering function upon temperature increase are described by corresponding changes in the long-time converged properties of the system.

Chapter 5 presents a theoretical analysis of the dynamical transition as measured by neutron scattering. Finite resolution of the instrument is taken explicitly into account. The frequency window model is shown to offer an alternative interpretation of the dynamical transition in terms of temperature dependent relaxation processes. The theoretical analysis is shown to reproduce quantitatively the temperature and timescale dependence of mean-square displacements calculated from MD simulations. The frequency-window interpretation is able to explain the experimentally observed timescale dependence of the transition temperature of a protein in solution.

Chapter 6 investigates the Gaussian approximation commonly made in analyses of neutron scattering spectra. Using neutron spectra calculated from molecular dynamics simulations it is possible to access the errors inherent in the analysis procedure and to test methods to improve the analysis. Furthermore, the origin of non-Gaussian scattering is investigated for internal protein dynamics and shown to arise mainly from heterogeneity of atomic mean-square displacements.

Finally, **Chapter 7** presents inelastic neutron scattering measurements on the protein dehydrofolate reductase (DHFR), an important enzyme in cancer research. A method is introduced to quantitatively derive the vibrational density of states. An estimate of the density of states changes upon ligand binding gives insight into the importance of internal vibrational degrees of freedom for the free energy of complex formation.

Chapter 1

Protein dynamics

> *. . . and if we were to name the most powerful assumption of all, which leads one on to and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms.*
> *R.P. Feynman, The Feynman lectures of physics (Feynman* et al.*, 1963)*

All things are made of atoms and these atoms are *jiggling and wiggling*. This seemingly naive statement is, as R.P. Feynman pointed out, the basic insight underlying modern structural biology. With the adoption of X-ray crystallography to the field of biology it became finally possible to look at the phenomena of life at atomic resolution. The determination of the structure of DNA still relied on the structural simplicity of the double helix and the ingenious guesswork of Watson and Crick. Soon afterwards, however, Perutz was able to solve the phase problem of protein crystallography and opened the door to the investigation of the structural aspects of life. The first protein structure was solved by Kendrew in 1957 (myoglobin) and Perutz was able to obtain an atomic model for haemoglobin two years later. By the year 2002 the Protein Data Bank (PDB) stored more than 15 000 structures of biological macromolecules (Berman *et al.*, 2000a).

One of the challenges of modern biology is to relate these structures to their function. This relation is still insufficiently understood and touches upon several open questions on the borderline between physics and biology (Frauenfelder *et al.*, 1999).

1.1 Structural diversity of proteins

Proteins are polypeptides consisting of a sequence of residues chosen from the set of 20 naturally occurring amino acids. For a protein like myoglobin with 153 residues there are already $20^{153} \approx 10^{199}$ possible sequences each bearing its own structural and dynamical peculiarities. Even for a given sequence of amino-acids the structural complexity is enormous. Assuming only two possible configurations per amino-acid, myoglobin can take on 2^{153} different conformations. This vast number of possible configurations lays at the heart of the folding problem of theoretical structural biology.

In spite of the huge number of possible configurations and the large amount of different proteins in living cells, the investigation of proteins revealed surprising similarities in the concepts underlying their structure. On different structural levels (secondary, tertiary, quaternary) proteins seem to use the same building blocks to realize their own specific fold. On the lowest level one can identify repeatedly use of secondary structural elements such as helices, β-sheets or turns. These structures themselves are often arranged in a higher level of order in so called motifs (Greek key, zinc-finger etc.). Similar structural arrangements can be found in proteins fulfilling similar functions, *e.g.* β-barrels or membrane-spanning arrangements of α-helices in membrane channels.

These striking findings strongly suggest that the three-dimensional structure of proteins holds the key to their function (Anfinsen, 1973). However, this insight must not be taken to an extreme that proteins are rigid molecules whose motions are merely thermal noise. Already at the very beginning, it became clear that the structure of proteins alone is not enough to understand the mechanisms by which proteins fulfill their function. Structural similar proteins, *e.g.* lysozyme and α-lactalbumin can serve rather different purposes. Sometimes the atomic structure even seems to forbid the protein to fulfill its task. The atomic structure of myoglobin for example does not reveal a permanent path connecting the solvent to the active center. Nevertheless it is experimentally known, that myoglobin can bind molecules as big as isocyanides ($\approx$ 10 atoms). It follows that a path a few Ångstrøm in diameter has to be transiently formed by myoglobin. It is the *jiggling and wiggling* of the atoms that allow ligands to enter proteins and perform their function (Frauenfelder & McMahon, 2000; Perutz & Mathews, 1966). To understand the molecular building blocks of life we therefore have to investigate the structure and the dynamics of biological macromolecules and learn how both aspects are connected to each other. The kind of motions and corresponding timescales of functional important dynamics are widely unknown. It is one of the prominent problems in biophysics to distinguish these *essential* motions from the background of thermal noise.

1.2 Flexibility-activity relationship

As already stated, proteins are flexible and have to be flexible to fulfill their function. This *flexibility-activity* relationship in proteins is widely accepted (for a review see Daniel *et al.* (2003a)). The actual motions required, their atomic description or the timescales of such *essential* motions, however, are widely unknown or under debate. One of the first concepts explicitly taking enzyme flexibility into account was the *induced fit* mechanism proposed by Koshland in 1958 (Koshland, 1958). Koshlands mechanism allowed the enzymes to undergo conformational changes during the binding process. The induced fit mechanism therefore implied that the same protein should adopt a different structure, whether a ligand is bound or not. This implicit prediction could be confirmed with the help of x-ray crystallography. For several ligand binding enzymes, such as insulin or dehydrofolate reductase, bound and unbound structures are known and shown to differ significantly. It is obvious that motions connecting these states are *essential* for function since they represent intermediate states in an catalytic cycle. Attempts have been made to classify these large-scale

motions (Gerstein *et al.*, 1994). However, their atomic description is usually unknown. The knowledge about this conformational flexibility stems from crystallography, providing only snapshots of rather stable intermediate conformations. The path between these intermediates, the processes and timescales involved in the necessary reorientations remain unsolved. Spectroscopic techniques such as Mössbauer-spectroscopy or neutron scattering on the other hand do provide such dynamic information, but the accessible timescales of such measurements are in the pico- to nanosecond range; more than three orders of magnitude faster than typical catalytic processes (microseconds to seconds). It is a surprising observation that for several enzymes a correlation between dynamical phenomena on the pico/nano-second timescale and enzyme activity has been observed. For these systems an intimate coupling between the fast degrees of freedom explored by spectroscopy and the slow biological important motions seems to exist.

Similar to structural biology, progress has been made in our understanding of the fast dynamics of biomolecules over the last decades. New concepts such as the rugged or complex energy-landscape or conformational substates have been developed or adjusted to account for the phenomena seen by spectroscopic techniques such as Mössbauer spectroscopy, X-ray or neutron scattering (Frauenfelder *et al.*, 1979, 1991). The observed phenomena show similarity to those attributed to complex systems such as glasses or polymer melts and make proteins a valuable testcase to investigate physical aspects of complex molecular systems (Angell, 1995; Green *et al.*, 1994; Iben *et al.*, 1989).

1.3 The energy landscape of proteins

The unifying concept connecting various aspects and fields of complex dynamical systems is that of a complex or rugged energy landscape. Introduced to the field of protein dynamics by Hans Frauenfelder (Frauenfelder *et al.*, 1991) in analogy to spin-glass systems, it serves as qualitative picture to interpret and connect experimental results and guide quantitative theoretical model building.

Generally speaking, a complex energy landscape is a high-dimensional[1] potential energy surface comprising a large number of distinct minima. Well known examples featuring such complex landscapes are spin-glasses; *i.e.* magnetic systems, where the interactions between the magnetic moments are *in conflict* with each other. A common example would be a dilute alloy of iron atoms in gold. The interaction between magnetic moments favor parallel or anti-parallel alignment depending on distance. As a consequence, there is not a unique ground state with all spins aligned in such a way as to locally minimize their own energy but a (high) number of similar minima with some spins (or clusters of spins) in unfavorable alignment with respect to each other. This phenomenon is known as "frustration". In proteins frustration arises from conflicting side-chain orientations, competing hydrogen bond donor-acceptor geometries and steric interactions (Stein, 1985).

[1]For a given number of atoms, the potential energy landscape has 3N-6 dimensions. Therefore *high-dimensional* is basically equivalent to *many-particle*. However, it should be noted that protein dynamics describes single proteins containing only a few thousand atom, much less than systems usually considered as complex or glassy systems (Angell, 1995; Green *et al.*, 1994)

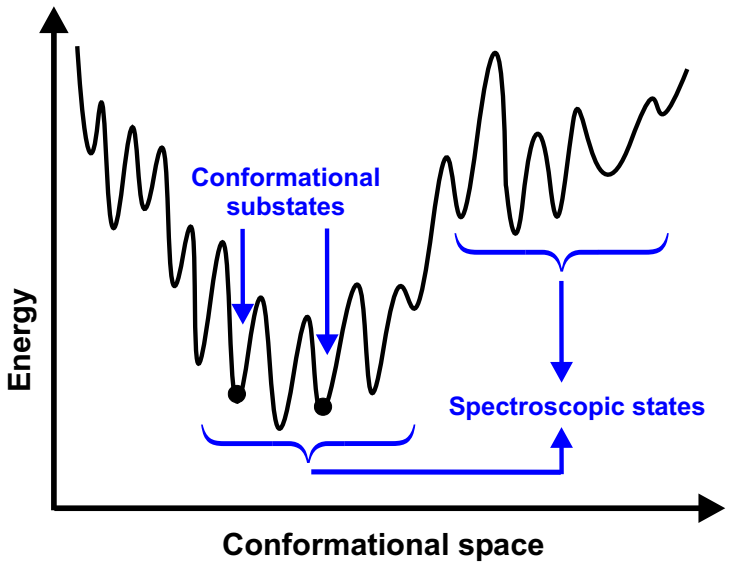


Figure 1.1: Energy landscape

Fig. 1.1 shows a schematic plot of a complex energy landscape. It is characteristic of complex systems that a large number of states with comparable energies exist separated by barriers of varying heights. The topology of this landscape determines the accessible conformational space as well as the dynamics exhibited to explore it. At sufficiently high temperatures, $k_B T \gg \langle H \rangle$, with $\langle H \rangle$ being the average barrier height, the system has enough energy to freely cross these barriers and *travel* through the landscape. All conformational space is accessible. With decreasing energy, however, higher energy states are no longer populated[2], the accessible conformational space decreases. Furthermore, it will take an increasing amount of time to overcome barriers between accessible states until the system finally drops out of ergodicity, *i.e.* until it is no longer able to sample the accessible conformational space within the timescale of observation. The glass transition is a well known example of such a drop out of ergodicity[3]. The topology of the energy landscape thus couples the dynamics of the system, mainly the realm of physical research, to the concept of conformational states used to explain biological function.

Dynamics in complex energy landscapes

In spite of the complexity and variability energy landscapes may show for different systems the dynamics exhibits characteristic features that find their substrate in heuristic laws applicable to most complex systems.

The dynamics of the system can be seen as traveling through the energy landscape, cross-

[2]Not with probability considerable greater than 0.

[3]The transition temperature (energy) from ergodic to non-ergodic, obviously depends on the timescale of observation. Such *timescale-dependence* is intrinsic to the definition of the glass transition

ing barriers to explore different minima. For a given barrier, the system will follow an exponential relaxation, determined by the barrier H between two minima:

$$f(t) \sim \exp(-\kappa t) \tag{1.1}$$

where $f(t)$, the relaxation function, describes the deviation of a system parameter from its initial value. In a complex landscape, however there will be a whole distribution of such barriers each leading to a slightly different time-dependence of relaxation processes. The relaxation function will thus be given by:

$$f(t) \sim \int_0^\infty \mathrm{d}H \; g(H) \exp(-\kappa(H)t) \tag{1.2}$$

where $g(H)$ denotes the distribution of barrier heights.

The correct form of $f(t)$ depends, of course, on the distribution $g(H)$, but it is possible to empirically represent the relaxation behavior of many complex systems by a stretched exponential or Kohlrausch-William-Watts (KWW) function:

$$f(t) = \exp\left[-(\kappa t)^\beta\right] \tag{1.3}$$

The whole complexity of the system condenses into two parameters, a characteristic frequency κ and a stretching factor β. The stretching factor is a measure of the deviation of the system from *normal* exponential behavior and can be seen as a measure of the systems complexity.

Although successful in reproducing the dynamical characteristics of relaxation processes, the KWW-function description is empirical. No strict theoretical relation between the energetics of a given system and the stretching factor β exists so far. A more rigorous, though more complicated approach to describe complex dynamics is the so called mode-coupling theory first proposed by Bengtzelius, Götze and Sjölander in 1984 (Bengtzelius *et al.*, 1984) (see Cummins (1999) for a recent review).

The relaxation frequency itself will depend not only on the barrier height H, but also on the temperature of the system, $\kappa = \kappa(H, T)$. For a single barrier κ is given by an Arrhenius law:

$$\kappa(H, T) = A \; \exp\left[-\frac{H}{k_B T}\right] \tag{1.4}$$

where A is the pre-exponential factor and k_B the Boltzmann constant.

For molecular reorientations the pre-exponential factor is typically of the order of $10^{-8} - 10^{-13} s^{-1}$. In complex systems trying to fit Eq. 1.4 to the temperature dependence of observed relaxation frequencies leads to unreasonable pre-exponential factors of $A \sim 10^{-30} - 10^{-40}$.

Again, agreement with experiment can be improved by empirical functions motivated by simple one-barrier dynamics. For many systems the temperature dependence of relaxation frequencies has been well reproduced by a Vogel-Fulcher law

$$\kappa(T) = A \ \exp\left[-\frac{C}{(T - T_0)}\right] \tag{1.5}$$

or equally well by a Ferry function

$$\kappa(T) = A \ \exp\left[-\left(\frac{E}{RT}\right)^2\right] \tag{1.6}$$

Conformational substates

A concept that distinguishes biological research from physics is the concept of *function*. Proteins are not merely macromolecules obeying the laws of quantum mechanics, but within this framework they fulfill a very specific task. Conformational space in physics is commonly described by the energy levels of the system where states with equal energy are considered as equal; *i.e.* the energy-levels are degenerate. For biological considerations, however, it is more important to distinguish those states of enzymes that are able to catalyse a given reaction from those that cannot. Those states may differ in energy, therefore being distinguishable in the above *physical* sense, but do not necessarily so. A mere change in the geometry or orientation, not necessarily connected to a (considerable) change in energy may well be enough to separate an enzyme from a *dead* macromolecule. A concept combining the energy landscape with biological function is that of conformational substates. In physical terms a conformational substate is a local minimum in the energy landscape. From a biological point of view such a state is either functioning or not. The concept of conformational substates therefore point out that a mere energetic point of view is not sufficient to characterize biological macromolecules. Further distinctions between states have to be made to explain the *biology* of these molecules.

Various theoretical and experimental studies have shown the existence and functional importance of conformational substates in biological macromolecules (Austin *et al.*, 1975; Elber & Karplus, 1987; Frauenfelder *et al.*, 1979; Schlichting *et al.*, 1994). Temperature dependent studies on the rebinding kinetics of ligands to the heme iron demonstrated that proteins do exist in, and are confined to conformational substates differing in their kinetic properties (Austin *et al.*, 1975; Frauenfelder, 1983). Direct measurements of atomic movements gave further proof of the existence of distinct substates in proteins and the underlying topology of the energy landscape. The most prominent phenomena in this respect, the dynamical transition will be discussed in more detail in the next section.

Hierarchy of conformational substates

The suggestion that proteins, just as other complex systems, should have a large number of similar states accessible at physiological conditions is counterintuitive. Although it was

recognized already in 1969 that a folded protein does not necessarily have to be in an absolute minimum of free energy (Levinthal, 1969), it still has to be in a *functioning state*. The lesson to learn from experimental and theoretical studies on protein dynamics is that the concept of a functioning state has to be replaced by a variety of heterogeneous functioning states (Frauenfelder *et al.*, 1991). The distinction between functioning and non-functioning can be made by the concept of an hierarchy of states separated by differences in free energy.

It is an intriguing question whether the energy landscape of proteins is a target of selection, *i.e.* whether the dynamics of proteins is in a non-trivial way *optimized* to its biological function. Suggestions were made that such a connection exists for thermophilic proteins, *i.e.* proteins adapted to function at high temperatures (Brunori *et al.*, 1999; Zaccai, 2000a). However, a precise connection between the energy landscape, its conformational substates and the biological function is not yet established.

1.4 The dynamical transition in proteins

A phenomenon showing the effect of conformational substates on the dynamics of proteins is the dynamical transition as shown in Fig. 1.2 (plot taken from Rasmussen *et al.* (1992)). Measuring the mean-square displacement of non-exchangeable hydrogens (neutron scattering), heme-iron atoms (Mössbauer spectroscopy) or heavy atoms (Debye-Waller factors in X-ray christallography) a linear increase in $\langle \Delta r^2 \rangle$ is observed up to a temperature $T_0 \sim 200$K. At the transition temperature T_0, the slope changes showing a sharp increase in atomic fluctuations. In analogy to a harmonic oscillator the low temperature linear behavior is called the harmonic regime. The sudden increase in fluctuation above T_0 is commonly seen as onset of anharmonic fluctuations. A qualitative connection to the energy landscape picture is readily made: below the transition temperature proteins are confined to a specific conformational substate, whereas above T_0 the protein starts to explore additional minima of the landscape thereby showing increased mean-square displacement.

The first protein shown to undergo a dynamical transition was myoglobin, sometimes called the "hydrogen atom" of biophysics (for a review of the work done on myoglobin see *e.g.* Frauenfelder & McMahon (2000); Frauenfelder *et al.* (1999)). Mössbauer spectroscopy revealed the transition at a temperature $T_0 \sim 180K$ by monitoring the mean-square displacement of the heme iron on a nano-second timescale (Bauminger *et al.*, 1983; Keller & Debrunner, 1980; Parak *et al.*, 1980, 1982). When neutron scattering became available for biological studies the dynamical transition could be demonstrated for a variety of globular proteins (Cusack & Doster, 1990; Daniel *et al.*, 1998; Doster *et al.*, 1989b; Fitter *et al.*, 1997), suggesting that the transition is a generic feature of proteins.

At the same time as neutron scattering experiments started to investigate protein dynamics, molecular dynamics simulations began to give insight into the atomic details of processes underlying the dynamical transition. Early simulations on BPTI suggested that dihedral transitions are correlated to the increase in mean-square displacement (Steinbach & Brooks, 1993) but are not their primary cause (Steinbach & Brooks, 1996). Kneller *et.al.* showed

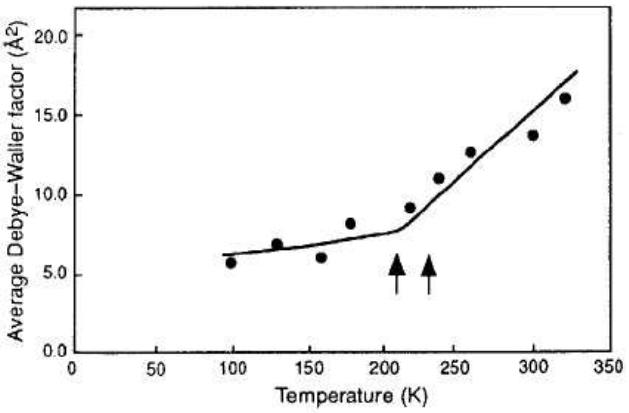


Figure 1.2: The dynamical transition; plot taken from Rasmussen *et al.* (1992).

the importance of the quasielastic scattering for a correct interpretation of the measured elastic spectrum and suggested a liquid like interpretation of protein picosecond dynamics (Kneller & Smith, 1994; Kneller *et al.*, 1992).

Activity and dynamical transition

Further interest in the dynamical transition arose from studies monitoring enzyme activity along with the mean-square displacement. These studies suggested a correlation between the onset of enzymatic activity and the increased fluctuations observed at the transition temperature T_0 (Austin *et al.*, 1975; Ferrand *et al.*, 1993; Parak *et al.*, 1980; Rasmussen *et al.*, 1992). The first hint towards such relations came from investigations on the temperature dependence of the rebinding kinetics of CO in myoglobin (Austin *et al.*, 1975). Here the transition temperature ($T_0 \sim 180$K) marks the onset of new processes in the binding kinetics. With the help of time resolved X-ray scattering, conformational substates corresponding to these processes could be identified (Ostermann *et al.*, 2000; Schlichting *et al.*, 1994).

In 1992 Rasmussen *et al.* reported the loss of activity below the transition temperature ($T_0 \sim 220$K) for the enzyme ribonuclease A (see Fig. 1.2)(Rasmussen *et al.*, 1992). A year later it was demonstrated that the completion of the bacteriorhodopsin photocycle is blocked upon cooling below the transition temperature ($T_0 \sim 200$K) (Ferrand *et al.*, 1993; Lehnert *et al.*, 1998).

Doubts on the generality of an activity-flexibility relationship were risen in Daniel *et al.* (1998) and Dunn *et al.* (2000). Here, the transition behavior of the proteins glutamate dehydrogenase and xylanase in solution were monitored along with activity measurements on the same system. Thus, whether the measured mean-square displacement, or more precisely the transition in the slope of $\langle \Delta r^2 \rangle$, is indeed related to enzyme activity remains an open question. Considering the variety of proteins and their functional mechanisms it

should not be surprising to find both kind of systems, those with functionality related to the transition and those with activity independent of T_0. The interesting question is, which underlying topology of the energy landscape couples the observed increase in mean-square displacement to the onset of function.

The suggested relation between activity and measured mean-square displacement is best discussed with the help of a simple two state model like the one depicted in Fig. 1.3. Two energies play an important role in this context. First, the difference ΔQ between timescale of transitions between these states. the states, determining the equilibrium distribution of conformational states and second, the energy barrier ΔH determining the

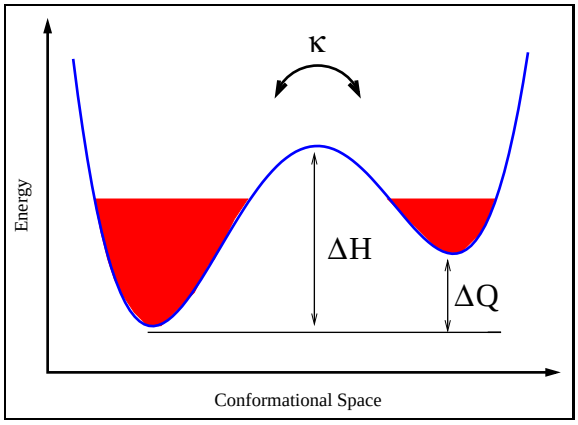


Figure 1.3: Asymmetric energy landscape model

The argument that the measured mean-square displacement is related to the activity goes as follows: The protein is in either of the two states. The lower state is the 'rigid', inactive state whereas the higher energy state is active. At temperatures $kT \ll \Delta Q$ the active state is not populated, the enzyme therefore inactive. With rising temperature the active state becomes populated, the enzyme starts to work[4].

The expected temperature dependence of the mean-square displacement for this system was shown in Ref. Keller & Debrunner (1980) to be consistent with Mössbauer spectroscopy data. The contribution of the change from one state to the other is proportional to a Boltzmann factor containing the activation enthalpy ΔQ:

$$\langle \Delta r^2 \rangle_c \sim e^{-\frac{\Delta Q}{k_B T}} \tag{1.7}$$

Here $\langle \Delta r^2 \rangle_c$ is commonly called the conformational contribution to the total mean-square displacement.

Since the increase in mean-square displacement in this model is caused by the change to an active state, this increase marks the onset of activity. Both, the onset of activity and the increase in mean-square displacement are controlled by the same temperature $T_0 \sim \frac{\Delta Q}{k_B}$. Measuring the temperature dependence of the mean-square displacement therefore gives access to the enthalpy difference ΔQ of functionally important states.

However, this does not take into account that spectroscopic techniques only detect motions in a given time-window determined by the instrument used. Whether or not a change in the mean-square displacement is seen for the system of Fig. 1.3 is determined not only by the occupation of the second state, but also by the timescale on which transitions between states occur. A given instrument will detect changes between states only if the barrier between

[4]The difference between the two states is not necessarily on/off. One can think of qualitatively different activity-temperature signature instead. This is the case for myoglobin, showing a transition from exponential to non-exponential rebinding kinetics

them is low enough that the rate of transitions is within the resolution of the instrument. Thus, ΔH controls whether a transition can be detected at a given instrument. Such an interpretation of the dynamical transition was suggested in Ref. Daniel *et al.* (2003b) based on measurements on proteins in a cryosolution showing no correlation between the dynamical transition and measured enzyme activity (Daniel *et al.*, 1998; Dunn *et al.*, 2000). A theoretical analysis of such an interpretation and consequences for the relation between protein function and the dynamical transition are discussed in Ch. 5.

Solvent dependence of protein dynamics

To understand protein dynamics one has to include the influence of the surrounding media. The effect of different hydration levels on the dynamical transition has been the subject of both experimental and theoretical investigations (Careri *et al.*, 1986; Diehl *et al.*, 1997; Ferrand *et al.*, 1993; Fitter *et al.*, 1996; Perez *et al.*, 1999; Smith *et al.*, 1987; Zanotti *et al.*, 1999a,b). Atomic fluctuations decrease significantly upon dehydration and neutron scattering results suggest that the dynamical transition is absent in dry protein powders (Doster *et al.*, 1990)[5]. Experimental studies on the dynamics of the hydration water demonstrated a dynamical transition of water dynamics at the same temperature as characteristic for proteins. A lively debate centers around the question whether the transition in hydration water dynamics *drives* the changes in protein internal dynamics, *i.e.* whether solvent dynamics *slaves* protein dynamics. This view is backed by recent results of computer simulations showing the strong influence of the solvent on internal dynamics of proteins (Tarek & Tobias, 2002; Tournier *et al.*, 2003; Vitkup *et al.*, 2000). It was suggested that the translational diffusion of water molecules on the surface of the protein is essential for the transition to occur. In contrast, earlier molecular dynamics simulations demonstrated the transition even for proteins in vacuum (Steinbach & Brooks, 1993). A solvent independent transition was experimentally observed by Mössbauer spectroscopy on oxymyoglobin solution compared to metmyoglobin crystals (Keller & Debrunner, 1980).

Theoretical considerations

Neutron scattering provides information on energetic and spatial properties of underlying relaxation processes. The topology of the energy landscape determines the scattering law. Comparing the measured incoherent spectra with theoretical considerations a direct connection to the underlying energy landscape can be made.

Several dynamical models have been proposed to explain neutron spectra of biological macromolecules. Doster *et al.* (1989b) proposed a two-state model as depicted in Fig. 1.3. This model is able to reproduce the elastic scattering function over a temperature range of $80 - 300K$, suggesting the existence of distinct substates separated by an activation enthalpy of $\Delta Q \sim 3$ kcal/mol. The energy barrier H between the two states was found to be surprisingly small (again ~ 3 kcal/mol). A similar model was proposed by Zaccai and Bicout, suggesting an *effective force constant* Bicout & Zaccai (2001). The protein is

[5] In agreement with the inactivity of dehydrated proteins

considered to change at the transition temperature T_0 from a *narrow* harmonic potential well to a *softer* energy well. In contrast to Doster *et al.* (1989b) the non-Gaussian wave vector dependence of the incoherent elastic scattering intensity is considered to be a result of a heterogeneous distribution of mean-square displacements (two regimes of $\langle \Delta r^2 \rangle$ in Lehnert *et al.* (1998)). The observed increase in quasielastic scattering at the transition has been explained by a simple diffusion model. Non-Gaussian scattering in this case can be assigned to a combined effect of wave vector dependence of a single sphere and the heterogeneity introduced by a distribution of accessible spheres (Dellerue *et al.*, 2001).

1.5 Protein association

So far mainly dynamic properties of isolated proteins were discussed. For biological function, however, it is common that proteins assemble to dimers, trimers or even oligomers. To understand these *binding* processes is of fundamental practical and theoretical importance in life sciences as well as medicine (Benkovic *et al.*, 1988; Gilson *et al.*, 1997; Klotz, 1985; Lamb & Jorgensen, 1997; Lian *et al.*, 1994; Wang & et al., 2001). As already mentioned binding processes between enzymes and substrate were among the first systems where dynamical aspects of proteins were recognized as being essential. Protein:ligand association has been assumed to be dominated by factors such as the hydrophobic effect, hydrogen bonding, electrostatic and van der Waals interactions. However, as early as 1963 it was suggested that an additional mechanism might exist: due to increased flexibility in the protein:ligand complex manifested by a change in the spectrum of vibrations due to formation of new, intermolecular interactions in the complex (Erickson, 1989; Finkelstein & Janin, 1989; Fischer *et al.*, 2001; Page & Jencks, 1971; Steinberg & Scheraga, 1963; Sturtevant, 1977; Tidor & Karplus, 1994). Theoretical normal mode analysis was used to estimate this vibrational change on insulin dimerization (Tidor & Karplus, 1994) and on water binding to bovine pancreatic trypsin inhibitor (Fischer *et al.*, 2001). These studies suggested that the effect is likely to be thermodynamically important. However, experimental determination of the vibrational change has been lacking. Inelastic neutron scattering, in which the dynamic structure factor $S(Q, \omega)$ is measured as a function of the scattering wave vector, Q, and energy transfer, $\hbar\omega$ has been used to determine the vibrational frequency distribution, $g(\omega)$, for several proteins (Cusack *et al.*, 1988; Doster *et al.*, 1989b; Smith *et al.*, 1986).

Chapter 2

Neutron scattering from proteins

Neutron scattering is a well established technique in solid state, polymer and recently also in biological physics (for reviews on neutron scattering on proteins see *e.g.* Smith (1991) and Gabe *et al.* (2002)).

Soon after the discovery of the neutron by Chadwick the theory of neutron scattering was well established (Van Hove, 1954). Although time- and lengthscale probed by neutron scattering techniques are well suited to examine biological systems, the first study dates back only to 1971 (Dahlborg & Rupprecht, 1971). In these first experiments Dahlborg and Rupprecht investigated DNA-fibers and one year later Middendorf and Willis presented experimental data on proteins and membranes (Middendorf & Willis, 1972).

Neutron scattering experiments were hampered by the high costs and the low flux of neutron sources at that time. The low flux demands high quantities of sample (∼300-600 mg) and long observation times (∼10 hours) to get sufficient counting statistics.

Only in the late 80's due to the advanced spectrometers build at the Institut Laue Langevin (ILL) biochemical inelastic neutron scattering (INS) and quasielastic neutron scattering (QENS) studies have been pursued more actively. Several groups started to investigate the dynamics of proteins with neutron spectroscopy (Cusack, 1986; Cusack *et al.*, 1986; Doster *et al.*, 1989a,b). At the same time computer simulations (Karplus & Petsko, 1990; McCammon *et al.*, 1977; Smith *et al.*, 1986, 1987) became an established tool in theoretical molecular biophysics (see Ch. 3).

This chapter starts with an introduction to neutron scattering theory as far as necessary to discuss specific problems in protein neutron scattering studies addressed in later chapters.

2.1 Theory of neutron scattering

Neutrons are uncharged particles of spin $\frac{1}{2}$ with mass $m = 1.922 \cdot 10^{-27} kg$. Having no charge, neutrons are unperturbed by the electric field of molecules and are scattered at the atomic core.

The energy, E, of a neutron is related to its wavevector $\mathbf{k}$ via:

$$E = \frac{\hbar^2 |\mathbf{k}|^2}{2m} \tag{2.1}$$

$$|\mathbf{k}| = \frac{2\pi}{\lambda} \tag{2.2}$$

where $\hbar$ is the Planck constant divided by 2π and λ is the wavelength associated with the wavevector $\mathbf{k}$.

For thermal and cold neutrons (see Table 2.1) Eqs. 2.1 & 2.2 yield a wavelength of $1-10$Å much larger than the typical lengthscale of nucleon-nucleon interactions ($\sim 10^{-3}-10^{-4}$Å). As a consequence, neutron-nucleon scattering is isotropic and can be described by a single parameter b, the so called scattering length. This is expressed by the use of a Fermi pseudo-potential, $\hat{V}(\mathbf{r})$, to describe the interaction of neutrons with a system of bound nuclei:

$$\hat{V}(\mathbf{r}) = \frac{2\pi\hbar}{m} \sum_l b_l \, \delta(\mathbf{r} - \mathbf{R}_l) \tag{2.3}$$

Here $\mathbf{R}_l$ denotes the position vector of atom l and b_l is its scattering length.

The scattering lengths b_l are in general complex numbers with the imaginary part representing neutron absorption. The real part of b_l can be either positive or negative and its value can differ largely from one isotope to another.

For a given element the mean-value taken over different isotopes and spin-states is referred to as the coherent scattering length b_{coh} whereas the variance of b is called the incoherent scattering length b_{inc} (see Sec. 2.3).

$$b_{coh} = \bar{b} \tag{2.4}$$

$$b_{inc} = \sqrt{\bar{b^2} - \bar{b}^2} \tag{2.5}$$

In App. 8.3 the scattering lengths and associated cross sections for the most important elements in biological applications of neutron scattering are listed.

The basic quantity measured in neutron scattering experiment is the double differential cross section, $\frac{\partial^2\sigma}{\partial\Omega\partial E}$, which is the number of neutrons scattered into the solid angle interval $[\Omega, \Omega + \Delta\Omega]$ with an energy transfer interval $[E, E + \Delta E]$.

$$\frac{d^2\sigma}{d\Omega dE} = \frac{\text{flux of neutrons in the solid angle d}\Omega \text{ per unit Energy}}{\text{flux of incoming neutrons}} \tag{2.6}$$

The scattering cross-section will now be derived for the simplified case where both, incident and scattered neutron state are described by plane wave solutions (contained in a large box of volume V):

	Energy [meV]
Cold	$0.1 - 10$
Thermal	$10 - 100$
Hot	$100 - 500$
Epithermal	≥ 500

Table 2.1: Neutron energy scales. Data taken from Lovesey (1987)

$$|k_i\rangle = \frac{1}{\sqrt{V}}\, e^{i\mathbf{k}_i\cdot\mathbf{r}} \tag{2.7}$$

$$|k_f\rangle = \frac{1}{\sqrt{V}}\, e^{i\mathbf{k}_f\cdot\mathbf{r}} \tag{2.8}$$

The derivation closely follows the one given in Lovesey (1987). A more detailed derivation in the framework of scattering theory can be found in Appendix A of the same book.

Fig. 2.1 shows a schematic description of the scattering event. An incident plane wave with wave vector $\mathbf{k}_i$ is scattered by the sample into the solid angle dΩ. The energy and momentum transfered to the sample during the scattering process are given by:

$$\hbar\mathbf{Q} = \hbar\mathbf{k}_i - \hbar\mathbf{k}_f \tag{2.9}$$

$$\hbar\omega = \frac{\hbar^2}{2m}(\mathbf{k}_i^2 - \mathbf{k}_f^2) \tag{2.10}$$

The cross-section for given initial and final states, $|k_i\lambda_i\rangle$ and $|k_f\lambda_f\rangle$, is given by:

$$\left(\frac{d^2\sigma}{d\Omega dE}\right)_{\lambda_i}^{\lambda_f} = \frac{W_{k_i,\lambda_i\rightarrow k_f,\lambda_f}}{j_0} \tag{2.11}$$

where $W_{k_i,\lambda_i\rightarrow k_f,\lambda_f}$ is the transition probability per unit time from state $|k_i\lambda_i\rangle$ to $|k_i\lambda_i\rangle$ and j_0 is the incoming flux:

$$j_0 = \frac{\hbar k_i}{Vm} \tag{2.12}$$

$W_{k_i,\lambda_i\rightarrow k_f,\lambda_f}$ can be calculated with Fermi's Golden Rule:

$$W_{k_i,\lambda_i\rightarrow k_f,\lambda_f} = \frac{2\pi}{\hbar}\ \left|\langle k_f\lambda_f|\hat{V}|k_i\lambda_i\rangle\right|^2\ \rho_{k_f}(E)\ \delta(\hbar\omega + E_{\lambda_i} - E_{\lambda_f}) \tag{2.13}$$

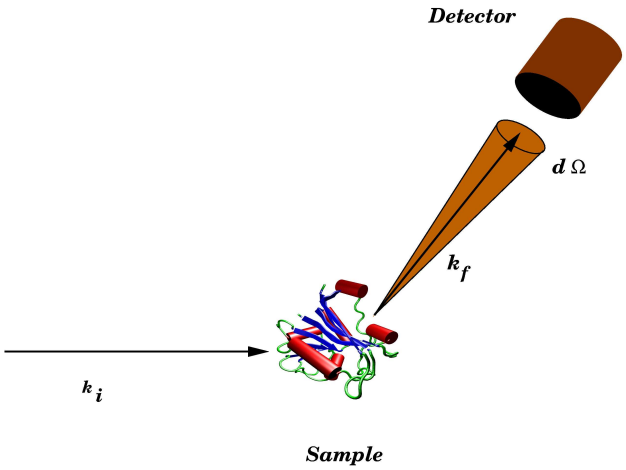


Figure 2.1: The scattering event

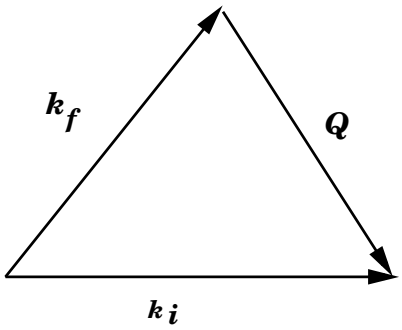


Figure 2.2: The scattering triangle

with $\rho_{k_f}(E)$ being the density of final neutron states.

$$\rho_{k_f}(E) = \frac{V}{(2\pi)^3}\frac{mk_f}{\hbar^2}d\Omega \tag{2.14}$$

$\hat{V}$ is the Fermi pseudo-potential Eq. 2.3. The delta function is put in Eq. 2.13 to ensure energy conservation for the scattering event.

Noting that

$$\langle k_f|\hat{V}|k_i\rangle = \sum_l b_l\ e^{i\mathbf{Q}\mathbf{R}_l} \tag{2.15}$$

the scattering cross-section is obtained by inserting Eqs. 2.13, 2.12, 2.14 and 2.15 into Eq. 2.11

$$\left(\frac{d^2\sigma}{d\Omega dE}\right)_{\lambda_i}^{\lambda_f} = \frac{k_f}{k_i}\left|\ \langle\lambda_f|\sum_l b_l\ e^{i\mathbf{Q}\mathbf{R}_l}|\lambda_i\rangle\ \right|^2 \delta(\hbar\omega + E_{\lambda_i} - E_{\lambda_f}) \tag{2.16}$$

Eq. 2.16 refers to given initial and final target states, $|\lambda_i\rangle$ and $|\lambda_f\rangle$. In a given experiment there will be a range of accessible initial states that can be assigned appropriate weights p_λ:

$$\sum_{\lambda_i} p_{\lambda_i} = 1 \tag{2.17}$$

The p_λ will be in general a thermodynamic factor $e^{-\frac{E_\lambda}{k_B T}}$ times a degeneration factor. In addition, there will be a range of accessible final states. Thus the observed partial differential cross-section finally reads:

$$\frac{d^2\sigma}{d\Omega dE} = \frac{k_f}{k_i} \overline{\sum_{\lambda_i \lambda_f} p_{\lambda_i} \left| \langle \lambda_f | \sum_l b_l \; e^{i\mathbf{Q}\mathbf{R}_l} | \lambda_i \rangle \right|^2 \delta(\hbar\omega + E_{\lambda_i} - E_{\lambda_f})} \tag{2.18}$$

The horizontal bar indicates an averages over all relevant parameters such as nuclear spin orientation, distribution of isotopes *etc.* Eq. 2.18 is known as the first Born approximation to the scattering cross-section.

