Faculty of Physics and Astronomy

Ruprecht-Karls-University of Heidelberg

Diploma thesis
in Physics
submitted by
Roland Schulz
born in Duisburg
2007

Effectively Filtered Instantaneous Normal Modes Reproduce Dynamical Transition of Proteins

This diploma thesis has been carried out by Roland Schulz at the
Interdisciplinary Center for Scientific Computing
under the supervision of
Prof. Dr. Jeremy C. Smith,
Dr. Isabella Daidone
and
Dr. Marimuthu Krishnan

Many thanks to Marimuthu Krishnan and Isabella Daidone for being very nice, supportive and helpful supervisors, my parents for being the best parents, David Foltin and Melanie Syha for being great friends, Jeremy C. Smith for the opportunity and the nice work environment, Jan-Hendrik Prinz and Mithun Biswas for proofreading, and to the whole Computational Molecular Biophysics Group.

Abstract

The Instantaneous Normal Mode theory predicts that the self-diffusion constant D is proportional to the fraction of diffusive modes f_u ($D \propto {f_u}^\alpha$). For a sample configuration the curvature of the potential energy is computed along all modes. The diffusive modes are a subset of those with a negative curvature. Different filtering methods were successfully used to compute this subset for simple liquids. We evaluated those methods on a protein over a temperature range from $20 - 300\,\mathrm{K}$ by testing the expected relation $D \propto {f_u}^\alpha$ with D computed from the slope of the mean square displacement of a molecular dynamics trajectory. The "escape mode" method supplied a good agreement over two orders of magnitude thus enabling the computation of temperature dependent properties of proteins from a few sampled configurations. The so computed self-diffusion constant reproduces the dynamical transition. For the present however, the computing time is too large for practical use.

Die Instantaneous Normal Mode Theorie sagt voraus, dass die Selbstdiffusionskonstante D proportional zum Anteil der zur Diffusion beitragenden Moden f_u ($D \propto {f_u}^\alpha$) ist. Für eine Beispielkonfiguration wird die Krümmung der potentiellen Energie längst aller Moden berechnet. Die beitragenden Moden sind eine Teilmenge derer mit negativer Krümmung.Verschiedene Filtermethoden wurden zuvor erfolgreich verwendet, um diese Teilmenge für einfache Flüssigkeiten zu berechnen. Wir haben diese Methoden an einem Protein im Bereich $20 - 300\,\mathrm{K}$ getestet, indem wir die vorhergesagte Relation $D \propto {f_u}^\alpha$ überprüft haben, wobei D aus der Steigung des mittleren quadratischen Abstandes einer Molekulardynamik berechnet wurde. Die "escape mode" Methode hat eine gute Übereinstimmung über zwei Größenordnungen geliefert und erlaubt somit, temperaturabhängige Größen ausgehend von wenigen Beispielkonfigurationen zu berechnen. Die so berechnete Selbstdiffusionskonstante zeigt den bekannten Verlauf der "dynamical transition". Die Rechenzeit ist jedoch vorerst zu groß für praktische Anwendungen.

Contents

Abstract **3**

1 Introduction **6**
1.1 Objective of thesis 6
1.2 Protein Structure and Dynamics 8
1.3 Simulation of Macromolecules 10

2 Theory **13**
2.1 Normal Mode Analysis 13
2.2 Instantaneous Normal Mode Theory 14
2.2.1 Derivation based on the Inherent Structure approach . 16
2.2.2 Derivation based on the Random Energy Model 20
2.3 Filtering Methods 21
2.3.1 Double-Well Modes 21
2.3.2 Escape Modes 23
2.3.3 Minimization based Filtering 24

3 Methods **26**
3.1 Molecular Dynamics and Force Fields 26
3.2 MD details for the sample systems 28
3.3 Computing Hessian Matrices for Large Systems with PME . . 29
3.4 Filtering Modes 30
3.4.1 Landscape based filtering 30
3.4.2 Minimization based filtering 32

4 Results **33**
4.1 Model systems 33
4.2 Results for Myoglobin 35
4.3 Results for the scorpion toxin 36
4.3.1 Comparison to MSD 36
4.3.2 Results for different filtering methods 37

4.3.3 Characteristics of escape filtering 46

5 Conclusion **51**

A CHARMM script for landscape based filtering **52**

B MPI program to run CHARMM distributed on a cluster **56**

Chapter 1

Introduction

1.1 Objective of thesis

The role of anharmonic dynamics in the function of proteins has been widely recognized in recent years [Daniel et al., 1998, 2003, Frauenfelder et al., 1988]. Proteins have a wide spectrum of dynamical processes ranging from femtosecond localized harmonic vibrations to collective diffusive movements with time scales up to the millisecond range. Despite of differences of many magnitudes in time scale the motions are often coupled, complicating the study by experiments and simulations.

Several experiments and simulations [Bizzarri et al., 2000, Doster et al., 1989, Dunn et al., 2000, Hayward and Smith, 2002, Parak et al., 1981, Smith et al., 1990, Tilton Jr et al., 1992] showed that at $180-220\,\mathrm{K}$ the mean-square displacement as a function of temperature, representing the dynamics, shows a change in the slope. This indicates large amplitude atomic displacements above this temperature. This slope change is called "Dynamical Transition" (DT) and is now understood as the activation of anharmonic dynamics. It also coincides with the onset of biological function [Ferrand et al., 1993, Parak et al., 1980]. Details like the role of hydration is an area of active research [Hayward and Smith, 2002, Kurkal et al., 2005, Reat et al., 2000].

To study those dynamical processes different analytical methods have been developed. Normal mode analysis (NMA) uses a second order harmonic

approximation of the potential energy around an equilibrium configuration. In the region of validity of this approximation, the dynamics are described completely by the second derivative matrix of this energy. This description includes the vibrational frequency of all possible independent motions, which are also called modes. These frequencies are typically visualized as a histogram and then called frequency spectrum.

The Instantaneous Normal Mode (INM) method [Keyes, 1994, 1997, Madan and Keyes, 1993, Seeley and Keyes, 1989] is also based on the second derivative of the potential energy. Unlike for the NMA the Hessian matrix is computed for sample configurations generated at a certain temperature instead of fully minimized configurations. Earlier INM studies were published mainly about simple supercooled liquids with a focus on glass transition, mode-coupling temperature, and crystallization [Keyes, 1995, Keyes et al., 1997, La Nave et al., 2000, Li et al., 1998, Madan and Keyes, 1993, Sciortino and Tartaglia, 1997]. Those results have demonstrated that it is possible to calculate temperature dependent dynamical properties, such as the self-diffusion constant, from a few sampled configurations. This is done by computing the fraction of unstable modes, which are a subset of the negative modes and are proportional to these dynamical properties. To reach a good agreement it was important to compute this subset accurately. For this task different filtering methods have been proposed.

Those types of studies, which connect these dynamical properties to the fraction of unstable modes have not been done for proteins so far. For proteins the INM method was only used for studies on energy barriers and vibrational relaxation. The first ones compute the temperature dependent barrier height distribution by using the fraction of unstable modes and a few assumptions leading to the functional form of the distribution [Straub and Choi, 1994, Straub and Thirumalai, 1993]. For the second ones, the whole INM spectrum was used similar to a NMA spectrum to identify modes and their frequency shift. Thus instead of using the fraction of unstable modes, which focuses on the difference between positive and negative eigenvalues, they ignored the negative part of the spectrum [Bu and Straub, 2003].

In this work the goal is to investigate this connection, meaning whether

computing the self-diffusion constant in configurational space using the INM method is also possible for proteins. This is done for temperatures around the dynamical transition to see if the INM method can reproduce this transition. As for liquids it is essential to accurately compute the subset of unstable modes out of all negative modes, so all filtering proposed for liquids have been evaluated for proteins.

The rest of the introduction will give a broad overview about proteins, their structure and dynamics and how to simulate them. More details can be taken from any standard textbook on biophysics e.g. [Lesk, 2004].

1.2 Protein Structure and Dynamics

Proteins are unbranched polymer chains of amino acids connected by peptide bonds. An amino acid has the general formula $NH_2C_\alpha RHCO_2H$ where R stands for the group which is specific for each of the 20 different naturally occurring amino acids. Proteins play a vital role in various biological processes ranging from structural to functional and thus are central building blocks of all cells.

The amino acid sequence, meaning the order of the residues along the chain, defines the structure of a protein together with the chemical and physical conditions. The sequence itself is called the primary structure. The native state of a protein is commonly a compact folded one. Subparts of proteins form reoccurring patterns like α-helices and β-strands. These patterns are called the secondary structure. The three dimensional arrangement of the whole protein, beyond these patterns, is called tertiary structure. It has been resolved for many proteins by X-ray scattering or nuclear magnetic resonance (NMR) experiments. The solved structures are available on the PDB database [Berman et al., 2000, 2003] and can be downloaded from there as a starting structure for simulations.

The question how proteins form the tertiary structure only based on the amino acid sequence is called the folding problem. Trying every possible configuration with one configuration every 100 fs would result in a time longer than the existence of the universe. This is known as the Levinthal paradox.

Overcoming this paradox is possible by using the fact that the free energy is lower near the native state and so can be interpreted as a folding funnel, enabling the protein to fold in biological time scales.

Because both the folding process and many functional processes are related to the possible dynamics of proteins, these dynamics are intensely studied. The time scale of these dynamics range over more than 12 orders of magnitude. Since a full characterization of dynamics on all time scales is difficult, it is common to define different groups of possible dynamics, such as harmonic and anharmonic. The distribution of possible dynamics is often represented as a frequency spectrum, with the number of possible vibrational modes versus their frequency.

Harmonic dynamics are all short range ($< 1\,\text{Å}$), small time scale ($< 20\,\text{fs}$) vibrations of bond lengths and angles around their equilibrium positions and are exited at all temperatures. They are exited numerously at high frequencies. Anharmonic dynamics are more diverse because they range from near zero frequency domain motions, which are motions of whole subparts, to dihedral angle bending, which are motions of whole amino acids, and to methyl group rotations. They are also related to conformational diffusion and biological function.