2.2 Response function

As shown by van Hove (Van Hove, 1954) the scattering cross-section derived in the last section can be written as thermal average of operators belonging solely to the target. In this section the necessary steps to reformulate Eq. 2.18 in form of a correlation function are sketched.

Starting with an integral representation of the delta function

$$\delta(\hbar\omega + E_{\lambda_i} - E_{\lambda_f}) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \mathrm{d}t \; e^{-it\frac{\hbar\omega + E_{\lambda_i} - E_{\lambda_f}}{\hbar}} \tag{2.19}$$

the sum in Eq. 2.18 can be transformed as follows:

$$\begin{aligned}
&\sum_{\lambda_i \lambda_f} p_{\lambda_i} \left| \langle \lambda_f | \sum_l b_l \; e^{i\mathbf{Q}\mathbf{R}_l} | \lambda_i \rangle \right|^2 \delta(\hbar\omega + E_{\lambda_i} - E_{\lambda_f}) \\
&= \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \mathrm{d}t \; e^{-i\omega t} \sum_{\lambda_i \lambda_f} p_{\lambda_i} \langle \lambda_i | \sum_l b_l^* \; e^{-i\mathbf{Q}\mathbf{R}_l} | \lambda_f \rangle \\
&\quad \times \langle \lambda_f | e^{it\frac{E_{\lambda_f}}{\hbar}} \sum_{l'} b_{l'} \; e^{i\mathbf{Q}\mathbf{R}_{l'}} \; e^{-it\frac{E_{\lambda_i}}{\hbar}} | \lambda_i \rangle \\
&= \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \mathrm{d}t \; e^{-i\omega t} \sum_{\lambda_i \lambda_f} p_{\lambda_i} \langle \lambda_i | \sum_l b_l^* \; e^{-i\mathbf{Q}\mathbf{R}_l} | \lambda_f \rangle \\
&\quad \times \langle \lambda_f | e^{it\frac{\hat{H}}{\hbar}} \sum_{l'} b_{l'} \; e^{i\mathbf{Q}\mathbf{R}_{l'}} \; e^{-it\frac{\hat{H}}{\hbar}} | \lambda_i \rangle
\end{aligned} \tag{2.20}$$

The last term in Eq. 2.20 is now written in the Heisenberg representation of the operator $e^{i\mathbf{Q}\mathbf{R}_l}$:

$$e^{i\mathbf{Q}\mathbf{R}_l(t)} = e^{it\frac{\hat{H}}{\hbar}}\ e^{i\mathbf{Q}\mathbf{R}_l}\ e^{-it\frac{\hat{H}}{\hbar}} \tag{2.21}$$

Together with the completeness relation of the states $|\lambda\rangle$,

$$\sum_{\lambda} |\lambda\rangle\langle\lambda| = 1 \tag{2.22}$$

the cross-section Eq. 2.18 becomes:

$$\begin{aligned}
\frac{d^2\sigma}{d\Omega dE} &= \frac{k_f}{k_i}\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} \mathrm{d}t\ e^{-i\omega t}\ \overline{\sum_{l,l'}\sum_{\lambda} p_\lambda\ b_l^* b_{l'} \langle\lambda|\ e^{-i\mathbf{Q}\mathbf{R}_l}\ \ e^{i\mathbf{Q}\mathbf{R}_{l'}(t)}|\lambda\rangle} \\
&= \frac{k_f}{k_i}\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} \mathrm{d}t\ e^{-i\omega t}\ \sum_{l,l'} \overline{b_l^* b_{l'}} \left\langle\ e^{-i\mathbf{Q}\mathbf{R}_l}\ \ e^{i\mathbf{Q}\mathbf{R}_{l'}(t)}\right\rangle
\end{aligned} \tag{2.23}$$

The angular brackets represent a thermal average $\sum_\lambda p_\lambda\langle\lambda|(\dots)|\lambda\rangle$ and the last equality expresses the independence of this thermal average and averaging over isotopic distribution and nuclei spin states (horizontal bar) for the Fermi pseudo-potential.

2.3 Coherent and incoherent scattering

For systems were the average $\overline{b_l^* b_{l'}}$ is independent of the type of nuclei, *i.e.* independent of l, Eq. 2.23 can be further simplified. This is appropriate for mono-atomic samples or samples for which the scattering is largely dominated by one species, *e.g.* hydrogen rich molecules.

Writing

$$\overline{b_{l'}^* b_l} = |\overline{b}|^2 + \delta_{l,l'}\left[\overline{|b|^2} - |\overline{b}|^2\right] = \frac{\sigma_c}{4\pi} + \delta_{l,l'}\frac{\sigma_{inc}}{4\pi} \tag{2.24}$$

Eq. 2.23 now reads

$$\begin{aligned}
\frac{d^2\sigma}{d\Omega dE} &= \left(\frac{d^2\sigma}{d\Omega dE}\right)_{coh} + \left(\frac{d^2\sigma}{d\Omega dE}\right)_{inc} \\
&= N\ \frac{k_f}{k_i}\ \frac{\sigma_c}{4\pi}\ S_{coh}(\mathbf{Q},\omega) + N\ \frac{k_f}{k_i}\ \frac{\sigma_i}{4\pi}\ S_{inc}(\mathbf{Q},\omega)
\end{aligned} \tag{2.25}$$

with the coherent and incoherent scattering functions

$$S_{coh}(\mathbf{Q},\omega) = \frac{1}{2\pi\hbar N}\sum_{l,l'}\int_{-\infty}^{\infty} \mathrm{d}t\ e^{-i\omega t}\ \langle\ e^{-i\mathbf{Q}\mathbf{R}_l}\ e^{i\mathbf{Q}\mathbf{R}_{l'}(t)}\ \rangle \tag{2.26}$$

$$S_{inc}(\mathbf{Q},\omega) = \frac{1}{2\pi\hbar N}\sum_{l}\int_{-\infty}^{\infty} \mathrm{d}t\ e^{-i\omega t}\ \langle\ e^{-i\mathbf{Q}\mathbf{R}_l}\ e^{i\mathbf{Q}\mathbf{R}_{l}(t)}\ \rangle \tag{2.27}$$

Eqs. 2.26 & 2.27 state the scattering law in the impulse-energy representation. Fourier transforming them with respect to time and position leads to the intermediate scattering function, $I(\mathbf{Q},t)$ and the van-Hove function, $G(\mathbf{r},t)$, respectively:
Intermediate scattering function:

$$I(\mathbf{Q},t)_{coh} = \frac{1}{N}\sum_{l,l'}\langle\exp(-i\mathbf{Q}\mathbf{r}_l(0))\exp(i\mathbf{Q}\mathbf{r}_{l'}(t))\rangle \tag{2.28}$$

$$I(\mathbf{Q},t)_{inc} = \frac{1}{N}\sum_{l}\langle\exp(-i\mathbf{Q}\mathbf{r}_l(0))\exp(i\mathbf{Q}\mathbf{r}_{l}(t))\rangle \tag{2.29}$$

Van-Hove function:

$$G(\mathbf{r},t) = \frac{1}{N}\sum_{l,l'}\int d\mathbf{r}'\langle\delta(\mathbf{r}-\mathbf{r}'+\mathbf{r}_l(0))\delta(\mathbf{r}'-\mathbf{r}_{l'}(t))\rangle \tag{2.30}$$

$$G(\mathbf{r},t)_s = \frac{1}{N}\sum_{l}\int d\mathbf{r}'\langle\delta(\mathbf{r}-\mathbf{r}'+\mathbf{r}_l(0))\delta(\mathbf{r}'-\mathbf{r}_{l}(t))\rangle \tag{2.31}$$

Whether to use the scattering function, $S(\mathbf{Q},\omega)$, the intermediate scattering function, $I(\mathbf{Q},t)$, or the van-Hove function, $G(\mathbf{r},t)$, is a matter of convenience, since either of them contain the whole information accessible via neutron scattering. The choice will usually be dictated either by experimental setup or the point of view of theoretical investigation.

Although very similar in structure, there is a profound difference between coherent and incoherent scattering. Coherent scattering is due to the superposition of the scattering amplitudes of all nuclei. Coherent scattering is therefore determined by interference of these amplitudes. Just as in X-ray scattering this leads to strong geometrical constrictions of the scattered intensity. No such interference exists in incoherent scattering leading to a scattering profile more or less continuously distributed.

In fact looking at the differential cross sections for a perfect Bravais lattice one finds:

$$\left(\frac{d^2\sigma}{d\Omega}\right)_{coh} = \frac{\sigma_c}{4\pi}\left|\sum_l \exp(i\mathbf{Q}\mathbf{r_l})\right|^2 \quad (2.32)$$

$$\left(\frac{d^2\sigma}{d\Omega}\right)_{inc} = N\frac{\sigma_c}{4\pi} \quad (2.33)$$

The coherent part thus represents the well known Bragg scattering whereas incoherent scattering provides an isotropic background.

This strict isotropy is no longer valid for real crystals or other molecular systems. Nevertheless, incoherent scattering remains to be continuously distributed in space whereas coherent scattering shows strong dependence on the scattering vector **Q**. This dependence can be exploited to gain structural information about the system or determine dispersion phenomena.

As can be seen in Eq. 2.31 incoherent scattering is determined by the density autocorrelation function. The scattering profile is determined by the probabilities of individual atoms moving in space. Therefore, incoherent scattering is determined by the average over atomic movements in the protein.

For biological samples it is mainly the incoherent part of the spectra that contributes to the measured intensity. As can be seen in App. 8.3 the incoherent cross section of hydrogen is about 8-10 times larger, than any other biological relevant scattering cross section. Since about half the atoms of a protein are hydrogens, measured intensities are clearly dominated by incoherent scattering of hydrogen atoms (Bee, 1988; Hayward *et al.*, 2002; Smith, 1991). Unless advanced deuteration techniques are exploited (Reat *et al.*, 2000) it follows that neutron scattering of biological molecules is dominated by incoherent scattering which in turn is dominated by the average dynamics of hydrogen atoms.

The measurable quantities analyzed and explored in this thesis, the mean-square displacement and the vibrational density of states, are both accessible via the incoherent scattering function. Further analysis will henceforth be restricted to incoherent scattering bearing in mind that the actual measured spectra will be a valid estimate of the corresponding scattering function.

2.4 Elastic, quasielastic and inelastic scattering

The spectrum of scattered neutrons is commonly divided into three distinct parts, elastic, quasielastic and inelastic. From a physical point of view it is easy to distinguish elastic from inelastic scattering, *i.e.* neutrons that do not change their energy during the scattering event and those which do. In this definition quasielastic scattering is part of the inelastic spec-

trum. Nevertheless, for practical purposes it is useful to further discriminate between neutrons that scatter 'nearly' elastic (quasielastic), *i.e.* that loose or gain only small amounts of energy, and those who change arbitrary amounts of energy. As is clear from this definition, there is no sharp distinction between these two parts, however, it proves helpful in practice.

In Fig. 2.3 an incoherent spectrum is sketched (black) showing schematically the different contributions according to their energy scale; elastic scattering (blue), quasielastic (red) and inelastic (green).

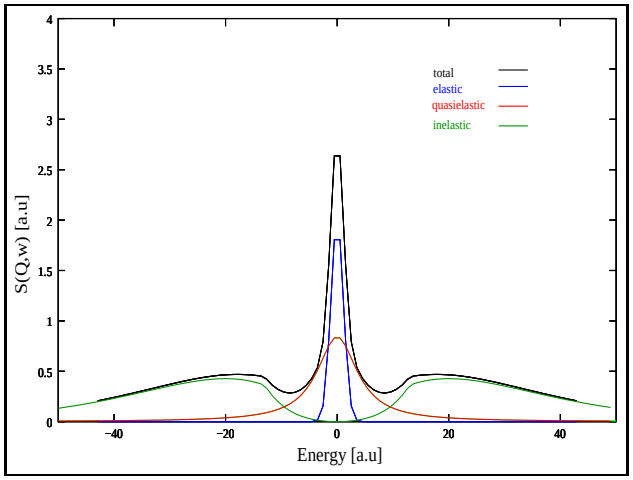


Figure 2.3: The scattering spectrum

In a strictly static system, all incoherently scattered neutrons would be scattered elastically. The more motions are present, the more neutrons are inelastically scattered. The ratio between elastically scattered neutrons to the total amount of scattered neutrons is therefore a measure of the overall amount of motions present in the sample. In biological systems the relatively high intensity of elastic neutron scattering is often used to obtain a first estimate of the dynamics present. The heights of the elastic peak can directly be related to the average atomic mean-square displacement (see Sec. 2.6) giving information about protein flexibility and dynamics. In any realistic experiment, the elastic scattering peak is broadened by the resolution of the instrument (as seen in Fig. 2.3).

Quasielastic scattering is usually considered as scattering with energy transfer of $\Delta E \leq 2\ meV$ ($\leq 16\ cm^{-1}$). This energy range corresponds to atomic motions over a time-scale of $10^{-10} - 10^{-12} s$, *i.e.* to the nano- to picosecond timescale. Motions in this timescale cause a broadening of the elastic peak. Processes leading to quasielastic scattering include rigid-body translational and rotational diffusion as well as sidechain reorientations such as methyl group rotation or diffusive motions of larger sidechains. For a thorough treatment of such processes and derivations of the associated scattering laws see Ref. Bee (1988). For biological samples quasielastic scattering is typically 3-4 orders of magnitude smaller than elastic scattering rendering a clear separation of both difficult. Ch. 5 will examine the consequences of this difficulty for the interpretation of the dynamical transition.

The inelastic region of the spectrum itself, *i.e.* energy transfer of $\Delta E \geq 2\ meV$ ($\geq 16\ cm^{-1}$), encodes the information about the energetics of atomic motions. At low temperatures these motions are given by harmonic vibrations around an equilibrium position. The scattering function itself is therefore determined by the phonon spectrum of proteins. In the *one phonon approximation* the inelastic part of incoherent neutron scattering can be directly related to the vibrational density of states (see Ch. 7). This relation can be used to access the thermodynamics of biologically relevant processes such as ligand binding.

2.5 Separation of motions

In complex systems various dynamical processes occur simultaneously. Different processes often have their own characteristic time- and/or lengthscales associated with them. Based on these time- and lengthscales different regions of the neutron spectra reveal information about different types of motions. As discussed in the previous section fast, vibrational motions are seen at corresponding high energies while slow, diffusive motions lead to a quasielastic broadening of the elastic peak. It is therefore good advice to separate the scattering law Eq. 2.27 into several components and study each of them separately. This can be achieved by splitting up the total displacement vector $\mathbf{r}$ according to contributions of different types of motions.

In regard to protein dynamics on the pico-nanosecond timescale, one usually addresses internal motions in contrast to rigid body motions of the molecule as a whole. This is expressed by writing the position vector $\mathbf{r}$ as

$$\mathbf{r}(t) = \mathbf{r}_{ext}(t) + \mathbf{r}_{int}(t) \tag{2.34}$$

where $\mathbf{r}_{ext}(t)$ and $\mathbf{r}_{int}(t)$ denote the external, rigid body and internal motions, respectively.

The external motions can be further divided into translations, *i.e.* motions of the center of mass, and rotations:

$$\mathbf{r}_{ext}(t) = \mathbf{r}_{trans}(t) + \mathbf{r}_{rot}(t) \tag{2.35}$$

The internal part is commonly subdivided into fast, vibrational motions and slower, *diffusive* dynamics:

$$\mathbf{r}_{int}(t) = \mathbf{r}_{vib}(t) + \mathbf{r}_{D}(t) \tag{2.36}$$

The term *diffusive* expresses that these types of motions lead to a quasielastic broadening of the elastic peak as described in the last section.

Assuming statistical independence of the external and internal motions the intermediate scattering function of a single scatterer factorizes into an external and internal part:

$$\begin{aligned} I(\mathbf{Q},t) &= \langle \exp(-i\mathbf{Q}\mathbf{r}(0)) \exp(i\mathbf{Q}\mathbf{r}(t)) \rangle \\ &= \langle \exp(-i\mathbf{Q}\mathbf{r}_{ext}(0)) \exp(i\mathbf{Q}\mathbf{r}_{ext}(t)) \rangle \langle \exp(-i\mathbf{Q}\mathbf{r}_{int}(0)) \exp(i\mathbf{Q}\mathbf{r}_{int}(t)) \rangle \\ &= I_{ext}(\mathbf{Q},t) \cdot I_{int}(\mathbf{Q},t) \end{aligned} \tag{2.37}$$

In the same fashion the internal scattering function can be factorized into fast, vibrational and slow, diffusive motions:

$$I_{int}(\mathbf{Q},t) = I_{vib}(\mathbf{Q},t) \cdot I_{D}(\mathbf{Q},t) \tag{2.38}$$

In frequency space the scattering function is then given by the convolution of the various scattering functions:

$$S(\mathbf{Q},\omega) = S_{ext}(\mathbf{Q},\omega) \otimes S_{vib}(\mathbf{Q},\omega) \otimes S_D(\mathbf{Q},\omega) \tag{2.39}$$

where the convolution of two functions is defined as:

$$g(\omega) \otimes h(\omega) = \int_{-\infty}^{\infty} d\omega' g(\omega') h(\omega - \omega') \tag{2.40}$$

2.6 The Gaussian approximation

Neutron scattering experiments can access the average atomic mean-square displacement of the sample via the Gaussian approximation. In this approximation the van-Hove self-correlation function, $G(\mathbf{r},t)_s$ is assumed to be a Gaussian function in the position vector, $\mathbf{r}$. This holds true for various limiting situations such as simple diffusion, Brownian dynamics, harmonic solids or a perfect gas. Given the variety of these limiting cases the Gaussian approximation is commonly taken as a reasonable first approximation.

In scattering studies on proteins and polymers it is used together with a classical limit of the scattering law, *i.e.* the position vectors, $\mathbf{r}(0)$ and $\mathbf{r}(t)$ are taken to commute at any time t.

Given both these assumptions, the incoherent intermediate scattering function Eq. 2.29 can be written as

$$I(\mathbf{Q},t) = I(\mathbf{Q},t)_{inc} = \frac{1}{N}\sum_l \langle \exp(-i\mathbf{Q}[\mathbf{r}_l(0) - \mathbf{r}_l(t)])\rangle = \frac{1}{N}\sum_l \exp(-\frac{1}{2}\langle \mathbf{Q}[\mathbf{r}_l(0) - \mathbf{r}_l(t)]\rangle^2) \tag{2.41}$$

where the last equality is the Gaussian approximation.

For isotropic samples this can be further simplified to yield

$$I(Q,t) = \sum_l \exp(-\frac{1}{6}Q^2[\mathbf{r}_l(0) - \mathbf{r}_l(t)]^2) \tag{2.42}$$

Taking the Fourier transform of Eq. 2.42 we obtain the Gaussian approximation for the elastic scattering amplitude:

$$S(Q,0) = \sum_l \exp(-\frac{1}{6}Q^2\langle \Delta r_l^2\rangle) \tag{2.43}$$

Thus, the elastic scattering is given by a sum of Gaussians with the atomic mean-square displacements as arguments. If the sample is homogeneous, *i.e.* if the atomic mean-square displacements can be considered equal, the averaging is trivial and the elastic scattering function reads

$$S(Q,0) = \exp(-\frac{1}{6}Q^2\langle\Delta r^2\rangle) \tag{2.44}$$

Eq. 2.44 is commonly taken as starting point to access the mean-square displacement, $\langle\Delta r^2\rangle$, from elastic neutron scattering. In Ch. 6 the influence of this last assumption on the measured mean-square displacement, $\langle\Delta r^2\rangle_{exp}$, will be examined.

Chapter 3

Computer simulations of proteins

Proteins are complex systems containing several hundreds to thousands of atoms. This excludes a quantum mechanical treatment of biological macromolecules[1] even with the high computational power nowadays available. To simulate large, complex systems, such as proteins, accuracy has to be sacrificed to a reasonable extend for computational efficiency. This is the aim of so called empirical force field methods or molecular dynamics. One of the basic assumptions underlying all molecular mechanics force fields is the validity of the Born-Oppenheimer approximation, *i.e.* the decoupling of electronic and nuclear degrees of freedom. The energy of the system is written as a function of the nuclear positions only. Furthermore the *true* energy function of the nuclei degrees of freedom is approximated by energy terms that can be more easily calculated, *i.e.* the force field is designed for computability.

The first molecular dynamics simulation dates back to the year 1957. Alder and Wainwright investigated phase transition behavior in a system of hard-spheres (Alder & Wainwright, 1957). It took, however, another twenty years for the first simulation of a protein to become feasible. In 1977 McCammon *et.al.* simulated the protein BPTI (Bovine pancreatic trypsin inhibitor, McCammon *et al.* (1977)) in vacuum for a timespan of $\sim 9ps$. Today molecular dynamics simulations are a common tool for theoretical studies on biological macromolecules.

The theory underlying molecular dynamics simulation can be divided into two main parts: the representation of the system under study and the description of its time evolution. In the following sections a short description of common methods and approximations suitable and/or necessary for simulations on large complex systems shall be given. A thorough treatment of the theory of molecular dynamics simulations would be beyond the scope of this thesis and can be found in several textbooks, *e.g.* Allen & Tildesley (1987); Leach (1996).

[1] Quantum mechanical treatment of parts of a protein, however, may well be feasible. The reaction center of an enzyme for example can be treated with higher accuracy, while the structural relaxation of the rest of the protein is modeled by empirical methods. This is the realm of MM/QM-Simulations (Molecular Mechanics/Quantum Mechanics) which will not be discussed here. A thorough treatment may be found in Leach (1996)

3.1 Molecular dynamics forcefield

Several force fields such as CHARMM (Brooks *et al.*, 1983) or Amber (Perlman *et al.*, 1995) have been designed to describe biological macromolecules. Although different in detail they share to a wide extent the basic structure of the energy terms considered as well as their functional form. The energy terms can be divided into two different classes; bonded and non-bonded interactions.

Bond stretching

The obvious bonded interaction term is the direct representation of covalent bonds between two atoms. The Morse potential (see Fig. 3.1) is a common form of this energy contribution, describing accurately the radial dependence up to the dissociation of the bonded atoms. In molecular dynamics simulations deviations from the minimum bond length are usually small, so that in most cases it is sufficient to account for covalent bonds by the harmonic approximation of the *true* potential:

$$V_{\text{bond}} = K_b(r - r_{eq})^2 \tag{3.1}$$

K_b is a force constant representing the strength of the bond and r_{eq} is often called the equilibrium bond length. Indeed, for an isolated pair of atoms the bond length vibrates with a frequency $\omega = \sqrt{\frac{K_b}{m}}$ around r_{eq}, m being the reduced mass of the system. However, for multi-atomic systems the equilibrium bond length is determined by the energy minimum of the total system which in turn is determined by a balance of all energy terms present, *i.e.* each individual energetic contribution is not necessarily at its minimum value.

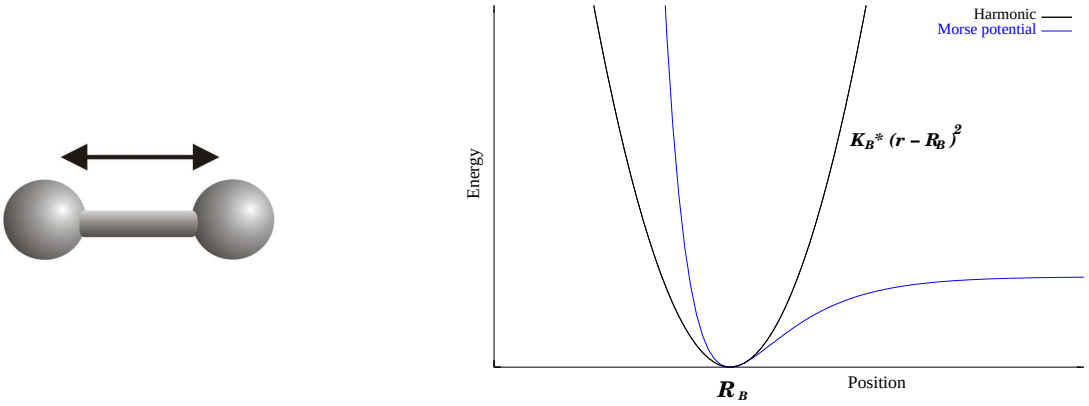


Figure 3.1: Bond stretching

In addition to the direct representation of covalent bonds a variety of other nearest neighbour interactions are included in the bonded energy terms, *e.g.* angle bending and torsional degrees of freedom.

Angle bending

Variations of binding angles (see Fig. 3.2) are represented by the harmonic approximation:

$$V_{\text{angle}} = K_\theta(\theta - \theta_{eq})^2 \tag{3.2}$$

Again K_θ is a measure of the energy required to deviate from the standard angle θ_{eq}. Force constants for angle bending are considerably smaller then those of bond stretching terms, reflecting the fact that the energy cost to bend angles away from their standard value is small compared to changes due to bond length vibrations. The remarks about equilibrium positions given for the bond stretching term are equally valid for angle bending.

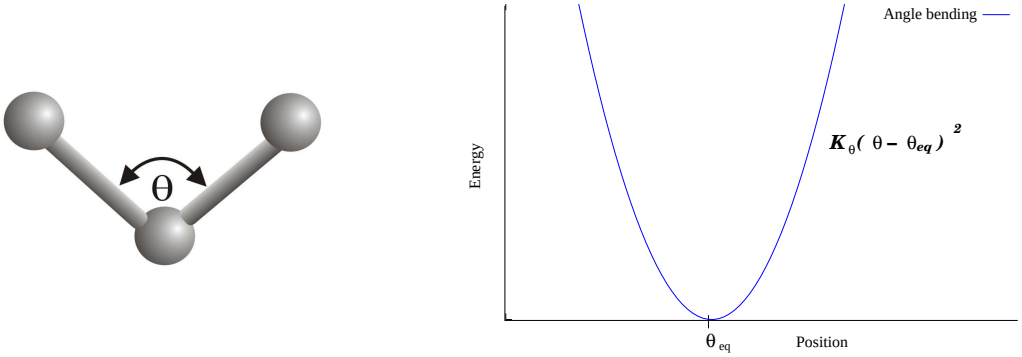


Figure 3.2: Angle bending

Torsion angles - Dihedrals

Bond stretching and angle bending are often referred to as hard degrees of freedom whereas torsion- and dihedral-terms constitute soft degrees of freedom. Despite their *softness* they are of utmost importance for a description of biological macromolecules. While the bond stretching and angle terms determine the local geometry of the system, the global three-dimensional structure and the slow, large-scale dynamics of proteins strongly depend on the balance of the torsional energy with the non-bonded interactions.

Torsional terms describe the rotation of atom groups around a given bond. The angle ϕ is given by the angle between the normals of the planes determined by atoms 1, 2 and 3 and 2, 3 and 4, respectively (see Fig. 3.3):

$$V_{\text{dihe}} = \sum_{\text{n}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \tag{3.3}$$

V_n is called the barrier of the torsion potential, n its multiplicity and γ its phase angle. In general one cosine function in Eq. 3.3 is sufficient to represent one torsional degree of freedom.

Additional torsional angles (see Fig. 3.4) are sometimes used to maintain the geometry of the system, *e.g.* the planarity of sp^2 hybridized carbon atoms. These *improper* dihedrals

are represented by a harmonic potential:

$$V_{\text{improp}} = K_\omega(\omega - \omega_{eq})^2 \tag{3.4}$$

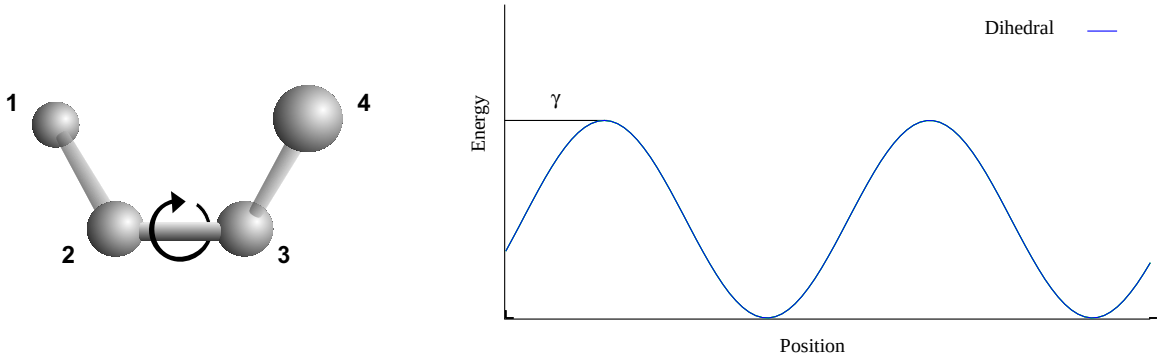


Figure 3.3: Torsional interactions

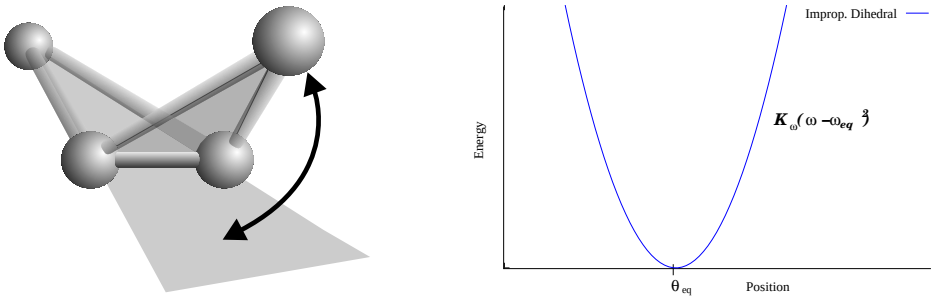


Figure 3.4: Improper dihedral

Van-der-Waals terms

Van-der-Waals or London forces are commonly described by a Lennard-Jones 12-6 potential, given by Eq. 3.5. Eq. 3.5 balances two terms, a short range repulsive and a slower decaying attractive force. The attractive R^{-6}-term arises from spontaneous dipoles inducing opposite dipoles in nearby atoms. The electric field E of a dipole varies as $E \sim R^{-3}$, which will in turn induce an dipole-moment $\mu \sim \alpha E$ at nearby atoms. The contribution to the potential energy of this induced dipole therefore varies as $\mu E \sim R^{-6}$.

No such direct physical justification can be given for the form of the short range repulsion. The origin of the repulsive interaction has its origin in exchange forces of overlapping electron wave functions. It is a consequences of the Pauli principle and its accurate treatment requires quantum mechanics. Indeed, quantum mechanical calculations suggest a form $\sim {}^1/_r$ for very short distances and $\sim \exp(-{}^{-2r}/_{a_0})$ for larger separations; a_0 is the Bohr radius. However, for force field methods an easy to calculate form of this repulsive interactions is needed. The R^{-12} dependence fulfills this requirement and has proven to give accurate

results in many circumstances.

$$V_{\text{vdw}} = 4\epsilon \left[\left(\frac{\sigma}{R}\right)^{12} - \left(\frac{\sigma}{R}\right)^{6} \right] \tag{3.5}$$

The Lennard-Jones potential is determined by two parameters: the collision parameter, σ, and the depth of the potential, ϵ. The collision parameter is the distance between two atoms at which the Van-der-Waals energy is zero.

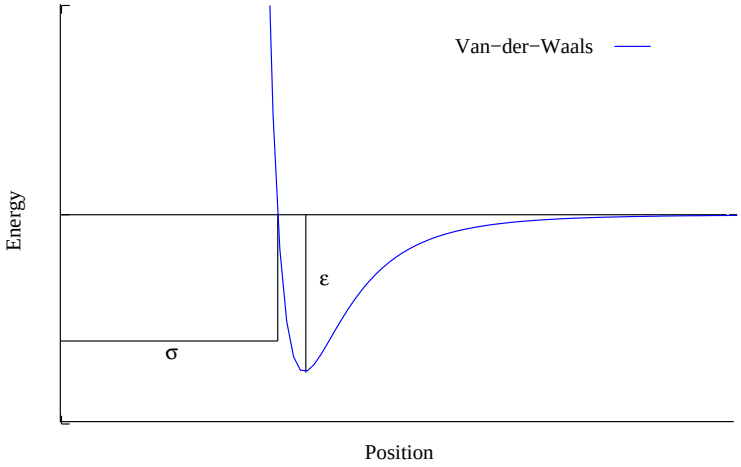


Figure 3.5: Van-der-Waals interactions

Electrostatic term

The forces between charged particles in the system are given by the Coulomb potential, Eq. 3.6. R_{ij} is the distance between atoms i and j, q_i, q_j the corresponding charges and ϵ_0 the dielectric constant in vacuum.

$$V_{\text{elec}} = \frac{q_i q_j}{4\pi\epsilon_0 R_{ij}} \tag{3.6}$$

The electrostatic forces are a textbook example of non-bonded interactions. From a computational point of view the main difference between the bonded and non-bonded interactions are the number of terms that have to be calculated for each atom. While there are only a few bonded terms to be calculated per atom, the non-bonded interactions in theory involve a sum over all atoms in the system. Therefore the total number of non-bonded terms to be calculated rises with the square of the system size ($\sim N^2$, where N is the number of atoms). However, the contribution of the non-bonded interactions to the energy function for atoms separated by large distances will be small, so that it is reasonable to introduce an artificial cutoff distance and calculate non-bonded interactions only for atoms within this distance. For long range potentials like the Coulomb law ($\sim \frac{1}{R}$) this may introduce artifacts according to the method used to cutoff the potential (Brooks, 1987; Brooks *et al.*, 1985). In Fig. 3.6 two such schemes are depicted. The *shift* method adds a constant to the electrostatic

potential, so that it becomes zero at a given distance r_{cutoff} and it is set to zero for longer distances. While computationally efficient this can introduce strong artificial forces at the cutoff distance. A second scheme is the *switch* mechanism. In this scheme to distances are specified r_{cuton} and r_{cutoff} and an interpolating function is introduced to smoothly change the potential such that potential and forces become zero at r_{cutoff}. The simulations analysed in this thesis used the second scheme for the non-bonded interactions.

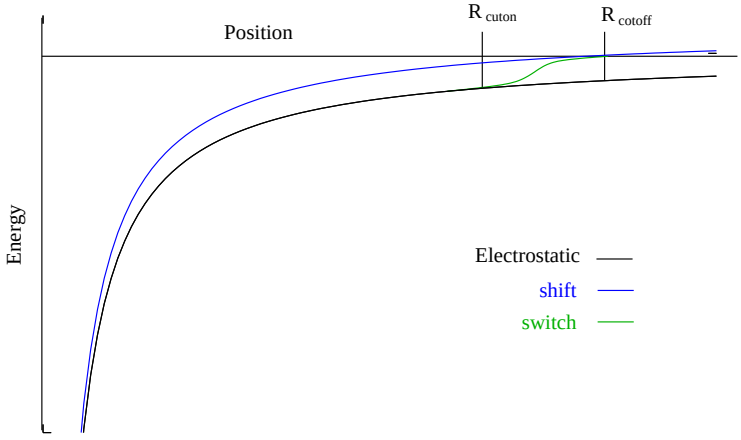


Figure 3.6: Electrostatic interaction

In summary, the total energy function for molecular dynamics simulations of proteins reads:

$$V_{\text{total}} = \sum_{\text{bonds}} K_b(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{\text{improper}} K_\omega(\omega - \omega_{eq})^2 + \sum_{i<j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}\right] + \sum_{i<j} \frac{q_i q_j}{\epsilon R_{ij}} \quad (3.7)$$

Depending on the package used or the system simulated additional terms may be added. Such terms can be used to assure correct geometry or to include additional effects, *e.g.* electrostatic multipoles to mimic polarization or Urey-Bradley terms to include coupling between angle bending and bond stretching.

3.2 Time evolution

In the previous section a potential energy function for biological macromolecules has been defined. In the next step the equations determining the time evolution of the system have to be chosen. In molecular dynamics simulations these are formulated within the realm of classical mechanics. The classical system can be characterized by its Hamiltonian function, $H(q_l, p_l)$ (Goldstein, 1991):

$$H(q_l, p_l) = K(q_l, p_l) + V(q_l, p_l) \quad (3.8)$$

where q_l, p_l are the generalized positions and momenta, respectively. $K(q_l, p_l)$ and $V(q_l, p_l)$ denote the kinetic and potential energy.

The equations of motions are given by[2]

$$\dot{q}_l = \frac{\partial}{\partial p_l} H(q_l, p_l) \tag{3.9}$$

$$\dot{p}_l = -\frac{\partial}{\partial q_l} H(q_l, p_l) \tag{3.10}$$

Extended systems molecular dynamics

Conventional molecular dynamics simulations produce trajectories in the micro-canonical ensemble, that is energy, volume as well as particle number are conserved. Experimental setup, however, most often demands the isobaric-isothermic ensemble, *i.e.* pressure and temperature kept constant. Andersen, Nosé and Hoover proposed methods to accomplish molecular dynamics simulation in this ensemble. The basic idea in their approach is to *extend* the system by additional degrees of freedom that represent the coupling to a heat or pressure reservoir.

As an example of this extended system approach the equations of motions for the isoenthalpic-isobaric ensemble are briefly discussed. Pressure control can be achieved by coupling the system to a fictitious piston and explicitly including the piston degree of freedom. This approach leads to a Hamiltonian of the form (Andersen, 1980):

$$H_{\mathrm{NHP}} = \sum_l V^{\frac{2}{3}} \frac{\tilde{p}^2}{2m} + V(V^{\frac{1}{3}}\tilde{r}) + \frac{1}{2} Q \dot{V}^2 + PV \tag{3.11}$$

where V, the Volume of the system, is introduced as an additional degree of freedom and Q represents the mass of the piston. $\tilde{p}$ and $\tilde{r}$ are scaled momenta and coordinates, respectively.

$$\tilde{r} = V^{-\frac{1}{3}} r \tag{3.12}$$

$$\tilde{p} = V^{-\frac{1}{3}} p \tag{3.13}$$

The corresponding equations of motion then read:

$$\ddot{\tilde{r}} = \frac{F}{m V^{\frac{1}{3}}} - \frac{2}{3} \dot{\tilde{r}} \frac{\dot{V}}{V} \tag{3.14}$$

$$\ddot{V} = \frac{P_{\mathrm{int}} - P_{\mathrm{ext}}}{Q} \tag{3.15}$$

[2]For Cartesian coordinates these equations reduce to Newtons equations of motions:

$$\ddot{\mathbf{r}} = -\frac{1}{m} \nabla_{\mathbf{r}} V(\mathbf{r})$$

The Hamiltonian point of view is taken here, since it is more common in the physical literature and incorporating constraints and temperature/pressure control is easier to accomplish.

where P_{ext} denotes the externally applied pressure and P_{int} is the instantaneous internal pressure of the system given by:

$$P_{\text{int}} = \frac{1}{\mathrm{d}V}\left[\sum_i \frac{\mathbf{p}_i^2}{m_i} + \sum_i \mathbf{r}_i\mathbf{F}_i - (\mathrm{d}V)\frac{\partial U}{\partial V}\right] \tag{3.16}$$

By coupling the system to a virtual piston continuous dynamical trajectories can thus be generated with a conserved Hamiltonian given by Eq. 3.11. The Hamiltonian can thus be used to check the stability of the simulation.