The mean square displacement (MSD) measured from neutron scattering and computed from a molecular dynamics (MD) trajectory for myoglobin is shown in fig. fig. 1.1. Similar results were obtained for other proteins and other experimental techniques like x-ray scattering and Mössbauer spectroscopy. One can clearly see the slope change at around 220K and the large MSD at higher temperatures. This can be understood by different models explaining this nonlinear temperature dependence. One model [Doster et al., 1989] consists of a two-state potential with a free-energy difference of ΔG and a distance d between the states. The increased population of the higher energy state increases the MSD above the DT. Another model [Bicout and Zaccai, 2001] assumes two potentials with different force constants. Above the DT the population of the lower force constant potential increases and thus also the MSD increases. Below the transition the MSD is proportional to the temperature, which is in agreement with the explanation of single

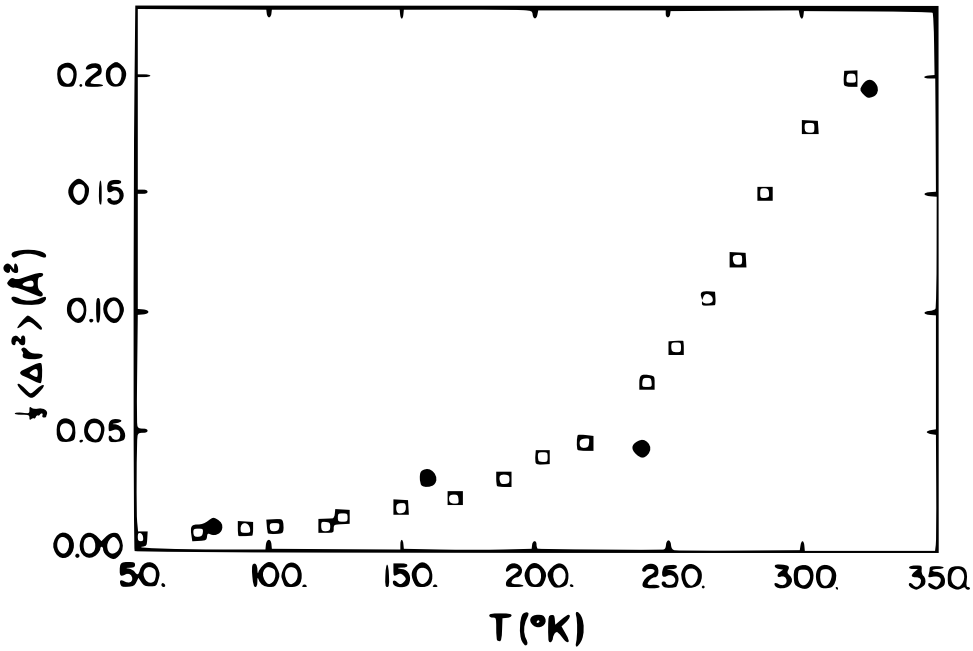

Figure 1.1: Temperature dependence of the isotropically averaged MSD of myoglobin, $1/3\langle\Delta r^2\rangle$, averaged over the exchangeable hydrogens. □, From the neutron experiment; •, from MD simulations [Smith et al., 1990].

potential harmonic dynamics at those temperatures.

Because experiments have shown a strong coupling of protein and water around the DT, dual-heat bath simulations were carried out [Smith et al., 2006], which set the protein and the water at different temperatures. They showed that the dynamical transition was suppressed up to 300K by either holding the protein or the water at or below 180K.

1.3 Simulation of Macromolecules

Macromolecules in general and biomolecules and proteins in particular can be simulated with different levels of detail and accuracy. This ranges from coarse grained simulations (with e.g. an amino acid or a whole protein represented as one beat) to quantum mechanics (QM) simulations. Since many biological processes require atomic detail and at least ns time scale, and QM simulations are computationally not feasible on this time scale, all-atomic MD simulation

is often used.

All-atomic simulations are possible because of the Born-Oppenheimer approximation. It assumes that the Schrödinger equation can be separated into an electron and a nuclei contribution because the electron motions are significantly faster. It thus neglects the terms in the equation which couple both types of motion and is so not valid for strong coupling (e.g. photochemistry). Permanent multipole moments are modeled by empirical partial charges assigned to all atoms. In the following we will always assume this approximation and thus only take the degrees of freedom of the nuclei into account. This allows to define a so called empirical force field, which specifies all interaction energy terms, and to compute the potential energy and forces for a certain configuration.

The so computed potential energy or the free energy are often visualized as an energy landscape, which is the energy as a function of the nuclei coordinates, to show e.g. folding funnels or transition states. Since every atom out of N atoms has three degrees of freedom, the potential energy is a $3N$ dimensional function and can be described as a hyper-surface in a $3N + 1$ dimensional space.

The complexity and limitation of macromolecule simulations arise mainly from the high dimensionality and the required time scale. An example for a value, which is difficult to compute because of the high dimensionality, are properties of a single minimum [Wales, 2003]. For a vast number of minima it is impossible to count or characterize all, so the practicability depends on this number. Although there is no analytical expression for the number of minima n_{min} as a function of the number of atoms N, also because it is system dependent, simple theories and observations show an exponential growth of $n_{min}(N)$. This can be understood by looking at the system subdivided into N_r equivalent subsystems of z atoms each. Then the number of minima can be written as

$$n_{min}(N_r z) = n_{min}(z)^{N_r} \tag{1.1}$$

because each minimum in one local region combined with any minimum

in another local region form a local minimum in the whole system. This exponential growth renders it impossible to analyze all minima for larger systems.

The second limitation, related to time scale, arises from the gap between the simulation time-step and biologically and chemically relevant time scales in the μs range. Because of accuracy the maximum simulation time-step is limited by the fastest vibration. For all-atomic MD of biomolecules, this is the covalent bond length vibration of hydrogen atoms, which is about 10fs, limiting the simulation time-step to 1fs. Thus 10^9 steps are required to reach the μs time scale and so even on the largest computer clusters it is only possible to compute a few hundreds ns long trajectory on one day of simulation [Fitch et al., 2006].

The harmonic dynamics can be investigated from a computational point of view by the NMA method, which is further explained in the theory part. Because this does not require any MD simulations, this is computationally very fast. The anharmonic dynamics can be investigated by analyzing a MD trajectory with principal component analysis, Langevin dynamics or MSD analysis. Since these methods require a molecular dynamics trajectory, the analysis is limited to time scales feasible for MD.

Chapter 2

Theory

2.1 Normal Mode Analysis

The Normal Mode Analysis (NMA) is a quadratic approximation of the potential energy around a local minimum. In the area of validity the dynamics can be described with just the second derivative matrix of the potential energy, which is also called the Hessian matrix.

The Taylor expansion of the potential energy $V(\boldsymbol{r})$ is calculated to the second order, where $\boldsymbol{r}$ is a vector in the $3N$ configurational space of a system with N atoms. By setting the potential energy to zero at the minimum ($V(\boldsymbol{r}_{min}) = 0\,{}^{\mathrm{kcal}}/_{\mathrm{mol}}$) the expansion becomes

$$V(\boldsymbol{r}) \;=\; \frac{1}{2}\boldsymbol{r}^{\dagger}\boldsymbol{H}\boldsymbol{r} \quad \text{with } H_{ij} = \frac{\partial^2 V}{\partial r_i \partial r_j}. \tag{2.1}$$

Writing Newton's Second Law in vector notation using the previous equa-

tion, results in

$$\boldsymbol{M}\ddot{\boldsymbol{r}} = -\boldsymbol{\nabla} V(\boldsymbol{r}) \tag{2.2}$$

$$\text{with } M = \begin{pmatrix} m_1 & & & 0 \\ & m_2 & & \\ & & \dots & \\ 0 & & & m_N \end{pmatrix} \tag{2.3}$$

$$\boldsymbol{M}\ddot{\boldsymbol{r}} = -\boldsymbol{H}\boldsymbol{r} \tag{2.4}$$

where m_i is the mass of atom i. This differential equation can easily be solved using mass-weighted coordinates $\boldsymbol{u} = \boldsymbol{T}^\dagger \boldsymbol{r}$, where $\boldsymbol{T}$ is the matrix transforming the mass-weighted coordinates to the original coordinates. Then $\boldsymbol{M}$ becomes $\mathbf{1}$ and it follows

$$\boldsymbol{\Lambda} = \boldsymbol{T}^\dagger \boldsymbol{H} \boldsymbol{T} \tag{2.5}$$

$$\ddot{\boldsymbol{u}} = -\boldsymbol{\Lambda}\boldsymbol{u} \tag{2.6}$$

This is a decoupled second order linear differential equation and thus has the solution

$$u_k = C_k \cos(\omega_k t + \phi_k) \quad \text{with } \omega_k = \sqrt{\lambda_k} \tag{2.7}$$

where λ_k is the k-th eigenvalue of $\boldsymbol{\Lambda}$ and ϕ_k the phase of the motion. Concluding it can be stated that only the eigenvalue problem of the Hessian matrix has to be solved to derive an analytical expression for the equation of motions in this approximation. Because the differential equations are decoupled the motions are independent and any linear combination is also a possible motion.

2.2 Instantaneous Normal Mode Theory

To also describe the more challenging anharmonic dynamics using the second derivative matrix, the Instantaneous Normal Mode (INM) approach com-

putes the Hessian matrices for a number of non-minimized sample configurations. Because they are not (fully) minimized the curvature and thus the eigenvalue can be negative. Dynamical properties can be computed from the Hessian matrices by linear relations, which were partly derived between the average fraction of diffusive modes on the one hand and on the other the self-diffusion constant or the configurational entropy [Keyes, 2000, 2006, 1994, 1997, La Nave et al., 2000, 2001, 2002]. They are either based on the random energy model or on Stillinger's Inherent Structure (IS) concept. These two will be explained in the next subsection. Equipped with this linear relation, INM is a non-equilibrium statistical theory connecting a dynamical property (e.g. diffusion constant) to an equilibrium average, in this case the curvature of the energy landscape of sample configurations. Despite much work on these derivations by different groups, no single comprehensive theory has been developed yet. The derivations relay on many approximations and have still open questions [Bembenek and Laird, 1995, Cho, 1994, Gezelter et al., 1997, Keyes, 1997, Keyes et al., 1997, Li and Keyes, 1997, 1999, Li et al., 1998, Madan and Keyes, 1993, Moore and Keyes, 1994, Sciortino and Tartaglia, 1997], some of which will be described after each derivation. The best evidence for the linear relation for liquids comes thus from several numerical results, which show a very good agreement over many decades [La Nave et al., 2000, Li and Keyes, 1999]. An intuitive connection is that the fraction of negative modes f_u is by definition equal to the fraction of directions in configurational space with negative curvature. Assuming the system is near a local minimum, the force is approximately zero ($F \approx 0$). Most of the time (larger Boltzmann weight) and for most of the system (see below for composite landscape) this is true. In this case directions along negative modes are unstable because the force is driving the system away from the local minimum and therefore these modes are called unstable modes. So it is reasonable to assume that diffusion will happen mainly in the f_u-dimensional subspace. By further assuming that the contribution of all possible diffusive directions is nearly the same, the self-diffusion constant D is proportional to f_u ($D \propto f_u$). In the following, different derivations will be presented.