In a similar fashion temperature control can be incorporated by coupling the system to a heat bath. The isobaric-isothermal ensemble was shown to be realized by the following set of equations (see Tuckerman & Martyna (2000)):

$$\begin{aligned}
\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} + \frac{p_\epsilon}{W}\mathbf{r}_i \\
\dot{\mathbf{p}}_i &= \mathbf{F}_i - \left(1+\frac{1}{N}\right)\frac{P_\epsilon}{W}\mathbf{p}_i - \frac{P_\eta}{Q}\mathbf{p}_i \\
\dot{V} &= \frac{dVp_\epsilon}{W} \\
\dot{p_\epsilon} &= dV(P_{\text{int}} - P_{\text{ext}}) + \frac{1}{N}\sum_i \frac{\mathbf{p}_i^2}{m_i} - \frac{p_\eta}{Q}p_\epsilon \\
\dot{\eta} &= \frac{P_\eta}{Q} \\
\dot{p_\eta} &= \sum_i \frac{\mathbf{p}_i^2}{m_i} + \frac{p_\epsilon^2}{W} - (dN+1)kT
\end{aligned} \tag{3.17}$$

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} and P_{int} are defined as above.

Although these equations of motions are not hamiltonian, they do preserve a hamiltonian function given by:

$$H_{Hoover} = H(p,r) + \frac{p_\epsilon^2}{2W} + \frac{p_\eta^2}{2Q} + P_{\text{ext}}V + (dN+1)kT\eta \tag{3.18}$$

Again this function can be used to test system stability.

The Leap-frog integrator

The equations of motions presented in the last section can be integrated numerically to generate a trajectory representing the time evolution of the system. The simulation package CHARMM uses a leap-frog scheme for constant temperature/pressure simulations. The leap-frog algorithm is a variant of the well known Verlet integrator and share with the latter the property of being symplectic, that is it preserves phase space density. Symplectic

integrators in general show good energy conservation thereby guaranteeing stability of the integration scheme.

The Verlet algorithm follows from a Taylor expansion of the position vector at times $t+\Delta t$ and $t-\Delta t$:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\ \Delta t^2 + \cdots \tag{3.19}$$

$$\mathbf{r}(t-\Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\ \Delta t^2 + \cdots \tag{3.20}$$

Addition of these equations directly leads to the Verlet algorithm:

$$\mathbf{r_i}(t+\Delta t) = 2\mathbf{r_i}(t) - \mathbf{r_i}(t-\Delta t) + \mathbf{a}_i(t)\Delta t^2 \tag{3.21}$$

$$\mathbf{a}_i(t) = -\frac{1}{m_i}\nabla_{\mathbf{r_i}} V(\mathbf{r_1}(t), \ldots, \mathbf{r_N}(t)) \tag{3.22}$$

It should be noted, that the velocity canceled out by adding Eqs. 3.19 & 3.20. The velocity is not needed to calculate the trajectory of the system, it is, however, useful to access the kinetic energy or calculate velocity correlation functions. In the Verlet-scheme velocities have to be estimated *aposteriori* from the positions at different times. Another drawback of the original Verlet integrator are numerical imprecisions due to the addition of large terms (order $O(\Delta t^0)$) to small terms (order $O(\Delta t^2)$)

The leap-frog algorithm avoids these problems by introducing an intermediate step for calculating the velocities. The equations read:

$$\mathbf{r_i}(t+\Delta t) = 2\mathbf{r_i}(t) + \mathbf{v}(t+\frac{1}{2}\Delta t)\Delta t \tag{3.23}$$

$$\mathbf{v}_i(t+\frac{1}{2}\Delta t) = \mathbf{v}_i(t-\frac{1}{2}\Delta t) + \mathbf{a}_i(t)\Delta t \tag{3.24}$$

$$\mathbf{a}_i(t) = -\frac{1}{m_i}\nabla_{\mathbf{r_i}} V(\mathbf{r_1}(t), \ldots, \mathbf{r_N}(t)) \tag{3.25}$$

3.3 Analysis of molecular dynamics simulations

System setup

Molecular dynamics simulations have been used to test theoretical analysis of the dynamical transition (Ch. 4 and Ch. 5) as well as to access errors inherent in experimental data analysis (Ch. 6). The atomic trajectories used for further analysis were created by Jennifer C. Hayward.

The model system chosen consists of one BPTI molecule in 70%v/v CD_3OD/D_2O in an orthorhombic box with periodic boundary conditions. BPTI has 892 atoms and 4 internal water molecules. There are 658 CD_3OD and 661 D_2O molecules in the solvent, providing at least three solvent shells around the protein. This is adequate for simulating a protein in a bulk solvent environment. As in typical neutron scattering experiments, the exchangeable BPTI hydrogen atoms were replaced by deuterium, leaving 324 hydrogen atoms. The concentration of BPTI in the simulation is 156mg/ml. This system was chosen to model a solution of a protein in 70%v/v CD_3OD/D_2O cryosolvent, as has been used in the neutron experiments (Daniel *et al.*, 1998, 1999; Reat *et al.*, 2000) further analyzed in the present work.

The system was simulated using CHARMM (Brooks *et al.*, 1983) version 27 with all-atom parameter set 22 (Mackerell *et al.*, 1998). All water molecules were represented by the TIP3P potential (Jorgensen *et al.*, 1983). The simulations were performed in the NPT ensemble with a reference pressure of 1 atm.. A timestep of $0.001ps$ was used with SHAKE (Ryckaert *et al.*, 1977) applied to constrain bonds containing hydrogen or deuterium atoms. Non-bonded and electrostatic interactions were truncated at 13.0Å using the shifting function (Steinbach & Brooks, 1994).

Trajectories were created at 15 different temperatures, starting from 80K, increasing in steps of 20K to 220K, and further increasing in steps of 10K to 280K and a final trajectory at 300K. The starting structure for the lowest temperature 80K was the energy-minimized BPTI crystal structure (Parkin *et al.*, 1996), Protein Data Bank reference (1BPI) (Berman *et al.*, 2000b) equilibrated for 500ps in a fully equilibrated box of solvent. The starting structures for the rest of the simulations at increasing temperatures were the final structures from the preceding temperature. The systems at each temperature were equilibrated for 150ps and then data collected every 0.1ps for 520ps

System properties

In the context of the dynamical transition and its relation to enzyme function, it is the internal dynamics of protein atoms that is of most interest. Consequently, the protein trajectories were decomposed into external (whole-molecule diffusion) and internal components. The internal motions were extracted by superimposing every frame from the atomic trajectory with an RMS fit onto the first frame (Kneller, 1991).

The result is a new atomic trajectory of internal motions. In the present work only these internal motions are considered, *i.e.* all quantities are calculated from these new trajectories. This reflects the separation of motions in external and internal motions as discussed in Ch. 2.

Mean-square displacement

The mean-square displacement, or more precisely the change in the slope of $\langle \Delta r^2 \rangle$, is the main indicator of the dynamical transition. From molecular dynamics simulations the mean-square displacement, $\langle \Delta r^2 \rangle_{Sim}$ is readily calculated via an running average:

$$\langle \Delta r^2 \rangle_{Sim}(t) = \frac{1}{N} \sum_{\alpha} \langle (\vec{r}_\alpha(m) - \vec{r}_\alpha(0))^2 \rangle \approx \frac{1}{N} \sum_{\alpha} \left\langle \frac{1}{N_t - m} \sum_{k=0}^{N_t - m - 1} (\vec{r}_\alpha(k+m) - \vec{r}_\alpha(k))^2 \right\rangle \tag{3.26}$$

where α indicates atoms of the system and m, k and N_t are integer numbers with $t = m \cdot \Delta t$, Δt being the timestep of the simulation of total length $N_t \cdot \Delta t$.

Auto-correlation functions

Incoherent neutron scattering functions, as seen in Ch. 2, can be written in terms of auto-correlation functions. As in the case of the mean-square displacement, such autocorrelation functions are calculated via a running average as follows:

$$C_{AA}(t) = \langle A(\tau) \cdot A(\tau + t) \rangle \tag{3.27}$$

$$= \frac{1}{\tau_{max}} \sum_{\tau=1}^{\tau_{max}} A(\tau) \cdot A(\tau + t) \tag{3.28}$$

In the analysis of molecular dynamics simulations presented in this work, these autocorrelation functions were calculated with help of the program package nMOLDYN (Kneller *et al.*, 1995).

Chapter 4

Protein dynamics and neutron scattering

The incoherent scattering function resulting from internal motions of the protein bovine pancreatic trypsin inhibitor (BPTI) is calculated from MD trajectories for temperatures ranging from 100-300K. On the ∼200 picosecond timescale it is shown that there is considerable freedom in the interpretation of the intermediate scattering function of internal protein dynamics. The temperature dependence of $I(Q,t)$ is well reproduced whether the elastic incoherent scattering function (EISF) is considered to be temperature dependent or independent. A single exponential relaxation is shown to be a valid approximation for times $10 \leq t \leq 200ps$.

The guiding picture in interpreting protein dynamics is that of an energy landscape. The shape of this energy landscape determines the associated microscopic dynamics. For proteins the landscape is complex and rugged with many local minima. This situation is depicted in Fig. 4.1. It shows a van-der-Waals plot of a small protein (BPTI) zooming in on a single side-chain methyl group (alanine). The dynamics of methyl groups, due to their threefold symmetry can be described by the energy barrier separating the minima and the distance between these minima. The form of the scattering function is well known and can be used to derive a complete description of the energy landscape of this simple system. However, as shown in Fig. 4.1 a multitude of such *simple* models combine to constitute the complete energy landscape of the protein.

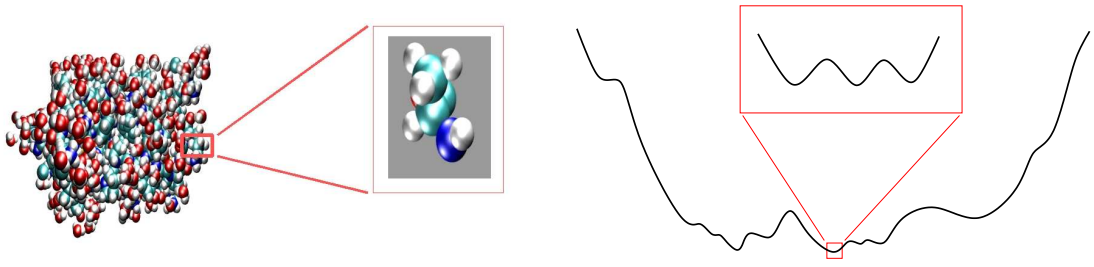


Figure 4.1: Dynamical Complexity of proteins

The heterogeneity of chemical compounds and the lack of symmetry in biological macromolecules lead to the presence of a wide range of vibrational and diffusive dynamical processes.

Naturally, the scattering law Eq. 2.23 remains valid, but the dynamics of different parts of the same system (atoms, residues, domains, *etc.*) can be qualitatively different from its neighbors, thereby rendering a *simple* formulation of the scattering law difficult. As mentioned in Ch. 1 this complexity is often condensed into heuristic laws containing only a few parameters, nevertheless yielding good agreement with experiment.

4.1 Internal protein dynamics

In Sec. 2.5 the scattering law for protein motions was split into different parts, rigid body rotation and translation, vibrational and quasielastic or diffusive. In this chapter the focus will be on the internal dynamics of proteins. Therefore the corresponding scattering law reads:

$$I(Q,t) = I_V(Q,t) \cdot I_D(Q,t) \tag{4.1}$$

The vibrational part, $I_V(Q,t)$ describes molecular vibrations of the system whereas the diffusive part contains relaxation processes.

The vibrational intermediate scattering function $I_V(Q,t)$ decays on timescales of $\sim 1ps$ or faster, so that for longer timescales one can approximate $I(Q,t)$ by

$$I(Q,t) = e^{-2W(Q)} I_D(Q,t) \tag{4.2}$$

where $e^{-2W(Q)}$ is the Debye-Waller factor.

The form of the diffusive part $I_D(Q,t)$ depends on the actual processes present but some general statements can be made. Since internal motions of a protein are restricted in space, $I_D(Q,t)$ decays to a finite value $I_D(Q,\infty) = A_0(Q) > 0$. Consequently Eq. 4.2 can be written as:

$$I(Q,t) = e^{-2W} \left(A_0(Q) + {I'}_D(Q,t) \right) \tag{4.3}$$

with ${I'}_D(Q,\infty) = 0$. $A_0(Q)$ is called the elastic incoherent scattering function and is given by:

$$A_0(Q) = \frac{1}{N} \sum_l \left| \langle \exp(i\mathbf{Q}\mathbf{r}_l) \rangle \right|^2 \tag{4.4}$$

The time dependent part ${I'}_D(Q,t)$ is subject to the condition

$$1 = I(Q,0) = A_0(Q) + I'_D(Q,0) \tag{4.5}$$

The complexity of internal protein dynamics leads to a non-exponential behavior of $I'_D(Q,t)$. This is taken into account by writing $I'_D(Q,t)$ as either a stretched exponential function or a sum of exponential relaxation processes.

The stretched exponential or Kohlrausch-William-Watts (KWW) function contains four parameters, the Debye-Waller term $\exp(-2W(Q))$, the EISF $A_0(Q)$, the stretching parameter β and one exponential parameter $\tilde{\kappa}$.

$$I(Q,t) = e^{-2W(Q)} \left(A_0(Q) + (1 - A_0(Q)) \ \exp[-(\tilde{\kappa}t)^{\beta}] \right) \tag{4.6}$$

The relaxation frequency is then given by:

$$\kappa = \left(\int_0^{\infty} \mathrm{d}t \exp[-(\tilde{\kappa}t)^{\beta}] \right)^{-1} = \tilde{\kappa} \frac{\beta}{\Gamma(\frac{1}{\beta})} \tag{4.7}$$

Here $\Gamma(x)$ dennotes the gamma function.

Describing $I(Q,t)$ as a sum of exponential relaxation functions introduces as parameters the Debye-Waller term $\exp(-2W(Q))$, the EISF $A_0(Q)$ and 2 parameters per exponential relaxation process, A_i and κ_i

$$I(Q,t) = e^{-2W(Q)} \left(A_0(Q) + \sum_i A_i(Q) \ \exp[-(\kappa_i t)] \right) \tag{4.8}$$

with the normalization condition

$$A_0(Q) = \sum_i A_i(Q) \tag{4.9}$$

Model descriptions with n exponential relaxation function thus provide $1 + 2n$ parameters.

4.2 Energy landscapes

Descriptions defined by either Eq. 4.6 or Eq. 4.8 pose several questions. The first question that arises is, whether these equations lead to a good description of the scattering profile. Once this question is answered in the confirmative, the next question would be whether the obtained parameters can be considered to be a faithful representation of the system, *i.e.* whether the parameters obtained are unique. Limitations in the accessible $(\mathbf{Q}, \omega)$ space, in both experiment and simulation, can make unequivocal interpretation of data using simplified descriptions difficult. It is therefore important to ask which alternative sets of parameters could describe the same set of data and how such alternative descriptions alter the interpretation of derived quantities like the mean-square displacement of the system.

It was pointed out in Ch. 1 that the concept of conformational substates forms a bridge between physical descriptions of protein dynamics in terms of energy landscapes and biological function of the enzyme. It is therefore important to understand how populations

of different substates and transitions between them manifest themselves and influence neutron scattering data. From Eq. 4.4 it can be seen that the elastic incoherent scattering function is determined by the equilibrium distribution of position vectors. The EISF is zero for liquids and nonzero for spatially constraint systems. Changes in the equilibrium distribution of coordinates, *e.g.* due to changes in the temperature, alter the average taken in Eq. 4.4 and therefore leave their traces in the EISF. In this sense it is a measure for the conformational space explored by the system. Considerations relating conformational substates to biological function are thus intimately related to the EISF and its changes with temperature (Doster *et al.*, 1989b; Frauenfelder & McMahon, 2000; Zaccai, 2000a,b).

In this respect the two qualitatively different situations depicted in Figs. 4.2 & 4.3 should be considered. The pictures show a symmetric and an asymmetric two-well potential each representing two conformational states separated by a barrier ΔB. The qualitative difference between both, the gap in energy between the substates in the asymmetric potential, shows up as qualitative difference in the EISF. The scattering function for these two energy landscapes is known and *e.g.* derived in Bee (1988)[1]:

$$I(Q,t) = e^{-2W(Q)} \left(A_0(Q) + A_1(Q) e^{-\kappa t} \right) \tag{4.10}$$

with

$$A_0(Q) = \frac{1}{(1+\rho)^2} \left[1 + \rho^2 + 2\rho j_0(Qr) \right] \tag{4.11}$$

$$A_1(Q) = \frac{2\rho}{(1+\rho)^2} \left[1 - j_0(Qr) \right] \tag{4.12}$$

$$\tag{4.13}$$

$j_0(Qr)$ is the zeroth order spherical Bessel function and ρ denotes the relative population of conformational substates with conformational coordinate $r_i; i = 1, 2$.

$$\rho = \frac{p(r_1)}{p(r_2)} = \exp(-\frac{\Delta G}{RT}) \tag{4.14}$$

where ΔG is the difference in free energy as shown in Fig. 4.3, R is the gas constant and T the temperature and $p(r_i)$ denotes the population of the energy well with conformational coordinate r_i.

In the symmetric case $\Delta G = 0$ and thus $\rho = 1$. Inspection of Eq. 4.11 shows that the symmetric potential has an EISF constant with respect to temperature, whereas the asymmetric landscape exhibits changes of the EISF upon temperature change. Even more, for the asymmetric landscape changes in the elastic incoherent scattering function should be expected at a temperature of $RT \approx \Delta G$. The gap in free energy thus defines a transition temperature intrinsic to the system. In contrast, no such distinct temperature exists for the symmetric landscape. Regarding the dynamical transition observed in protein dynamics (cf Ch. 1) it is tempting to use this qualitative difference to explain and interpret experimental

[1] The scattering function is known in the limit of jump models. Here the timescale of crossing the barrier is assumed to be fast compared to the average time between successive crossing events.

data (for recent reviews see Gabe *et al.* (2002); Parak (2003)). However, it will be shown in this chapter that neutron scattering data calculated from MD simulation of internal motions of the globular protein BPTI can equally well be described within the symmetric scenario, *i.e.* without changes in the EISF. The consequences of this ambiguity for the interpretation of the dynamical transition as measured by neutron scattering will be discussed in Ch. 5.

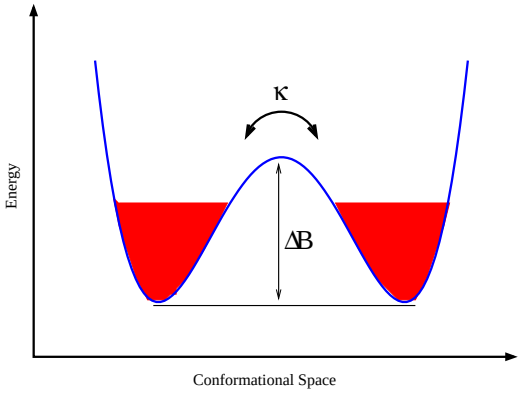


Figure 4.2: Symmetric double well potential

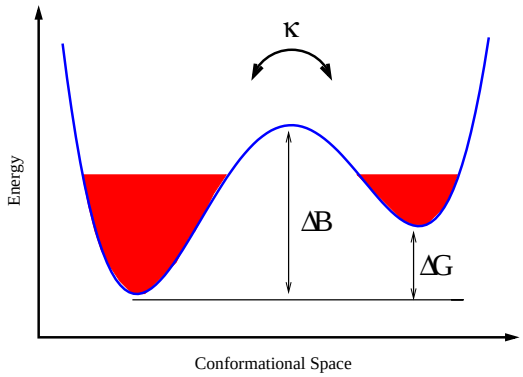


Figure 4.3: Asymmetric double well potential

4.3 Stretched exponential relaxation

The most prominent difference between models derived from the landscapes shown in Figs. 4.2 & 4.3 is the temperature dependence of the EISF. To test whether an analysis in the low Q-range, determining the mean-square displacement of the system, is able to distinguish both landscapes, both scenarios were used to analyze neutron spectra calculated from MD simulations.

The two columns of Fig. 4.4 show a comparison of a description of the spectra by a Kohlrausch-Williams-Watts function with a) all four parameters allowed to vary and b) $A_0(Q)$ kept fixed.

The intermediate scattering function, $I(Q,t)$, was calculated from MD trajectories of internal protein motions as described in Sec. 3.3. Each graph shows $I(Q,t)$ at the indicated temperatures for Q-values ranging from $0.2 - 1.6$Å (symbols).

The left column shows fits of Eq. 4.6 with all four parameters allowed to vary whereas the right column shows the resulting fits of the restraint model.

The fixed amplitude $\bar{A}_0(Q)$ was determined from the amplitudes of the *unrestraint* fit as follows: the decay of the intermediate scattering function is more pronounced at high temperatures than at low temperatures, allowing a better extraction of $A_0(Q)$ at high temperatures. The fixed EISF for the second model was therefore obtained by averaging $A_0(Q)$ over the temperatures above 250K (see Fig. 4.5). $\bar{A}_0(Q)$ thus represents an estimate of the long-time, converged EISF.

The timerange of $1 \leq t \leq 200ps$ is equally well reproduced by both models. Keeping the EISF of the system constant with temperature has no effect on the reproducibility of obtained data. Their interpretation, however, bears some important differences.

To understand these differences it is instructive to look at the two limiting situations, $t \to 0$ and $t \to \infty$. $t \to 0$ here means times short compared to the lowest characteristic relaxation time, $1/\kappa$, but still long enough for vibrational motions to be resolved. In this limit, the vibrational motions determine the intermediate scattering function and $I(Q,t)$ is given by the Debye-Waller factor, $e^{-2W(Q)}$. For times long enough to sample all relaxation processes of the system ($t \to \infty$), $I(Q,t)$ is given by $I(Q, t \to \infty) = e^{-2W(Q)} A_0(Q)$ *i.e.*, the product of the Debye-Waller factor and the EISF.

Looking now at the temperature dependence of the scattering function both models agree on the fast timescale, where this dependence is solely determined by the Debye-Waller factor. However, in the limit $t \to \infty$ both models differ. While in the constraint model (EISF fixed) the temperature dependence is still determined by the Debye-Waller factor, an additional dependency is introduced for the free model via the temperature dependence of the EISF. Both models therefore disagree in there long time behavior.

The question then arises, why both models describe the data in Fig. 4.4 accurately? The scattering function clearly shows temperature dependence beyond that given by the Debye-Waller factor, nevertheless this dependency is well reproduced in both models. The additional information needed to distinguish between both descriptions is the timescale of motions present in the system. For complex systems like biological molecules we cannot easily decide whether simulated or observed timespans are long enough to justify a description of the system in the limit $t \to \infty$. Looking again at Eq. 4.6 reveals that at intermediate timescales the interpretation of the temperature dependence of $I(Q,t)$ is complicated by the fact that the relaxation frequency κ is expected to depend on temperature as well. Thus, at intermediate timescales two scenarios lead to a temperature dependency of $I(Q,t)$. First, the EISF itself can be temperature dependent. As discussed, this is related to an increase in conformational space explored. This possibility will be called the 'equilibrium model'.

Alternatively, the temperature dependence of $I(Q,t)$ can be reproduced by temperature dependent relaxation frequencies. The scattering function exhibits notable decrease due to a given relaxation process only on timescales longer than their typical relaxation time. In this scenario, changes in $I(Q,t)$ upon temperature thus depend on the interplay of observation time and relaxation frequency. This scenario will henceforth be called the 'frequency window' interpretation.

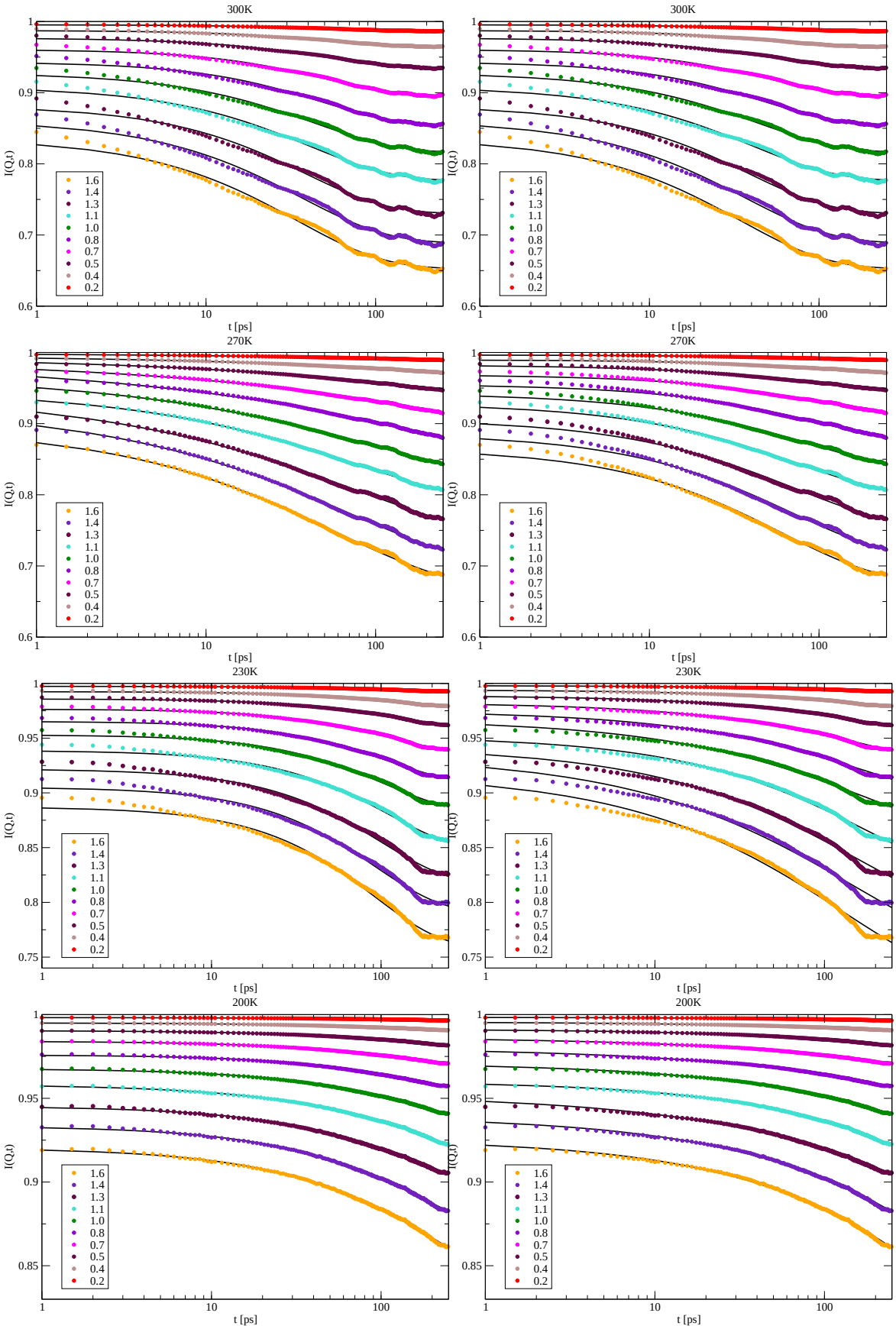


Figure 4.4: The intermediate scattering function, $I(Q,t)$, calculated from internal motions of BPTI over a Q-range of $0.2 - 1.6$Å (symbols). solid lines: fits of Eq. 4.6 to $I(Q,t)$ with a) all parameters free to vary (left) and b) $A_0(Q)$ fixed (right)

System parameters

The three free parameters of the frequency window model reproduce the calculated spectra as well as the four parameters of the full KWW function. Moreover, a comparison of the remaining three parameters reveals that the effect of the choice of the frequency window model is mainly to restrict the observed variation of the resulting three parameters, *i.e.* in the frequency window model the remaining parameters are better defined. The qualitative behavior of these parameters does not change and even quantitively both models are comparable.

Fig. 4.5 depicts the amplitudes $A_0(Q)$ obtained from Eq. 4.6 for temperatures $T \gtrsim 250$K (symbols). The black curve marks $\bar{A}_0(Q)$ as described above. For temperatures between $200 - 240$K $A_0(Q)$ is plotted in Fig. 4.6. The errorbars in both plots are given by the statistical errors of the fitting procedures. Inspection of Figs. 4.5 & 4.6 reveals that the EISF as fitting parameter is well defined for temperatures $T \gtrsim 250K$, however, errorbars become large at $200K$ rendering the amplitudes meaningless for temperatures $T \lesssim 180$K.

Fig. 4.5 suggests a general decrease of $A_0(Q)$ with temperature with 250K being an exception. Such trend would be expected from an equilibrium scenario since with rising temperature, more of conformational space is explored. No such trend, however, can be seen in Fig. 4.6 for temperatures below 250K. Considering the large errorbars at low temperatures this suggests that $A_0(Q)$ is not unambiguously defined by the calculated spectra.

The reason for this ambiguity in the amplitude can be understood by inspection of Fig.

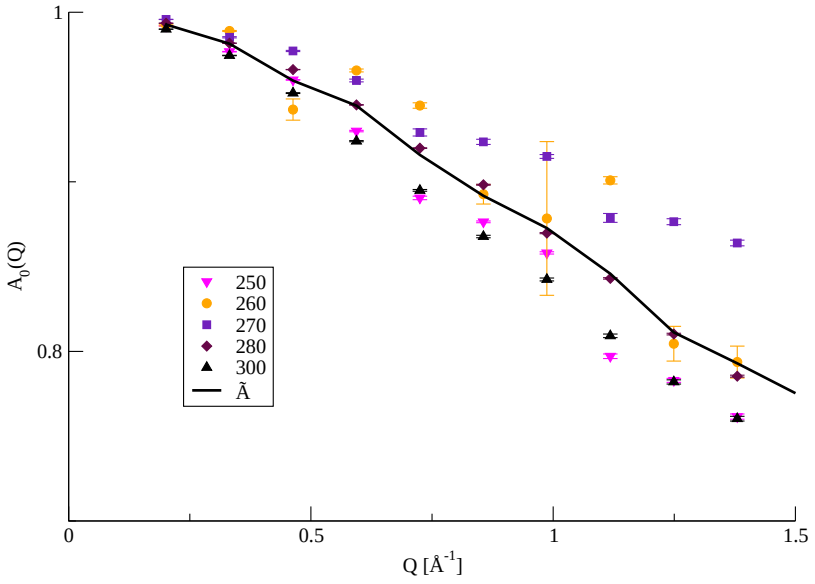


Figure 4.5: Elastic incoherent scattering factor, $A_0(Q)$, determined for $T \geq 250$K by fitting Eq. 4.6 to $I(Q,t)$ (symbols). The black line depicts the average amplitude, $\bar{A}_0(Q)$

4.7. Fig. 4.7 shows a comparison of the relaxation frequency κ for both models. κ was

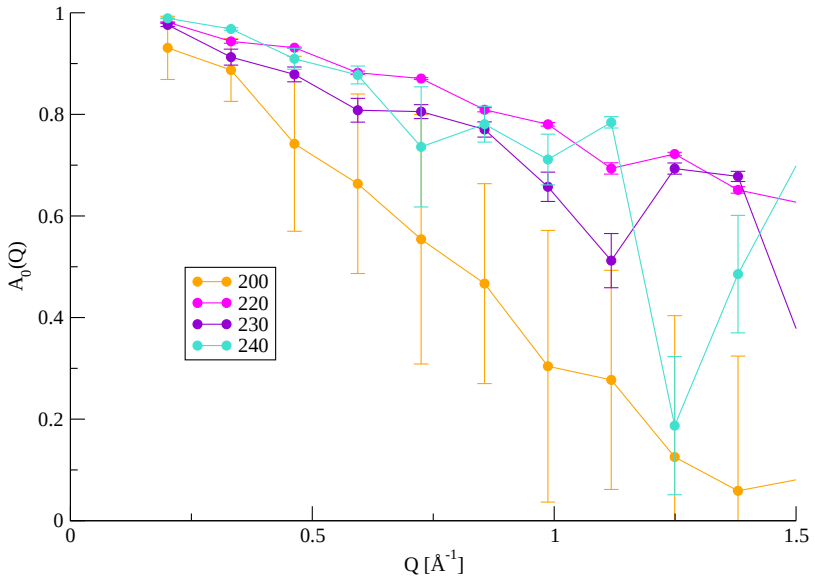


Figure 4.6: Elastic incoherent scattering factor,$A_0(Q)$, determined for $200 \leq T \leq 240$K by fitting Eq. 4.6 to $I(Q,t)$ obtained from MD simulations.

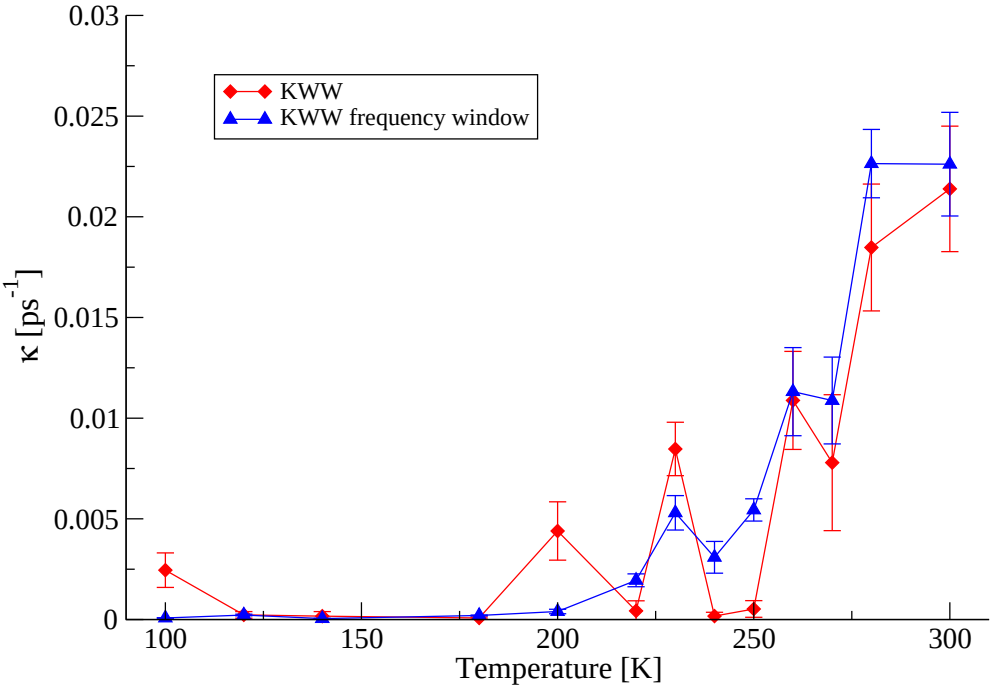


Figure 4.7: The relaxation frequency κ as determined by Eq. 4.7. (○) κ obtained by varying all four parameters. (△) κ obtained for the frequency window model ($A_0(Q)$ fixed).

calculated from Eq. 4.7 and averaged over all Q values. The errorbars were calculated from the variance of κ with respect to different Q-values. Both models draw the same qualitative picture of the temperature dependence of relaxation processes. At low temperatures, $T \leq 200K$, κ is small, $\kappa \leq 0.0025ps^{-1}$, thus the typical relaxation timescale $1/\kappa$ is of the same magnitude as the total simulation time. Relaxation due to processes related to $A_0(Q)$ are to slow to be observed in the simulations. Therefore, the amplitudes are indetermined in the fitting process. Above 200K there is a general sharp increase in κ for both models indicating either the onset (equilibrium) or speeding up (frequency window) of relaxation processes. As fitting parameter, however, the relaxation frequency, κ, and the amplitude $A_0(Q)$ are not independent of each other. A given value of $I(Q,t)$ can be reproduced in two ways, either by adjusting the amplitude making sure κ is large enough for the amplitude to be resolved, or by chosing $A_0(Q)$ and adjusting the frequency such that the amplitude is only partly resolved. For example, the larger decay of $A_0(Q)$ at $T = 250K$ compared to higher temperatures, is compensated for by a small relaxation frequency κ. The resulting intermediate scattering function, $I(Q,t)$ thus shows in effect the expected decrease with increasing temperature. Fixing the amplitude in the frequency window scenario removes this ambiguity and thus leads to a smoother temperature dependence of κ as seen in Fig. 4.7.

A similar *improvement* of parameters in the frequency model can be observed for the stretching parameter, β (Fig. 4.8), as well as for the vibrational mean-square displacement derived from the Debye-Waller factor, $\langle \Delta r^2 \rangle_{Vib}$ (Fig. 4.9). The stretching parameter in Fig. 4.8 was obtained by averaging over all Q values. The errorbars, as in Fig. 4.7 indicate the variation of the fitting parameter with Q. In agreement with experimental values, β varies between 1 and 0.4 (Nienhaus *et al.*, 1997). As mentioned in Ch. 1, β is a heuristic factor. Its dependence on temperature is in general not known. In the analysis here β fluctuates around a mean value of 0.7 for both models, the fluctuations are considerable smaller in the frequency window model. Again this has to be attributed to the number of free parameters. Neither model shows systematic variation with temperature.

The vibrational mean-square displacement depicted in Fig. 4.9 were calculated from the Debye-Waller factors in the Gaussian approximation, *i.e.* by fitting a straight line to the plot of $-2W(Q)$ *vs* Q^2 (compare Ch. 2). The errorbars represent the statistical error of the fit. Representing fast, vibrational motions of the system, $\langle \Delta r^2 \rangle_{Vib}$ is expected to rise linearly with temperature. As seen in Fig. 4.9 this is fulfilled for both descriptions up to 280K, again smoother for the frequency window scenario. The increased fluctuations between 220K and 260K can once more be attributed to the interdependencies of the fitting parameters contributing to $I(Q,t)$ in Eq. 4.6.

The stretched exponential function accurately reproduces the scattering function. However, the four fitting parameters already show some interdependence and even redundancy. There is enough leeway in the choice of parameters to hold the EISF constant with temperature and furthermore the stretching parameter β shows no systematic temperature dependence. It can be asked whether a single exponential relaxation function would be sufficient to represent the data. In addition it is interesting to see whether system properties like the relaxation frequency or $\langle \Delta r^2 \rangle_{Vib}$ do depend on the description chosen. These questions will

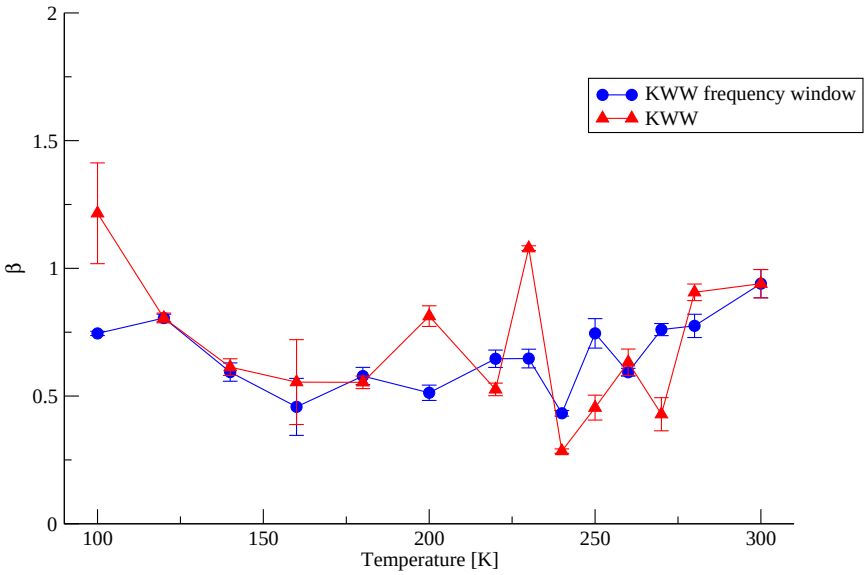


Figure 4.8: The stretching parameter β determined by fitting Eq. 4.6 to $I(Q,t)$ obtained from MD simulations. ($\triangle$) β obtained by varying all four parameters. ($\circ$) β obtained for the frequency window model ($A_0(Q)$ fixed).