Afterwards different filtering methods will be discussed. They compute

the subset of unstable modes out of all negative modes. For liquids it was shown that the linearity to the number of unstable modes only holds after this filtering and thus is essential for applying the INM approach and most filtering methods were material specific [Keyes, 2006]. Because of the importance of the filtering method to the approach and its material dependence, we focused on evaluating these methods for biomolecules. Another problem is the choice of the coordinate system. Since the system is not minimized, the results are not independently of it. It is still unresolved which coordinate system has to be used. We assume in the following that the Cartesian coordinate system is adequate.

2.2.1 Derivation based on the Inherent Structure approach

The Inherent Structure (IS) defines a basin as the region of the energy landscape, for which steepest descent minimizations started at any point converges to the same minimum [Stillinger and Weber, 1982, 1984]. Stillinger and Webber introduced it to understand the dynamics during melting. The participation function $Q(T)$ as a function of temperature T can be subdivided onto all minima. Written as an integral over all possible minimum energies, one gets

$$Q(T) = \int dU_m \Omega(U_m) \langle Q_m(U_m, T) \rangle \mathrm{e}^{-\beta U_m} \tag{2.8}$$

where U_m is the energy of a certain local minimum, $\Omega(U_m)$ is the distribution of those energies of all local minima, $\langle Q_m(U_m, T) \rangle$ is the participation function inside each IS basin averaged over all basins with the same minimum energy and $\beta = \frac{1}{kT}$ with k being the Boltzmann constant.

The INM derivation, based on the IS notion, connects the self-diffusion constant $D(T)$ to the escape rate $R(T)$ out of one basin which in turn is connected to the fraction of diffusive modes. This is based on the observation that diffusion requires the system to visit different IS basins, which are connected by barriers and these contribute to the fraction of negative modes.

This can be written as

$$
\begin{aligned}
D(T) &= \langle(\delta R)^2\rangle R(T)/6N \qquad \text{(Markov Approximation)} && (2.9)\\
R(T) &\propto N_r S(T) && (2.10)\\
f_u(T) &= \frac{1}{3z} S(T) && (2.11)\\
\Rightarrow D(T) &\propto f_u(T) && (2.12)
\end{aligned}
$$

where R is the rate the system is visiting different IS, δR is the displacement of successive IS, N_r is the number of local regions, S is the saddle sum and z is the number of atoms per local region. Eq. 2.10 and 2.11 will be derived next.

By decomposing the participation function not only by the minima as for IS but for all stationary points, the Q_m in eq. 2.8 can be written as

$$
Q_m = Q_{mm} + \sum_{s\in m} \frac{Q_{ms}}{m(s)} \mathrm{e}^{-\beta(U_s - U_m)} \tag{2.13}
$$

where Q_{mm} or Q_{ms} are computed in the subbasin belonging to a minimum or a saddle and the latter is divided by the number of minimum basins $m(s)$ sharing one saddle. Using this Q_m eq. 2.8 becomes

$$
Q(T) = \int dU_m \Omega(U_m) Q_{mm}(U_m, T)\,(1 + S(U_m, T))\,\mathrm{e}^{-\beta U_m} \tag{2.14}
$$

with

$$
S(U_m, T) = \left\langle \sum_{s\in m} \frac{Q_{ms}}{m(s)} \mathrm{e}^{-\beta \Delta U_s} \right\rangle / \, Q_m(U_m, T) \tag{2.15}
$$

where we assumed that $S(U_m, T) \ll 1$. f_u is now written in a form similar to $Q(T)$

$$
f_u(T) = \int dU_m \Omega(U_m) \mathrm{e}^{-\beta \Delta U_m} \bar{f}_u(U_m, T)/Q(T) \tag{2.16}
$$

with

$$\bar{f}_u(U_m, T) = \left\langle \sum_{s \in m} \frac{K_s}{3Nm(s)} Q_{ms} \mathrm{e}^{-\beta \Delta U_s} \right\rangle \tag{2.17}$$

where K_s is the saddle order. We now assume that K_s is self-averaging and so we extract a mean, effective saddle order K^e, it follows

$$\bar{f}_u(U_m, T) = \frac{K^e(U_m, T)}{3N} S(U_m, T) Q_{mm}(U_m, T). \tag{2.18}$$

In the thermodynamic limit the most probable minimum energy $U_m(T)$ dominates the U_m integrals and so eq. 2.14 becomes $Q(T) = Q_{mm}(U_m, T)$ for $S(U_m, T) \ll 1$. Also integration eq. 2.16 with $\bar{f}_u(U_m, T)$ from eq. 2.18 becomes

$$f_u(T) = \frac{K^e(T)}{3N} S(T) \tag{2.19}$$

dropping the $U_m(T)$ argument for functions of both $U_m(T)$ and T.

We now assume the system consists out of a number of local weakly interacting regions N_r with each $z \sim O(1)$ atoms. We further assume the local regions are moving independent and dominant saddles in the local regions are ordinary transition states, $K = 1$. Thus the mean effective saddle order becomes equal to the number of local regions, $K^e(T) = N_r$. Replacing $K^e(T)$ by $N_r = \frac{N}{z}$ in eq. 2.19 we have derived eq. 2.11. This result becomes plausible by noticing that f_u is proportional to the average number of regions above the inflection point in the approximation of the weakly interacting regions.

Next we will derive eq. 2.10. The average rate of escape from a basin is $R(T) = N_r R_{loc}(T)$ with $R_{loc}(T)$ computed for a single local region. Using the transition state theory the rate r of crossing a specific barrier can be computed as

$$r = \kappa(\omega_s) \left(\frac{\omega_m}{2\pi} \right) \mathrm{e}^{-\beta U_s} \tag{2.20}$$

where w_m and w_s are the normal mode frequencies computed at the minimum

and the saddle. This can be generalized for many dimensions by multiplying with the cross chapter for the saddle divided by the one for the minimum. Those cross chapters can be computed as the configurational integral divided by the length along the reaction coordinate. So r becomes

$$r = \frac{l_m}{l_s}\frac{Q_{ms}}{Q_{mm}}\kappa(\omega_s)\left(\frac{\omega_m}{2\pi}\right)\mathrm{e}^{-\beta U_s} \tag{2.21}$$

where l_m and l_s are those lengths. A harmonic approximation for l_m gives

$$l_m = (2\pi k_b T/m_e\omega_w^2)^{1/2} \tag{2.22}$$

with m_e being an effective mass. The full rate can now be computed by integrating over frequency

$$R(T) = N_r(2k_bT/\pi m_e)^{1/2}\int \mathrm{d}w\frac{\kappa(\omega)}{l_s(\omega)}S(\omega,T) \tag{2.23}$$

with

$$S(\omega,T) = 2\langle Q_{ms}\mathrm{e}^{-\beta\Delta U_s}\rangle(\omega,T)s(T)\langle n(\omega,T)\rangle/Q_m(T) \tag{2.24}$$

where $s(t)$ is the average number of transition states connected to one minimum and $n(\omega,T)$ is the total distribution of unstable frequencies of all the saddles connected to one minimum. Integrating $S(\omega)$ we see that

$$\int S(\omega,T)\mathrm{d}\omega = S(T). \tag{2.25}$$

By factoring out $S(T)$ from the integral above $R(T)$ becomes

$$R(T) = N_r(2k_bT/\pi m_e)^{1/2}\left[\int \mathrm{d}w\frac{\kappa(\omega)}{l_s(\omega)}S(\omega,T)/S(T)\right]S(T). \tag{2.26}$$

The ratio $S(\omega,T)/S(T)$ can also be rewritten as normalized distribution of the unstable frequencies [Keyes, 2006]. We now assume that the integral containing this ratio and $\kappa(\omega)/l_s(\omega)$ has only a weak T dependence. Thus we can write $R(T) \propto N_r S(T)$ which is eq 2.10.

Together with the derivation of f_u we thus derived that both the rate R and f_u are proportional to the saddle sum S. Using also the Markov approximation for the self-diffusion constant written as the eq. 2.9, we now derived the proportionality between the self-diffusion constant D and the fraction of negative modes f_u ($D(T) \propto f_u(T)$).

The derivation is limited to lower temperatures because both the Markov and the $S(U_m, T) \ll 1$ assumption are not valid for higher temperatures. This makes the derivation questionable above the dynamical transition temperature.

2.2.2 Derivation based on the Random Energy Model

The random energy model (REM) was originally developed for spin glasses [Derrida, 1980, 1981]. It was then adapted to biomolecules [Bryngelson and Wolynes, 1987]. Because of the large number of interactions and the central limit theorem, one expects a Gaussian distribution for the sum of all interaction energies in the case of a random amino acid sequence. In this model, a local region has a ground state with a fixed energy and a number of exited states. The interaction energy of neighboring regions is a constant in case both regions are in the ground state and a random value otherwise. Every REM state is identified with a Kth order stationary point with K neighbors with lower energy and $N_{cn} - K$ neighbors with higher energy, where N_{cn} is the number of connected neighbors. Thus f_u can be computed as

$$f_u(U) = p_c^<(U) = \int_{-\infty}^{U} \mathrm{d}U' G_C(U', U) \tag{2.27}$$

where $G_C(U', U)$ is the energy distribution of states connected to a state with energy U and $p_c^<(U)$ is the probability that a connected state has a lower energy.

The contributions to the rate R of transitions is computed independently for neighbors with lower and higher energy, each with a Metropolis ansatz. For the neighbors with lower energy there is no Boltzmann weight and thus $R^< = R_0 N_{cn} f_u$. For those with higher energy it can be shown that the

rate is also proportional to f_u [Keyes et al., 2002]. By again using the proportionality between D and R, also the REM model predicts $D \propto f_u$.

2.3 Filtering Methods

From the beginning of the INM method, it was noticed that some negative modes are not connected to diffusion and thus should be filtered out and the property linear to the self-diffusion constant is then the number of modes in this subgroup of diffusive modes. A first study showing that was on a Lennard-Jones liquid [Madan and Keyes, 1993]. In this case the filtering was done by subtracting from the number of negative modes for the liquid the the number of negative modes observed for the crystal. It was found out that this our cut-off based filtering methods were material specific and subsequently different filtering methods were developed to separate the diffusive modes out of all negative modes. With this filtering the linear relationship was found for many materials. The later methods are either based on minimization prior to computing the Hessian matrix or use the energy landscape view.

2.3.1 Double-Well Modes

For the landscape based methods double-well (DW) modes are found by calculating the energy profile along the negative modes [Bembenek and Laird, 1995]. While doing this the Hessian matrix is not recomputed, even though the direction of the mode associated to a fixed frequency changes quickly while moving away from the instantaneous configuration. This causes the energy to raise quickly along the mode [La Nave et al., 2001] because of Van-der-Waals repulsion. This energy difference can be seen in fig. 2.1. This error makes it impossible to find more distant saddles, and hence the number of DW modes are connected to the number of close-by saddles.

Unfortunately the proportionality of the number of DW modes compared to all negative modes is not better for all materials [Riberio and Madden, 1998]. Also for some DW modes the two one-dimensional configurations along the mode do not belong to different inherent structure (IS) basins [Gezelter

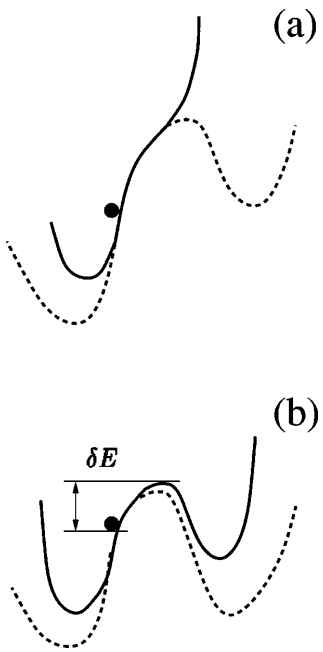

Figure 2.1: A double-well profile with the error caused by neglecting the change of direction of a mode (solid line) with respect to one with a fixed frequency and thus a changing direction (dotted line). [La Nave et al., 2001]

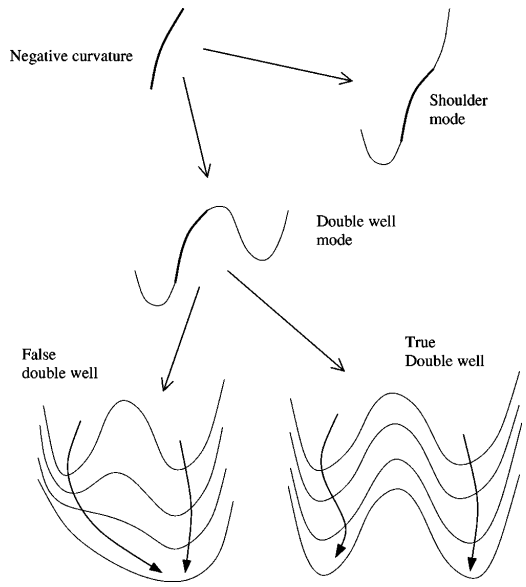

Figure 2.2: A DW profile and the escape filtering [La Nave et al., 2001]

et al., 1997] proving that the one-dimensional saddle found this way is not a saddle in all dimensions.

2.3.2 Escape Modes

The finding that DW modes not always minimize to different IS basins, lead to the definition of escape modes [Donati et al., 2000, La Nave et al., 2001]. For the escape filtering the two one-dimensional minima along the mode on both sides of the saddle are minimized. These minimizations, by definition of the IS basin, check whether the two one dimensional minima belong to the same IS basin. Those modes for which the minima don't belong to the same basin are escape modes or true DW modes and the others are false DW modes. This is illustrated in fig. 2.2. For water the number of escape modes represented the self-diffusion constant very well [Cho, 1994, La Nave et al., 2000].

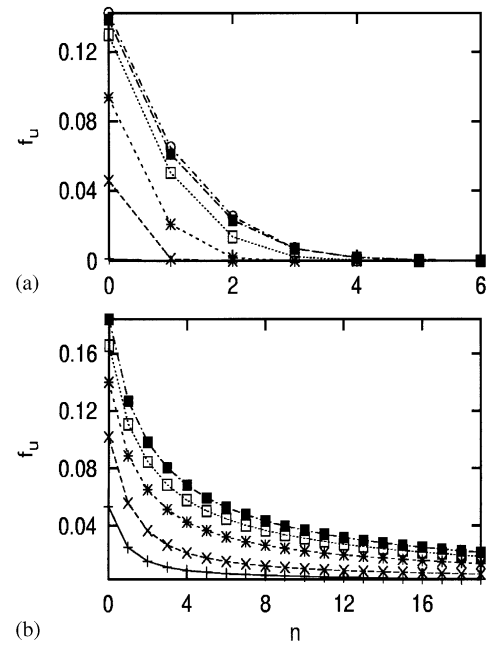

Figure 2.3: Fraction of unstable modes computed with partial minimization for Lennard-Jones liquid (b) and crystal (a) vs. minimization steps for different temperatures increasing from bottom to top [Chowdhary and Keyes, 2002].

2.3.3 Minimization based Filtering

Besides those two energy landscape based methods there are two minimization based methods. By minimizing the squared gradient $|\nabla U|^2$ the system is driven to the next stationary point and the order of the stationary point is then used as the number of diffusive modes.

The partial minimization method uses a small number of minimization steps compared to that required for a full minimization. For Lennard-Jones liquid and crystal, it was found that the negative non-diffusive modes vanished with the first minimization steps and only the diffusive modes persisted [Angelani et al., 2000]. This result is shown in fig. 2.3 and can be understood because the gradient along the diffusion path is small. The Hessian matrix is thus computed after a specific number of minimization steps and the fraction of negative modes of this matrix is used as fraction of diffusive modes.

For all those four filtering methods linear proportionality to the self-

diffusion constant was found at least for some materials. The computational methods used to test the proportionality for these methods for proteins will be discussed in the next chapter.

Chapter 3

Methods

3.1 Molecular Dynamics and Force Fields

The most frequently used methods for all-atomic simulations are Monte Carlo (MC) and molecular dynamics (MD). Both methods need a force field (FF) to compute the energy (and forces) of a certain configuration. In case one wants to simulate with all-atomic details, this FF specifies the interaction constants for all pairs of atoms. Both methods can also be applied to coarse grained simulations.

All-atomic MD was used for this project so I will only explain this method. For the simulation the time is discretized in time steps Δt ($\Delta t \approx 1\,\mathrm{fs}$). For each time step the force acting on each atom $F_i = -\frac{\partial V}{\partial r_i}$ is computed by summing up all FF terms. The so computed acceleration a_i ($F_i = ma_i$) is integrated numerically, most often using the formula [Verlet, 1967, 1968]

$$x(t_0 + \Delta t) = 2x(t_0) - x(t_0 - \Delta t) + a\Delta t^2 \tag{3.1}$$

A popular FF named CHARMM was used in its version 22 and 27 [Foloppe and MacKerell, 2000]. It uses two-, three- and four-body terms

between bonded atoms and the formula is

$$
\begin{aligned}
V &= \sum_{\text{bonds}} k_b \left(b - b_0\right)^2 + \sum_{\text{angles}} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} k_\phi \left[1 + \cos\left(n\phi - \delta\right)\right] \\
&+ \sum_{\text{impropers}} k_\omega \left(\omega - \omega_0\right)^2 + \sum_{\text{Urey-Bradley}} k_u \left(u - u_0\right)^2 \\
&+ \sum_{\text{nonbonded}} \epsilon \left[\left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon r_{ij}}
\end{aligned}
\tag{3.2}
$$

The two-body term depends on the distance b between the atoms and is harmonic. All harmonic terms have two constants: a spring constant and an equilibrium value. Of the two harmonic three-body terms one depends on the angle θ and one on the distance u. The latter is the Urey-Bradley term. The four-body term depends on the dihedral angle and can be either harmonic or periodic. The harmonic term is called improper. These terms are illustrated with animations at [Stote et al., 1999].

The last two terms are the nonbonded terms representing the Van-der-Waals and electrostatic interaction. Because the electrostatic interactions are shielded significantly under physiological conditions by the water, which is a polar solvent, an explicit or implicit solvent is required for biological simulations. In our case the solvent is explicitly modeled, meaning the simulation box is filled with simulated water molecules.

For larger molecules more advanced algorithms are needed for computing the long range interactions because directly computing all pairs has a $O(N^2)$ complexity. The easiest method, a cut-off radius, has been shown to introduce a large error for explicit solvent simulations. In our case the Particle-Mesh Ewald method [Darden et al., 1993] was used. It sums the short range interaction in real space and the long range interaction in Fourier space. Because each summation is converging fast in their respective space, the computation can be truncated with a good accuracy and bringing the computational complexity to $O(N \log N)$.

3.2 MD details for the sample systems

To apply the INM method and to compute the INM frequencies sample configurations for the two proteins, myoglobin and the scorpion toxin, are needed. These sample configurations were taken from molecular dynamics trajectories. Though it is not necessary to take these configurations from a MD trajectory, doing so helps in comparing the results to other methods (in this case the MSD). For both proteins the CHARMM program [Brooks et al., 1983] with the CHARMM all-atom force field, explicit solvent in a rectangular box, Particle Mesh Ewald (PME) electrostatics, NPT ensemble and the Verlet integrator with a step-size of 1fs was used.

The trajectory for myoglobin was computed by K. Moritsugu [Moritsugu and Smith, 2005]. He used the CHARMM version 30b2 with the CHARMM 22 force-field with a box size of (57.5 ± 1.5)x(53.5 ± 1.5)x$(42 \pm 1)\,\text{Å}^3$ with 3090 water molecules and a total of 11780 atoms. The PME had a real-space cut-off of $12\,\text{Å}$ and Gaussian width $\kappa = .34\,\text{Å}^{-1}$. Trajectories were computed for temperatures from 120K to 300K in steps of 10K. The trajectories were calculated by K. Moritsugu for 1ns for all temperatures and I. Daidone has computed up to a total of 5ns for T=190,200,..,250K.

MD simulations of a crystal with 4 scorpion toxin proteins were carried out by M. Krishnan (not published yet) with CHARMM 27 force-field for a range of temperatures from 20 to 300 K in steps of 20 K. Including the water the system has 5116 atoms and the box size is 43.8x39.6x27.7 Å^3. At each temperature, the system was equilibrated for 1 ns followed by a 1 ns production run during which atomic coordinates were stored at every 50 fs. 40 sample configurations were chosen for each temperature each separated by 5 ps from the end of the trajectory.

3.3 Computing Hessian Matrices for Large Systems with PME

Computing the Hessian Matrix for the whole system including the water with the correct electrostatic terms is computational intensive. For the program to be capable to compute the Hessian matrix analytically an analytical expression of the second derivative for all terms in the energy function is needed. A number of programs (GROMACS [Lindahl et al., 2001], NAMD [Phillips et al., 2005], LAMMPS [Plimpton, 1995], CHARMM [Brooks et al., 1983], VMD [Humphrey et al., 1996], BALL [Moll et al., 2006]) were analyzed on whether they can compute the matrix analytically with PME electrostatics and whether they can compute it in parallel on several processors. None is able to do either.

So we were forced to compute the matrix with a finite difference method and in serial on a single processor. For a system with N atoms, the finite difference method computes the first derivative of the energy function (the force) for a small displacement to both sides for each of the $3N$ directions. Thus $6N$ force calculations are needed. We used CHARMM because already the trajectories were done with it and no other program had a significant advantage. In case one could compute the second derivative analytically, either because it is added for PME or one is not using PME, these computations would be significantly faster.