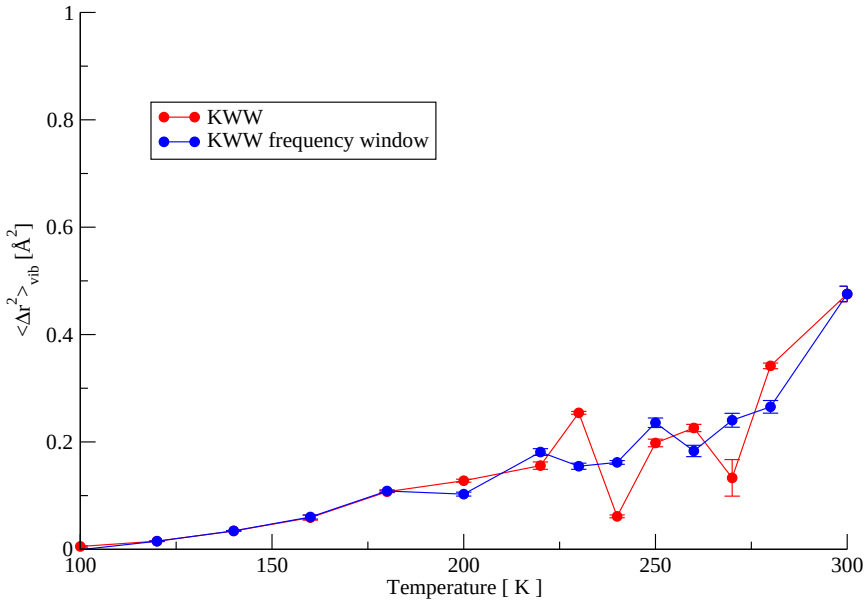


Figure 4.9: The vibrational mean-square displacement determined from the Debye-Waller factors in the Gaussian approximation. Debye-Waller factors were obtained by fitting Eq. 4.6 to $I(Q,t)$ obtained from MD simulations.

be adressed in the next section.

4.4 Single exponential relaxation

$I(Q,t)$ decays approximately exponentially for $t \geq 10ps$ and a single exponential function provides the same number of fitting parameters as the stretched 'frequency window' model. Therefore, it was decided to model $I(Q,t)$ with a single exponential that reproduces the long-time behavior. Consequently, the equation fitted to the simulation-derived $I(Q,t)$ is the following

$$I(Q,t) = e^{-\frac{1}{6}Q^2\langle\Delta r^2\rangle_{fast}} \left[\bar{A}_0(Q) + (1-\bar{A}_0(Q))e^{-\kappa_{FW}t}\right] \tag{4.15}$$

Here, the Debye-Waller factor representing vibrational motions in Eq. 4.8 has been replaced by the term $e^{-\frac{1}{6}Q^2\langle\Delta r^2\rangle_{fast}}$. $\langle\Delta r^2\rangle_{fast}$ includes all dynamics of the system faster than $\sim 10ps$, comprising not only vibrations but also fast diffusive motions.

The extraction of the parameters of Eq. 4.15 was again performed in two steps. First, the equation was fitted to the simulation-derived $I(Q,t)$ with all parameters allowed to vary freely with temperature, including the amplitudes, A_0.

In analogy to the previous section $\bar{A}_0(Q)$ was obtained by averaging $A_0(Q)$ over the temperatures above 250K. $\bar{A}_0(Q)$ again represents an estimate of the long-time, converged elastic incoherent scattering function, and was kept invariant with temperature in all further analyses, as required by the frequency window hypothesis. In a second fitting step Eq. 4.15 was fitted to the simulation-derived $I(Q,t)$ with $\bar{A}_0(Q)$ kept fixed.

Fig. 4.10 shows a comparison of the complete (left) and frequency window (right) model fits with the simulation-derived $I(Q,t)$. Both models reproduce the temperature-dependent decay of $I(Q,t)$ over the range $10 \leq t \leq 300ps$. Motions on timescales faster than 10ps are subsidised in the Debye-Waller factor according to the redefinition given above (compare Fig. 4.14). For temperatures between $180-240K$ the frequency window scenario does not reproduce the levelling off of the scattering funtion around $t = 200ps$. This might point to the fact, that indeed in this temperature range changes in the EISF are present. On the other hand inspection of the related relaxation frequency κ indicates that this temperature region marks the transition of motions into the 200ps range. Given a total simulation time of 500ps it can not be excluded that 200ps motions are not completely sampled in these trajoctories.

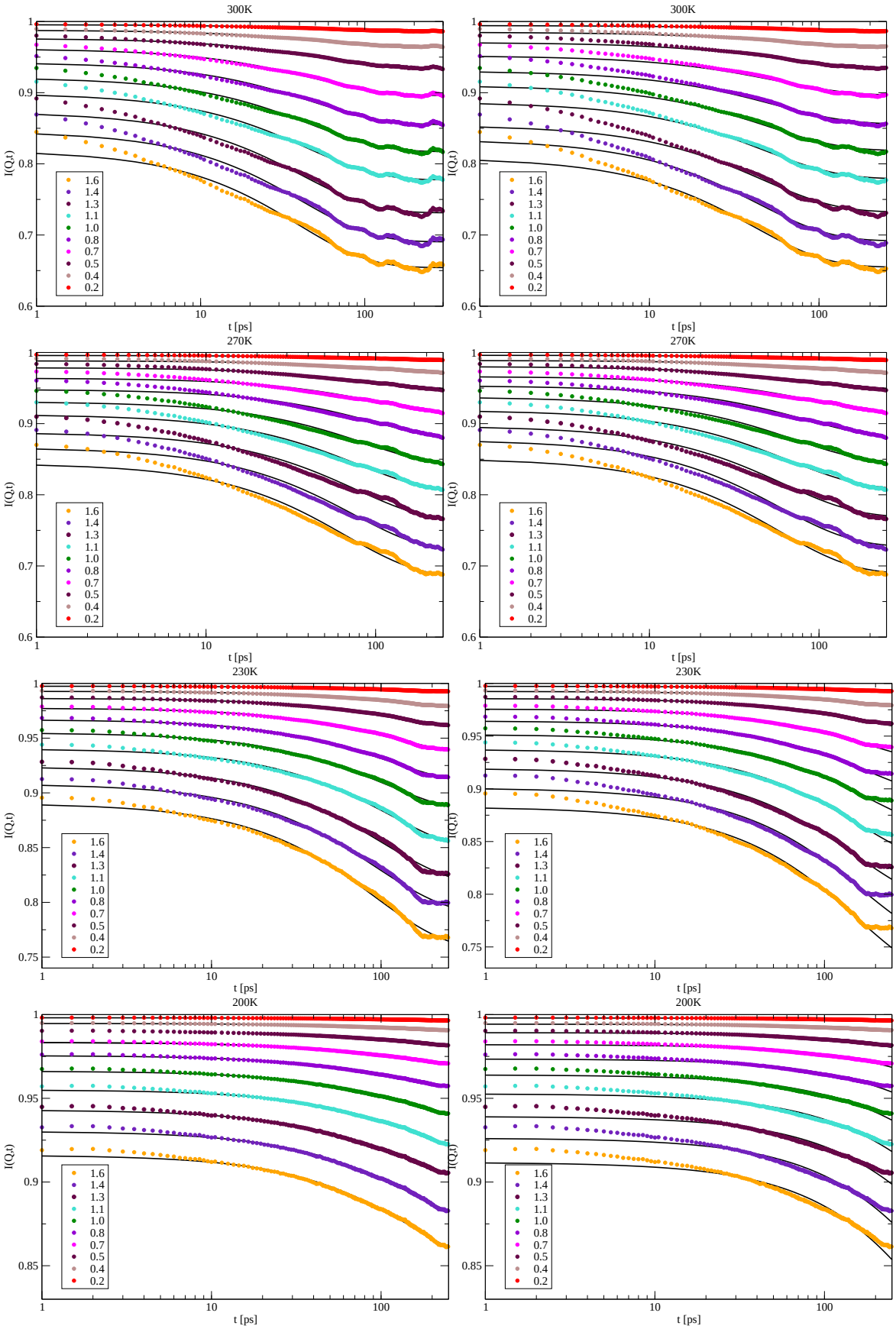


Figure 4.10: The intermediate scattering function, $I(Q,t)$, calculated from internal motions of BPTI over a Q-range of $0.2 - 1.6$Å (symbols). solid lines: fits of Eq. 4.15 to $I(Q,t)$ with a) all parameters free to vary (left) and b) $A_0(Q)$ fixed (right)

System parameters

Fig. 4.11 depicts the amplitudes $A_0(Q)$ obtained from Eq. 4.15 for temperatures $T \geq 250$K (symbols). The blue curve marks $\bar{A}_0(Q)$ (compare previous section). For temperatures below 250K $A_0(Q)$ is plotted in Fig. 4.12. The errorbars in both plots are given by the statistical errors of the fitting procedures. Figs. 4.11 & 4.12 exhibit a general decrease of $A_0(Q)$ with temperature as expected from an equilibrium scenario. There are, however, several exceptions reflecting the interdependency of the fitting parameters. In contrast to the previous section, the EISF as fitting parameter is now well defined for all temperatures.

Inspection of Fig. 4.13 shows that the relaxation frequency obtained for the full exponential model has values above $0.0025ps^{-1}$ even in the low temperature range. In these fits $A_0(Q)$ thus contributes in a non-neglectable amount. In consequence the errorbars remain small for all temperatures. Fig. 4.13 shows a comparison of the relaxation frequencies κ for both models. For comparison κ is plotted for the stretched frequency model as well. κ was again averaged over all Q values. The errorbars were calculated from the variance of κ with respect to different Q-values. Above 250K both models agree quantitatively with each other and with the relaxation frequency obtained for the KWW function (frequency window model) in the previous section. The exponential frequency scenario closely follows the KWW frequency for all temperatures, the full exponential fits led to consistently higher values for κ for temperatures below 250K. Again, this demonstrates the relation between equilibrium and frequency model at lower temperatures. The changes in the EISF in the full exponential fits *allow* the frequency to remain large enough for the amplitude to be resolved. Fixing the EISF instead *forces* κ to be small enough for the amplitude not to be completely resolved. Both possibilities are flexible enough to reproduce the spectra here obtained equally well.

Fig. 4.14 depicts $\langle \Delta r^2 \rangle_{fast}$ together with the vibrational mean-square displacement obtained in the previous section (frequency window model). $\langle \Delta r^2 \rangle_{fast}$ was obtained by fitting a straight line to the low Q-region ($Q^2 \leq 1\text{Å}^{-2}$) of $ln(a)$ *vs* Q^2, $a = \exp(-1/6 Q^2 \langle \Delta r^2 \rangle_{fast})$ being a fitting parameter of Eq. 4.15 for both scenarios.

$\langle \Delta r^2 \rangle_{fast}$, as discussed, includes all motions on timescales faster than $10ps$. Consequently both models lead to values of $\langle \Delta r^2 \rangle_{fast}$ consistently higher than $\langle \Delta r^2 \rangle_{Vib}$ representing only vibrational motions. Both models show an approximately linear dependence of $\langle \Delta r^2 \rangle_{fast}$ on temperature.

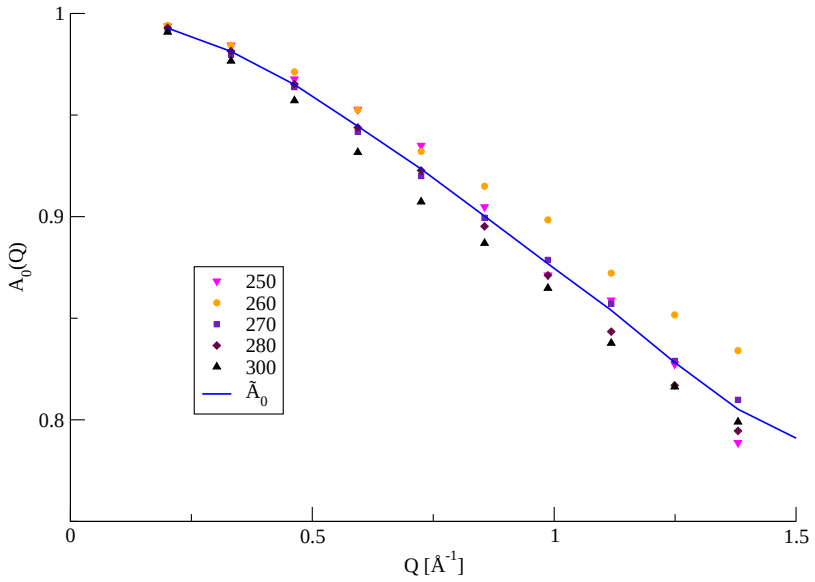


Figure 4.11: Elastic incoherent scattering factor, $A_0(Q)$, determined for $T \geq 250$K by fitting Eq. 4.15 to $I(Q,t)$ obtained from MD simulations (symbols). The blue line depicts the average amplitude, $\bar{A}_0(Q)$.

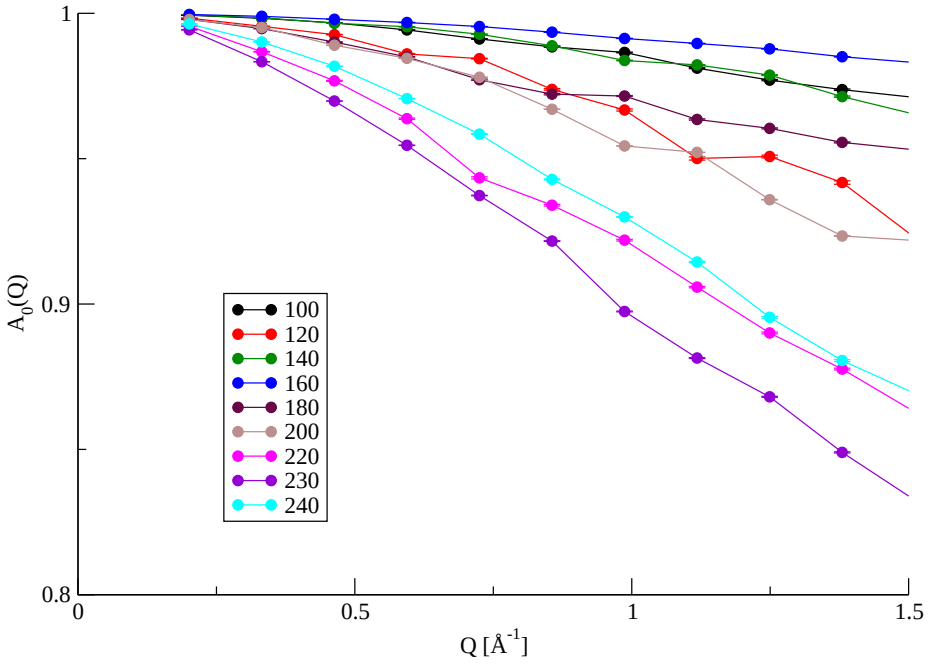


Figure 4.12: Elastic incoherent scattering factor ($A_0(Q)$) determined for $T \leq 240$K by fitting Eq. 4.15 to $I(Q,t)$ obtained from MD simulations.

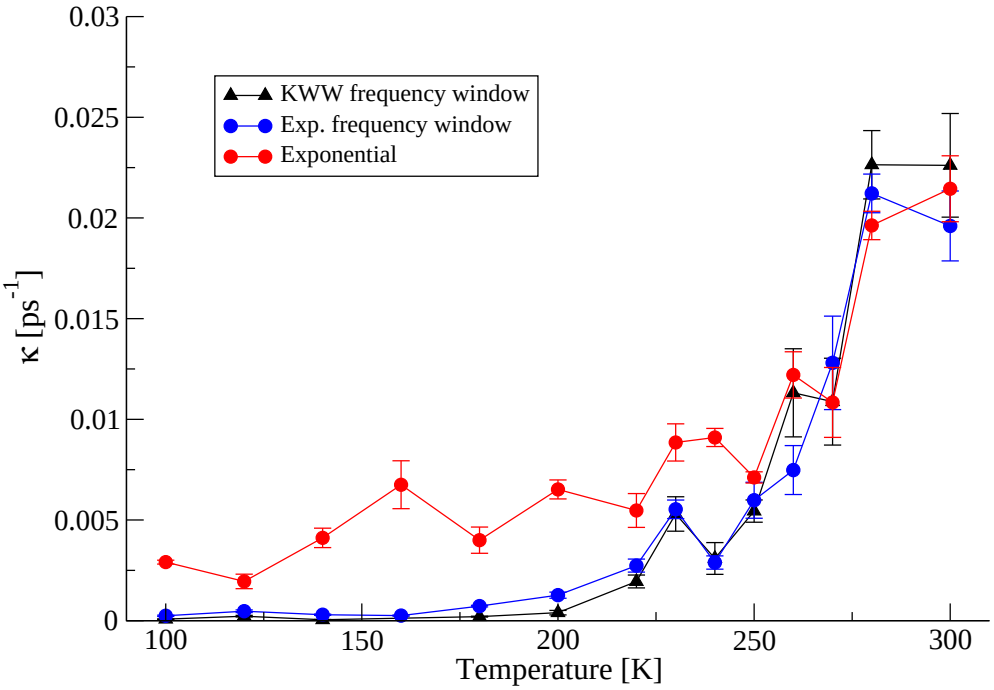


Figure 4.13: Relaxation frequency κ determined by fitting Eq. 4.15 to $I(Q,t)$ obtained from MD simulations. ($\circ$) κ obtained by varying all three parameters. ($\circ$) κ obtained for the frequency window model ($A_0(Q)$ fixed). ($\triangle$) for comparison: κ obtained for a KWW relaxation function.

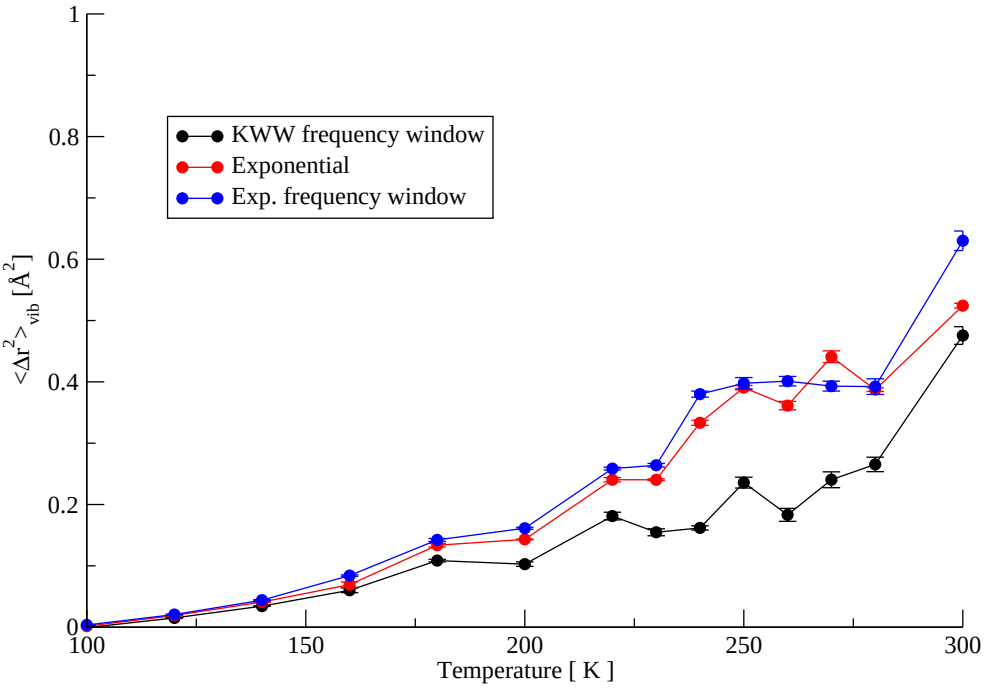


Figure 4.14: The vibrational mean-square displacement determined from the Debye-Waller factors in the Gaussian approximation. Debye-Waller factors were obtained by fitting Eq. 4.15 to $I(Q,t)$ obtained from MD simulations ($\circ$) for comparison: $\langle \Delta r^2 \rangle_{Vib}$ obtained for a KWW relaxation function.

4.5 Conclusions

The temperature dependence of the incoherent intermediate scattering function, $I(Q,t)$, was shown to be well reproduced by a four parameter stretched exponential function. However, keeping the elastic incoherent scattering function of the system fixed does not affect the quality of data reproduction. The remaining three parameters show qualitatively the same behavior and are even quantitatively in good agreement. Whereas the temperature dependence of both, the Debye-Waller factor and the *average* frequency κ is well established in all models used in this chapter, there is no clear pattern visible for the stretching parameter β. The stretching parameter, although within the range commonly found in complex systems ($0.4 \leq \beta < 1$), does not show a systematic dependence on temperature.

Describing the system by a single exponential relaxation function fails to reproduce $I(Q,t)$ on timescales shorter than 10ps. The long-time decay due to slower relaxation processes, however, is still well reproduced. Again, fixing the scattering amplitude with respect to temperature leads to a description representing the spectra equally well. The parameters obtained from both scenarios resemble each other indicating that the frequency window model is a real alternative allowed by the ambiguity of incomplete data.

Even more, the relaxation frequency κ shows qualitatively the same behavior for all models, a rather flat region at low temperatures followed by a sharp increase as the temperature rises. Especially there is no qualitative difference in κ whether the spectra are described by a single exponential or a KWW function. For the frequency window model both are in good quantitative agreement.

Quantitavely similar results are also obtained for an averaged EISF $\bar{A}_0(Q)$. This is of importance with hindsight to the dynamical transition since it will be shown that the low Q-regime of the EISF determines the converged mean-square displacement (see Ch. 5). Thus the frequency window models discussed here agree quantitatively with respect to contributions of the EISF to the total mean-square displacement.

Consequently, the scattering spectra calculated from internal protein motions of BPTI does not allow to rule out a pure frequency window scenario, *i.e.* it can not be decided whether there is at all a change of the elastic incoherent scattering factor. Both, the KWW relaxation model and the approximation by a single exponential relaxation process can reproduce the intermediate scattering function in a pure frequency window scenario without loss of accuracy. It should be stressed, however, that the frequency window scenario here is proposed as a *limiting* model that is still able to reproduce scattering data. The frequency windows of neutron scattering experiments are in the same range as characteristic frequencies of molecular motions, *i.e.* $\sim 1 - 100\mu$eV. Therefore, the frequency window scenario is likely to play a role in determining the dynamics observed using this technique. Whether it plays a dominant role in the description of protein dynamics or contributes only corrections to the picture provided by the equilibrium model cannot be decided on the basis of the present analysis.

In a complex energy landscape, such as that explored by a solvated protein, the dynamics observed by neutron scattering is likely to involve a combination of both the frequency-

window and equilibrium scenarios. That is, the true description of the intermediate scattering function will require both, changes of the EISF with temperature as well as an EISF that becomes *visible* upon temperature rising due to an increase in κ.

Hence regarding neutron scattering a pure frequency window scenario has to be considered as a possible, limiting description of internal protein dynamics. The consequences of this finding with hindsight to the dynamical transition and its relation to enzymatic activity will be discussed in the next chapter.

Chapter 5

The dynamical transition in proteins

The leeway in model descriptions of neutron scattering data is shown to allow for two limiting interpretations of the dynamical transition as measured with neutron scattering, an equilibrium and a frequency window interpretation. In the former the transition is due to the occupation of higher energy conformational substates whereas in the latter interpretation the observed transition in the measured mean-square displacement is caused by a the temperature dependence of relaxation times crossing the instrumental resolution.

The dynamical transition, as explained in Ch. 1, is one of the prominent features encountered in pico- to nanosecond dynamics of proteins. The observed connection to enzyme activity makes it a valuable phenomenon to investigate the relation between the energy landscape of enzymes and their biological function. In this chapter the dynamical transition measured by neutron scattering is discussed. Bearing in mind the structural and dynamical complexity of proteins, the focus of attention is not to impose an energy landscape model of protein dynamics, but rather to reduce the observed phenomena to basic features of the energy landscape.

Systems for which such a dynamical transition, *i.e.* a deviation from harmonic, linear dependence of the average atomic mean-square displacement, $\langle \Delta r^2 \rangle$ with increasing temperature, has been observed include glass-forming liquids, polymers and proteins (Bizzarri *et al.*, 2000; Cohen *et al.*, 1981; Doster *et al.*, 1989b; Frauenfelder *et al.*, 1979; Keller & Debrunner, 1980; Knapp *et al.*, 1982; Rasmussen *et al.*, 1992). For proteins, suggestions have been made that the additional motions present at temperatures above the transition may be important in biological function (Ding *et al.*, 1994; Ferrand *et al.*, 1993; Fitter *et al.*, 1997; Lehnert *et al.*, 1998; Rasmussen *et al.*, 1992). Several models have been used to describe the dynamics activated at the dynamical transition, including continuous diffusion (Kneller & Smith, 1994), jumping between minima (Doster *et al.*, 1989b; Elber & Karplus, 1987; Frauenfelder *et al.*, 1979, 1991; Lamy *et al.*, 1996), mode-coupling theory (Doster & Settles, 1999; La Penna *et al.*, 1999; Perico & Pratolongo, 1997), stretched-exponential behavior (Dellerue *et al.*, 2001) and 'effective force constants' (Zaccai, 2000a).

The common theme connecting these model descriptions is that they explain the dynamical transition via changes in the long-time, converged properties of proteins. Translated to the

language of neutron scattering these models relate to changes in the elastic incoherent scattering function, EISF of the system. As seen in the last chapter, neutron scattering spectra of internal protein motions can be explained without such changes. The dynamical transition is thus not necessarily related to changes in equilibrium properties of proteins. The mean-square displacement as measured by neutron scattering techniques will now be derived. It will be shown how the dynamical transition can be explained within the context of the frequency window scenario.

5.1 Equilibrium and frequency window scenario

The experimental detection of the dynamical transition depends on the relationship between the timescale of the characteristic relaxation processes leading to the increased mean-square displacement and the time resolution of the experimental technique employed. Following Ch. 4 two contrasting scenarios for the dynamical transition can be envisaged. In the first, 'equilibrium' scenario all motions in the system are resolved by the instrument at all temperatures examined. In other words, the characteristic relaxation frequencies of the dynamics are all within the energy resolution of the instrument used; the EISF of the system is a quantity accessible by experiment. In this case, an observed dynamical transition results from a change with temperature of the long-time probability distribution of single atom displacements. Analysis of the dynamical transition can in principle lead to a characterization of energy levels occupied by different conformational substates (Doster & Settles, 1999; Doster *et al.*, 1989b).

In the alternative, 'frequency window' scenario, there is no change in the time-converged atomic probability distribution with temperature. Apparent dynamical transition behavior can be observed in this scenario if two criteria are fulfilled: a) below the transition, the relaxation frequencies of the dynamics determining the mean-square displacement are too slow to be detected by the finite energy-resolution instrument, and b) if the frequencies increase with temperature such that they pass into the frequency window of the instrument. This description of the dynamical transition has recently been suggested to explain the timescale dependence of the mean-square displacement observed for an enzyme solution (Daniel *et al.*, 2003b). For systems in which the frequency window scenario dominates, the dynamical transition reveals information about the timescales of motions crossing the resolution window of the instrument. If these motions involve activated dynamics, then the barriers concerned can be determined.

5.1.1 Finite Energy Resolution and the Mean-Square Displacement.

The frequency window scenario leads to a new interpretation of the dynamical transition as measured by neutron scattering. For this purpose, the experimental derivation of mean-square displacements has to be reexamined. $\langle \Delta r^2 \rangle$ is accessible via neutron scattering experiments with the help of the Gaussian approximation (see Ch. 2). From Eq. 2.44 it can be seen that $\langle \Delta r^2 \rangle$ is determined by the initial slope of the logarithm of the elastic scattering *vs* Q^2.

Fitting the low Q-region of experimentally-obtained elastic scattering data with a linear function yields the 'measured' mean-square displacement, which we call $\langle \Delta r^2 \rangle_{Exp}$. The Gaussian approximation gives:

$$\langle \Delta r^2 \rangle_{Exp} = -6 \frac{\partial}{\partial Q^2} \ln(S_{Exp}(Q,0))|_{Q^2=0} \tag{5.1}$$

For any given instrument $\langle \Delta r^2 \rangle_{Exp}$ may not be the time-converged mean-square displacement, but may have contributions due to finite energy resolution. We will now analyze the components of $\langle \Delta r^2 \rangle_{Exp}$ using Eq. 4.3 as a starting point.

Inspection of Eq. 4.3 suggests that the elastic scattering can be obtained by neglecting the quasielastic term, $I'_D(Q,t)$. However, in a real experiment due to finite energy resolution the intensity under the elastic peak can contain contributions from quasielastic scattering. To represent this analytically we write the measured elastic scattering, $S_{Exp}(Q,0)$ as:

$$S_{Exp}(Q,0) = \int_{-\infty}^{\infty} d\omega \; R(\omega) S(Q,\omega) \tag{5.2}$$

$$= e^{-2W} \left(A_0(Q) + I'_D(Q,\Delta\omega) \right) \tag{5.3}$$

where $R(\omega)$ is the instrumental energy resolution function and
$I'_D(Q,\Delta\omega) = \int_{-\infty}^{\infty} d\omega R(\omega) I'_D(Q,\Delta\omega)$.

We thus obtain from Eq. 5.3 for the measured mean-square displacement:

$$\langle \Delta r^2 \rangle_{Exp} = -6 \frac{\partial}{\partial Q^2} \ln(S_{Exp}(Q,0))|_{Q^2=0} \tag{5.4}$$

$$= -6 \frac{\partial}{\partial Q^2} \ln(e^{-2W(Q)}) \tag{5.5}$$

$$-6 \frac{\partial}{\partial Q^2} (\ln[A_0(Q) + I'_D(Q,\Delta\omega)) \tag{5.6}$$

$$\langle \Delta r^2 \rangle_{Exp} = \langle \Delta r^2 \rangle_{Vib} + \langle \Delta r^2 \rangle_{EISF} - \langle \Delta r^2 \rangle_{Res} \tag{5.7}$$

$$= \langle \Delta r^2 \rangle_{Conv} - \langle \Delta r^2 \rangle_{Res} \tag{5.8}$$

$\langle \Delta r^2 \rangle_{Exp}$ in Eq. 5.8 contains two contributions, $\langle \Delta r^2 \rangle_{Conv}$ and $\langle \Delta r^2 \rangle_{Res}$. One of these, $\langle \Delta r^2 \rangle_{Conv}$, is the long-time, converged $\langle \Delta r^2 \rangle$, consisting of the vibrational contribution, $\langle \Delta r^2 \rangle_{Vib}$ and the diffusive contribution associated with the EISF, $\langle \Delta r^2 \rangle_{EISF}$.

The second term in Eq. 5.8, $\langle \Delta r^2 \rangle_{Res}$ is the contribution to $\langle \Delta r^2 \rangle_{EISF}$ due to relaxation processes too slow to be resolved by the instrument. Its subtraction means that it reduces the observed $\langle \Delta r^2 \rangle$. Neglecting this second term corresponds to making the assumption that all motions in the system are fast enough to be detected.

$\langle \Delta r^2 \rangle_{Res}$ is generally given by

$$\langle \Delta r^2 \rangle_{Res} = 6 \frac{\partial}{\partial Q^2} I'_D(Q,\Delta\omega) \tag{5.9}$$

Further simplification can be achieved by choosing the form of $I'_D(Q,\omega)$. An approximate analytical solution for $\langle \Delta r^2 \rangle_{Res}$ shall now be derived for the case that the diffusive scattering can be described as a sum of exponential relaxation processes (compare Eq. 4.8):

$$I'_D(Q,\omega) = \sum_l A_l(Q) \frac{1}{\pi} \frac{\kappa_l}{\kappa_l^2 + \omega^2} \tag{5.10}$$

From Eq. 5.9 and 5.3 $\langle \Delta r^2 \rangle_{Res}$ is then given by $\langle \Delta r^2 \rangle_{Res} \equiv \sum_l a_l \Delta\omega_l$ with

$$a_l = 6 \frac{\partial}{\partial Q^2} A_l(Q)|_{Q^2=0} \tag{5.11}$$

$\Delta\omega_l$, the width of the Lorentzian function, is here assumed to be independent of Q. This is true for most analytical models of spatially-confined systems (*e.g.,* jumping between a finite number of states, rotational or confined diffusion). However, in heuristic approaches to experimental data treatment the width is often taken to be Q-dependent, in which case $\langle \Delta r^2 \rangle_{Res}$ would obtain one additional term.

Since

$$\int_{-\infty}^{\infty} d\omega \; \frac{1}{\pi} \frac{\kappa_l}{\kappa_l^2 + \omega^2} = 1, \tag{5.12}$$

inspection of Eq. 5.7 shows that the maximal contribution of the relaxation mode $A_l(Q)$ to $\langle \Delta r^2 \rangle_{Res}$ is given by a_l. This is intuitive for systems in which a real physical process can be related to a given relaxation mode, l. If a certain mode l describes, for example, jumping between two minima, this maximal contribution a_l is given by the distance between these minima, and Eq. 5.8 is simply the statement that these jumps will be detected only if they occur on timescales accessible to the instrument. However, for models where the physical process involved cannot be described by a single relaxation mode, no direct physical interpretation of a_l is possible.

Finally, we derive an approximate functional form of $\Delta\omega_l$. To do this we make an approximation to the resolution function, $R(\omega)$ that allows $\Delta\omega_l$ to be expressed analytically *i.e.*, $R(\omega)$ is assumed to be a rectangular function of width $\Delta\omega$:

$$R(\omega) = \begin{cases} 1 & -\Delta\omega \leq \omega \leq \Delta\omega \\ 0 & \text{other} \end{cases} \tag{5.13}$$

This allows us to write:

$$\langle \Delta r^2 \rangle_{Exp} = \langle \Delta r^2 \rangle_{Conv} - \sum_l a_l \frac{2}{\pi} \arctan \frac{\Delta\omega}{\kappa_l} \tag{5.14}$$

Eq. 5.14 will be used to examine dynamical transition behavior in the context of the frequency window scenario in both MD simulations and experiment (Sec. 5.2 and Sec. 5.3). First, however, a relation between transition temperatures measured at different instruments and their respective energy resolutions will be derived.

5.1.2 The transition temperature

The dynamical transition temperature, T_0 can be operationally defined as the temperature at which the measured mean-square displacement, $\langle \Delta r^2 \rangle_{exp}$, starts to deviate from a straight line. According to Eq. 5.8 $\langle \Delta r^2 \rangle_{exp}$ consists of two parts: $\langle \Delta r^2 \rangle_{Conv}$ and$\langle \Delta r^2 \rangle_{Res}$. In principle changes in either of these can produce a change in $\langle \Delta r^2 \rangle_{exp}$. A change in $\langle \Delta r^2 \rangle_{Conv}$ corresponds to a change in the time-converged properties of the protein. Such changes could be due, for example, to the presence of distinct minima in the energy landscape of a protein. However, if these minima are of the same energy they will not produce changes in the EISF as their relative populations will not change with temperature. Rather, the energies of the minima must be different so that the increased relative occupation of the higher energy state with increasing temperature leads to an increase in $\langle \Delta r^2 \rangle_{EISF}$ and hence in $\langle \Delta r^2 \rangle_{Conv}$. The exact relation between T_0 and the difference in energy ΔG will depend on the underlying landscape model. However, the increase in $\langle \Delta r^2 \rangle$ will be approximately proportional to the Boltzmann factor $\exp(-\Delta G/RT)$. The transition temperature will thus be determined by $RT_0 \approx \Delta G$. Since the changes occur in time-converged properties of the protein the transition will be independent of instrumental resolution provided that all associated motions are resolved. An example of a model based on a change in $\langle \Delta r^2 \rangle_{EISF}$ is that of Ref. Doster *et al.* (1989b).

In the frequency window model the experimentally-observed transition is an effect of finite energy resolution alone with no change in the time-converged dynamics. Since $\langle \Delta r^2 \rangle_{Conv}$ is now independent of instrumental resolution, any difference in the measured mean-square displacement between two instruments must arise from contributions of the second term in Eq. 5.8, *i.e.* $\langle \Delta r^2 \rangle_{Res}$.

Comparing transition temperatures, T_0, due to changes in $\langle \Delta r^2 \rangle_{Res}$, leads to a relation between the instrumental resolutions and the relaxation frequencies. Each relaxation frequency will lead to an observed dynamical transition once it is high enough to make a measurable contribution to the total mean-square displacement, $\langle \Delta r^2 \rangle_{exp}$, that is, once the change in $\langle \Delta r^2 \rangle_{Res}$ is large enough to lead to a detectable deviation of $\langle \Delta r^2 \rangle_{exp}$ from linearity. For two instruments of different energy resolutions, $\hbar\Delta\omega_1$ and $\hbar\Delta\omega_2$, this will happen at different temperatures T_1 and T_2 respectively. We consider the case where $\delta\langle \Delta r^2 \rangle_{Res}(T, \Delta\omega)$ the measured excess mean-square displacement for instrument 1 at T_1 (*i.e.* , the portion of $\langle \Delta r^2 \rangle_{exp}$ that is above linear behavior), is equal to the measured excess mean square displacement for instrument 2 at T_2, that is

$$\delta\langle \Delta r^2 \rangle_{Res}(T_1, \Delta\omega_1) = \delta\langle \Delta r^2 \rangle_{Res}(T_2, \Delta\omega_2) \tag{5.15}$$

If we assume that these two transitions are due to the same dynamical process characterized by the relaxation frequency $\kappa(T)$ we can then write:

$$\arctan \frac{\Delta\omega_1}{\kappa(T_1)} = \arctan \frac{\Delta\omega_2}{\kappa(T_2)} \tag{5.16}$$

or, simplified:

$$\frac{\kappa(T_2)}{\kappa(T_1)} = \frac{\Delta\omega_2}{\Delta\omega_1} \tag{5.17}$$

Further information can be obtained if the functional form for $\kappa(T)$ is known. Assuming an Arrhenius temperature dependence, $\kappa(T) = \kappa_0 e^{\Delta B/RT}$, gives:

$$\frac{1}{T_2} - \frac{1}{T_1} = \frac{R}{\Delta B} ln(\frac{\Delta\omega_2}{\Delta\omega_1}) \tag{5.18}$$

where ΔB is the barrier height (see Fig. 4.2) and R is the gas constant.

In this case, then, comparing the transition temperatures measured on different instruments can lead to direct determination of the barrier height, ΔB.