After the Hessian matrix is computed it has to be diagonalized. Diagonalizing the whole matrix has a $O(N^3)$ complexity. Because only the negative modes are required using an iterative eigenvector algorithm the problem could be reduced by the factor $1 - f_u$, with the fraction of negative modes $f_u \ll 1$. CHARMM has one iterative method implemented, the Diagonalization In a Mixed Basis [Mouawad and Perahia, 1993], but this implementation can not use the finite differences, thus we had to diagonalize the whole matrix.

The non-iterative method was not only less efficient, but also required more memory. Even though it is possible to diagonalize in place with the Householder reduction followed by the QR algorithm [Press, 1992], which is

task	total time
Hessian matrices (6.5 hours each)	179 days
Energy profiles (0.5 minutes each)	150 days
Total: DW filtering (600 matrices, profiles)	313 days
Minimization (15 minutes each)	2 years
Total: escape filtering (DW, minimization)	3 years
Total: partial minimization (45 matrices)	12 days
Total: gradient minimization	123 days

Table 3.1: Computation time for scorpion toxin for current CPUs (Opteron, Athlon MP, Pentium 4) with 1.8-3GHz

also more efficient than the Jacobian transformation for larger matrices, the non-iterative diagonalization in CHARMM is not in place with one matrix in single precision and thus requires 14 bytes per matrix element in memory. The space of one matrix is $(3N)^2 * 14\,\text{byte}$ so that the scorpion toxin (5116 atoms) requires for both matrices 3.1GB and myoglobin (11780 atoms) requires even 14.1GB, limiting the number of available computer systems able to do these calculations. In total we computed about 660 Hessian matrices, and for each the process of computing it with finite differences and then diagonalizing it took about 6.5 hours for the scorpion toxin and 3.2 days for myoglobin. All compute times are summarized in table 3.1.

3.4 Filtering Modes

3.4.1 Landscape based filtering

The landscape based filtering methods were implemented as a CHARMM input script. CHARMM has the advantage that all required functions are implemented and can be executed with a single command. The used functions are: reading the CHARMM computed eigenvector files and coordinate files, evaluating the energy for a certain configuration and energy minimizing the system. The disadvantage of using the CHARMM syntax is that it is limited compared to a regular programming language. E.g. it has only the IF and GOTO commands as control structures. This makes the program less

readable compared to languages with loops [Dijkstra, 1968].

The script, which is printed in the appendix A, reads the selected sample configuration and the computed eigenvectors. It then evaluates the energy along one eigenvector until an energy threshold is reached and saves this energy profile to a text-file. Using the profile it is evident whether the profile along this mode has one or two minima. By applying this procedure to all negative modes of all sample configurations, we have computed the DW filtered average fraction of unstable modes. In total there were about 400,000 negative modes and each took about half a minute totaling to about 150 days.

To also identify escape modes the DW modes are additionally examined by energy minimization. As starting structures for these minimizations the prior identified 1-dimensional minima were used. As minimization method the adopted basis Newton-Raphson method (ABNR) was used with 1000 steps. This minimization method was found out to be the fastest. The number of steps we decided after some test minimizations were run. Those are described in the results section. After the minimization the distance between the two minimized configurations were computed and assumed to have converged to one minimum below a certain threshold. The method, the number of minimization steps and the distance cut-off were chosen after convergence tests. Since the minimizations are the most computing time consuming part, a more efficient minimization scheme would be most important for speeding up the escape filtering. In total about 80,000 minimizations were done with 15 minutes for each totaling to about 800 days.

A similar procedure as the escape filtering was also applied to all modes including the single wells. Since it is not possible to use the 1-dimensional minima in this case, the starting structures were chosen with an energy or a distance criterion equally to both sides of the sample configuration. The rest of the procedure was identical, so a very similar CHARMM script was used.

The three long compute task, Hessian matrices, energy profiles and minimizations had to be run in parallel, in order to finish them within the time of the thesis. Instead of parallelizing a single task because for the matrices this was not achievable within the available time and for the minimization it would have been inefficient, the numerous tasks were run independently

in parallel. Therefore a very small C program using MPI was written which executes the CHARMM script for a group of tasks once for every task in parallel on a cluster. The source is printed as appendix B.

3.4.2 Minimization based filtering

The minimization based methods were implemented with small CHARMM scripts. After reading the sample configuration, it was minimized and then the Hessian matrix was computed and diagonalized to compute the fraction of unstable modes. Because only very few minimization steps were needed for the partial minimization method, the minimization was much faster than the computation of the diagonalized Hessian matrix. The Hessian matrices had to be recomputed because the earlier computed ones for landscape based filtering were of course without the minimization before. This took again 6.5 hours per matrix.

The gradient minimization requires the second derivative matrix of the energy function because to minimize a function its derivative is needed and the derivative of the squared gradient as a function of the second derivative matrix. Here the same problem as with the iterative diagonalization arose. The Hessian matrix is not available in CHARMM with a PME term and the gradient minimization was not implemented with the finite difference method. Because the second derivative matrix is needed for every minimization step and the finite difference method is computational intensive it is not a very practical method. In order to compare also this filtering method, this option was added to CHARMM. For our system it took about 8 days for one configuration to be minimized by 60 steps. This would also be faster using an analytical expression of the second derivative for the PME term.

Chapter 4

Results

4.1 Model systems

We used two proteins for this thesis. Myoglobin is a cytoplasmic hemoprotein and was the first crystallographically resolved protein. It is shown in fig. 4.1. We used the protein database (PDB) [Berman et al., 2000, 2003] structure 1A6G [Vojtechovsky et al., 1999], which is myoglobin from the sperm whale. It consists of a single polypeptide chain of 154 amino acids and a heme group. It is named after the structurally and functionally similar and evolutionary related hemoglobin and its main location in the cardiac myocytes. It can bind O_2 and can therefore store and transport it.

Different from hemoglobin it does not transport the O_2 in the red blood cells but from the red blood cells to the mitochondria and it does not have a sigmoid-shaped but a logarithmic saturation function of the oxygen binding affinity as a function of the concentration [Ordway and Garry, 2004].

The secondary structure is composed of eight α-helices and the heme group is fixed by two histidine residues, His64 and His93. The central iron ion is bound to five nitrogen atoms, part of four pyroles and the His93. The sixth position binds the O_2 or other ligands as NO or CO. It is of great interest because it is the simplest protein capable of reversible oxygenation.

The second protein studied is a scorpion toxin, the Toxin II from the Scorpion Androctonus australis Hector, which has the protein database id 1ptx

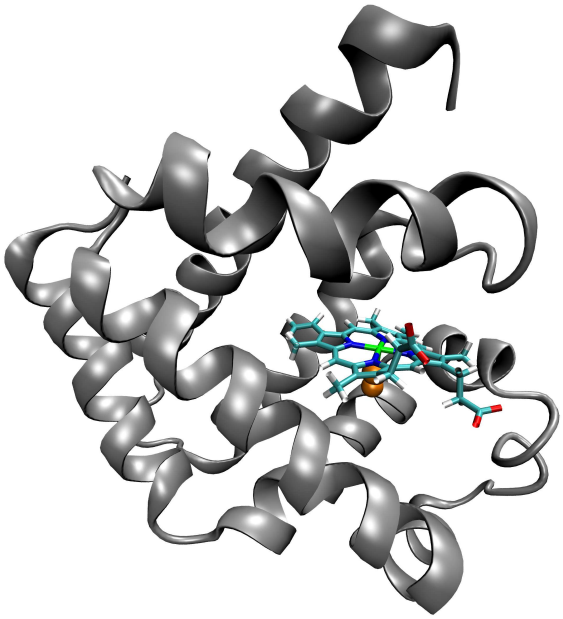

Figure 4.1: Sperm whale myoglobin protein. For the heme group the bonds are shown and the CO molecule is shown as orange Van-der-Waals spheres.

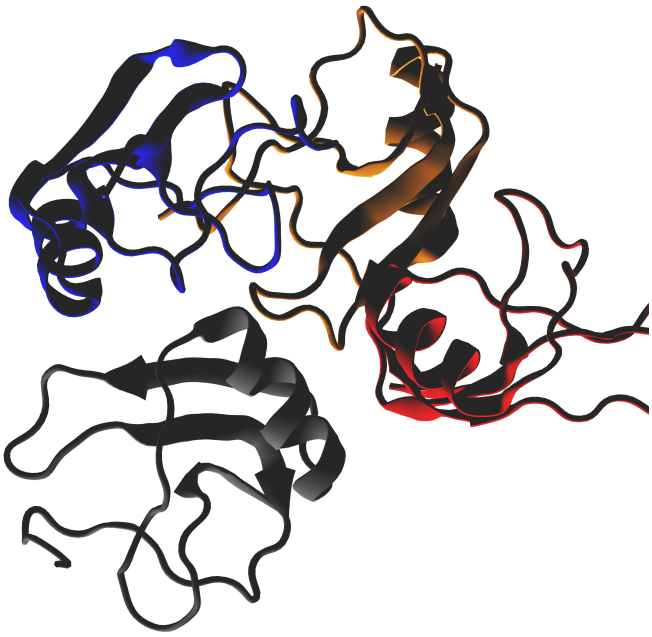

Figure 4.2: The crystalline unit cell of the scorpion toxin (1PTX)

[Housset et al., 1994]. The structure was resolved using X-ray diffraction and is shown in fig. 4.2. It is neurotoxic and is classified as a long α-toxin. It has 64 residues and so medium sized for a long toxin (60 to 70 residues). Scorpion toxins are classified depending on their binding to brain synaptosomes and their electrophysiological effect as either α or β type. These long toxins bind to NA^+ channels and can be specific to mammals or insects. The secondary structure is composed of one α-helix and one β-sheet. The Lys58 amino-acid is the most reactive.

4.2 Results for Myoglobin

As first step of the INM method the Hessian matrices for two sample configurations and four temperatures were computed. The sample configurations were taken from the MD trajectory and the system was unchanged, thus the same full hydration was included. Only very few matrices could be computed because of the computation time discussed in the last chapter. The so computed frequencies give the frequency spectrum of myoglobin, which are shown for two temperatures in fig. 4.3. For the configuration taken at 280K temperature, the spectrum is lower at around $500\,\mathrm{cm}^{-1}$ and higher for imaginary frequencies (the imaginary frequencies are always shown as negative in the graphs) compared to the configuration taken at 150K. The spectra for the not shown temperatures (190K, 220K) show the same trend. The two spectra for the two configurations taken at same temperature are very similar. Even though a temperature dependence of f_u is very likely from this result, a thorough analysis is not possible because of the limited number of spectra computable in the available time.