5.2 Dynamical transition in MD Simulations

In Ch. 4 it was shown that the longtime relaxation of the scattering function $I(Q,t)$ is well reproduced by a single exponential relaxation function with temperature independent EISF. We now examine whether the frequency window scenario can describe the simulation-derived $\langle \Delta r^2 \rangle_{Sim}$.

Assuming the presence of a single relaxation frequency, the mean-square displacement from the frequency window model, $\langle \Delta r^2 \rangle_{FW}$, is given by (see Eq. 5.14):

$$\langle \Delta r^2 \rangle_{FW} = \langle \Delta r^2 \rangle_{fast} + \langle \Delta r^2 \rangle_{\bar{A}_0} (1 - \frac{2}{\pi} \arctan \frac{\Delta\omega}{\kappa_{FW}}) \tag{5.19}$$

where $\langle \Delta r^2 \rangle_{fast}$ and $\langle \Delta r^2 \rangle_{\bar{A}_0}$ are the fast and slow contributions to the mean-square displacement, respectively. $\langle \Delta r^2 \rangle_{fast}$, as in Eq. 4.15, subsidizes all motions faster than 10ps.

The temperature dependence of the mean-square displacement was determined directly from the set of MD simulations, using Eq. 3.26. $\langle \Delta r^2 \rangle_{Sim}$ was calculated for four different values of t: 20, 50, 100 and 200 ps. Fig. 5.1 shows the resulting mean-square displacements, $\langle \Delta r^2 \rangle_{Sim}$. Below $T \approx 180$K $\langle \Delta r^2 \rangle_{Sim}(t)$ is the same on all timescales, indicating that no processes are activated with timescales between 20-200ps at these temperatures. Raising the temperature further, a dynamical transition, *i.e.* a deviation from linearity of $\langle \Delta r^2 \rangle_{Sim}$, is observed on all timescales. However, the transitions on different timescales differ in two aspects.

First, T_0, the temperature at which $\langle \Delta r^2 \rangle_{Sim}$ deviates from the straight line, shows a timescale dependence. This is illustrated in Fig. 5.2, in which $\langle \Delta r^2 \rangle_{Sim}$ is plotted for 20ps and 200ps together with a linear fit over the temperature range $100 \leq T \leq 180$K. Whereas $\langle \Delta r^2 \rangle_{Sim}$ deviates from linearity at $\sim$180K for the 200*ps* data, a deviation appears only at $\sim$220K on the 20*ps* timescale. This reflects the fact that for temperatures between 170-220K the intermediate scattering function in Fig. 4.10 exhibits an additional decrease on the 200*ps*, but not on the 20*ps* timescale. The presence of a timescale dependence of the

transition temperature is a characteristic feature of the frequency window scenario. Since in the frequency window picture no new motions are activated at T_0, but rather existing motions become faster and drift into the timescale accessible to an instrument, T_0 depends on the instrumental resolution.

A second difference between $\langle \Delta r^2 \rangle_{Sim}$ on different timescales is the magnitude of $\langle \Delta r^2 \rangle_{Sim}$ for $T > T_0$. This magnitude increases with timescale such that $\langle \Delta r^2 \rangle_{Sim}(200ps) \approx 2\langle \Delta r^2 \rangle_{Sim}(20ps)$ at 300K. Again, this reflects the presence of motions on the 20-200ps timescale.

To see whether a frequency window model is able to reproduce these differences, the mean-square displacement, $\langle \Delta r^2 \rangle_{FW}$ was calculated using Eq. 5.19 at different values of the resolution: $\Delta\omega = {}^1\!/_{20ps}, {}^1\!/_{50ps}, {}^1\!/_{100ps}$ and ${}^1\!/_{200ps}$. The results thus obtained are compared with $\langle \Delta r^2 \rangle_{Sim}$ in Fig. 5.1. The only difference between the four $\langle \Delta r^2 \rangle_{FW}$ shown is the factor $\Delta\omega$ representing different instrumental resolutions. Nevertheless, $\langle \Delta r^2 \rangle_{FW}$ reproduces the simulation data on all timescales (resolutions). As expected, the agreement for the lowest resolution (timescale 20ps) is slightly worse than at other resolutions, since the frequency-window model used here (Eq. 4.15) was parametrized to reproduce the simulation-derived $I(Q,t)$ only for longer times (see Ch. 4).

The inset to Fig. 5.1 depicts $\langle \Delta r^2 \rangle_{fast}$, the fast contribution to $\langle \Delta r^2 \rangle$ together with the time-converged slow contribution due to diffusive motions, $\langle \Delta r^2 \rangle_{\bar{A}_0}$. Also shown is a linear fit to $\langle \Delta r^2 \rangle_{fast}$ over the temperature range $100 < T < 180$K. Below $T \approx 220$K the linearity of $\langle \Delta r^2 \rangle_{fast}$ is consistent with the motions being harmonic, given by the Debye-Waller factor $I_{fast}(Q,t) \approx e^{-2W}$. The linear fit suggests that there is a small transition in $\langle \Delta r^2 \rangle_{fast}$ at $T \approx 200 - 220$K. This might indicate the onset of fast anharmonic motions on timescales faster than 10ps.

$\langle \Delta r^2 \rangle_{Conv}$ is given by the sum of $\langle \Delta r^2 \rangle_{fast}$ and $\langle \Delta r^2 \rangle_{\bar{A}_0}$ (see Eqs. 5.7 & 5.8), where $\langle \Delta r^2 \rangle_{\bar{A}_0}$ is constant with respect to temperature. The high value of $\langle \Delta r^2 \rangle_{\bar{A}_0}$ at low temperature in the inset to Fig. 5.1 might be surprising at first but can be understood if one recalls that in the frequency-window model motions do not cease to exist at a given temperature, but simply shift to longer timescales. Given enough time, these motions still occur, leading to a constant mean-square displacement even at low temperatures. Taking the symmetric two-state model as an example again, this reflects that barrier crossings do occur at very low temperatures, with the timescale of these crossings tending to infinity.

In Fig. 5.3 the relaxation frequency, κ_{FW}, is plotted obtained by the fit shown in Fig. 4.10. For $T \leq 180$K the relaxation time, ${}^1\!/_{\kappa_{FW}}$ is indeed much longer than the 200ps timescale, *i.e.* κ_{FW} is sufficiently small not to lead to substantial decrease in $I(Q,t)$ for $t \leq 200ps$. At the onset of the measured transition, at $T \approx 220$K, it is $\sim 200ps$, and becomes faster with increasing temperature, reaching a timescale of $\sim 50ps$ at 300K.

The frequency window model does not suppose any particular form of the dynamics associated with the relaxation frequency, κ_{FW}. For example, *a priori*, the dynamics could be continuous or jump diffusion. Distinguishing between these types of dynamics is beyond the remit of the present analysis method. Further analysis can involve fitting an Arrhenius function, $\kappa(T) = ae^{-\frac{\Delta B}{RT}}$, to κ_{FW}. $\kappa(T)$ in Fig. 5.3 indeed follows approximately

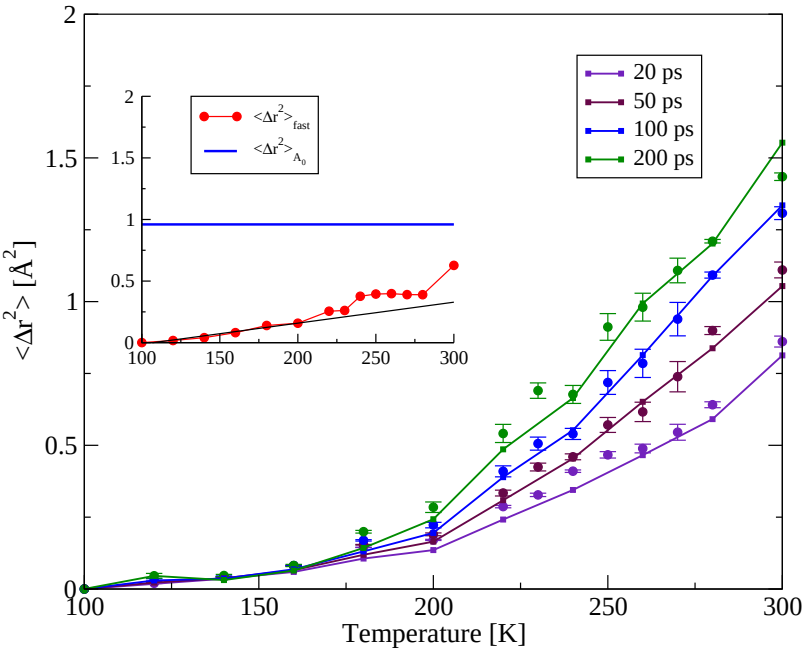


Figure 5.1: Energy resolution and timescale-dependence of $\langle \Delta r^2 \rangle$. Solid lines: $\langle \Delta r^2 \rangle_{Sim}$ calculated from MD trajectories using Eq. 3.26 with t=20,50,100 and 200ps. Symbols: $\langle \Delta r^2 \rangle_{FW}$ calculated using Eq. 5.19 with the instrumental energy resolution, $\Delta\omega$, set to $1/20, 1/50, 1/100$ and $1/200$ ps^{-1}. Inset: The fast component, $\langle \Delta r^2 \rangle_{fast}$ of $\langle \Delta r^2 \rangle_{FW}$ (red) and $\langle \Delta r^2 \rangle_{A_0}$ (blue). The straight line represents a linear fit to the temperature range $100 - 180$K.

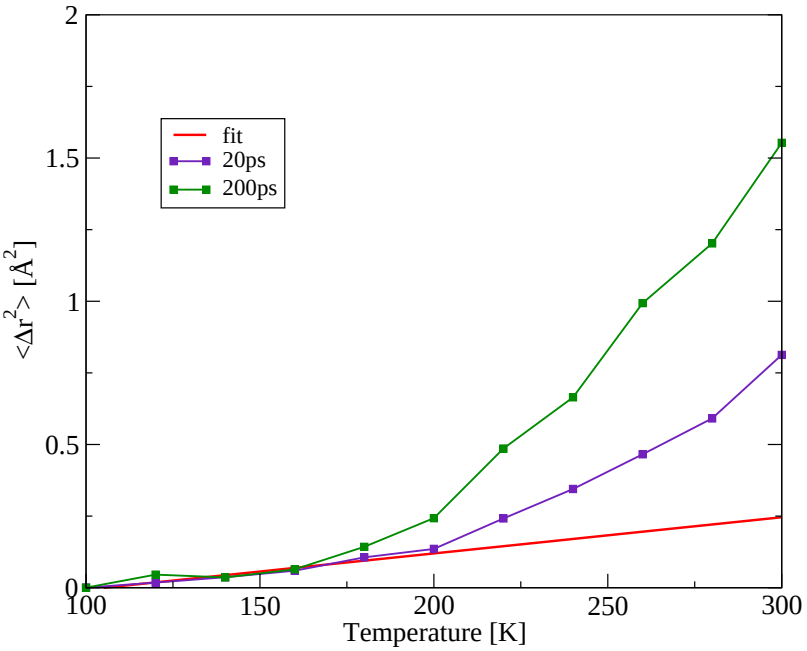


Figure 5.2: $\langle \Delta r^2 \rangle_{Sim}$ calculated from the MD trajectories for two values of t, $20ps$ and $200ps$. Also shown is a linear fit to the temperature range $100 - 180$K.

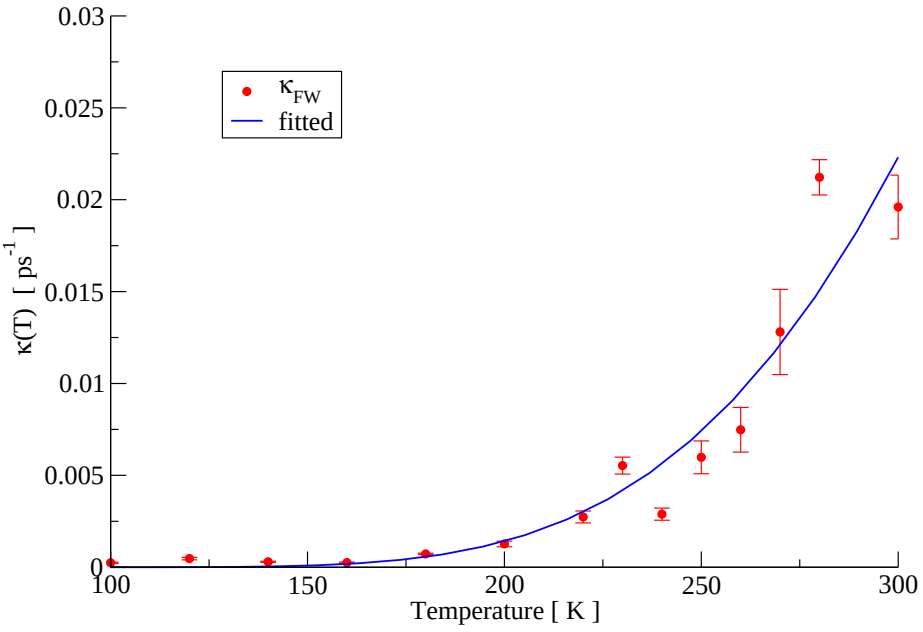


Figure 5.3: Symbols: Long-time relaxation frequency κ as obtained by fitting Eq. 4.15 to $I(Q,t)$. The solid line shows the result of fitting an Arrhenius function to κ.

Arrhenius behavior. The resulting parameters determining the slow dynamics in $\langle \Delta r^2 \rangle_{FW}$ are $\langle \Delta r^2 \rangle_{A_0}$, the pre-exponential a and ΔB, the activation free energy. $\langle \Delta r^2 \rangle_{A_0}$ is 0.96Å, ΔB is 3.0kcal/mol, and the pre-exponential factor is $a \sim 10^{12} s^{-1}$. These values are typical for barrier crossing in condensed-phase molecular systems (see Bee (1988)). It must be stressed, however, that overinterpretation of κ_{FW} is dangerous and that attribution of the dynamics to any single process is unwarranted. Indeed, for a heterogeneous system such as a protein in solution it is unlikely that simple types of motions determine the scattering profile of the system. Further analysis of the simulations and quasielastic scattering would be necessary to disentangle these different contributions and to obtain a detailed picture of the underlying dynamics at atomic resolution (Tournier & Smith, 2003).

5.3 Dynamical transition in experiment

The frequency window model is now used to analyze experimentally-derived mean-square displacements obtained from glutamate dehydrogenase in a cryosolution (Daniel *et al.*, 1999). The analysis again uses Eq. 5.19 as the fitting function. The fast component, $\langle \Delta r^2 \rangle_{fast}$ was assumed to depend linearly on temperature, $\langle \Delta r^2 \rangle_{fast} = \alpha T$. Eq. 5.19 was fitted simultaneously to the mean-square displacements determined using the instruments IN6 and IN16 with $\Delta\omega_{IN6} = 50\mu eV = 50\Delta\omega_{IN16}$, reflecting the fifty times higher energy resolution of IN16.

Fig. 5.4 shows the result of the least-squares fit of Eq. 5.19 to the experimental data sets. The model is able to reproduce the most prominent features of the experimental data, namely the pronounced shift of the transition temperature, T_0, from 150K to 220K between the two instruments, and the leveling off of $\langle \Delta r^2 \rangle_{exp}$ at higher temperatures seen on the higher-resolution instrument. The shift in T_0 with resolution is in harmony with the shift seen in the simulation data, $\langle \Delta r^2 \rangle_{Sim}$, in Fig. 5.2. Motions being fast enough to be detected at 150K with the instrument IN16 become fast enough to be seen on IN6 only at 220K. The fact that the shift in the simulation data in Fig. 5.2 is smaller than that seen experimentally in Fig. 5.4 is consistent with the fact that the difference in timescale between the two simulation data sets is smaller than that between the two experimental data sets.

The leveling off of $\langle \Delta r^2 \rangle_{exp}$ with T at high resolution is a consequence of the finite amount that the relaxation process contributes to $\langle \Delta r^2 \rangle_{Conv}$. According to Eq. 5.7 each process l contributes at most a_l to the converged mean-square displacement. This can be illustrated, again, with the two-state model. Jumping between the minima makes a contribution to $\langle \Delta r^2 \rangle_{Conv}$. This contribution will not be seen in $\langle \Delta r^2 \rangle_{exp}$ at low temperature since the timescale of these jumps is too slow. At higher temperatures the jumps contribute more and more, thereby leading to a transition in $\langle \Delta r^2 \rangle_{exp}$. If the timescale of these jumps is much faster than the timescale accessible to the instrument *i.e.*, $\tau_{\text{jump}} \ll \tau_{\text{res}}$, all of the jumps are seen leading to no further increase in the observed mean-square displacement with temperature.

At higher temperatures the frequency window model fails to obtain quantitative agreement

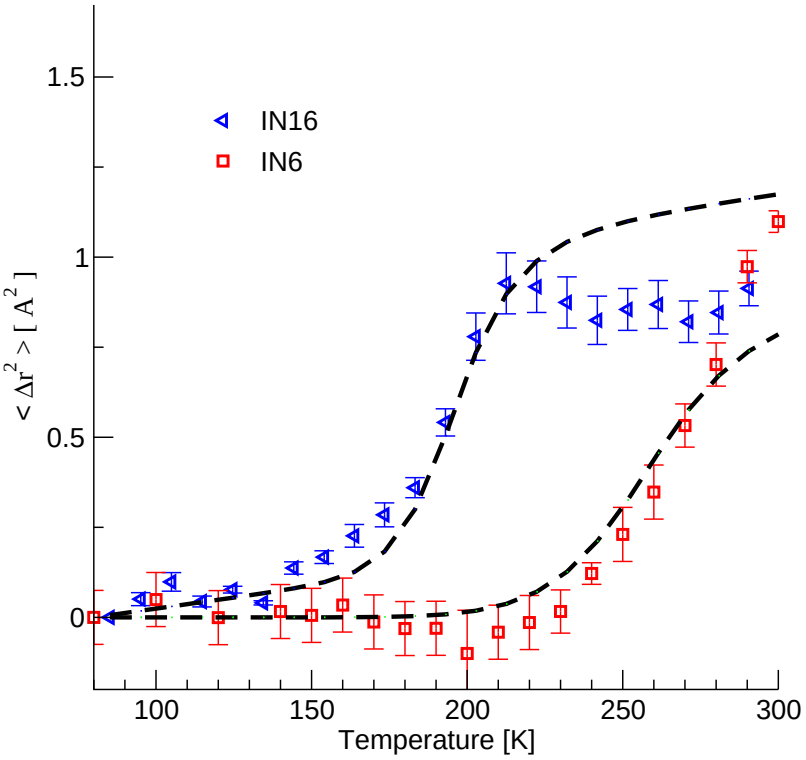


Figure 5.4: $\langle \Delta r^2 \rangle_{exp}$ determined by two different instruments (IN6, IN16) fitted using equation 5.19. Experimental data were taken from Daniel *et al.* (1999).

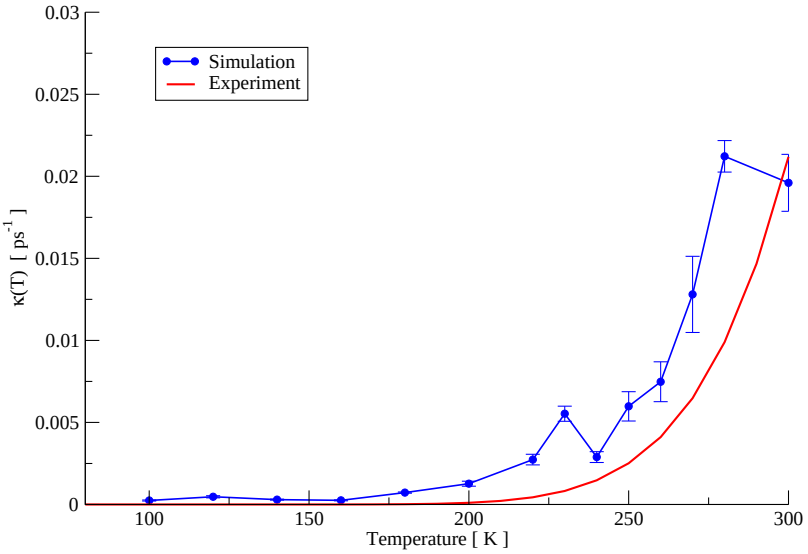


Figure 5.5: Characteristic relaxation time as a function of temperature. $\frac{1}{\kappa(T)}$ was determined by fitting Eq. 5.19 to $\langle \Delta r^2 \rangle_{exp}$

with experiment (Fig. 5.4). In particular, the continued increase in $\langle \Delta r^2 \rangle_{exp}$ above 270K determined using the lower-resolution instrument is not reproduced. However, experimental errors may be significantly higher than estimated at high $\langle \Delta r^2 \rangle$, due to the correspondingly low elastic peak intensities. Furthermore, additional solvent-driven processes are likely to be activated at $\sim$270K due to solvent melting. The assumption of a single relaxation process is unlikely to be valid at $T \geq$270K.

The temperature dependence of the relaxation frequency, κ (red curve) is shown in comparison to the data of Fig. 4.13 in Fig. 5.5. In the temperature range 150K-280K, the relaxation time changes from ca. $100ns$ to $\sim 10ps$. Thus the motion passes through the approximate time resolution windows of the two instruments, which are $5ns$ for IN16 and $100ps$ for IN6. κ obtained from the analysis of the experimentally derived mean-square displacements is in reasonable quantitative agreement with the frequency derived from the simulated intermediate scattering function. $\langle \Delta r^2 \rangle_{A_0}$ is 0.81Å^2, similar to the value obtained in the MD analysis.

5.4 Transition-function relationship

The intriguing aspect of the dynamical transition from a biological point of view is the observed relation to protein function. Various proteins show a pronounced change in their functional characteristics or even cease to be active upon cooling down below the transition temperature (Austin *et al.*, 1975; Rasmussen *et al.*, 1992; Réat *et al.*, 1997). In this chapter a new interpretation of the dynamical transition in proteins has been suggested, its prominent feature being a timescale dependent transition temperature. It shall now be

discussed whether and in which way both transition mechanisms lead to different expectations regarding the temperature dependence of enzyme activity and its relation to the transition in the measured mean-square displacement.

This relation between activity and measured mean-square displacement can be discussed within the framework of two state models like the ones depicted in Figs. 4.2 & 4.3. Two energies play an important role in this context. First, the energy difference ΔG between the states, determining the population probability of these states and the energy barrier ΔB determining the timescale of these transitions.

The argument that the measured mean-square displacement is related to the activity goes as follows: The protein is in either of the two states. The lower state is the 'rigid', inactive state whereas the higher energy state is active. At temperatures $kT \ll \Delta G$ the active state is not populated, the enzyme therefore inactive. With rising temperature the active state becomes populated, the enzyme starts to work.

The expected temperature dependence of the mean-square displacement for this system was shown in Ref. Keller & Debrunner (1980). The contribution of the change from one state to the other is proportional to a Boltzmann factor containing the activation enthalpy ΔG:

$$\langle \Delta r^2 \rangle_c \sim e^{-\frac{\Delta G}{k_B T}} \tag{5.20}$$

Here $\langle \Delta r^2 \rangle_c$ is commonly called the conformational contribution to the total mean-square displacement (Frauenfelder & McMahon, 2000).

Since the increase in mean-square displacement in this model is caused by the change to an active state, this increase marks the onset of activity. Both, the onset of activity and the increase in mean-square displacement are controlled by the same temperature $T_0 \approx \frac{\Delta G}{k_B}$. Measuring the temperature dependence of the mean-square displacement therefore gives access to the enthalpy difference ΔG of functionally important states.

However, this does not take into account that timeresolved scattering techniques only detect motions in a timewindow determined by the instrument used. Whether or not a change in the mean-square displacement is seen for the system of Fig. 4.3 is not only determined by the occupation of states, but also by the timescale on which transitions between these states occur. A given instrument can detect changes between states only if the barrier between these states is low enough that the rate of transitions is within the resolution of the instrument.

This kind of interpretation of neutron scattering data is given by the frequency window point of view suggested here. In the frequency window scenario the increase in mean-square displacement detected by neutron scattering is not the onset of new anharmonic motions due to new states that become occupied, but the transition is rather due to motions becoming fast enough to be detected. The temperature at which motions become observable depends on the resolution of the instrument, so the transition temperature should depend on the instrument used. This interpretation has been shown to explain the more prominent features of the data presented in Ref. Daniel *et al.* (1999) (Fig. 5.1). Regarding enzyme activity, since no 'fixed' transition temperature exists, no correlation should be expected.

It should be mentioned, that this does not imply that these motions are not required for function. Since in this picture the dynamical transition does not mark the onset of new motions but a decrease of their typical timescale only a corresponding increase in the activity should be expected. The interpretation thus is consistent with the Arrhenius behavior of the activity seen in Ref. Daniel *et al.* (1998) even if these motions are required for function.

In accordance with these considerations it was found in Daniel *et al.* (1998) & Dunn *et al.* (2000) that enzyme activity does not show a deviation from Arrhenius behavior down to 180K. The experimental data in Fig. 5.4, however exhibits several transitions. The mean-square displacement as measured by the instrument IN6 shows a transition at ∼220K. For the IN16 data at least two transitions are seen, a non-linear increase in $\langle \Delta r^2 \rangle_{exp}$ at ∼150K and a leveling-off at ∼220K. The absence of a measurable deviation of activity from Arrhenius behavior down to 180K suggests that neither the increase of $\langle \Delta r^2 \rangle_{exp}$ as seen with the instrument IN6 nor the leveling-off seen with IN16 are related to enzyme activity. Since the enzyme activity can be measured down to only 180K the relationship of the ∼150K transition to enzyme activity cannot be addressed.

5.5 Conclusions

The present theoretical analysis shows that the dynamical transition as observed by molecular dynamics simulation and neutron experiment can be described without invoking a temperature dependent change in the long-time, equilibrium atomic dynamics. In the 'frequency window' model used to analyze the data the temperature dependence of dynamical relaxation processes leads to the appearance of dynamical transition behavior in the measured signal as, with increasing temperature, the processes become fast enough to be resolved by the instrument used. The frequency window model is a true dynamical model insofar as the change in slope of the observed $\langle \Delta r^2 \rangle$ at the dynamical transition is due to the temperature dependence of timescales of motions. Assuming activated dynamics the transition would thus be determined by the barriers between energy minima rather than by differences in energies between the minima.

The frequency window model reproduces the broad features of the experimental mean-square displacement results for a protein in a cryosolution and is also in good quantitative agreement with MD simulation data on a smaller protein in the same solvent. The observed timescale dependence of the mean-square displacement rules out an interpretation of the present molecular dynamics or neutron experimental results based solely on the 'equilibrium' model. However, the analysis does not rule out a combination of the frequency window and equilibrium models in which the transition involves both barrier crossing and populating of higher energy states. Indeed, it is highly unlikely that the minima of a protein free energy landscape are equienergetic. Allowing the equilibrium dynamics to vary (by allowing A_0 to vary as in Fig. 4.11) led to a fit to the simulation-derived $I(Q,t)$ that is as good as, though not better, than that achieved with the frequency window model (see Ch. 4). Moreover, some experimental work on proteins has not produced clear evidence for resolution-dependent transition temperatures, although the difference in resolution of the instruments used may not have been sufficient to detect an effect (Doster & Settles, 1999;

Réat *et al.*, 1997).

Work remains to be done in order to tease apart the equilibrium and frequency window contributions to the temperature dependence of atomic fluctuations in proteins and their surrounding solvent, as observed using neutron scattering. One possible way of doing this would be to use single instruments with variable elastic resolution that can then effectively access I(Q,t) at different times (Doster *et al.*, 2001). Alternatively, different instruments with different elastic resolutions can be used, as in Fig. 5.4. Varying the elastic resolution is a procedure akin to examining MD simulations on different timescales. Extending MD timescales and improving instrumental energy resolutions (for example, by using the spin-echo technique (Bellissent-Funel *et al.*, 1998; Dellerue *et al.*, 2000)) will also be important in this regard, as will be experiments with techniques sensitive to longer timescales, such as Mössbauer absorption and NMR relaxation (Parak, 2003).

While spectroscopic techniques are time resolved, X-ray scattering accesses in principle the converged mean-square displacement: $\langle \Delta r^2 \rangle_{Conv}$. If the frequency window model holds there should be no transition due to internal protein dynamics observable in X-ray B-factors. Extraction of protein mean-square displacements *via* X-ray scattering is hampered by the fact that crystallographic B-factors cannot distinguish between dynamic and static contributions to $\langle \Delta r^2 \rangle$ (Frauenfelder *et al.*, 1979). Notwithstanding, there is presently an enthralling debate as to whether X-ray crystal diffraction does indeed detect a temperature dependent transition in internal protein dynamics (Chong *et al.*, 2001; Halle, 2004; Joti *et al.*, 2002; Teeter *et al.*, 2001). A transition in crystalline crambin has been reported (Teeter *et al.*, 2001) while X-ray crystallographic evidence for the absence of a dynamical transition in internal displacements in myoglobin has also been presented (Chong *et al.*, 2001; Joti *et al.*, 2002). A challenge for the future is to understand the temperature dependence of protein dynamics in the framework of the equilibrium and frequency window descriptions and in the context of typical timescales for protein activity.

Chapter 6

Gaussian approximation of the elastic scattering function

Non-Gaussian scattering is shown to cause systematic underestimation of measured protein mean-square displacements. Non-Gaussian scattering below $Q^2 < 6\AA^{-2}$ is shown to arise mainly from the heterogeneity of atomic mean-square displacements.

An analysis of the mean-square displacement as measured by elastic incoherent neutron scattering was given in Ch. 5. Central to the derivation of $\langle \Delta r^2 \rangle_{exp}$ is the Gaussian approximation (see Sec. 2.6). In this approximation, the Q-dependence of $S(Q,0)$ is given by the function $\exp(-1/6 Q^2 \langle \Delta r^2 \rangle)$, *i.e.* plotting the natural logarithm of the elastic intensity *vs* Q^2 shows a straight line with slope $-1/6\langle \Delta r^2 \rangle$. In general it can not be expected that the Gaussian approximation holds true over a large range of Q-values. Only in the limit $Q \to 0$ the Gaussian approximation is exact and as a rule of thumb it is often assumed to be valid as long as $\langle \Delta r^2 \rangle Q^2 \lesssim 1$. With increasing values of Q, $ln[S(Q,0)]$ starts to deviate from the straight line expected from Gaussian scattering. This non-Gaussian scattering is the topic of this chapter.

Non-Gaussian scattering may arise from two aspects, dynamic or static. First, looking at a single atom, there can be non-Gaussian scattering due to dynamics of this atom, *i.e.* the motions of the atom do not lead to a Gaussian distribution of displacements. In contrast, even in a system of Gaussian scatterers, non-Gaussian scattering arises if there is heterogeneity in the mean-square displacements of single atoms. In such systems the elastic scattering is described by a sum of Gaussians (see Eq. 2.43) which itself is not Gaussian anymore.

It is an ongoing debate in the context of the dynamical transition which of the two possibilities dominate the Q-dependence of the elastic scattering function $S(Q,0)$ (Doster & Settles, 1999; Doster *et al.*, 1989b; Gabe *et al.*, 2002; Parak, 2003; Réat *et al.*, 1997; Zaccai, 2000b).

The dynamical point of view is taken for example in Doster & Settles (1999); Doster *et al.* (1989b). Inspired by work on the rebinding kinetics of myoglobin a two-state model for the dynamics of proteins was proposed. The full elastic scattering function of such a model is

used to fit experimental spectra in the range $0 \leq Q^2 \leq 20\text{Å}^{-2}$ and derive the dynamical transition behavior. The deviation from Gaussian scattering is thus assumed to origin from dynamical aspects of the system.

The opposite point of view has been put forward in Fitter *et al.* (1997); Gabe *et al.* (2002); Zaccai (2000b). Two linear regions in the spectrum of bacteriorhodopsin are identified and interpreted as two distinct populations of high and low $\langle \Delta r^2 \rangle$, respectively. Further experiments on partly deuterated bacteriorhodopsin suggested to identify these two populations with the switch region around the retinal (low $\langle \Delta r^2 \rangle$) and the surrounding protein (high $\langle \Delta r^2 \rangle$) (Réat *et al.*, 1997). Interestingly, Engler *et al.* (2003) reexamined the data of Doster *et al.* (1989b) on myoglobin and showed that it can be as well interpreted, assuming three different populations of mean-square displacements.

Non-Gaussian scattering, unless negligible in magnitude, will influence experimentally obtained mean-square displacements in two ways. First, for systems with strong non-Gaussian scattering contributions, the Q-range over which the Gaussian approximation is valid, will be small. Typically, the experimentally accessible low Q-range is limited either by the small-angle coherent scattering region or by the geometrical arrangements of detectors. If the Q-range allowed by the Gaussian approximation is small, there may not be enough data points to accurately determine $\langle \Delta r^2 \rangle_{exp}$. Second, even in regions where $ln[S(Q,0)]$ is approximately linear with Q^2, $\langle \Delta r^2 \rangle_{exp}$ can be quantitatively wrong due to non-Gaussian scattering. In Ref. Hayward & Smith (2002) it was shown, that even in the low Q-region, experimental analysis systematically underestimates $\langle \Delta r^2 \rangle$. Thus, one task in experimental analysis is to determine the initial slope of $S(Q,0)$ as accurately as possible even in situations where non-Gaussian scattering contributes considerably to $ln[S(Q,0)]$.

Computer simulations can be used to access the range of validity of assumptions inherent in experimental analysis. The analysis of calculated spectra allows to compare directly with the underlying system properties accessible *via* the atomic detailed description offered by molecular dynamics simulation. In Hayward & Smith (2002) mean-square displacements calculated directly from MD simulations were compared to those derived from simulations following the protocol of experimental data analysis. A reduction of $\langle \Delta r^2 \rangle_{exp}$ of up to 30% was observed for instrumental resolutions available in current spectrometers. Comparing $\langle \Delta r^2 \rangle_{exp}$ derived from spectra without convolution of a resolution function suggested that motional heterogeneity itself leads to considerable underestimation of mean-square displacements.

6.1 Q-dependence of $S(Q,0)$

The Q-dependence of the scattering function contains information about spatial characteristics of dynamical processes. Thus, in principle analysis of this dependence enables to distinguish between different dynamical models, such as diffusion models, often strictly obeying the Gaussian approximation, and jump models showing considerable deviations at higher values of Q (see Bee (1988)). For complex, heterogeneous systems this dynamic information is intermingled with contributions reflecting the variety of components contributing

to the total spectrum. Thus, analysis of the Q-dependence of $S(Q,0)$ is hampered by the necessity to estimate and disentangle contributions of both, static and dynamic origin.

Here, both these contributions will be expanded around the Gaussian approximation, *i.e.* the deviations from Gaussian scattering are formally written in a power series of the scattering vector Q. Both expansions are formally the same, thus without additional information on the observed system, observation of the Q-dependence of elastic scattering does not distinguish between both sources of non-Gaussian scattering. In other words, to every dynamical process leading to a certain shape of $S(Q,0)$, a static distribution of mean-square displacements can be defined resulting in the same $S(Q,0)$.

Though static and dynamic contributions might be difficult to disentangle, the power series expansion does lead to an approximation of elastic scattering around the Gaussian contribution that allows for a quantitatively improved derivation of $\langle \Delta r^2 \rangle$ (see Sec. 6.2).

6.1.1 Non-Gaussian single-atom scattering

Non-Gaussian behavior can arise from single atoms when their dynamics does not lead to a Gaussian distribution of displacements. This case was thoroughly treated already in Ref. Rahman *et al.* (1962) and is briefly summarized here. For single atoms the intermediate scattering function, $I(Q,t)$ can be written as:

$$I(Q,t) = e^{-\sum_{l=1}^{\infty}(Q^2)^l \gamma_l(t)} \tag{6.1}$$

where the $\gamma_l(t)$ are defined by l-point velocity autocorrelation functions. In Ref. Rahman *et al.* (1962) it is shown that if $\gamma_l(t) \rightarrow C_l$ rapidly enough with time then the elastic scattering can be written as:

$$S(Q,0) = e^{-\sum_{l=1}^{\infty}(Q^2)^l C_l} \tag{6.2}$$

Explicit calculation or comparison with the Gaussian approximation shows that the first constant is given by $C_1 = \frac{\langle \Delta r^2 \rangle}{6}$

As mentioned, it is central to experimental analysis that Eq. 6.2 remains close to a Gaussian function determined by the mean-square displacement of the system. With that in mind the RHS of Eq. 6.2 can be expanded as:

$$\begin{aligned} S(Q,0) &= e^{-\sum_l (Q^2)^l C_l} = e^{-Q^2 C_1} e^{-\sum_{l>1}(Q^2)^l C_l} \\ &= e^{-\frac{1}{6}Q^2\langle \Delta r^2 \rangle} \left(1 + \sum_{m=2}^{\infty} b_m \cdot (-Q^2)^m\right) \end{aligned} \tag{6.3}$$

where the last equality uses the relation between C_1 and $\langle \Delta r^2 \rangle$. The power series expansion in Q^2 expresses the idea, that corrections to the Gaussian approximation are small for low Q. The parameters b_m are combinations of the C_l (*e.g.* $b_2 = C_2$) and are therefore determined by the dynamics of the system.

6.1.2 Dynamical heterogeneity

Dynamical heterogeneity, *i.e.* a distribution of mean-square displacements also leads to corrections to the Gaussian approximation. Fig. 6.1 shows the average mean-square displacement per residue obtained from molecular dynamics simulations of the system described in Sec. 3.3. Different colors indicate different temperatures. For low temperatures

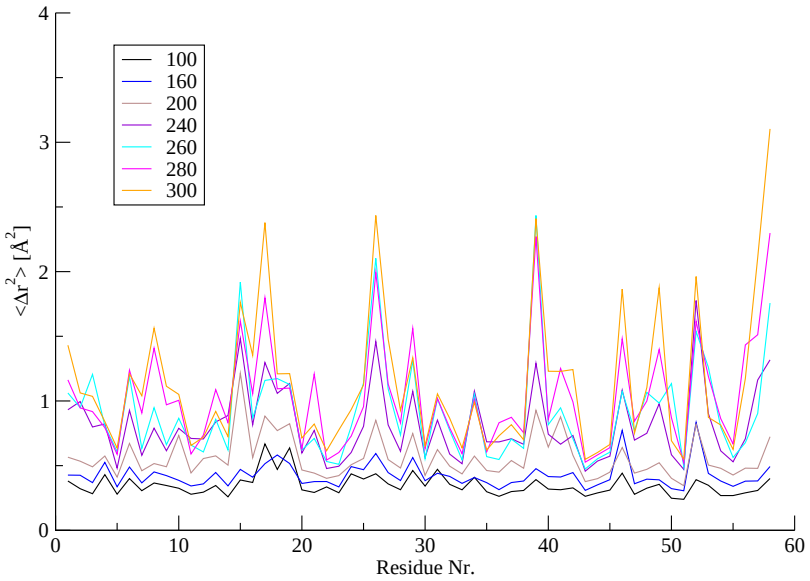


Figure 6.1: Temperature dependence of motional heterogeneity. Mean-square displacementa were calculated per residue from internal motions of BPTI in solution

($T \lesssim 180K$) mean-square displacements are small and comparable throughout the sample. At higher temperatures, however, considerable differences in $\langle \Delta r^2 \rangle$ can be observed from residue to residue, expressing the heterogeneity of dynamical processes of different constituents of the protein. Thus, non-Gaussian scattering arising from heterogeneity can be expected to increase with temperature.

This contribution can be accessed by assuming Gaussian single-atom scattering. The elastic scattering function then reads:

$$S(Q,0) = \frac{1}{N} \sum_{i}^{N} e^{-\frac{1}{6}Q^2 \langle \Delta r_i^2 \rangle} \tag{6.4}$$

where i denotes atoms $i = 1, \ldots, N$.

Although a sum of Gaussians is not itself Gaussian, Eq. 6.4 can again be rewritten to

express the proximity to a Gaussian function. Formally, we rewrite Eq. 6.4 as follows:

$$S(Q,0) = e^{-\frac{1}{6}Q^2\langle\Delta r^2\rangle}\left(\frac{1}{N}\sum_i^N e^{-\frac{1}{6}Q^2(\langle\Delta r_i^2\rangle-\langle\Delta r^2\rangle)}\right) \tag{6.5}$$

$$= e^{-\frac{1}{6}Q^2\langle\Delta r^2\rangle}\left(\sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{-Q^2}{6}\right)^m\mu(m)\right) \tag{6.6}$$

$$\tag{6.7}$$

Here, $\langle\Delta r^2\rangle$ is the average mean-square displacement of the system and $\mu(m)$ is the m^{th} central moment of the distribution of $\langle\Delta r^2\rangle$. Thus, in systems where heterogeneity is the dominating contribution to non-Gaussian behavior, the elastic scattering can, in principle, be used to obtain experimentally the variance and higher statistical moments of the distribution of mean-square displacements.

It should be noted here, that the first central moment $\mu(1) = \frac{1}{N}\sum_i^N(\langle\Delta r_i^2\rangle - \langle\Delta r^2\rangle) = 0$, thus the first correction to the Gaussian approximation is of the order Q^4. Thus the expansions Eqs. 6.3 & 6.6 are formally equivalent.

6.2 Non-Gaussian scattering in MD Simulations

Theoretical analysis of the scattering law elucidates the possibility and form of non-Gaussian scattering in complex systems. Whether and to what extent non-Gaussian scattering is present in or even dominates elastic scattering of a complex system can only be addressed by experiment or computational studies.

Here, molecular dynamics simulations were used to address two questions related to the existence of non-Gaussian behavior in elastic scattering spectra of proteins. The first question examined is, whether including corrections to the Gaussian approximation leads to substantial improvement of $\langle\Delta r^2\rangle_{exp}$. The second question asks about the origin of non-Gaussian scattering in the simulated system. Two opposing possibilities were introduced and it will be shown, that for values of $Q^2 \leq 6\text{Å}^{-2}$ non-Gaussian scattering is dominated by the observed heterogeneity of the atomic mean-square displacements (see Fig. 6.1).

6.2.1 The measured mean-square displacement

The measured mean-square displacement does depend on the energy resolution of the instrument. This dependence may lead to a substantial underestimation of the *true* $\langle\Delta r^2\rangle$ (Hayward & Smith, 2002). The errors introduced by finite resolution can not be solved by any analysis method. Only the expected error in the analysis may be estimated. However, in addition to this instrument inherent error, there is a second aspect in the derivation of mean-square displacements that leads to an underestimation of $\langle\Delta r^2\rangle_{exp}$. The way $\langle\Delta r^2\rangle_{exp}$ is derived from experimental data, itself results in a systematically underestimated mean-square displacement due to non-Gaussian scattering. The linear function commonly

fitted to the low Q-region of $ln[S(Q,0)]$ best approximates all data points in the chosen Q-range, thereby compromising between the *true* initial slope of elastic scattering and the non-Gaussian contributions at higher Q-values.

To estimate the error associated with the Gaussian approximation and to test whether a description of $S(Q,0)$ along the lines of Eqs. 6.3 & 6.6 leads to a better determination of the *true* initial slope, the following strategy was applied. First, the elastic scattering function $S(Q,0)$ was calculated from the trajectories described in Sec. 3.3 with the software package nMOLDYN (Kneller *et al.*, 1995). In Ref. Hayward & Smith (2002) it was shown, that instrumental resolution alone already leads to a reduction of $\langle \Delta r^2 \rangle_{exp}$ compared to $\langle \Delta r^2 \rangle$ directly calculated from simulations. To access the error related to an underestimation of the initial slope alone, the reference value for comparison with experimentally derived mean-square displacements was determined from the elastic scattering function, $S(Q,0)$. To obtain a reference $\langle \Delta r^2 \rangle$, the very low Q-region, $0 \leq Q^2 \leq 0.2\text{Å}^{-2}$ was fitted with a linear function. This very low region, not accessible by experimental analysis, serves as a good estimate of the initial slope of $ln[S(Q,0)]$. The calculated spectra and the resulting fits are shown in Fig. 6.2. Equipped with this reference, possible analysis schemes can be examined.

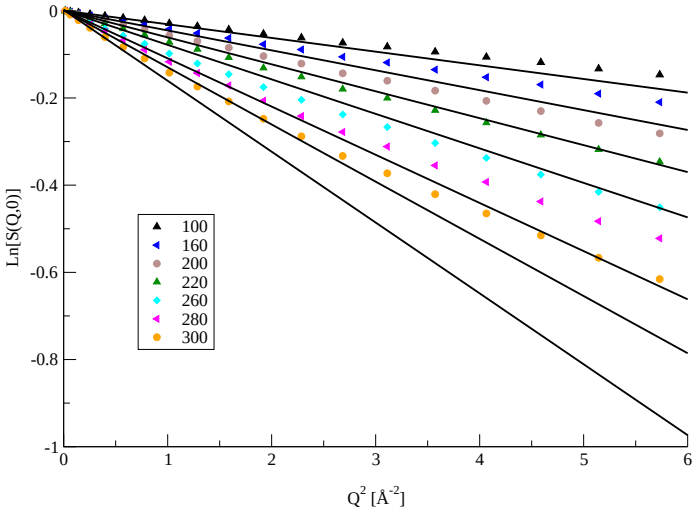


Figure 6.2: Determination of the reference $\langle \Delta r^2 \rangle$. Symbols: Ln[S(Q,0)] calculated from MD-simulations. Lines: Linear fit to the very low Q-range, $0 \leq Q^2 \leq 0.2\text{Å}^{-2}$.

Here, we concentrate on the behavior of the mean-square displacement determined *via* the Gaussian approximation (linear regression) compared to a non-linear fit of functions of the form of Eq. 6.6 with various numbers of correction terms. The accuracy of these models for different accessible Q-ranges is studied as well as the increase in accuracy upon including higher order terms in Eq. 6.6.

Two different Q-ranges were examined here. A low Q-region, $0 \leq Q^2 \leq 1.4\text{Å}^{-2}$, which is commonly used to derive $\langle \Delta r^2 \rangle_{exp}$ from instruments like IN6 or IN16 and a high Q-range, $0.2 \leq Q^2 \leq 3\text{Å}^{-2}$, used for data analysis on the instrument IN13. Figs. 6.3 & 6.4 show

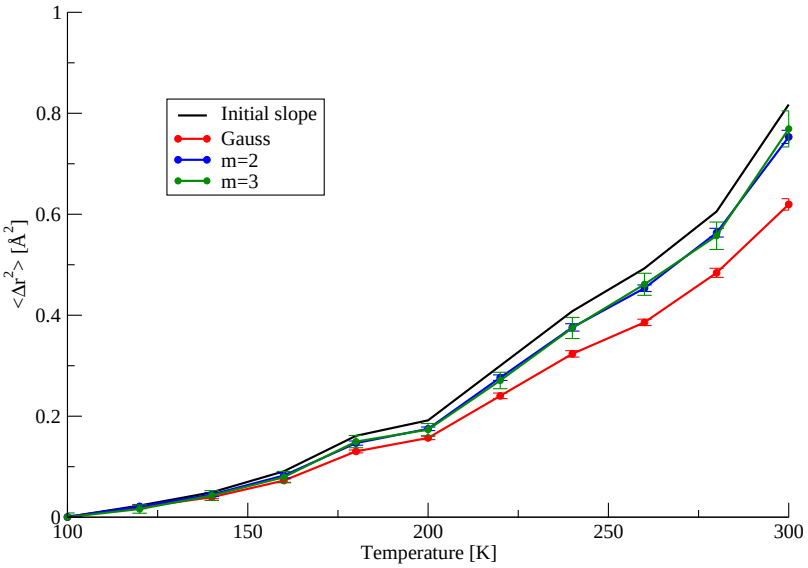


Figure 6.3: $\langle \Delta r^2 \rangle_{exp}$ as measured by instrument IN6 compared to the reference values obtained from Fig. 6.2 (black). (○) linear fit. (○) incorporating one correction term in Eq. 6.6 (m=2). (○) incorporating two correction terms (m=3)

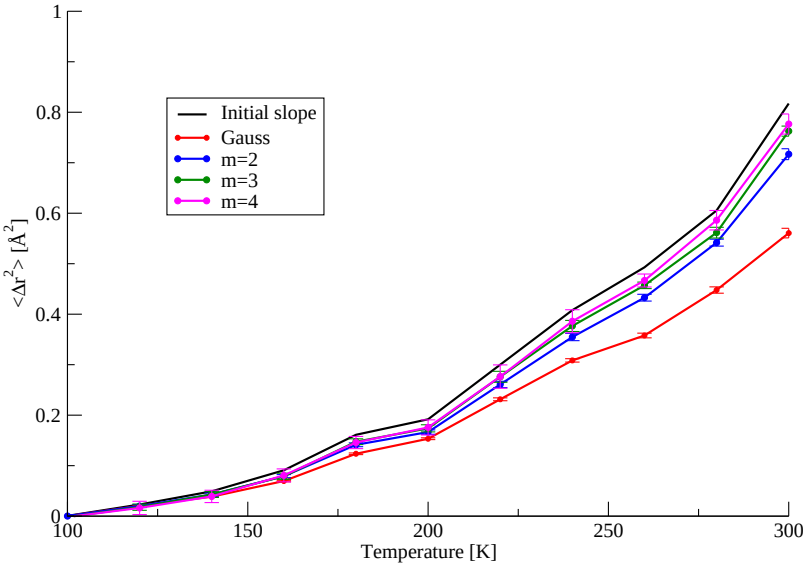


Figure 6.4: $\langle \Delta r^2 \rangle_{exp}$ as measured by instrument IN13 compared to the reference values obtained from Fig. 6.2 (black). (○) linear fit. (○) incorporating one correction term in Eq. 6.6 (m=2). (○) incorporating two correction terms (m=3) (○) three correction terms (m=4)

the results of fitting various corrections to the Gaussian approximation to the calculated spectra.

In Fig. 6.3 $\langle \Delta r^2 \rangle_{exp}$ determined by fitting various approximations to the low Q-range is shown. Three fitting models were tested. The Gaussian approximation (red) underestimates the initial slope (black) by up to 20% (300K). Introducing a single correction term of Eq. 6.6 (blue), *i.e.* a correction term with Q^4 dependence, reduces the error to 6%. Including higher order corrections (green) does not further improve the estimation of the initial slope. From an experimental point of view this is rather satisfactory since this low Q-region hardly encompasses enough data points to allow for more then 3 fitting parameters. Only in computer simulations, not restricted by any detector geometry, enough data points can be calculated to render fits to higher order corrections well defined.

Fig. 6.4 depicts the corresponding results for the Q-range accessible to the instrument IN13. $\langle \Delta r^2 \rangle_{exp}$ derived by the Gaussian approximation (red) underestimates the initial slope by up to 25% (300K). Again substantial improvement can be achieved by taking into account the first correction in Eq. 6.6. Thereupon the error drops down to 6%. Thus, although fitting to the large Q-range, the error of $\langle \Delta r^2 \rangle_{exp}$ obtained for one correction term did not increase compared to the low Q-region. Without losing accuracy one can thus take advantage of the increase in data points to reduce additional errors related to the statistics of available data points.

The increase in data points may renders the introduction of higher order corrections feasible. The incorporating of a second correction (green) improved the obtained $\langle \Delta r^2 \rangle_{exp}$, the corresponding error now being as small as 4%. Further correction terms (magenta) did not lead to substantial improvement.

Keeping in mind the limited amount of data points available in experimental analysis, corrections to the Gaussian approximation should be restricted to one or two terms for the low and high Q-region, respectively. Including these corrections leads to considerable improvement of measured mean-square displacements.

6.2.2 Dynamic vs Heterogeneity – Q-dependence of S(Q,0)

Non-Gaussian scattering has profound effects on the results of elastic incoherent neutron scattering experiments. This alone draws attention to the question about the origin of non-Gaussian scattering in proteins. A second aspect of interest is added, again in the context of the dynamical transition. Both possible origins, dynamics and heterogeneity, have been exploited to derive information from the Q-dependence of elastic scattering. The dynamic point of view suggests to use the additional information stored in the Q-dependence of elastic scattering to adjust parameters of a given dynamical model, *e.g.* the asymmetric energy landscape depicted in Fig. 4.3 (Doster *et al.*, 1989b). Heterogeneity of mean-square displacements was assumed to be responsible for non-Gaussian scattering for example in Réat *et al.* (1997) and a population of high and of low mean-square displacement was introduced to account for the observed non-Gaussian scattering.

As shown, the two scenarios can not be easily distinguished by the shape of the scattering

function (compare Eqs. 6.3 & 6.6). Further information about the system is required to unambiguously identify the main contributions. To address this question in the present system the calculated spectra of Fig. 6.2 were compared to spectra of a heterogeneous ensemble of perfect Gaussian scatterers with the same distribution of $\langle \Delta r^2 \rangle$ as the present system. Therefore, spectra were directly calculated from Eq. 6.4 with the simulation-derived distribution of mean-square displacements.

The distribution of mean-square displacements depends on the timescale of motions investigated. To compare to elastic scattering arising in a system of Gaussian scatterers, it is therefore important to access the distribution of $\langle \Delta r^2 \rangle$ for the timescale observed in the comparison spectra. This was achieved by adjusting the timescale such that the average mean-square displacement reproduces the initial slope of the comparison spectra. This timescale was then used to calculate the full distribution of $\langle \Delta r^2 \rangle$ and determine $S(Q,0)$ according to Eq. 6.4. Fig. 6.5 shows a comparison of the resulting spectra. The *true* elastic

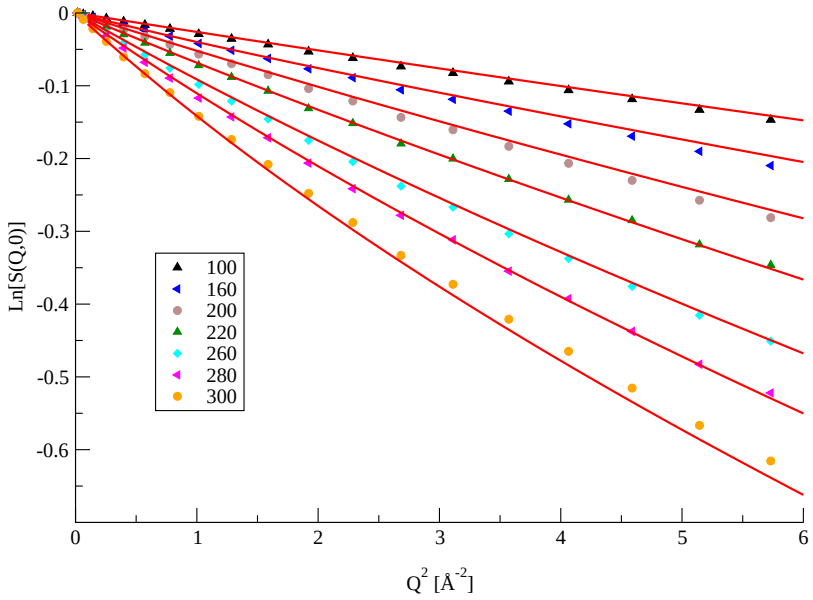


Figure 6.5: Non-Gaussian scattering for $0 \leq Q \leq 6 \text{Å}^{-2}$. Comparison of simulated system (symbols) with a heterogeneous system of Gaussian scatterers (lines).

scattering function is plotted as symbols (same data as in Fig. 6.2), the corresponding spectra of a heterogeneous sample of Gaussian scatterers is plotted as lines. For all temperatures a remarkable agreement between both spectra is obtained, thus suggesting that the non-Gaussian scattering in the present system is clearly dominated by contributions due to the heterogeneity of mean-square displacements.

In the context of Fig. 6.1 it was deduced that non-Gaussian scattering arising from heterogeneity should increase in temperature. Fig. 6.6 further supports this statement. Shown is the distribution of mean-square displacements leading to the graphs (solid lines) in Fig. 6.5. Starting from a narrow distribution concentrated around the corresponding mean-value

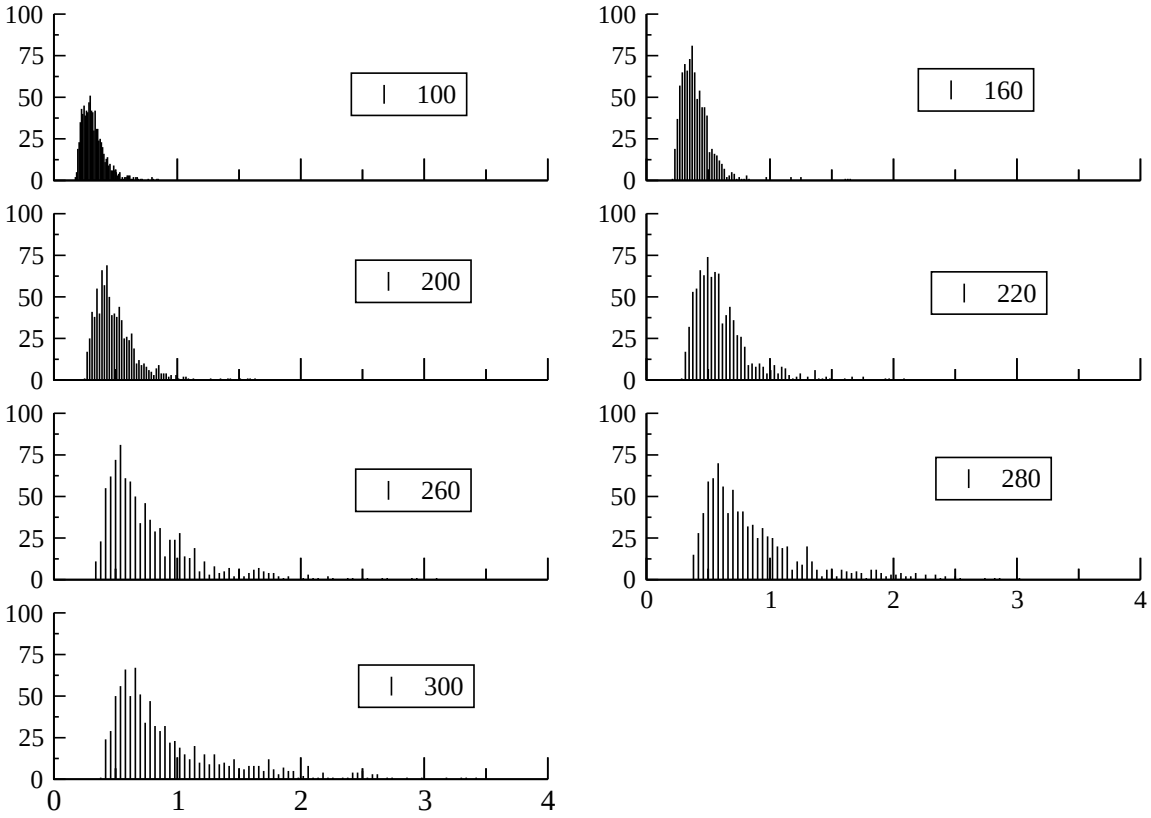


Figure 6.6: Temperature dependence of the distribution of mean-square displacements. Atomic mean-square displacements were binned with a bin-width of 0.02Å.

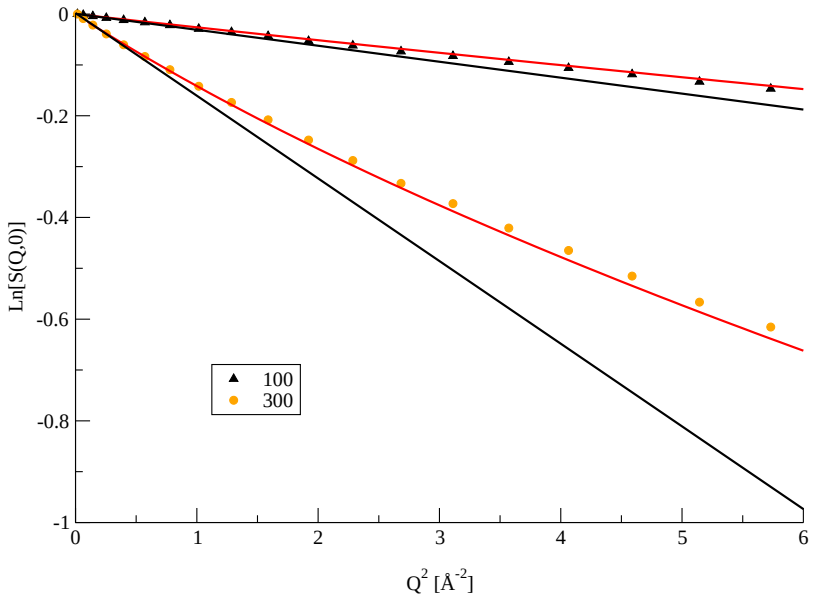


Figure 6.7: Temperature dependence of non-Gaussian scattering. Gaussian scattering (black) compared to a distribution of Gaussian scatterers (red) and the full trajectory (symbols).

($T = 100K$), the distribution clearly broadens upon temperature increase. The peak of the distribution slightly moves to higher values and dissolves into a tail of increasing length. For the amount of non-Gaussian scattering it is not the shift in the peak position that is important but the formation of the broad tail upon temperature rising. The mean-value of the distribution determines the initial slope of the elastic scattering function, the width however, determines the amount of non-Gaussian scattering present in the system.

Fig. 6.7 indeed shows a considerable increase of non-Gaussian contributions. For 100K and 300K the data of Fig. 6.5 are plotted again together with a Gaussian scattering function of the same average mean-square displacement. Whereas the Gaussian function can be considered a reasonable approximation to the *true* scattering function for 100K it clearly fails to reproduce the spectra at 300K.

6.3 Conclusions

In this chapter it was demonstrated that computer simulations are a valuable tool to test and examine experimental methods and the errors inherent in the underlying assumptions. The atomic detailed description of molecular dynamics simulations may serve as *experimental* data set as well as it provides reference values. References may be derived either directly from trajectories (Hayward & Smith, 2002) or use can be made of the precision and range accessible by simulations.

In the present work, based on theoretical considerations on the Q-dependence of the elastic incoherent scattering function, the origin and effects of non-Gaussian scattering were discussed. A formal expansion of the scattering function around the Gaussian contribution suggests a functional form to include corrections to the Gaussian approximation in experimental analysis.

The deviation of the scattering function from linear behavior leads to an underestimation of $\langle \Delta r^2 \rangle$. Applying the Gaussian approximation in the Q-range accessible by current instruments led to an underestimation of $\langle \Delta r^2 \rangle$ by 20% (IN6) or even 25% (IN13). A non-linear analysis of elastic scattering, incorporating corrections to the Gaussian approximation, substantially improved the derived mean-square displacements. Incorporating only the lowest order correction, a reduction of the error down to 6% could be achieved for both Q-ranges investigated. Including further corrections did improve results in the high Q-range. However, whether or not it is feasible to incorporate higher corrections in experimental analysis critically depends on the amount of data points available and their respective quality. Including higher order corrections inevitably leads to a better reproduction of data points not necessarily wanted in noisy spectra.

The improvement in $\langle \Delta r^2 \rangle_{exp}$ here achieved is independent of the origin of non-Gaussian scattering since both possible sources lead to the same fitting functions. However, the Q-dependence of $S(Q,0)$ contains more information than the mean-square displacement of the system. Non-Gaussian behavior may arise if the dynamics of single atoms itself leads to non-Gaussian scattering, or it has its origin in the heterogeneity of the system investigated. Spatial characteristics of dynamical processes are thus intermingled with information about the distribution of mean-square displacements.

Here, the origin of non-Gaussian scattering of a small protein (BPTI) in solution was examined with the help of computer simulations. Spectra directly calculated from trajectories were compared to those calculated from a system of Gaussian scatterers possessing the same distribution of $\langle \Delta r^2 \rangle$ as the simulated system. The spectra are in good agreement for all temperatures. The increase in non-Gaussian scattering upon temperature rising, expected from the increase of width in the distribution of $\langle \Delta r^2 \rangle$, is reflected in the scattering spectra directly calculated from molecular dynamics trajectories.

Chapter 7

Protein association - a neutron scattering study

A method to derive the vibrational density of states on an absolute scale from low-temperature inelastic incoherent neutron scattering is suggested and examined for ligand binding of the protein dihydrofolate reductase (DHFR). The change in the vibrational density of states of DHFR upon binding a ligand methotrexate (MTX) is determined. The vibrations of the complex soften significantly relative to the unbound protein. The resulting free energy change, which is directly determined by the density of states change, is found to contribute significantly to the binding equilibrium.

An understanding of how ligands bind to proteins is of fundamental importance in biology and medicine (Benkovic *et al.*, 1988; Gilson *et al.*, 1997; Klotz, 1985; Lamb & Jorgensen, 1997; Lian *et al.*, 1994; Wang & et al., 2001). Protein:ligand association has been assumed to be dominated by factors such as the hydrophobic effect, hydrogen bonding, electrostatic and van-der-Waals interactions. However, as early as 1963 it was suggested that an additional mechanism might exist, due to increased flexibility in the protein:ligand complex manifested by a change in the spectrum of vibrations due to formation in the complex of new, intermolecular interactions (Erickson, 1989; Finkelstein & Janin, 1989; Fischer *et al.*, 2001; Page & Jencks, 1971; Steinberg & Scheraga, 1963; Sturtevant, 1977; Tidor & Karplus, 1994). Theoretical normal mode analyses, used to estimate this vibrational change on insulin dimerization (Tidor & Karplus, 1994) and on water binding to bovine pancreatic trypsin inhibitor (Fischer *et al.*, 2001), have suggested that the effect is likely to be thermodynamically important. However, experimental determination of the vibrational change has been lacking. Inelastic neutron scattering, in which the dynamic structure factor $S(Q, \omega)$ is measured as a function of the scattering wavevector, Q, and energy transfer, $\hbar\omega$ has been used to determine the vibrational frequency distribution, $g(\omega)$ for several proteins (Cusack *et al.*, 1988; Doster *et al.*, 1989b; Smith *et al.*, 1986). Here, we present a method to experimentally determine the change in $g(\omega)$ accompanied with association processes of biological molecules. This determination allows thermodynamic quantities associated with the vibrational change to be derived. The enzyme chosen is dihydrofolate reductase (DHFR), an important target for anticancer and antibacterial drugs (Epstein *et al.*, 1995;

Howell *et al.*, 1986; Kamiyama & Gekko, 2000; Sawaya & Kraut, 1997; Stone & Morrison, 1984). DHFR catalyzes the reduction of dihydrofolate to tetrahydrofolate in the presence of nicotinamide adenine dinucleotide phosphate (NADPH) cofactor.

7.1 Thermodynamics of association processes

From a thermodynamic point of view biological association processes are determined by the free energy of association.

$$\Delta G^0 = -k_B \log(K_{eq}) \tag{7.1}$$

where ΔG^o is the standard free energy and K_{eq} the equilibrium constant of the association process $A + B \rightleftarrows AB$.

The free energy itself can be decomposed into two parts, an enthalpic and an entropic part.

$$\Delta G = \Delta H - T\Delta S \tag{7.2}$$

$$= -k_B T \ln\left(\frac{Z_{Dim}}{Z_{Mon}}\right) \tag{7.3}$$

Z_i denotes the partition function of the dimer or monomer, respectively. Eq. 7.3 relates the macroscopic observable free energy to the microscopic states of the system.

Usually the enthalpic part ΔH is considered to favor protein dimerization whereas the entropic part is considered unfavorable due to loss of three translational and three rotational degrees of freedom:

- **unfavorable:**
  - loss of 3 rotational degrees of freedom
  - loss of 3 translational degrees of freedom
- **favorable:**
  - hydrophobic effects
  - hydrogen bonding
  - electrostatic/van der Waals interactions
  - gain of six internal vibrational degrees of freedom

From Table 7.1, however, we see that the loss of translational and vibrational degrees of freedom is counterbalanced by an increase in internal degrees of freedom. In addition the vibrational spectrum might change upon complex formation; sloppy speaking the system might get stiffer or more flexible. *A priori*, it is not clear what the net entropic contribution to the free energy of complex formation is, since positive and negative contributions

might cancel out. Indeed, normal mode calculations on insulin suggested that entropic contributions due to changes in the vibrational spectrum of an insulin dimer compared to an isolated insulin are of the same order of magnitude as contribution due to the loss of translational/rotational degrees of freedom (Tidor & Karplus, 1994).

Knowing the vibrational spectra of a system one can calculate the partition function and therefore the vibrational contributions to entropy and free energy via Eq. 7.3:

$$\Delta A_{vib} = \sum_{l=1}^{3N-6} \left\{ \left[\frac{1}{2}\hbar\omega + \frac{\hbar\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1} \right] - \left[\frac{\hbar\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1} - k_B T \ln(1 - e^{-\frac{\hbar\omega}{k_B T}}) \right] \right\} \tag{7.4}$$

$$\Delta H_{vib} = \sum_{l=1}^{3N-6} \left[\frac{1}{2}\hbar\omega + \frac{\hbar\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1} \right] \tag{7.5}$$

$$T\Delta S_{vib} = \sum_{l=1}^{3N-6} \left[\frac{\hbar\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1} - k_B T \ln(1 - e^{-\frac{\hbar\omega}{k_B T}}) \right] \tag{7.6}$$

Neutron scattering, as will be shown in the next section, provides the facility to measure the vibrational density of states of the scattering system. Hence, it is possible to experimentally access the entropy contribution to the free energy of complex formation due to the increase in internal degrees of freedom and changes in vibrations upon protein association.

7.2 Neutron scattering and vibrational density of states

The scattering function $S(\mathbf{Q}, \omega)$ can be divided according to the energy change of the neutron into an elastic and an inelastic part as follows

$$S(\mathbf{Q}, \omega) = S_{el/qe}(Q, \omega) + S'_{inelas}(\mathbf{Q}, \omega) \tag{7.7}$$

With regard to entropy changes due to dimerization it is the inelastic part that is of most interest, since it gives access to the vibrational density of states in the limit of the one-phonon approximation discussed in the next section.

7.2.1 One-phonon approximation

The one-phonon approximation includes two independent assumptions. First the harmonic approximation, *i.e.* the system is assumed to vibrate harmonically around its minimum position. This approximation can be expected to hold at low temperatures. The second approximation assumes that the interaction of the sample and the neutron is dominated by one-phonon exchange, that is during the scattering event the sample either gains or looses the energy of one phonon. This approximation can be considered to be reasonable at moderate interaction energies due to thermal neutrons (meV-range).

The inelastic incoherent scattering function $S'(\mathbf{Q},\omega)$ can be written:

$$\begin{aligned} S'(\mathbf{Q},\omega) &= \frac{1}{2\pi\hbar N}\int_{-\infty}^{\infty} dt\ e^{-i\omega t}\sum_l b_l^2 \exp(-2W_l(\mathbf{Q})) \\ &\quad \times \sum_d \left[\exp(< \mathbf{Qu}(l,d,0)\mathbf{Qu}(l,d,t) >) - 1\right] \end{aligned} \tag{7.8}$$

Here $W_l(\mathbf{Q})$ is the Debye-Waller factor of atom l and $\mathbf{u}(l,d,t)$ denotes the displacement of atom l from its equilibrium position in the d^{th} unit cell.

The one-phonon approximation we obtain by expanding the exponential and keeping only the quadratic term in $\mathbf{Q}$:

$$\begin{aligned} S'(\mathbf{Q},\omega) &= \frac{1}{2\pi\hbar N}\int_{-\infty}^{\infty} dt\ e^{-i\omega t}\sum_l \exp(-2W_l(\mathbf{Q})) \\ &\quad \times \sum_d < \mathbf{Qu}(l,d,0)\mathbf{Qu}(l,d,t) > \end{aligned} \tag{7.9}$$

Using the harmonic assumption the correlation function $< \mathbf{Qu}(l,d,0)\mathbf{Qu}(l,d,t) >$ can be calculated to yield:

$$< \mathbf{Qu}(l,d,0)\mathbf{Qu}(l,d,t) >= \frac{\hbar}{2m_l}\sum_{j,\mathbf{q}} \frac{|\mathbf{Q}\sigma_l^j(\mathbf{q})|^2}{\omega_j(\mathbf{q})}\left[n_j(\mathbf{q})e^{-it\omega_j(\mathbf{q})} + (n_j(\mathbf{q})+1)e^{it\omega_j(\mathbf{q})}\right] \tag{7.10}$$

where m_l is the mass of atom l and the sum is taken over all possible lattice modes. Inserting Eq. 7.10 into Eq. 7.9 finally yields:

$$S(\mathbf{Q},\omega) = \sum_l b_l^2 \frac{e^{-2W_l(\mathbf{Q})}}{2Nm_l}\sum_{j,\mathbf{q}} \frac{|\mathbf{Q}\sigma_l^j(\mathbf{q})|^2}{\omega_j(\mathbf{q})}\left[n_j(\mathbf{q})\delta(\omega-\omega_j(\mathbf{q})) + (n_j(\mathbf{q})+1)\delta(\omega+\omega_j(\mathbf{q}))\right] \tag{7.11}$$

In general, the polarization vectors $\sigma_l^j(\mathbf{q})$ are not known and therefore Eq. 7.11 cannot be further simplified.

However, in a powder sample the lattice vectors $\mathbf{q}$ can be considered to be isotropically distributed, *i.e.*

$$< |\mathbf{Q}\sigma_l^j(\mathbf{q})|^2 >= \frac{1}{3}|Q|^2|\sigma_l^j|^2. \tag{7.12}$$

Furthermore, in the limit of low Q,[1] Eq. 7.11 can be related to a weighted density of states (Cusack *et al.*, 1988; Smith *et al.*, 1986):

[1] Q small enough that only terms quadratic in Q have to be considered.

$$G(\omega) = \lim_{Q\to 0} \frac{6\omega}{Q^2}\left(\exp(\frac{\omega}{k_B T}) - 1\right) S(Q,\omega) \tag{7.13}$$

$$G(\omega) = \sum_{j,l} \frac{b_l^2 |\sigma_l^j|^2}{m_l} \delta(\omega - \omega_j) \tag{7.14}$$

Compared to the *real* density of states, defined as

$$g(\omega) = \sum_j \delta(w - w_j) \tag{7.15}$$

$G(\omega)$ is weighted by the factor $\sum_l \frac{b_l^2 |\sigma_l^j|^2}{m_l}$. For proteins it has been shown by computer simulations that this factor is to a good approximation independent of the vibrational mode j (Smith, 1991). In that case, $G(\omega)$ is proportional to the true density of states and it remains to estimate the proportionality factor. A method to scale the weighted density of states in experimental neutron scattering studies will be presented in the following sections.

7.3 Experimental data analysis

In the previous section it was shown how the scattering function $S(Q,\omega)$ can be related to a weighted density of states. Thus, if the scattering function is accessed by experiment, the density of states is a measurable quantity in inelastic incoherent neutron scattering. However, the scattering function of the sample is not directly observable by experiment. As mentioned at the beginning of Sec. 7.2 the observable quantity in neutron scattering is the number of neutrons scattered in a solid angle Ω, having lost or gained an amount of energy ΔE. The conversion to the dynamic structure factor of the measured sample is performed by standard software packages at ILL such as SQW or INX. Since experimental errors in the measured scattering function determine the error in the measured density of states, a short summary of the corrections involved in primary data processing will now be given.

The subject of interest is the dynamics of the sample alone, thus several corrections to account for the number of neutrons absorbed or scattered by the sample holder have to be applied. The absorption and scattering contributions of the sample holder can be measured and are thus well known. The absorption of the sample is taken into account *via* an attenuation factor $\alpha = \exp(-\sigma_{\text{abs}}\rho d)$, where σ_{abs} is the absorption cross section, ρ denotes the density of the powder sample, and d is the thickness of the sample. Thus to accurately estimate the absorption by the sample, the true density of the powder inside the neutron beam has to be known which can differ considerable from the average density of the whole sample. Of all the standard corrections this is the most critical, fortunately however, it contributes to the measured $S(Q,\omega)$ an overall correction factor A_{att}. The measured dynamic structure factor is thus proportional to the *true* $S(Q,\omega)$.

Further corrections have to be applied to account for detector efficiency. The efficiency of each detector is accessed by measuring the spectra of a pure vanadium sample, known to be a nearly pure elastic scatterer.

With these corrections applied, the accessible experimental value, $Y(Q,\omega)$, relates to the scattering factor as follows:

$$Y(Q,\omega) = A_{\text{att}} \frac{N_S \sigma_S}{N_V \sigma_V} \cdot S(Q,\omega) \tag{7.16}$$

where $N_S, \sigma_S$and N_V, σ_V are the number of molecules and their incoherent scattering cross section for the sample and Vanadium, respectively.

Although both, A_{att} and N_S, are difficult to estimate, $Y(Q,\omega)$ is proportional to the scattering function $S(Q,\omega)$. Thus, the proportionality of the experimentally determined, weighted density of states, $G(\omega)$ to the true $g(\omega)$ persists (compare Eq. 7.13). The proportionality factor, however, is no longer solely dependent on properties of the sample alone, but accounts for experimental errors as well.

7.4 From $G(\omega)$ to $g(\omega)$.

Incoherent inelastic scattering on protein powders provides an observable proportional to the incoherent scattering function, $S(Q,\omega)$. $S(Q,\omega)$ allows to derive $G(\omega)$, the weighted density of states which in turn can be considered to be proportional to the true vibrational density of states , $g(\omega)$.

Taken together neutron scattering provides a quantity $\tilde{G}(\omega)$ which is proportional to the vibrational density of states :

$$\tilde{G}(\omega) = \alpha g(\omega) \tag{7.17}$$

The proportionality factor α, as discussed in the previous section, contains the theoretical factor $\sum_l \frac{b_l^2 |\sigma_l^j|^2}{m_l}$ (see Eq. 7.14) as well as contributions due to the experimental setup.
Since it is difficult to access both these contributions individually, a different approach shall be taken here. To derive the weighted density of states from experimental data, the fact was used that protein dynamics at low temperatures can be considered to be harmonic. In this approximation the vibrational density of states determines the scattering function $S(Q,\omega)$ and therefore all experimentally accessible quantities derived from it. Thus, any quantity derived from $S(Q,\omega)$ and known on an absolute scale, can be used to adjust $\tilde{G}(\omega)$ such that the same quantity calculated from $\tilde{G}(\omega)$ agrees with experiment. The quantity suggested here is the mean-square displacement of the system. As explained in detail $\langle \Delta r^2 \rangle$ can be measured from elastic scattering within the Gaussian approximation (see *e.g.* Ch. 6).