To test one possibility to speed up the computation of the Hessian matrix, the degrees of freedom of the solvent were decreased, to see whether one of these reduced systems would resemble the whole system. This could then speed up the computation significantly because the computation time for the diagonalization of the matrix is $O(N^3)$ with N being the degrees of freedom. The degrees of freedom were reduced by either fixing the water, using smaller amounts of solvation (as box or shell), or using no solvation. The

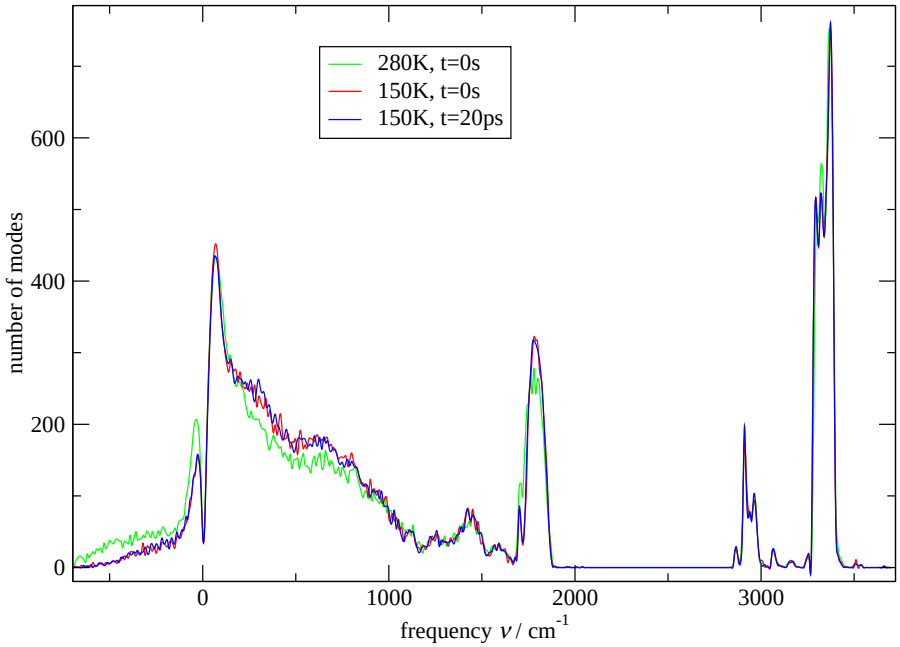

Figure 4.3: Comparison of frequency spectra for different temperatures and time steps

comparison is shown in fig. 4.4 and shows that only the shell of water gave a roughly similar spectrum to the one obtained with full hydration. Because it is difficult to estimate the effect of this difference and the computation time is still very large, we decided that a smaller model system should be used and chose the scorpion toxin for the rest of the simulations.

4.3 Results for the scorpion toxin

4.3.1 Comparison to MSD

As a benchmark for the INM method the self-diffusion constant was computed from the Mean Square Displacement (MSD) of a MD trajectory. The self-diffusion constant can be computed from the slope of MSD as a function of time. The MSD was computed after a translation and rotation with least-square-fit relative to the starting structure. In figure 4.5 the resulting self-diffusion constant is shown as a function of temperature for the scorpion

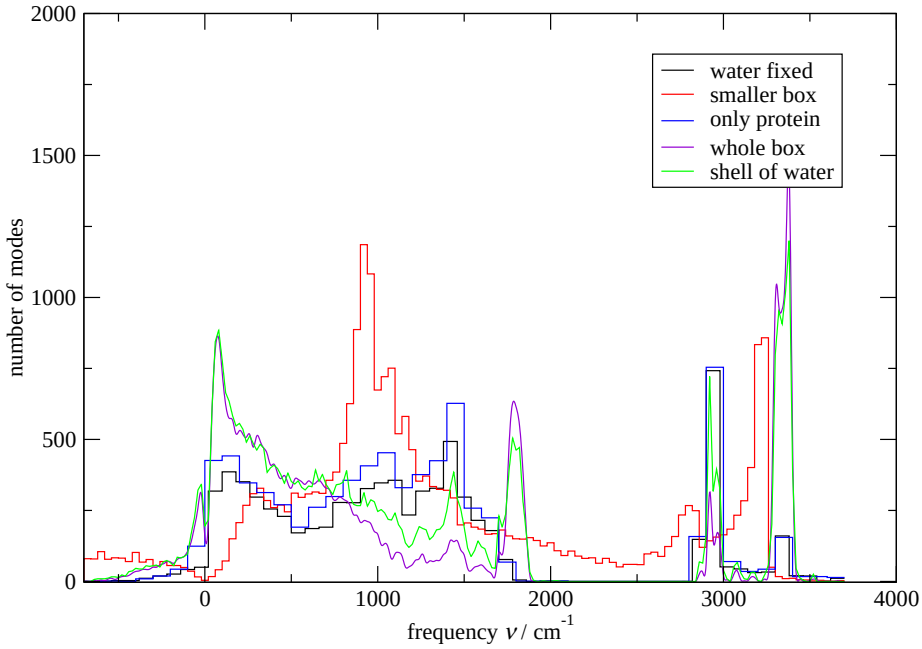

Figure 4.4: Comparison of frequency spectra for different solvation (bin size scaled for better comparison)

toxin, for both the whole system including the solvent and only the protein by itself. It shows clearly the dynamical transition at around 220K.

4.3.2 Results for different filtering methods

As the first necessary step for the INM method, the Hessian matrices were computed for 15 temperatures and 40 configurations each. The configurations were taken from the last 200ps of the MD production run in steps of 5 ps. The number of negative modes without filtering are shown in fig. 4.6. One can see that f_u is almost proportional to the temperature, but different to the result obtained from MSD, with a small decrease of slope for higher temperature and no slope change at the dynamical transition temperature.

For the landscape based filtering methods, the energy profiles were computed, by evaluating the potential energy in steps along the modes. A few sample DW profiles are shown in fig. 4.7. The step-size is automatically adjusted during the computation, so that the number of steps, and thus

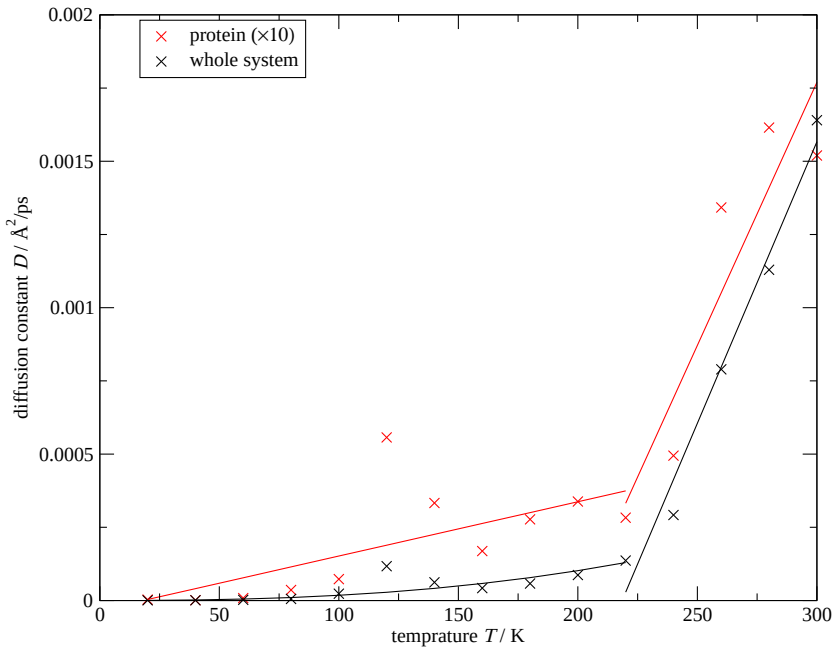

Figure 4.5: Diffusion constant computed from MSD for scorpion toxin alone and including hydration for temperatures around dynamical transition

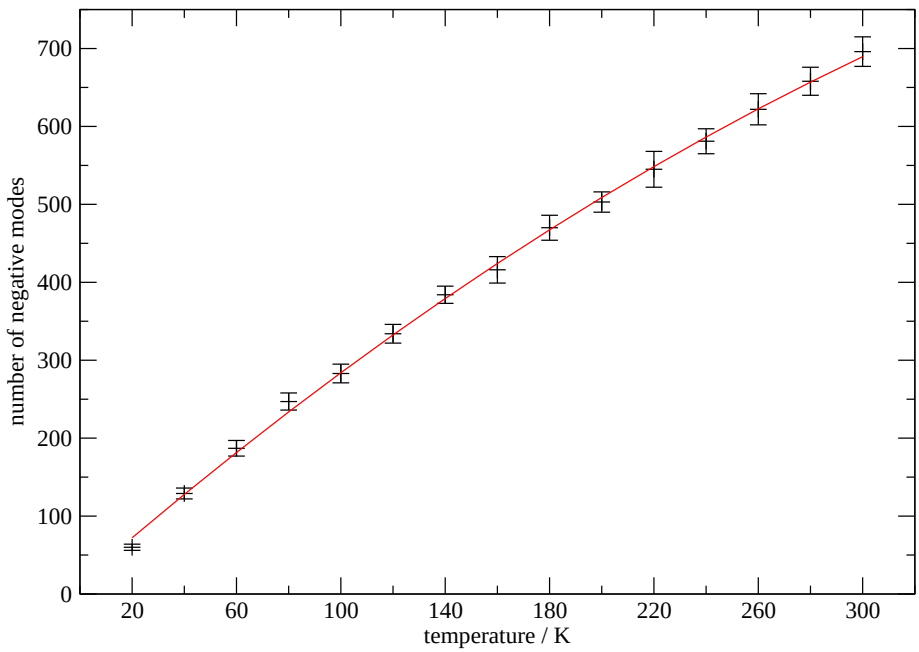

Figure 4.6: Number of negative modes for the scorpion toxin over temperature

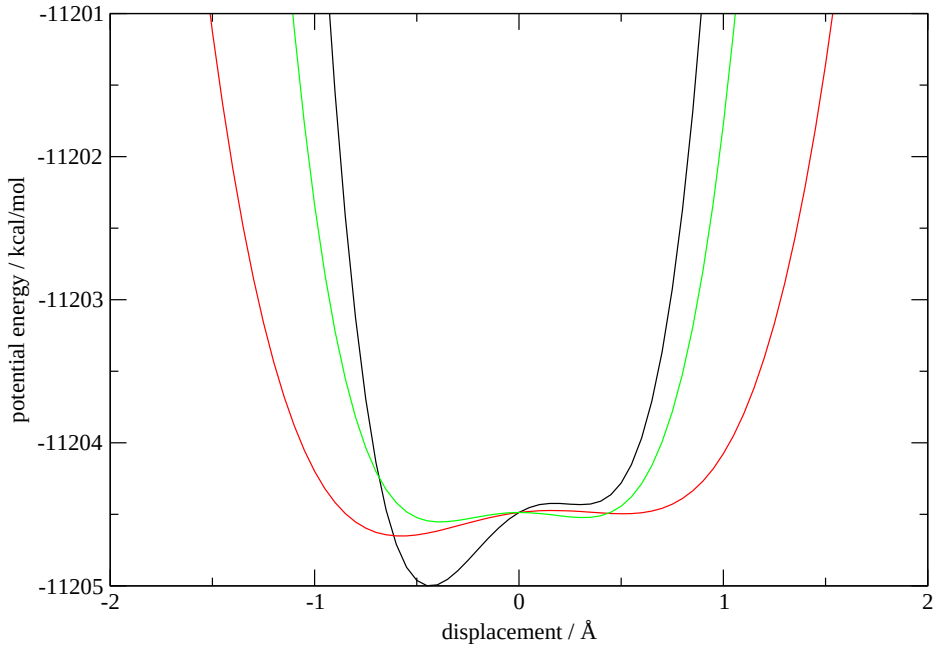

Figure 4.7: Sample DW profiles for a configuration computed at 300K

precision, is similar for all modes.