On the other hand, the mean-square displacement of a harmonic system of equal masses m

can be calculated from its density of states as follows:

$$\langle \Delta r^2 \rangle = \sum_{l=1}^{3N-6} \frac{3\hbar}{4N \cdot m} \coth(\frac{\hbar\omega_l}{2k_BT}) \tag{7.18}$$

$$\approx \int_0^\infty d\omega \; g(\omega) \frac{3\hbar}{4N \cdot m} \coth(\frac{\hbar\omega}{2k_BT}) \tag{7.19}$$

$$\tag{7.20}$$

where $g(\omega)$ is the density of states and N equals the number of atoms. The mean-square displacement of the system can thus be calculated from $\tilde{G}(\omega)$ interpreting m as the average mass of a protein atom.

The *true* vibrational density of states can be determined along the following lines. In a first step the mean-square displacement of the sample, $\langle \Delta r^2 \rangle_{exp}$, is determined using the methods discussed in Ch. 5 & Ch. 6. In a second step the weighted density of states, $\tilde{G}(\omega)$, is derived from the inelastic part of the incoherent scattering function. $\tilde{G}(\omega)$ is then used to calculated the mean-square displacement

$$\langle \Delta r^2 \rangle_{\tilde{G}} = \int_0^\infty d\omega \; \tilde{G}(\omega) \frac{3\hbar}{4N \cdot m} \coth(\frac{\hbar\omega_l}{2k_BT}) \tag{7.21}$$

$\langle \Delta r^2 \rangle_{\tilde{G}}$ is the mean-square displacement of a system whose true vibrational density of states is given by $\tilde{G}(\omega)$.

The scaling factor necessary to obtain $g(\omega)$ such that the vibrational density of states is consistent with the measured mean-square displacement can thus be determined by dividing the experimental mean-square displacement by the calculated $\langle \Delta r^2 \rangle_{\tilde{G}}$:

$$g(\omega) = \frac{\langle \Delta r^2 \rangle_{exp}}{\langle \Delta r^2 \rangle_{\tilde{G}}} \tilde{G}(\omega) \tag{7.22}$$

That $g(\omega)$ of Eq. 7.22 indeed accounts for the right mean-square displacement is easily checked by inserting Eq. 7.22 into Eq. 7.19.

Reexamining the explicit relation between the weighted density of states and $g(\omega)$ shows that the conversion factor, in general, does depend on the frequency ω. As mentioned, the proportionality between both weighted and true vibrational density of states is only approximate. Although this proportionality has been observed in computer simulations, and thus, given this proportionality, $g(\omega)$ should be equally well approximated for all frequency ranges, it is interesting to ask, which frequency range contributes most to the scaling factor here obtained. $\tilde{G}(\omega)$ will be scaled such that the vibrational density of states is best reproduced in the frequency range contributing most to the scaling. In case the proportionality between $\tilde{G}(\omega)$ is insufficiently fulfilled, this frequency range is expected to be truthfully reproduced.

The mean-square displacement of a harmonic system is dominated by contributions of low frequency. This can be seen by inspection of Eq. 7.18 and noticing that $\coth(\frac{\hbar\omega}{2k_BT}) \sim \frac{k_BT}{\hbar\omega}$ for low ω. Therefore, it can be expected that the obtained scaling factor is biased towards the low frequency region. For calculating thermodynamic quantities like the free energy contribution of internal vibrations this is favorable since it is the low frequency region that dominates these quantities (compare Eq. 7.4).

This method to obtain the vibrational density of states on an absolute scale will be used in Sec. 7.6 to estimate the contribution of changes in the vibrational spectrum to the free energy change upon binding of a ligand. First, however, an alternative method to access $\langle \Delta r^2 \rangle$ in association measurements shall be discussed. The method uses the additional information contained in the inelastic part of the spectra to derive self-consistent relative mean-square displacements of monomer and complexed system.

7.5 Absolute vs relative mean-square displacements

A method to use relative mean-square displacements to obtain the looked for scaling factor shall be illustrated by working through an example. The spectra plotted in Figs. 7.1 & 7.2 show a typical example of the scattering function obtained in incoherent inelastic neutron scattering measurement on proteins. The systems are DHFR/MTX monomers and a DHFR/MTX dimerized system. Dimerization was achieved by connecting MTX molecule pairs *via* a chain of ten carbon atoms, see Carlson *et al.* (2003) for details. The methods discussed here are, however, in no way particular to the presented system but rather illustrate generic questions common to neutron scattering studies of biological association processes. The spectra show the typical peak around $3-4$meV seen in low temperature

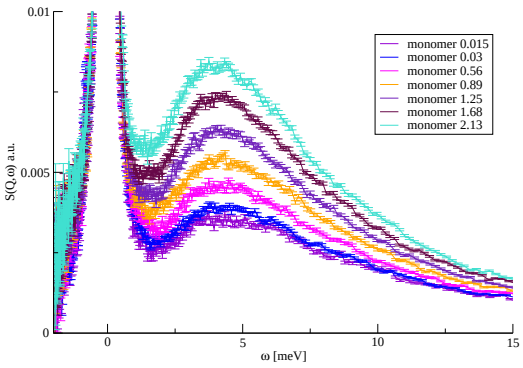


Figure 7.1: Inelastic scattering: Monomer

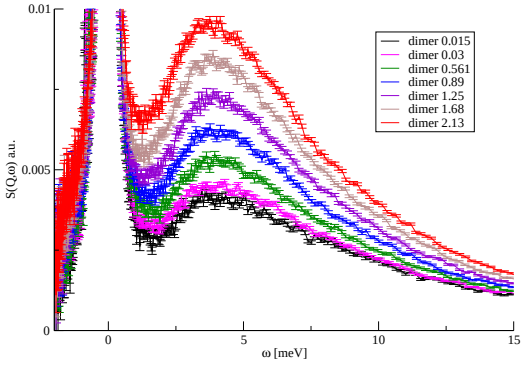


Figure 7.2: Inelastic scattering: Complex

scattering experiments on proteins (120K in the present systems). The peak in Fig. 7.1 (monomer) is slightly lower in height than the corresponding peak in Fig. 7.2 (complex) indicating an increase of inelastic scattering intensity in the complexed system. However, both spectra are on a relative scale, thus this increase cannot readily be quantified.

7.5.1 Measured mean-square displacements

The straightforward method to get the mean-square displacement from the spectra plotted in Figs. 7.1 & 7.2 has already been intensively discussed in previous chapters (*e.g.* see Ch. 6). The common approach is to plot the integrated elastic peak intensity vs Q^2 and derive the initial slope of the resulting curve.

Fig. 7.3 shows a plot of $ln\,[S(Q,0)]$ versus Q^2, $S(Q,0)$ being the integrated elastic intensities obtained from the spectra in Figs. 7.1 & 7.2. The straight lines are linear fits to the data in the range $0 < Q^2 < 1.5\text{Å}^{-2}$. The obtained mean-square displacements for the monomer and complex are listed in Tab. 7.1. Fig. 7.3 reveals on inspection two seemingly odd

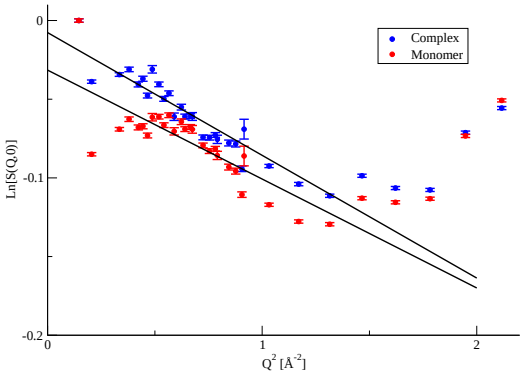


Figure 7.3: Inelastic peak height

	$\langle \Delta r^2 \rangle$ [Å^2]
Monomer	0.38 ± 0.05
Complex	0.49 ± 0.04
Average	0.44 ± 0.04

Table 7.1: $\langle \Delta r^2 \rangle_{exp}$

facts. First, the spectra that is lower in absolute value (monomer) is assigned the smaller mean-square displacement. Second, the data in the very low region seems to be noisy. The first point is in obvious contradiction to the observation that the monomer has a lower inelastic intensity indicating that this shift in the elastic intensity may be an artifact of the measurement.

The temperature dependent studies of mean-square displacements often are able to reduce the noise using the lowest temperature as a normalization spectra. The resulting mean-square displacements are then no longer on an absolute scale but refer to the lowest temperature as origin.

Obviously this possibility is not given here. However, the information present in the elastic peak can be used to derive the ratio of mean-square displacements between the two systems. Together with an best estimate of the average $\langle \Delta r^2 \rangle$ a scaling of the changes in the vibrational density of states can be achieved that is consistent with the relative ratios of elastic and inelastic intensity of both systems.

7.5.2 Inelastic peak height

In cases where the absolute value of the mean-square displacement is difficult to determine with high enough accuracy, the relative value might be accessible more readily and used together with an average mean-square displacement. The main idea is not to use values of

the mean-square displacements independently from each other as in the previous case, but to maintain the ratio between them.

The ratio of $\langle \Delta r^2 \rangle$ between monomer and complex can be obtained from the inelastic peak height, since for low Q values the peak height is proportional to the mean-square displacement and the square of the scattering vector Q (Smith, 1991):

$$S(Q, \omega_{\max}) \sim \langle \Delta r^2 \rangle Q^2 \tag{7.23}$$

and hence:

$$\frac{S_{\mathrm{C}}(Q, \omega_{\max})}{S_{\mathrm{M}}(Q, \omega_{\max})} = \frac{\langle \Delta r^2 \rangle_{\mathrm{C}}}{\langle \Delta r^2 \rangle_{\mathrm{M}}} \tag{7.24}$$

Measuring the elastic peak height in the limit of low Q thus allows to access the ratio of mean-square displacements between the complexed and monomeric system. Along with an average $\langle \Delta r^2 \rangle$ the scaling factor for both systems can be determined along the lines described in Sec. 7.4

The proportionality is given only in the limit of very low Q. For Q *low enough*, Eq. 7.23 can be considered to provide a good approximation. The meaning of *low enough*, *i.e.* the range of Q values for which Eq. 7.23 can be expected to hold true, has to be accessed experimentally. A glance at Eq. 7.23 suggests a straightforward criteria. The proportionality of the peak height and Q^2 holds true at very low Q, thus plotting the inelastic peak-height *vs* Q^2 has to be linear at low Q. If the proportionality of Eq. 7.23 breaks down, a deviation from the initial linear increase is expected. Thus the range of Q for which Eq. 7.24 can be expected to hold true is given by the linear regime of a plot of the inelastic peak height *vs* Q^2. These plots for the monomeric and complexed system are shown in Figs. 7.4 & 7.5. Both samples show a linear increase in peak-height over a

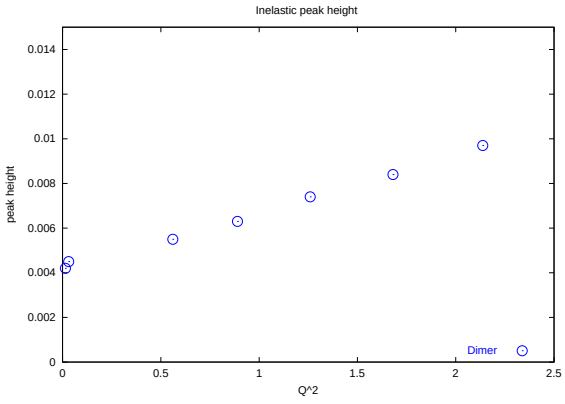


Figure 7.4: Inelastic peak height:Complex

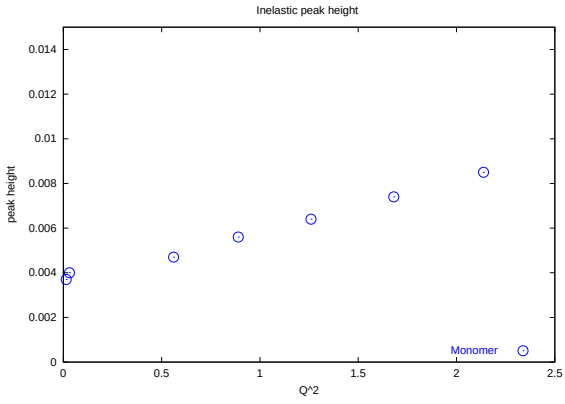


Figure 7.5: Inelastic peak height:Monomer

Q-range of $0.5 < Q^2 < 2\text{Å}^{-2}$. Thus, although strictly valid only in the limit of very low Q (Eq. 7.24) can be considered to hold true over a finite Q range easily accessible by neutron scattering instruments.

As mentioned in Sec. 7.3 the scattering function $S(Q,\omega)$ is known only up to a scaling factor. In comparing inelastic peak heights directly in Eq. 7.24 it was assumed that this scaling factor is the same for both samples, *i.e.* that it cancels out in the ratio. However, in the present example the complexed system exhibits the larger inelastic peak (Figs. 7.1 & 7.2) as well as a higher elastic intensity (Fig. 7.3), in contrast to the assumption that both spectra are scaled by the same factor. Thus, before applying Eq. 7.24 the spectra have to be set on the same scale. This can be achieved, by adjusting the spectra in a self-consistent manner as described below.

7.5.3 Scaling of spectra

To minimize the effect of the scaling factor discussed in the last section the spectra were scaled to give a self-consistent ratio between complex and monomer for both, elastic and inelastic peak height.

The inelastic peak height is proportional to the mean-square displacement (Eq. 7.24), whereas the elastic peak intensity is proportional to the exponential of $\langle \Delta r^2 \rangle$ (gaussian approximation). A different scaling factor for the two samples therefore leads to inconsistent inelastic peak heights compared to the corresponding elastic peak intensities. To obtain consistency, a difference in inelastic peak height of for example 10% has to correspond to a certain difference in elastic peak intensity, which is in general not equal to 10%. The corresponding difference in elastic peak intensity depends on the absolute magnitude of the mean-square displacement, in contrast to the dependency on the relative magnitudes for the inelastic peak. The elastic peak is given by:

$$\frac{I_{\mathrm{Dim}}}{I_{\mathrm{Mon}}} \sim \frac{e^{-Q^2\langle \Delta r^2 \rangle_{\mathrm{Dim}}}}{e^{-Q^2\langle \Delta r^2 \rangle_{\mathrm{Mon}}}} = \exp\left[-Q^2\langle \Delta r^2 \rangle_{\mathrm{Mon}}\,(\kappa - 1)\right] \tag{7.25}$$

with $\kappa = {}^{\langle \Delta r^2 \rangle_{\mathrm{Mon}}}\!/_{\langle \Delta r^2 \rangle_{\mathrm{Dim}}}$ This additional information can be used to scale the spectra to obtain self-consistency. To achieve this the following procedure was iteratively applied:

- For a given value of Q^2 both spectra, monomer and complex, were scaled to give an elastic peak intensity corresponding to a chosen mean-square displacement $\langle \Delta r^2 \rangle_{\mathrm{M}}$.
- The spectra of the complex was multiplied by a factor $\alpha < 0$.
- The new ratio $\kappa = {}^{\langle \Delta r^2 \rangle_{\mathrm{M}}}\!/_{\langle \Delta r^2 \rangle_{\mathrm{C}}}$ was calculated from the inelastic peak height and the ratio of elastic peak intensity was compared to Eq. 7.25

Fig. 7.6 shows the change in the dimer spectra for three different values of Q^2. The changes needed to obtain self-consistent peak heights are small, even within errorbars.

Although the changes in the spectra are small, the influence on the relative mean-square displacement can be considerable. Fig. 7.7 shows the ratio ${}^{\langle \Delta r^2 \rangle_{\mathrm{C}}}\!/_{\langle \Delta r^2 \rangle_{\mathrm{M}}}$ before scaling whereas Fig. 7.8 shows the corresponding ratio after scaling.

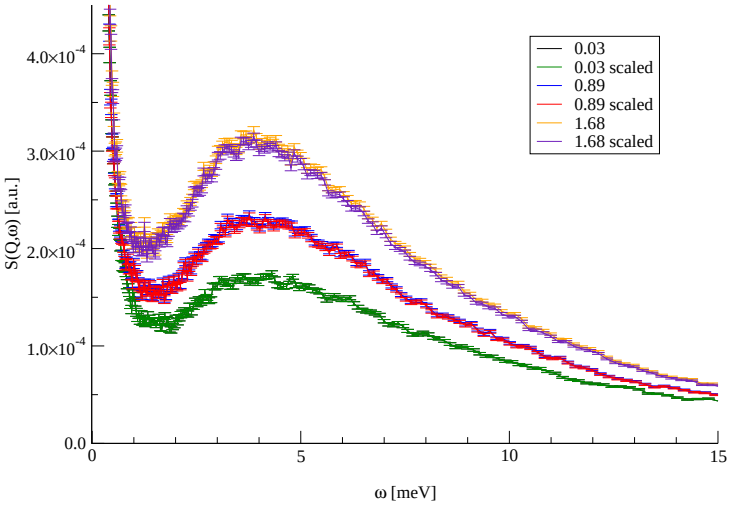
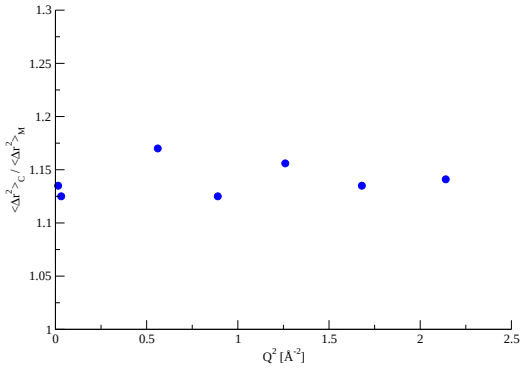
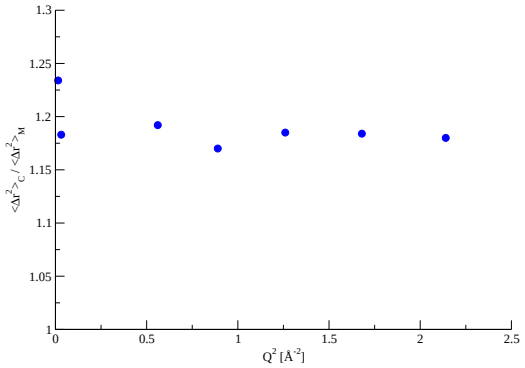


Figure 7.6: $S(Q,\omega)$ scaled with mean $\langle \Delta r^2 \rangle$



Q^2 [Å^{-2}]	$\langle \Delta r^2 \rangle_C / \langle \Delta r^2 \rangle_M$
0.0148	1.135
0.0314	1.125
0.561	1.170
0.89	1.125
1.26	1.156
1.68	1.135
2.14	1.141

Figure 7.7: Ratio of inelastic peak heights before scaling



Q^2 [Å^{-2}]	$\langle \Delta r^2 \rangle_C / \langle \Delta r^2 \rangle_M$
0.0148	1.234
0.0314	1.183
0.561	1.192
0.89	1.170
1.26	1.185
1.68	1.184
2.14	1.180

Figure 7.8: Ratio of inelastic peak heights after scaling

7.6 Ligand binding of dihydrofolate reductase

The theory developed in the previous section, especially the scaling method explained in Sec. 7.4, will now be used to experimentally determine changes in free energy due to changes in the vibrational spectrum upon ligand binding. The enzyme chosen is dihydrofolate reductase (DHFR), an important target for anticancer and antibacterial drugs, catalyzing the reduction of dihydrofolate to tetrahydrofolate in the presence of nicotinamide adenine dinucleotide phosphate (NADPH) cofactor. The ligand used is methotrexate (MTX), a folate antagonist of DHFR that has been used effectively as a cytotoxic agent in the treatment of cancers Huennekens (1994).

7.6.1 Experimental setup and Materials

Neutron scattering spectra were taken for the uncomplexed system DHFR/NADPH and the complex DHFR/NADPH + MTX. To minimize scattering from solvent molecules both systems were exchanged with D_2O. To do this, lyophilized DHFR from E.coli was dissolved in D_2O equilibrated at 4°C overnight and freeze dried. NADPH and NADPH+MTX were added in equimolar ratios to the enzyme. As the dissociation constants of DHFR with NADPH and MTX are low ($K^d_{\mathrm{NADPH}} = 0.01M, K^d_{\mathrm{MTX}} = 0.15M$ (Basran *et al.*, 1995)) it can be assumed that the ligands bind quantitatively to the enzyme. D_2O was added, the solutions equilibrated for 4h at 4°C and freeze dried, and a final exchange step involving dissolution in D_2O and lyophilization was performed. The uncomplexed enzyme (DHFR+NADPH) and the complexed enzyme (DHFR+NADPH+MTX) were hydrated to a degree of 30% i.e. 30 mg of D_2O per 100 mg of dry weight protein. This was performed by equilibrating the samples in a saturated solution atmosphere (KBr in D2O) at 20°C for 3 days. The system data relevant for the analysis of neutron scattering spectra of these samples are listed in Tabs. 7.2 & 7.3. For the neutron scattering experiments the samples were contained in an aluminium sample holder. Sample amounts were 98.1 mg (uncomplexed) and 108.6 mg (complexed). The measurements were performed on the time-of-flight spectrometer IN6 at the Institute Laue-Langevin (ILL), Grenoble with an incident neutron beam wavelength of 5.12 Å. The scattering experiments were performed at 120K to ensure all dynamics present is harmonic. Accumulation times were 24 hours/sample. Samples were oriented at 135° with respect to the incident beam. Both sample transmissions were 97.7%, high enough to ensure the absence of multiple scattering effects. The raw data were corrected using the INX program at ILL (Rieutord, 1990). INX normalizes the detector responses with respect to an angle-independent standard vanadium sample, normalizes each data collection run to the number of incident neutrons, performs slab corrections for self-shielding and subtracts the cell scattering from sample + cell scattering. To improve the statistics, the spectra derived from 122 scattering angles (from 10° to 114°) were binned into 11 constant-angle spectra. In order to compare the relative differences in the inelastic part of the spectra of the two samples dynamical structure factors were normalized to the elastic peak height. The errors in $S(q,\omega)$ (see Fig. 7.9) are statistical errors of the measurement. The errors in deduced quantities were determined by error propagation.

	DHFR/NADPH	D_2O
Mass [mg]	98.1	29
σ_{scatt} [barn]	90247,5	19,5
σ_{abs} [barn]	2110,1	0.0
AtMass [g/mol]	14921,3	20,0
Density [g/cm^3]	1,285	-

Table 7.2: Scattering cross sections, molar masses and concentrations of the uncomplexed system consisting of DHFR/NADPH protein in D_2O

	DHFR/NADPH + MTX	D_2O
Mass [mg]	108.6	32
σ_{scatt} [barn]	91743.0	19,5
σ_{abs} [barn]	2221.1	0.0
AtMass [g/mol]	19520.8	20,0
Density [g/cm^3]	1,282	-

Table 7.3: Scattering cross sections, molar masses and concentrations of the uncomplexed system consisting of MTX bound to DHFR/NADPH in D_2O

7.6.2 Results

Fig. 7.9 shows the dynamic structure factor $S(Q,\omega)$, of the 'uncomplexed' (DHFR+NADPH) and 'complexed' (DHFR+NADPH+MTX) forms of the enzyme. A significant difference can be seen in the low-frequency region of the spectra ($\omega \lesssim 40cm^{-1}$). The fact that $S(Q,\omega)$ is higher for the complex in the low-frequency region implies that the complexed form of the enzyme has a higher number of low-frequency modes and thus is more flexible than the uncomplexed form.

The experimental weighted frequency distribution, $\tilde{G}(\omega)$ was obtained from the spectra using Eq. 7.13. As discussed $\tilde{G}(\omega)$ is on a relative scale. To determine the absolute vibrational density of states , $g(\omega)$, experimental mean-square displacements, $\langle \Delta r^2 \rangle_{exp}$, of the samples were determined using the Gaussian approximation. $\langle \Delta r^2 \rangle_{\tilde{G}}$ was calculated from Eq. 7.21 and the experimentally derived $\tilde{G}(\omega)$ were finally scaled *via* Eq. 7.22.

The increase of low frequency modes in the complexed form is confirmed by the derived vibrational density of states . Fig. 7.10 shows that below $\sim 15 - 20cm^{-1}$ $g(\omega)$ of the complexed form is significantly higher than that of the uncomplexed protein.

For the protein to soften on binding there must be a shift in $g(\omega)$ from higher frequencies (in the uncomplexed protein) to lower frequencies (in the complex). Inspection of Fig. 7.9 does not reveal significant differences between both systems at higher frequencies. One possibility consistent with this finding would be that softened modes may be of too-high frequency in the uncomplexed protein to be detectable by the instrument. However, there is evidence of a frequency shift occurring within the frequency range shown in Fig. 7.10. Integration of the difference between $g(\omega)$ of the two samples shows that the net equivalent

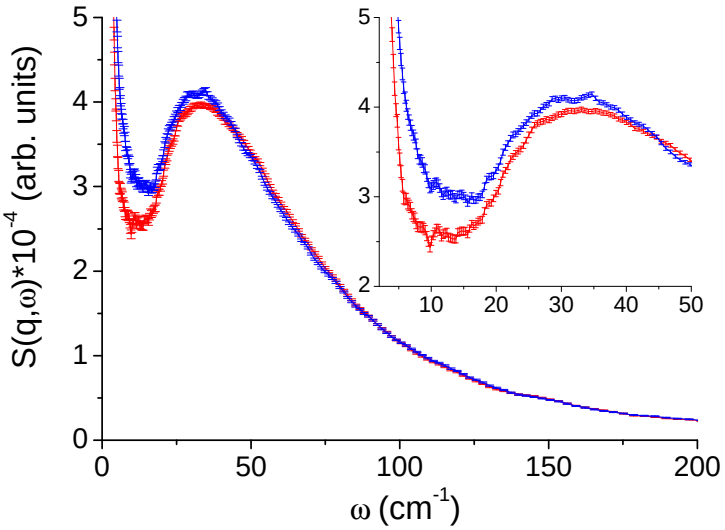


Figure 7.9: Dynamic structure factor versus frequency for uncomplexed DHFR (red) and complexed with methotrexate form of DHFR at 120K. Data from all scattering angles are summed. Both spectra are normalized to the elastic peak height. Inset: the low frequency region of the spectra.

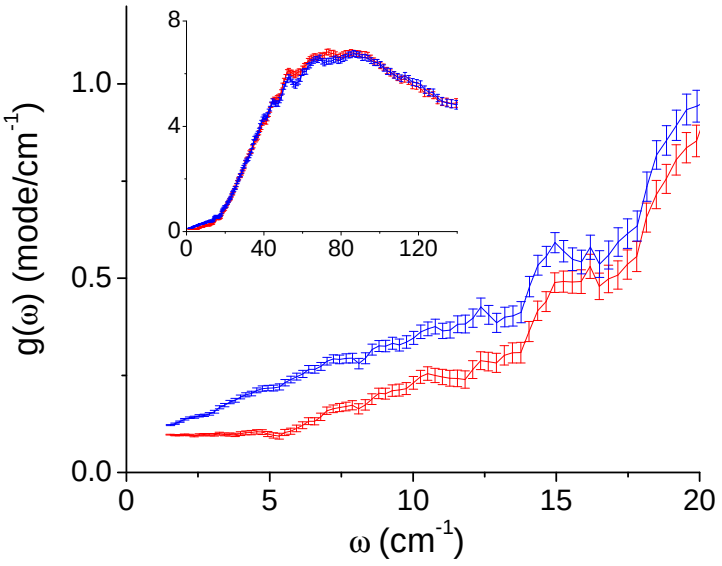


Figure 7.10: Vibrational density of states $g(\omega)$ for uncomplexed (red) and complexed (blue) forms of DHFR. Inset: enlarged frequency region of the spectra.

of about 2 modes are transferred from the 40-80cm^{-1} region in the uncomplexed protein to the region below 20 cm^{-1} in the complex.

The change in $g(\omega)$ over the range 40-80cm^{-1} is not accompanied by a statistically significant change in the $S(Q,\omega)$ in Fig. 7.9. The reason for this is that $S(q,\omega)$ is amplitude weighted. Thus, removal of a low-amplitude, higher-frequency mode from the spectrum leads to only a relatively small intensity loss whereas adding a high-amplitude, low-frequency mode adds a relatively large intensity. Consequently, the shifted modes scatter much more strongly in the complex than in the uncomplexed molecule, with the net result that the spectral change appears as an increase in intensity at low frequencies in the complex.

The vibrational partition function, Z_{vib},*i.e.* the partition function associated with the vibrational energy levels in a harmonic system, is determined by $g(\omega)$ as follows McQuarrie (1976):

$$Z_{\text{vib}} = \prod_{i}^{3N-6} \frac{e^{-\beta\hbar\omega_i/2}}{1-e^{-\beta\hbar\omega_i}} = \prod_{\omega=\omega_{\min}}^{\infty} \left(\frac{e^{-\beta\hbar\omega/2}}{1-e^{-\beta\hbar\omega}} \right)^{g(\omega)} \tag{7.26}$$

where $\omega_{\min}$ is the lowest frequency in the system and $\beta = 1/kT$. The associated Helmholtz vibrational free energy, A_{vib} is given by

$$A_{\text{vib}} = -kTln[Z_{\text{vib}}] \tag{7.27}$$

the energy[2] by

$$\bar{E}_{\text{vib}} = -\frac{\partial}{\partial\beta} ln[Z_{\text{vib}}] \tag{7.28}$$

and finally the entropy by

$$S_{\text{vib}} = k_B \left[ln[Z_{\text{vib}}] + \beta\bar{E} \right]. \tag{7.29}$$

The free energy change, ΔG calculated from the density-of-states change in Fig. 7.10 is $-17 \pm 4^{kJ}/_{mol}$ at 300K, favoring binding. The free-energy change contains a large entropic contribution $-TS = -25 \pm 6^{kJ}/_{mol}$ that is partially compensated by the energy term $\Delta E = 8 \pm 2^{kJ}/_{mol}$ that opposes binding.

[2]In association processes of biomolecules, the change in volume can often be neglected. In these cases, the energy here calculated is equivalent to the enthalpy of the system (Hill, 1994)

7.7 Conclusions

The thermodynamic results here presented are subject to errors associated with the use of the harmonic approximation (Sec. 7.2) and the necessity of determining the dynamic structure factor on an absolute scale (Sec. 7.4). The scattering measurements were performed at 120K. At this temperature the protein exhibits only harmonic vibrational dynamics (Doster *et al.*, 1989b), thus permitting the direct determination of the vibrational density of states.

The 300K vibrational density of states of trehalose-coated myoglobin has been found to be identical to that at 100K, indicating harmonic behavior (Cordone *et al.*, 1999). For hydrated proteins at physiological temperatures a significant contribution to the atomic mean-square displacements arises from anharmonic dynamics (Cordone *et al.*, 1999; Smith *et al.*, 1990; Tournier & Smith, 2003). These anharmonic motions, together with overdamping of the low-frequency modes, result in quasielastic scattering that renders difficult the derivation of model-independent thermodynamic quantities from experimental data (Smith *et al.*, 1990). It is conceivable that anharmonic degrees of freedom might also be modified on ligand binding. However, molecular dynamics simulations suggest that $\sim 99.5\%$ of the modes in a protein are effectively harmonic at 300K (Kitao *et al.*, 1998). Assuming the protein does remain harmonic, the density-of-states change will be the same at 300K as it is at 120K.

The dynamical change seen here indicates softening of the protein on complexation. Normal mode calculations on proteins have indicated that the vibrational modes in the frequency range at which the change is seen here involve mainly collective displacements of groups of atoms distributed throughout the protein (Brooks & Karplus, 1983). These modes may be modified by environmental effects, although in the present case the sample preparation is consistent with the environments of the complexed and uncomplexed form being the same. Normal mode calculations concur with the present findings in that they also indicate an increase in flexibility on ligand binding to proteins and protein:protein association (Fischer *et al.*, 2001; Tidor & Karplus, 1994). Furthermore, increased backbone conformational flexibility on binding a hydrophobic ligand to mouse major urinary protein has been detected using nuclear magnetic resonance (NMR) relaxation experiments (Zidek *et al.*, 1999). However, these results contrast with other observations using NMR and crystallography of a flexibility decrease on protein binding to small organic ligands (Cheng *et al.*, 1994; Fushman *et al.*, 1994; Rischel & et al., 1994). NMR and crystallographic measurements are likely to be dominated by changes in anharmonic degrees of freedom not present under the conditions studied here. Moreover, the vibrational changes seen in the present work are not detectable using NMR or crystallography (Fischer *et al.*, 2001; Tidor & Karplus, 1994).

There remains much to learn about the various contributions to ligand binding free energies. For example, although hydrogen bonds normally must be satisfied, whether they bring any net contribution to the binding free energy in aqueous solution is debatable. For the present system, although significant ligand:protein hydrogen bonding exists, free energy perturbation calculations suggest that hydrophobic interactions contribute as much to the binding free energy as do hydrophilic (Singh & Benkovic, 1988). Another effect that remains to be accurately quantified is the loss on binding of the whole-molecule translational and rotational entropic free energies of the protein and ligand (Karplus & Janin, 1999; Tamura

& Privalov, 1997; Tournier & Smith, 2003; Yu *et al.*, 2001; Zaman *et al.*, 1999).

The contribution elucidated here, that from vibrational changes, has been hitherto largely neglected in the consideration of binding equilibria, due largely to the absence of an experimental technique for determining it. The estimation of $-17 \pm 4^{kJ}/_{mol}$ for the vibrational free energy change can be compared with the experimentally estimated total binding free energy in the present system which is $39^{kJ}/_{mol}$ (Basran *et al.*, 1995). This suggests that the vibrational change is thermodynamically significant and would contribute a factor of $\sim 10^3$ to the binding constant. This illustrates how modifications of equilibrium internal fluctuations in a protein can have consequences for binding equilibria of biomedical importance. Recently, the vibrational effect has been included together with electrostatic, van der Waals and nonpolar terms in an empirical theoretical analysis of ligand binding free energies (Schwarzl *et al.*, 2002). Whether the vibrational softening effect seen here is general for protein:ligand interactions remains to be determined. The direct access to the vibrational density of states provided by inelastic neutron scattering holds promise for the study of vibrational thermodynamic changes in many biomolecular association processes.

Chapter 8

Final Remarks

Everything that living things do can be understood in terms of the jigglings and wigglings of atoms. This quotation of R.P. Feynman preceded the first chapter of this thesis. One hundred pages later it might be useful to finish with a short summary of what has been achieved in the previous chapters. Therefore, this last chapter tries to put the results here obtained in a broader context, to discuss their relevance for the field of protein dynamics and finally to give an outlook on possible future directions of research.

This thesis mainly divides in two distinct parts. From a physicists point of view that division can be drawn between elastic/quasielastic scattering on the one hand and inelastic scattering on the other, each bearing quite distinct information on the system. From a more biological point of view these two parts correspond to internal motions of single proteins on one hand (elastic/quasielastic) and the investigations of association processes between biological molecules on the other.

8.1 Protein dynamics

The interdisciplinary field of protein dynamics resides on the border between physics and biology. The laws of physics govern the motions of atoms constituting a protein. A protein itself, however, serves a biological purpose, *e.g.* it catalyzes a chemical reaction inside a living cell. It is the aim of the field of protein dynamics to shed light on the relationship between the dynamics or flexibility of proteins and their biological function. A prominent and intriguing phenomenon bridging between both aspects is the dynamical transition in proteins. The dynamical transition is a generic feature of globular proteins and its relation to enzyme function promises to reveal insight into the flexibility-function relationship fundamental to protein dynamics.

The Dynamical Transition

The dynamical transition,*i.e.* the sharp increase in atomic fluctuations at a transition temperature T_0 has been observed by a wide range of experimental techniques as well as computer simulations (Cohen *et al.*, 1981; Doster *et al.*, 1989b; Keller & Debrunner, 1980; Knapp *et al.*, 1982; Rasmussen *et al.*, 1992). A relation to the function of enzymes has been established for various enzymes such as myoglobin (Austin *et al.*, 1975), ribonuclease A (Rasmussen *et al.*, 1992) or Bacteriorhodopsin (Lehnert *et al.*, 1998). Inspired by spectroscopic work on myoglobin (Austin *et al.*, 1975; Frauenfelder *et al.*, 1979) energy landscape models have been suggested to account for both, the experimentally observed transition as well as the onset of function at T_0. The transition temperature in theses models reflects the energetics of different conformational substates.

However, contrasting experimental evidence was reported for solutions of the proteins xylanase (Dunn *et al.*, 2000) and glutamate dehydrogenase (Daniel *et al.*, 1998, 1999). In these systems the dynamical transition seen on a 100ps timescale is neither related to the onset of activity nor is the transition temperature T_0 a property of the protein itself. T_0 was observed to significantly shift to lower temperatures upon increasing energy resolution.

The main achievement of the first part of this thesis is to offer an alternative interpretation of the dynamical transition seen in these systems that explains the prominent experimental findings. This 'frequency window' interpretation proposes that the transition seen by elastic neutron scattering is not due to populating an energetically higher conformational substate, but rather due to the process of barrier crossing itself that becomes fast enough to be detected by the instrument.

To test this interpretation the intermediate scattering function, $I(Q,t)$, was calculated from molecular dynamics simulations for a small protein (BPTI) in a cryosolution. For a temperature range of 100-300K the frequency window scenario was shown to reproduce the calculated spectra over a time range of $1 \leq t \leq 200ps$ (stretched relaxation) or $10 \leq t \leq 200ps$ (exponential relaxation). The frequency window scenario was thus shown to be a valid alternative to the conventional interpretation of neutron scattering data in the context of the dynamical transition.

Explicitly including finite energy resolution of the instrument in a theoretical analysis of the scattering function, $S(Q,\omega)$, a description of the measured mean-square displacement, $\langle \Delta r^2 \rangle_{exp}$ was derived and compared to both experimentally derived mean-square displacements and simulation results.

The theoretical description of the dynamical transition, based on a frequency window interpretation, is in quantitative agreement with simulation results. Both, the timescale dependence of $\langle \Delta r^2 \rangle$ as well as the temperature dependence are well reproduced fostering the frequency window interpretation to account for the observed transition behavior.

Reanalyzing the experimentally obtained mean-square displacements of Ref. Daniel *et al.* (1999) in the context of the frequency window interpretation led to qualitative agreement with experiment. Both the prominent features observed in experiment, the shift in the transition temperature T_0 as well as the leveling off of $\langle \Delta r^2 \rangle_{exp}$ around 220K, is well reproduced

by the theoretical analysis.

Another interesting aspect of the frequency window interpretation concerns the relation between the transition temperature T_0 and enzyme activity. Since models of the dynamical transition so far attributed the transition temperature to the population of new, energetically higher substates, this temperature was a property of the system determined by the energy landscape of the protein alone. If these new substates are involved in enzyme activity a clear relation between protein function and the transition temperature can be observed.

In contrast, the frequency window scenario attributes the transition to dynamical processes crossing the resolution window of the instrument. The transition temperature is no longer determined by the system alone, but depends on the relation between the relaxation frequencies of the system and the resolution of the instrument. Consequently, no relation between the onset of activity and the transition temperature should be expected. Thus, in addition to the ability to reproduce experimentally observed mean-square displacements, the frequency window interpretation explains the absence of a relation between T_0 and enzyme activity observed in Refs. Daniel *et al.* (1998); Dunn *et al.* (2000).

As mentioned, the interpretation suggested here is an alternative to the conventional view of energetically distinct substates. Although the presence of a timescale dependence of T_0 in both simulation and experiment, is enough to rule out an explanation of the transition behavior solely in terms of different populations of substates, it can not be concluded that the frequency windows are the only factors contributing to the observed transition. Furthermore, the relation between T_0 and activity for various enzymes might well suggest that in these systems indeed populations of activity related substates are populated at T_0. It will be a task for future research to disentangle these opposing scenarios and estimate their relative contributions (see 8.3)

8.2 Association processes – ligand binding

Association processes are common and crucial in biochemical systems. Enzymes either have to temporarily bind their substrate to catalyze specific reactions or proteins have to form complexes to be functioning. Protein:ligand association has been assumed to be dominated by factors such as the hydrophobic effect, hydrogen bonding, electrostatic and van der Waals interactions. However, as early as 1963 it was suggested that an additional mechanism might exist: due to increased flexibility in the protein:ligand complex manifested by a change in the spectrum of vibrations in the complex due to formation of new, intermolecular interactions (Erickson, 1989; Finkelstein & Janin, 1989; Fischer *et al.*, 2001; Page & Jencks, 1971; Steinberg & Scheraga, 1963; Sturtevant, 1977; Tidor & Karplus, 1994). Theoretical normal mode analysis, used to estimate this vibrational change on insulin dimerization (Tidor & Karplus, 1994) and on water binding to bovine pancreatic trypsin inhibitor (Fischer *et al.*, 2001), have suggested that the effect is likely to be thermodynamically important. Inelastic incoherent neutron scattering allows to experimentally access contributions to the binding free energy due to changes in the vibrational density of states upon association.

The second part of this thesis thus concentrated on the information contained in the inelas-

tic part of incoherent neutron scattering. In the one-phonon approximation the inelastic scattering function can be related to a weighted vibrational density of states. To experimentally derive the *true* vibrational density of states, and thus to experimentally determine thermodynamic contributions of changes in the vibrational density of states upon ligand binding, a method was suggested to scale the weighted density of states. The extra information needed to obtain the scaling factor resides in the elastic intensity. Mean-square displacements are derived from this intensity and subsequently the weighted density of states is scaled such as to be consistent with the measured mean-square displacement.

This scaling method was applied to neutron scattering experiments on dihydrofolate reductase, an important target for anticancer and antibacterial drugs, catalyzing the reduction of dihydrofolate to tetrahydrofolate in the presence of nicotinamide adenine dinucleotide phosphate cofactor. The ligand used is methotrexate, a folate antagonist of DHFR that has been used effectively as a cytotoxic agent in the treatment of cancers.

The present experimental analysis suggests a *softening* of the protein upon binding a ligand, *i.e.* a shift of frequencies to lower energy. Thermodynamically this resulted in a contribution to the free energy of association of $-17 \pm 4^{kJ}/_{mol}$. Compared to the total binding free energy in the present system which is $39^{kJ}/_{mol}$ (Basran *et al.*, 1995), this fosters the importance of changes in the vibrational spectrum for the thermodynamics of ligand binding.

The contribution elucidated here, that from vibrational changes, has been hitherto largely neglected in the consideration of binding equilibria, due largely to the absence of an experimental technique for determining it. The scaling method here proposed might encourage further experiments to be done to access these contributions.

The analysis here presented can only be a first step towards an understanding of vibrational changes upon binding and their contribution to biological function. Work has to be done to confirm the present results as well as to test the assumptions and possible sources of error in the presented scaling method.

8.3 Outlook

The work presented is neither the beginning nor the end of necessary and interesting research in the field of protein dynamics. The questions addressed and the answers given might, however, encourage further work on the topics presented in this thesis. As one of the achievements of this work is the suggestion of an alternative interpretation of the dynamical transition, it is a natural wish of the author that further work is initiated to test and hopefully discriminate between the alternative scenarios here presented.

Timescale dependence of T_0

As explained in Ch. 5, a characteristic feature of the frequency window scenario is the expected timescale dependence of the transition temperature T_0. Experimental studies testing the dependency of T_0 on the timescale explored are scarce. Combined studies on

instrument of different energy resolution could be designed to investigate the timescale dependence of T_0. It would be interesting to know whether the timescale dependence of the transition temperature observed in Ref. Daniel *et al.* (1999) can be observed for systems other than cryosolutions as well. For example, for a myoglobin powder, T_0 was found to be unchanged upon increasing the resolution of the instrument by a factor of 10 (Doster & Settles, 1999). It has yet to be seen whether a timescale dependence of T_0 is common to most neutron scattering studies on proteins or whether it is a peculiarity of the protein cryosolutions.

With regard to the complexity of biological macromolecules it should not come as a surprise if the answer here is neither yes nor no. It may well turn out that T_0 is indeed inherent to certain systems like myoglobin powders whereas for other systems it is dependent on instrumental resolution. Investigating this aspect further promises to shed new light on the underlying energy landscape of proteins. As discussed in Ch. 5, T_0 is intimately related to the energy landscape of proteins. Thus an understanding of the transition temperature in various systems can be used to quantify the energy landscape probed by pico- to nanosecond dynamical processes. To accomplish this however, the features of the energy landscape determining a given transition have to be identified, *i.e.*, the contribution of both, equilibrium and frequency window scenario have to be distinguished.

Computer simulation will play an important role in this respect. The fast increase in computational power of modern computers allow already to access protein dynamics on the nanosecond timescale. This timescale will be continuously expanded in the near future. The detailed description of the time evolution of the system allows to freely vary the timescales to be accessed. The analysis of $\langle \Delta r^2 \rangle$ presented in chapters 4 & 5 may be extended in the timescales explored. The structure of most proteins investigated in the context of the dynamical transition are known. Molecular Dynamics simulations may thus be used to investigate the role of the frequency window scenario in different proteins.

Solvent dependence

Along the same lines of thought it is important to understand the influence the surrounding has on protein dynamics. The timescale dependence of T_0 is so far unambiguously observed only for protein cryosolutions. It may be asked whether this holds true, *i.e.* whether other environments indeed assure a system inherent transition temperature. It is in any case interesting to ask in which way the environment alters the energy landscape of proteins. The transition behavior in the context of its two opposing interpretations can shed light on this aspect.
Again, a most promising approach would be to combine experimental analysis on various timescales with computer simulations of the same system.

Describing dynamical processes in proteins

The motions and processes leading to the dynamical transition are still insufficiently understood. Recent progress in the field showed that the transition is due to the change of just a

few principal modes of protein dynamics (Tournier & Smith, 2003). This promising result indicates that it might be possible to condense the complexity of protein motions into a few degrees of freedom and thus getting a firmer grip on the principle processes necessary for function. The vast amount of experimental data on the dynamical transition will serve as a guide to model building. Thus, the theoretical analysis here presented may help to quantify and access the transition behavior of simplified models of protein dynamics. The presented relations between transition temperature and features of the underlying energy landscape, such as barrier heights, allow to derive experimental observables from the model chosen.

The ultimate test for all simplified theoretical models, of course, is experiment. In spite of the vast amount of data, especially in the context of the dynamical transition, it is my conviction that above all experimental advances have to be made to flourish theoretical model building. Atomic detailed descriptions, as obtained by X-ray crystallography, provide *only* a static picture of proteins. Spectroscopic techniques, such as neutron scattering or Mössbauer spectroscopy provide dynamic information that is either related to a single atom within a protein (Mössbauer) or given by an average over large parts of the protein (hydrogen atoms in neutron scattering). Computer simulations are not yet capable to fully bridge this gap although they do already serve as guiding instrument in theoretical model building. What is needed is a more detailed and selective picture of the dynamical processes taking place in biological macromolecules. Two experimental techniques may play a major role in this respect, incoherent neutron scattering and time-resolved X-ray crystallography.

The dominance of hydrogen atoms in the neutron scattering profile of proteins can be used to selectively investigate the dynamics of specific parts of proteins, thus delivering a more detailed picture of protein dynamics. This can be achieved by selectively deuterating all those parts of the protein not of interest in the current study, thus significantly increasing the relative signal of the desired focus of interest. The Institute Laue-Langevin already dedicated a new labratory to the deuteration of molecular systems. The required deuteration techniques should be further developed to provide a more detailed picture of a moving protein needed for theoretical model building. Promising first results have been obtained by selectively deuterating Bacteriorhodopsin (Réat *et al.*, 1997).

A second key technique in this respect will be time-resolved X-ray crystallography (Cruickshank *et al.*, 1992). With this technique snapshots of proteins are taken at short time intervals, thus providing time information along with atomic detailed description of X-ray crystallography. This techniques already gave important insight in biological processes and promises to shorten the gap between structural and dynamical aspects of proteins (Schlichting *et al.*, 1994, 2000; Teng *et al.*, 1994).

Flexibility-activity relationship

Independent of the question whether T_0 is related to the activity of a certain enzyme, it is important to further investigate the relationship between flexibility and activity of enzymes. As discussed in Ch. 1 this relationship bridges the gap between a physical description of protein dynamics and the biological relevant function of an enzyme. To establish and test

such relations expertise in biochemical as well as physical methods is required. So far, most work concentrated on one aspect, either flexibility or activity, and correlated afterwards to relevant data of the other aspect often obtained for slightly different systems. However, direct transfer of results from one system to another is a difficult task in biological systems. It would be important to combine measurements on physical and biological aspects of proteins more closely to be able to establish the looked for relation between dynamical processes and protein function. The experimental data reexamined in this thesis took such an approach and raised doubts on an assumed general relationship between fast picosecond dynamics and the activity of enzymes. With further investigations combining expertise in both fields new, interesting aspects will be revealed.

Association processes

The proposed method to scale experimentally derived density of states and thus access thermodynamical quantities related to vibrational degrees of freedom of proteins needs to be thoroughly tested. Computer simulations are well adapted to provide such tests. Normal mode calculation, due to the large memories available today, can be performed even on large proteins providing the vibrational density of states of the system. Molecular Dynamics simulations can be used to calculate the scattering function $S(Q, \omega)$. The scattering function can then be analysed along the lines here proposed and the derived vibrational density of states can be compared to the *true* vibrational density of states obtained by normal mode analysis. It would thus be possible to direectly access the quality of the scaling method

In addition, changes in the vibrational density of states upon binding can directly be accessed with normal mode calculations if the atomic structure of both, protein and complex is known. These simulations promise to further elucidate the role of flexibility for association processes of proteins.

Neutron scattering cross sections

Isotope	conc	b_{coh}	b_{inc}	σ_{coh}	σ_{inc}	σ_{scatt}	σ_{abs}
H	—	-3.7390	—	1.7568	80.26	82.02	0.3326
1H	99.985	-3.7406	25.274	1.7583	80.27	82.03	0.3326
2H	0.015	6.671	4.04	5.592	2.05	7.64	0.000519
3H	(12.32 a)	4.792	-1.04	2.89	0.14	3.03	0
C	—	6.6460	—	5.551	0.001	5.551	0.0035
12C	98.9	6.6511	0	5.559	0	5.559	0.00353
13C	1.1	6.19	-0.52	4.81	0.034	4.84	0.00137
N	—	9.36	—	11.01	0.5	11.51	1.9
14N	99.63	9.37	2.0	11.03	0.5	11.53	1.91
15N	0.37	6.44	-0.02	5.21	0.00005	5.21	0.000024
O	—	5.803	—	4.232	0.0008	4.232	0.00019
16O	99.762	5.803	0	4.232	0	4.232	0.0001
17O	0.038	5.78	0.18	4.2	0.004	4.2	0.236
18O	0.2	5.84	0	4.29	0	4.29	0.00016
P	100	5.13	0.2	3.307	0.005	3.312	0.172
S	—	2.847	—	1.0186	0.007	1.026	0.53
32S	95.02	2.804	0	0.988	0	0.988	0.54
33S	0.75	4.74	1.5	2.8	0.3	3.1	0.54
34S	4.21	3.48	0	1.52	0	1.52	0.227
36S	0.02	3.(1.)	0	1.1	0	1.1	0.15

Table 1: Scattering cross sections of biological relevant molecules. All cross sections are given in barns. Data were taken from Sears (1992)

References

Alder, B.J. & Wainwright, T.E. (1957). Phase transition for a hard-sphere system. *J. Chem. Phys.*, **27**, 1208.

Allen, M.P. & Tildesley, D.J. (1987). *Computer Simulation of Liquids*. Oxford University Press, Oxford.

Andersen, H. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, **72**, 2384–2393.

Anfinsen, C. (1973). Principles that govern the folding of protein chains. *Science*, **181**, 223–230.

Angell, C. (1995). Formation of glasses from liquids and biopolymers. *Science*, **267**, 1924–1934.

Austin, R., Beeson, L., K.W.and Eisenstein, Frauenfelder, H. & Gunsalus, I. (1975). Dynamics of ligand binding to myoglobin. *Biochemistry*, **13**, 5355–5373.

Basran, J., Casarotto, M., Barsukov, I. & Roberts, G.C.K. (1995). Role of the active-site carboxylate in dihydrofolate reductase: Kinetic and spectroscopic studies of the aspartate 26 .fwdarw. asparagine mutant of the lactobacillus casei enzyme. *Biochemistry*, **34**, 2872–2882.

Bauminger, E., Cohen, S., Nowik, I., Ofer, S. & Yariv, J. (1983). Dynamics of heme iron in crystals of metmyoglobin and deoxymyoglobin. *Proc. Natl. Acad. Sci. USA*, **80**, 736–740.

Bee, M. (1988)). *Quasielastic Neutron Scattering: Principles and Applications in Solid-state Chemistry, Biology and Materials Science*. (Hilger, Bristol.

Bellissent-Funel, M.C., Daniel, R., Durand, D., Ferrand, M., Finney, J., Pouget, S., Réat, V. & Smith, J.C. (1998). Nanosecond protein dynamics: First detection of a neutron incoherent spin-echo signal. *J. Am. Chem. Soc.*, **120**, 7347–7348.

Bengtzelius, U., Götze, W. & Sjölander, A. (1984). Dynamics of supercooled liquids and the glass transition. *J. Phys. C*, **17**, 5915–5934.

Benkovic, S., Fierke, C. & Naylor, A. (1988). Insights into enzyme function from studies on mutants of dihydrofolate reductase. *Science*, **239**, 1105–1110.

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000a). The protein data bank. *Nucleic Acids Research*, **28**, 235–242.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000b). The protein data bank. *Nucleic Acids Research*, **28**, 235–242.

Bicout, D.J. & Zaccai, G. (2001). Protein flexibility from the dynamical transition: a force constant analysis. *Biophys. J.*, **80**, 1115–1123.

Bizzarri, A.R., Paciaroni, A. & Cannistraro, S. (2000). Glasslike dynamical behavior of the plastocyanin hydration water. *Phys. Rev. E*, **62**, 3991–3999.

Brooks, B. & Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA*, **80**, 6571–6575.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). CHARMM. *J. Comput. Chem*, **4**, 187–217.

Brooks, C.I. (1987). The influence of long-range force truncation on the thermodynamics of aqueous ionic solutions. *J. Chem. Phys.*, **86**, 5156–5162.

Brooks, C.I., Pettitt, B. & Karplus, M. (1985). Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *J. Chem. Phys.*, **83**, 5897–5908.

Brunori, M., Cutruzzolà, F., Savino, C., Travaglini-Allocatelli, C., Vallone, B. & Gibson, Q. (1999). Does picosecond protein dynamics have survival value. *Trends Biochem Sci.*, **24**, 253–255.

Careri, G., Giansanti, A. & Rupley, J. (1986). Proton percolation on hydrated lysozyme powders. *Proc. Natl. Acad. Sci. USA*, **83**, 6810–6814.

Carlson, J.C.T., Kanter, A., Thuduppathy, G.R., Cody, V., Pineda, P.E., McIvor, R.S. & Wagner, C.R. (2003). Designing protein dimerizers: The importance of ligand conformational equilibria. *J. Am. Chem. Soc.*, **125**, 1501–1507.

Cheng, J.W., Lepre, C.A. & Moore, J.M. (1994). 15n nmr relaxation studies of the fk506 binding protein: Dynamic effects of ligand binding and implications of calcineurin recognition. *Biochemistry*, **33**, 4093–4100.

Chong, S.H., Joti, Y., Kidera, A., Go, N., Ostermann, A., Gassmann, A. & Parak, F. (2001). Dynamical transition of myoglobin in a crystal: comparative studies of x-ray crystallography and mössbauer spectroscopy. *Eur. Biophys. J.*, **30**, 319–329.

Cohen, S., Bauminger, E., Nowik, I., Ofer, S. & Yariv, J. (1981). Dynamics of the iron-containing core in crystals of the iron-storage protein, ferritin, through mössbauer spectroscopy. *Phys. Rev. Lett.*, **46**, 1244–1248.

Cordone, L., Ferrand, M., Vitrano, E. & Zaccai, G. (1999). Harmonic Behavior of Trehalose-Coated Carbon-Monoxy-Myoglobin at High Temperature. *Biophys. J.*, **76**, 1043–462.

Cruickshank, D., Helliwell, J. & Johnson, L., eds. (1992). *Time-resolved Macromolecular Crystallography*, Oxford U. Press, New York.

Cummins, H. (1999). The liquid-glass transition: a mode-coupling perspective. *J. Phys.: Condens. Matter*, **11**, A95–A117.

Cusack, S. (1986). *Comm. molec. cell. Biophys.*, **3**, 243–271.

Cusack, S. & Doster, W. (1990). Temperature dependence of the low frequency dynamics of myoglobin. Measurement of the vibrational frequency distribution by inelastic neutron scattering. *Biophys. J.*, **58**, 243–422.

Cusack, S., Smith, J.C., Finney, J., Karplus, M. & Trewhella, J. (1986). Low frequency dynamics of proteins studied by neutron time-of-flight spectroscopy. *Physica B*, **136**, 256–259.

Cusack, S., Smith, J.C., Finney, J., Tidor, B. & Karplus, M. (1988). Inelastic neutron scattering analysis of picosecond internal protein dynamics: comparison of harmonic theory with experiment. *J. Mol. Biol.*, **202**, 903–908.

Dahlborg, U. & Rupprecht, A. (1971). Hydration of dna. a neutron scattering study of oriented nadna. *Biopolymers*, **10**, 849.

Daniel, R., Smith, J.C., Ferrand, M., Hery, S., Dunn, R. & Finney, J. (1998). Enzyme activity below the dynamical transition at 220K. *Biophys. J.*, **75**, 2504–2507.

Daniel, R., Finney, J., Reat, V., Dunn, R., Ferrand, M. & Smith, J.C. (1999). Enzyme dynamics and activity: Time-scale dependence of dynamical transitions in glutamate dehydrogenase solution. *Biophys. J.*, **77**, 2184–2190.

Daniel, R., Dunn, R., Finney, J. & Smith, J.C. (2003a). The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 69–92.

Daniel, R., Finney, J. & Smith, J.C. (2003b). The dynamic transition in proteins may have a simple explanation. *Faraday Discuss.*, **122**, 163–169.

Dellerue, S., Petrescu, A., Smith, J.C., Longeville, S. & Bellisent-Funel, M.C. (2000). Collective dynamics of a photosynthetic protein probed by neutron spin-echo spectroscopy and molecular dynamics simulation. *Physica B*, **276-278**, 514–515.

Dellerue, S., Petrescu, A.J., Smith, J.C. & Bellissent-Funel, M. (2001). Radially Softening Diffusive Motions in a Globular Protein. *Biophys. J.*, **81**, 1666–1676.

Diehl, M., Doster, W., Petry, W. & Schober, H. (1997). Water-coupled low-frequency modes of myoglobin and lysozyme observed by inelastic neutron scattering. *Biophys. J.*, **73**, 2726–422.

Ding, X., Rasmussen, B., Petesko, G.A. & Ringe, D. (1994). Direct structural observation of an acyl-enzyme intermediate in the hydrolysis of an ester substrate by elastase. *Biochemistry*, **33**, 9285–9293.

Doster, W. & Settles, M. (1999). The dynamical transition in proteins: The role of hydrogen bonds. In Bellisent-Funel, M-C & Teixera, J., eds., *Hydration Processes in Biology*, vol. 305 of *NATO Science Series: Life Sciences*, IOS Press.

Doster, W., Cusack, S. & Petry, W. (1989a). Dynamic instability of liquidlike motions in a globular protein observed by inelastic neutron scattering. *Phys. Rev. Lett.*, **65**, 1080–1083.

Doster, W., Cusack, S. & Petry, W. (1989b). Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature*, **337**, 754–756.

Doster, W., Schirmacher, W. & Settles, M. (1990). Percolation model of ionic channel dynamics. *Biophys. J.*, **57**, 681–422.

Doster, W., Diehl, M., Petry, W. & Ferrand, M. (2001). Elastic resolution spectroscopy: a method to study molecular motions in small biological samples. *Physica B*, **301**, 65–68.

Dunn, R.V., Reat, V., Finney, J., Ferrand, M., Smith, J.C. & Daniel, R. (2000). Enzyme activity and dynamics: xylanase activity in the absence of fast anharmonic dynamics. *Biochem. J.*, **346**, 355–358.

Elber, R. & Karplus, M. (1987). Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science*, **235**, 318–321.

Engler, N., Ostermann, A., Niimura, N. & Parak, F.G. (2003). Hydrogen atoms in proteins: Positions and dynamics. *PNAS*, **100**, 10243–10248.

Epstein, D., Benkovic, S.J. & Wright, P.E. (1995). Dynamics of the dihydrofolate reductase-folate complex: catalytic sites and regions known to undergo conformational change exhibit diverse dynamical features. *Biochemistry*, **34**, 11037–48.

Erickson, H. (1989). *J. Mol. Biol.*, **296**, 465.

Ferrand, M., Dianoux, A.J., Petry, W. & Zaccai, G. (1993). Thermal motions and function of bacteriorhodopsin in purple membranes: effects of temperature and hydration studied by neutron scattering. *Proc. Natl. Acad. Sci. USA*, **90**, 9668–9672.

Feynman, R., Leighton, R. & Sands, M. (1963). *The Feynman Lectures on Physics*, vol. 1. Addison Wesley.

Finkelstein, A. & Janin, J. (1989). The price of lost freedom: entropy of bimolecular complex formation. *Protein Eng.*, **3**, 1–3.

Fischer, S., Smith, J.C. & Verma, C.S. (2001). Dissecting the vibrational entropy change on protein/ligand binding : burial of a water molecule in bovine pancreatic trypsin inhibitor. *J. Phys. Chem. B*, **105**, 8050–8055.

Fitter, J., Lechner, R., Bueldt, G. & Dencher, N. (1996). Internal molecular motions of bacteriorhodopsin: Hydration-induced flexibility studies by quasielastic incoherent neutron scattering using oriented purple membranes. *Proc. Natl. Acad. Sci. USA*.

Fitter, J., Lechner, R.E. & Dencher, N.A. (1997). Picosecond molecular motions in bacteriorhodopsin from neutron scattering. *Biophys. J.*, **73**, 2126–2137.

Frauenfelder, H. (1983). In *Stucture and Dynamics: Nucleic Acids and Proteins*, 369–376, New York: Adenine.

Frauenfelder, H. & McMahon, B. (2000). Energy landscape and fluctuations in proteins. *Ann. Phys.*, **9**, 655–667.

Frauenfelder, H., Petsko, G. & Tsernoglou, D. (1979). Temperaure-dependent x-ray diffraction as a probe of protein structural dynamics. *Nature*, **280**, 558–563.

Frauenfelder, H., Sligar, S. & Wolynes, P. (1991). The energy landscapes and motions of proteins. *Science*, **254**, 1598–1603.

Frauenfelder, H., Wolynes, P. & Austin, R. (1999). Biological physics. *Rev. Mod. Phys.*, **71**, 419–430.

Fushman, D., Ohlenschlager, O. & Ruterjans, H. (1994). *J. Biomol. Struct, Dyn.*, **11**, 1377.

Gabe, F., Bicout, D., Lehnert, U., Tehei, M., Weik, M. & Zaccai, G. (2002). Protein dynamics studied by neutron scattering. *Quat. Rev. Biophys.*, **35**.

Gerstein, M., Lesk, A.M. & Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry*, **33**, 6739–6749.

Gilson, M., Given, J., Bush, B. & McCammon, J. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.*, **72**, 1047–1069.

Goldstein, H. (1991). *Klassische Mechanic*. AULA-Verlag Wiesbaden.

Green, J., Fan, J. & Angell, C. (1994). The protein-glass analogy: some insights from homopeptide comparison. *J. Phys. Chem.*, **98**, 13780–13790.

Halle, B. (2004). Biomolecular cryocrystallography: Structural changes during flash-cooling. *PNAS*, **101**, 4793–4798.

Hayward, J. & Smith, J.C. (2002). Temperature dependence of protein dynamics: Computer simulation analysis of neutron scattering properties. *Biophys. J.*, **82**, 1216–1225.

Hayward, J., Becker, T. & Smith, J.C. (2002). The glass transition in proteins. In Krause, E. & Jäger, W., eds., *High Performance Computing in Science and Engineering'02*, 503–511, Springer.

Hill, T.L. (1994). *Thermodynamics of small systems*. Dover publications.

Howell, E., Villafranca, J., Warren, M., Oatley, S. & Kraut, J. (1986). Functional role of aspartic acid-27 in dihydrofolate reductase revealed by mutagenesis. *Science*, **231**, 1123–8.

Huennekens, F. (1994). The methotrexate story: A paradigm for development of cancer chemotherapeutic agents. *Adv. Enzyme Regul.*, **34**, 397–402.

Iben, I.E.T., Braunstein, D., Doster, W., Frauenfelder, H., Hong, M.K., Johnson, J.B., Luck, S., Ormos, P., Schulte, A., Steinbach, P.J., Xie, A.H. & Young, R.D. (1989). Glassy behaviour of a protein. *Phys. Rev. Lett.*, **62**, 1916–1919.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

Joti, Y., Nakasako, M., Kidera, A. & Go, N. (2002). Nonlinear temperature dependence of the crystal structure of lysozyme: correlation between coordinate shifts and thermal factors. *Acta Cryst. D*, **58**, 1421–1432.

Kamiyama, T. & Gekko, K. (2000). Effect of ligand binding on the flexibility of dihydrofolate reductase as revealed by compressibility. *Biochim Biophys Acta*, **1478**, 257–266.

Karplus, M. & Janin, J. (1999). Comment on: 'the entropy cost of protein association'. *Protein Eng.*, **12**, 185–186.

Karplus, M. & Petsko, G. (1990). Molecular dynamics simulations in biology. *Nature*, **347**, 631–639.

Keller, H. & Debrunner, P. (1980). Evidence for conformational and diffusional mean square displacements in frozen aqueous solution of oxymyoglobin. *Phys. Rev. Lett.*, **45**, 68–71.

Kitao, A., Hayward, S. & Go, N. (1998). Energy landscape of a native protein: Jumping-among-minima model. *Proteins*, **33**, 496–517.

Klotz, I. (1985). Ligand–receptor interactions: facts and fantasies. *Quart. Rev. Biophys.*, **18**, 227–259.

Knapp, E., Fischer, S. & Parak, F. (1982). Protein dynamics from Mößbauer spectra. the temperature dependence. *J. Phys. Chem.*, **86**, 5042–5047.

Kneller, G.R. (1991). Superposition of molecular structures using quaternions. *Molecular Simulations*, **7**, 113–119.

Kneller, G.R. & Smith, J.C. (1994). Liquid-like side-chain dynamics in myoglobin. *J. Mol. Biol.*, **242**, 181–185.

Kneller, G.R., Doster, W., Settles, M., Cusack, S. & Smith, J.C. (1992). Methyl group dynamics in the crystalline alanine dipeptide: A combined computer simulation and neutron scattering analysis. *J. Chem. Phys.*, **97**, 8864–8879.

Kneller, G.R., Keiner, V., Kneller, M. & Schiller, M. (1995). nMOLDYN: A program package for a neutron scattering oriented analysis of molecular dynamics simulations. *Comp. Phys. Commun.*, **91**, 191–214.

Koshland, D. (1958). Application of a theory of enzyme specifity to protein synthesis. *Proc. Natl. Acad. Sci. USA*, **44**, 98–105.

La Penna, G., Pratolongo, R. & Perico, A. (1999). Mode-coupling smoluchowski dynamics of polymers in the limit of rigid structures. *Macromolecules*, **32**, 506–513.

Lamb, M. & Jorgensen, W. (1997). Computational approaches to molecular recognition. *Curr. Opin. Chem. Biol.*, **1**, 449–457.

Lamy, A.V., Souaille, M. & Smith, J.C. (1996). Simulation evidence for experimentally detectable low-temperature vibrational inhomogeneity in a globular protein. *Biopolymers*, **39**, 471–478.

Leach, A.R. (1996). *Molecular modelling: principles and applications*. Addison Wesley Longman.

Lehnert, U., Reat, V., Weik, M., Zaccai, G. & Pfister, C. (1998). Thermal Motions in Bacteriorhodopsin at Different Hydration Levels Studied by Neutron Scattering: Correlation with Kinetics and Light-Induced Conformational Changes. *Biophys. J.*, **75**, 1945–462.

Levinthal, C. (1969). How to fold graciously. In *Mössbauer spectroscopy in biological systems*.

Lian, L., Barsukov, I., Sutcliffe, M., Sze, K. & Roberts, G. (1994). Protein-ligand interactions: Exchange processes and determination of ligand conformation and protein-ligand contacts. *Methods. Enzymol.*, **239**, 657–700.

Lovesey, S. (1987)). *Theory of Neutron Scattering from Condensed Matter, Vol.1*. (Oxford University Press.

Mackerell, A.D., Bashford, D., Bellot, M., Dunbrack, J.R., Evenseck, R.L., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, I.W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, J. & Karplus, M. (1998). All-atom empirical potential for molecular modelling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.

McCammon, J.A., Gelin, B.R. & Karplus, M. (1977). Dynamics of folded proteins. *Nature*, **267**, 585–90.

McQuarrie, D.A. (1976). *Statistical mechanics*. Harper&Row.

Middendorf, H. & Willis, B. (1972). *Fourth International Liquid Crystal Conference - Kent Ohio*.

Nienhaus, U.G., Müller, J.D., McMahon, B.H. & Frauenfelder, H. (1997). Protein dynamics; conformational substates; energy landscape; myoglobin; infrared spectroscopy. *Physica D*, **107**, 297–311.

Ostermann, A., Waschipky, R., Parak, F.G. & Nienhaus, G.U. (2000). Ligand binding and conformational motions in myoglobin. *Nature*, **404**, 205–208.

Page, M. & Jencks, W. (1971). Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. USA*, **68**, 1678–1683.

Parak, F. (2003). Physical aspects of protein dynamics. *Rep. Prog. Phys.*, **66**, 103–129.

Parak, F., Frolov, E.N., Kononenko, A.A., Mössbauer, R.L., Goldanskii, V.I. & Rubin, A.B. (1980). Evidence for a correlation between the photoinduced electron transfer and dynamic properties of the chromatophore membranes from Rhodospirillum Rubrum. *FEBS Letters*, **117**, 368–372.

Parak, F., Knapp, E. & Kucheida, D. (1982). Protein dynamics: Mößbauer spectroscopy on deoxymyoglobin crystals. *J. Mol. Biol.*, **161**, 177–194.

Parkin, S., Rupp, B. & Hope, H. (1996). Structure of bovine pancreatic trypsin inhibitor at 125K: Definition of carboxyl-terminal residues Gly57 and Ala58. *Crystallogr. D. Biol. Crystallogr.*, **52**, 18–29.

Perez, J., Zanotti, J.M. & Durand, D. (1999). Evolution of the internal dynamics of two globular proteins from dry powder to solution. *Biophys. J.*, **77**, 454–469.

Perico, A. & Pratolongo, R. (1997). Maximum-Correlation Mode-Coupling Approach to the Smoluchowski Dynamics of Polymers. *Macromolecules*, **30**, 5958–5969.

Perlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham III, T.E., DeBold, S., Ferguson, D., Seibel, G. & Kollmann, P. (1995). Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structure and energetic properties of molecules. *Comp. Phys. Commun.*, **91**, 1–41.

Perutz, M. & Mathews, F. (1966). An x-ray study of azide methaemoglobin. *J. Mol. Biol.*, **21**, 199–202.

Rahman, A., Singwi, K.S. & Sjölander, A. (1962). Theory of slow neutron scattering by liquids. i. *Phys. Rev.*, **126**, 986–996.

Rasmussen, B., Stock, A., Ringe, D. & Petsko, G.A. (1992). Crystalline ribunuclease a loses function below the dynamical transition at 220K. *Nature*, **357**, 423–424.

Réat, V., Zaccai, G., Ferrand, M. & Pfister, C. (1997). Functional dynamics in purple membrane. In Biological Macromolecular Dynamics, 117–122, Adenine Press, Schenectady, NY.

Reat, V., Dunn, R., Ferrand, M., Finney, J., Daniel, R. & Smith, J.C. (2000). Solvent dependence of dynamic transitions in protein solutions. *Proc. Natl. Acad. Sci. USA*, **97**, 9961–66.

Rieutord, F. (1990). Tech. Rep. 90RI17T, ILL, Grenoble.

Rischel, C. & et al. (1994). Comparison of backbone dynamics of apo- and holo-acyl-coenzyme a binding protein using 15n relaxation measurements. *Biochemistry*, **33**, 13997 – 14002.

Ryckaert, J.P., Ciccotti, G. & Berendsen, H.J.C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes (SHAKE). *J. Comput. Phys.*, **23**, 327–341.

Sawaya, M. & Kraut, J. (1997). Loop and subdomain movements in the mechanism of escherichia coli dihydrofolate reductase: Crystallographic evidence. *Biochemistry*, **36**, 586–603.

Schlichting, I., Berendzen, J., Phillips, G.N. & Sweet, R.M. (1994). Crystal structure of photolysed carbonmonoxy-myoglobin. *Nature*, **371**, 808 – 812.

Schlichting, I., Berendzen, J., Chu, K., Stock, A.M., Maves, S.A., Benson, D.E., Sweet, R.M., Ringe, D., Petsko, G.A. & Sligar, S.G. (2000). The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution. *Science*, **287**, 1615–1622.

Schwarzl, S., Tschopp, T., Smith, J.C. & Fischer, S. (2002). Can the calculation of ligand binding free energies be improved with continuum solvent electrostatics and an ideal-gas entropy correction? *J. Comp. Chem.*, **23**, 1143–1149.

Sears, V.F. (1992). Neutron scattering lengths and cross-sections. *Neutron News*, **3**, 26–37.

Singh, U. & Benkovic, S. (1988). A free-energy perturbation study of the binding of methotrexate to mutants of dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA*, **85**, 9519–23.

Smith, J.C. (1991). Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. *Quart. Rev. Biophys.*, **24**, 227–291.

Smith, J.C., Cusack, S., Pezzeca, B., U.and Brooks & Karplus, M. (1986). Inelastic neutron scattering analysis of low frequency motion in proteins: A normal mode study of the bovine pancreatic trypsin inhibitor. *J. Chem. Phys.*, **85**, 3636–3654.

Smith, J.C., Cusack, S., Poole, P. & Finney, J. (1987). Direct measurement of hydration-related dynamic changes in lysozyme using inelastic neutron scatterin spectroscopy. *J. Biomol. Struct. Dyn.*, **4**, 583–588.

Smith, J.C., Cusack, S., Tidor, B. & Karplus, M. (1990). Inelastic neutron scattering analysis of low-frequency motions in proteins: Harmonic and damped harmonic models of bovine pancreatic tryspin inhibitor. *J. Chem. Phys.*, **93**, 2974.

Stein, D. (1985). A model of protein conformational substates. *Proc. Natl. Acad. Sci. USA*, **82**, 3670–3672.

Steinbach, P.J. & Brooks, B.R. (1993). Protein hydration elucidated by molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA*, **90**, 9135–9139.

Steinbach, P.J. & Brooks, B.R. (1994). New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comp. Chem.*, **15**, 667–683.

Steinbach, P.J. & Brooks, B.R. (1996). Hydrated myoglobin's anharmonic fluctuations are not primarily due to dihedral transitions. *Proc. Natl. Acad. Sci. USA*, **93**, 55–59.

Steinberg, I. & Scheraga, H. (1963). *J. Biol. Chem.*, **238**, 172.

Stone, S. & Morrison, J.F. (1984). Catalytic mechanism of the dihydrofolate reductase reaction as determined by ph studies. *Biochemistry*, **23**, 2753–8.

Sturtevant, J. (1977). Heat capacity and entropy changes in processes involving proteins. *Proc. Natl. Acad. Sci. USA*, **74**, 2236–2240.

Tamura, A. & Privalov, P.L. (1997). The entropy cost of protein association. *J. Mol. Biol.*, **273**, 1048–1060.

Tarek, M. & Tobias, D.J. (2002). Role of protein-water hydrogen bond dynamics in the protein dynamical transition. *Phys. Rev. Lett.*, **88**, 138101.

Teeter, M., Yamano, A., Stec, B. & Mohanty, U. (2001). On the nature of a glassy state of matter in a hydrated protein: Relation to protein function. *Proc. Natl. Acad. Sci. USA*, **98**, 11242–11247.

Teng, T.Y., Srajer, V. & Moffat, K. (1994). Photolysis-induced structural changes in single crystals of carbonmonoxy myoglobin at 40 k. *Nature Structural Biology*, **1**, 701 – 705.

Tidor, B. & Karplus, M. (1994). The contribution of vibrational entropy to molecular association. *J. Mol. Biol.*, **238**, 405–414.

Tournier, A., Xu, J. & Smith, J.C. (2003). Translational Hydration Water Dynamics Drives the Protein Glass Transition. *Biophys. J.*, **85**, 1871–1875.

Tournier, A.L. & Smith, J.C. (2003). Principal components of the protein dynamical transition. *Phys. Rev. Lett.*, **91**.

Tuckerman, M.E. & Martyna, G.J. (2000). Understanding modern molecular dynamics: Techniques and applications. *J. Phys. Chem.*, **194**, 159–178.

Van Hove, L. (1954). Correlations in space and time and born approximation scattering in systems of interacting particles. *Phys. Rev.*, **95**, 249–262.

Vitkup, D., Ringe, D., Petsko, G.A. & Karplus, M. (2000). Solvent mobility and the protein glass transition. *Nature Structural Biology*, **7**, 34–38.

Wang, W. & et al. (2001). Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 211–243.

Yu, Y.B., Privalov, P.L. & Hodges, R.S. (2001). Contribution of translational and rotational motions to molecular association in aqueous solution. *Biophys. J.*, **81**, 1632–1642.

Zaccai, G. (2000a). How Soft Is a Protein? A Protein Dynamics Force Constant Measured by Neutron Scattering. *Science*, **288**, 1604–1607.

Zaccai, G. (2000b). Moist and soft, dry and stiff: a review of neutron experiments on hydrations-dynamics-activity relations in the purple membrane of *halobacterium salinarum*. *Biophys. Chem.*, **86**, 249–257.

Zaman, M.H., Berry, R.S. & Sosnick, T.R. (1999). Entropic benefit of a cross-link in protein association. *Proteins*, **48**, 341–351.

Zanotti, J., Bellissent-Funel, M. & Parello, J. (1999a). Hydration-Coupled Dynamics in Proteins Studied by Neutron Scattering and NMR: The Case of the Typical EF-Hand Calcium-Binding Parvalbumin. *Biophys. J.*, **76**, 2390–1676.

Zanotti, J.M., Bellissent-Funel, M.C. & Parello, J. (1999b). Dynamics of a globular protein as studied by neutron scattering and solid-state nmr. *Physics B*, **234**, 228–230.

Zidek, L., Novotny, M.V. & Stone, M.J. (1999). Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat. Struct. Biol.*, **6**, 1118–1121.