The contribution to the number of negative modes by the protein is computed using a participation ratio, which is defined for each negative mode via its eigenvector. The normalized eigenvector of the mode, in the configurational space of the system, is projected onto the protein subspace. The length of this projection is the participation ratio.

It was found, similar to the liquid studies that there are no double well modes with positive frequency for any of the analyzed configurations. This can be understood by the difference between the real energy profile and the computed one without recomputing the Hessian matrix, as described in ch. 2.3.1. The temperature dependence of the number of diffusive modes after DW filtering are shown in fig. 4.8 and the temperature dependence of the average negative frequency of DW modes $\langle \omega_u \rangle$ in fig. 4.9. One sees that there is no negative slope change as for the unfiltered case but the large slope change seen for the self-diffusion constant around the dynamical transition is not reproduced. There is only a very small slope change for those contributing to the motion of the water, but none for those for the protein.

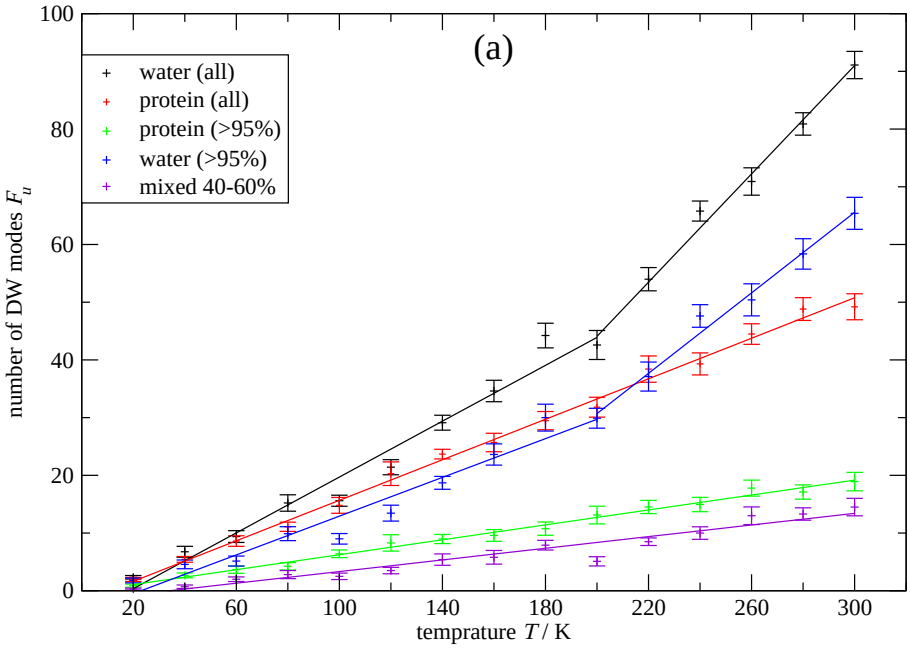

Figure 4.8: Number of DW modes as a function of temperature

The average frequency, which was suggested to be the reason for the small T dependence of the proportional coefficient between D and f_u [Keyes, 2006], is proportional to the temperature without a slope change.

As discussed in sec. 3.4.1, for the escape filtering the two one-dimensional minima are checked whether they belong to the same IS basin. This is done by minimization and for this a minimization scheme is needed. In order to make a decision on this scheme, configurations along some modes were minimized with different methods and number of steps. We concluded that the fastest method is the Adopted Basis Newton-Raphson (ABNR) and the results with this method for different number of steps for two configurations are shown in fig. 4.10.

Because of the large number of minimizations required for the escape filtering, a compromise between a well converged minimization and an affordable computation time is needed. According to the results shown in fig. 4.10, 1000 steps were considered to be a good compromise. Nonetheless, because the minimization is not fully converged after 1000 steps, a cut-off for the euclidean distance between the two minima is needed, to de-

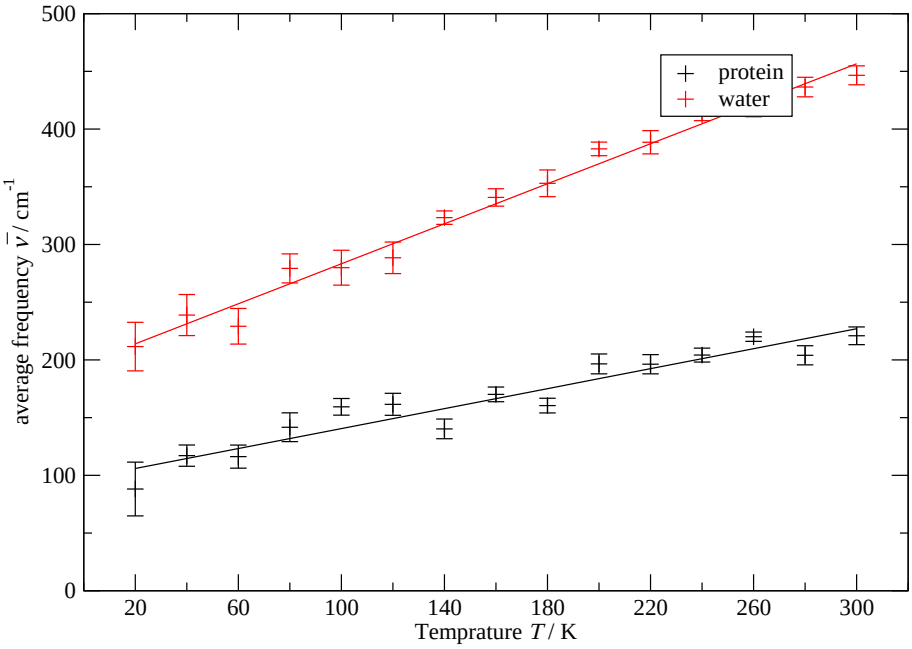

Figure 4.9: Average frequency $\langle \omega_u \rangle$ of DW modes as a function of temperature

cide whether these minimizations would converge to the same point after a full minimization. To estimate this cut-off, we used for one configuration a similar approach as for the escape filtering itself. Instead of taking the one-dimensional minima as starting structures for the minimization we took two configurations to both sides for each mode with a distance of 0.014 Å. This distance was chosen because it is the average of the distances at which the energy is $kT = 0.6\,\mathrm{kcal/mol}$ (with $T = 300\,\mathrm{K}$) higher than at the starting configuration. A histogram of the distances between these minima is shown in fig. 4.11. One can see that the first minimum is at 0.01 Å.

With this minimization scheme of 1000 ABNR-steps we analyzed all double wells. The resulting number of escape modes as a function of temperature computed with a cut-off of 0.01 Å are shown in fig. 4.12. The escape modes obtained with a cut-off of 0.01 Å and 0.02 Å are compared to the self-diffusion constant computed from the MSD in fig. 4.13 and 4.14. These results with escape filtering show a very good agreement with the self-diffusion constant computed from MSD compared to that obtained with DW filtering. In the double logarithmic plots a power fit is shown. One can see that the coefficient

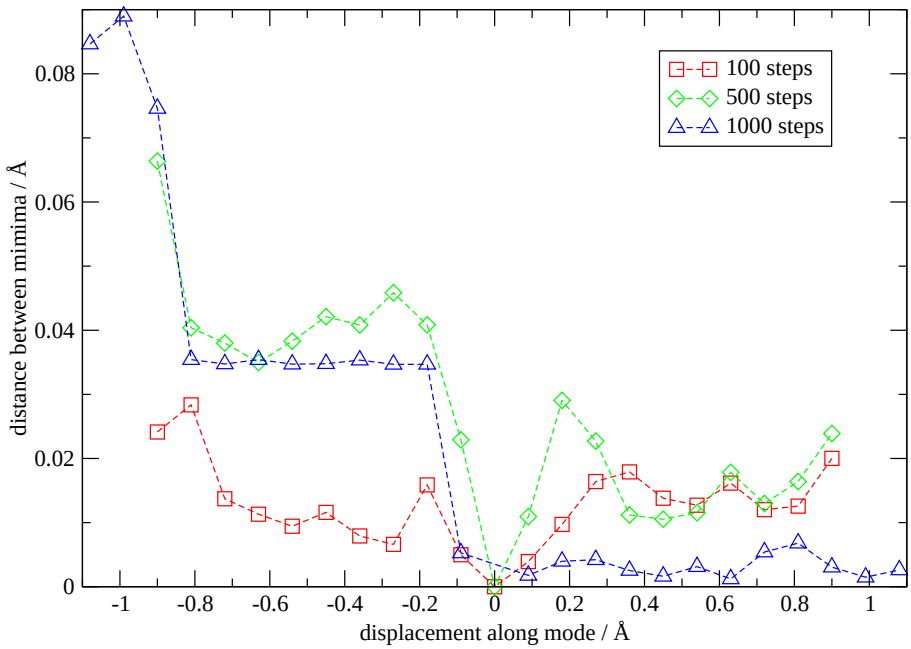

Figure 4.10: Distance between minima as a function of the displacement along a mode

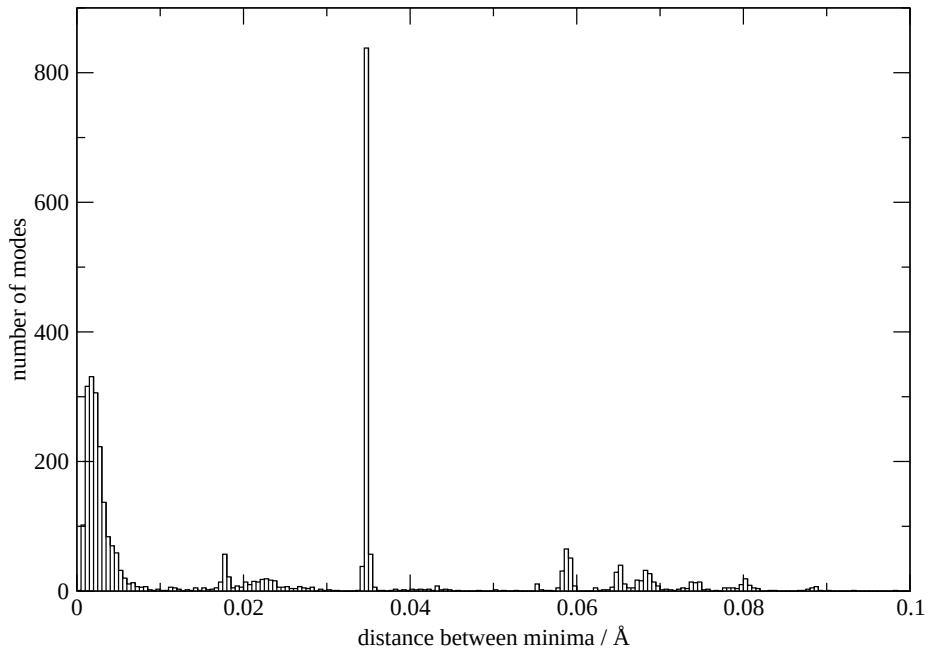

Figure 4.11: Histogram of distance between minima of configurations generated by displacing along mode

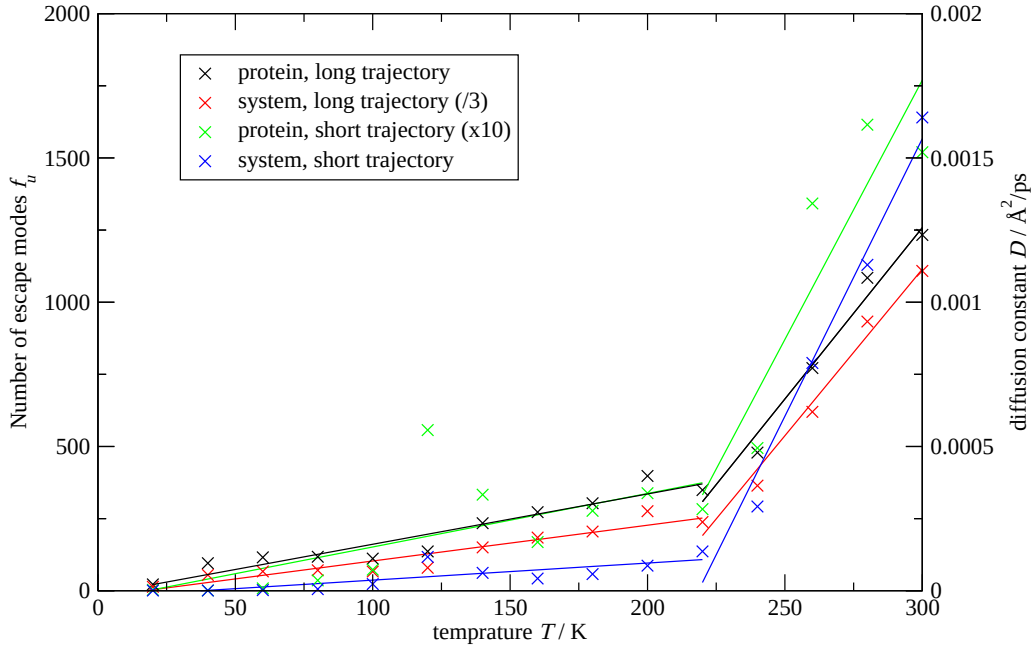

Figure 4.12: Diffusion constant computed from MSD and number of escape modes with a 0.01 Å cut-off for configurations from 800-1000ps as a function of temperature

depends on the choice of the cut-off, thus the uncertainty introduced by the cut-off prohibits a result for the coefficient.

In fig. 4.15 a comparison of the number of escape modes for 30 configurations taken from 800-950ps and for 10 configurations from 950-1000ps is shown. This is very similar for a 0.02 Å cut-off and thus not shown separately. For the shorter trajectory the fluctuation around the linear fit is larger showing that this number of configurations is not enough for a converged number of diffusive modes. The fluctuation is very similar for the protein and the whole system, suggesting a strong coupling between them.

Besides these fluctuations being very similar for the protein and the whole system, also the participation ratio itself suggests that there is a strong coupling in the motions described by the escape modes. The protein contribution represented by the participation ratio is shown in fig. 4.16, as the number of modes with a specific contribution in 1% steps. Even though the number of modes in a 1% bin are little above average near 0, 60 and 100%, the contri-

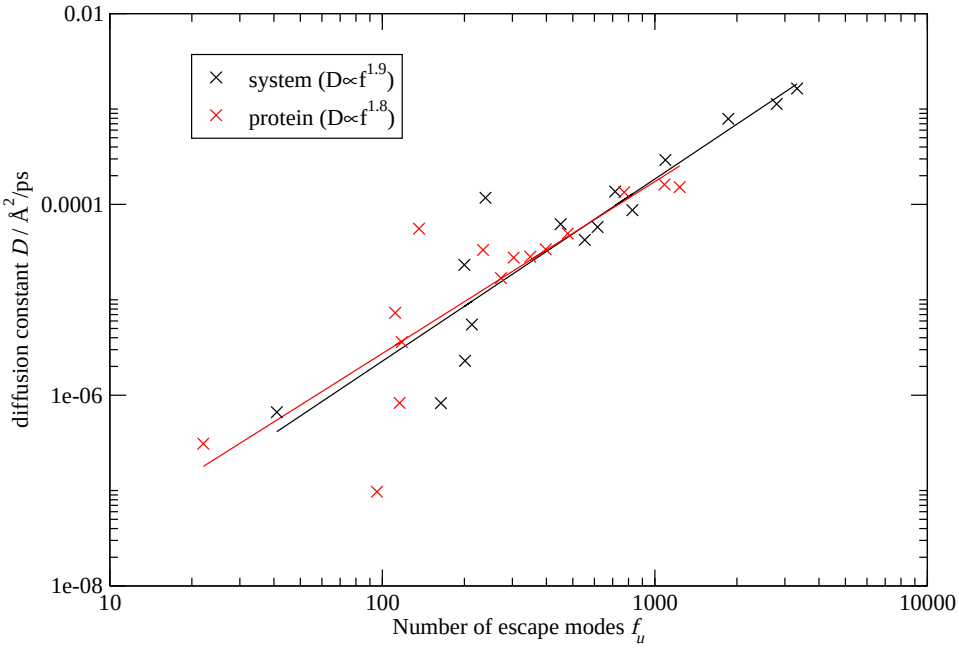

Figure 4.13: Diffusion constant computed from MSD as a function of the number of escape modes with a 0.01 Å cut-off for configurations from 800-1000ps

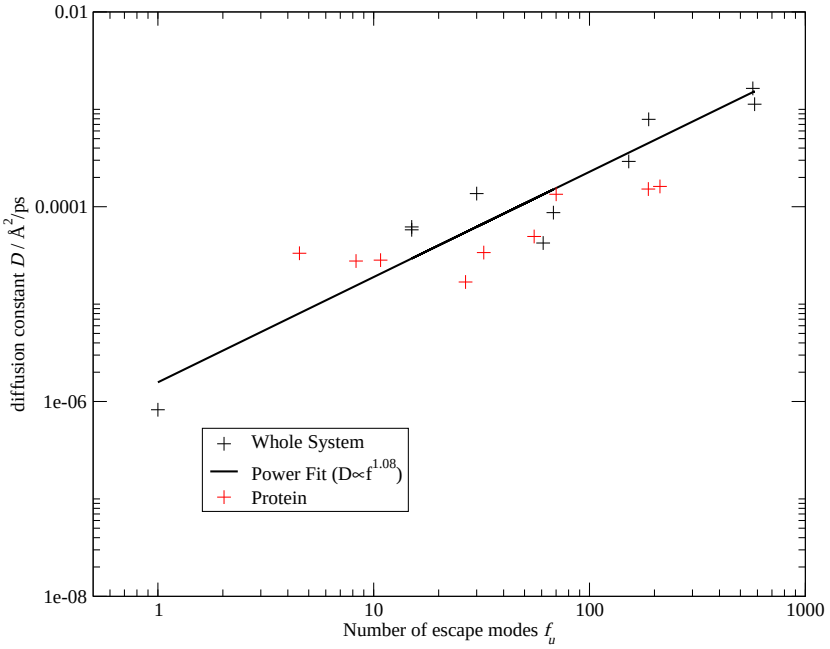

Figure 4.14: Diffusion constant computed from MSD as a function of the number of escape modes with a 0.02 Å cut-off for configurations from 950-1000ps

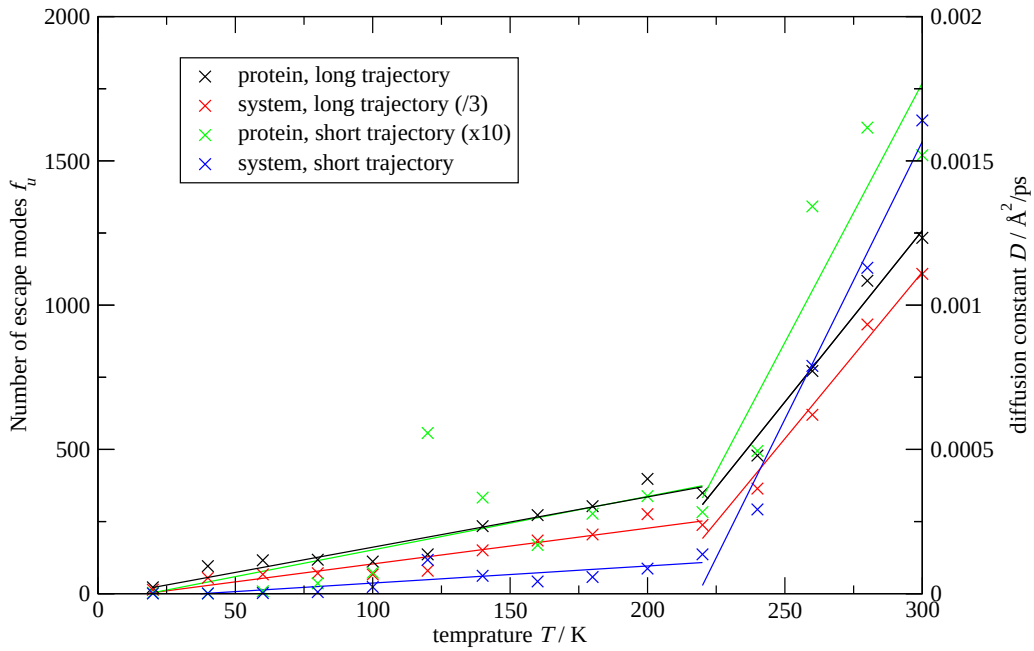

Figure 4.15: Number of escape modes with a 0.01 Å cut-off for configurations from 800-950ps and 950-1000ps as a function of temperature

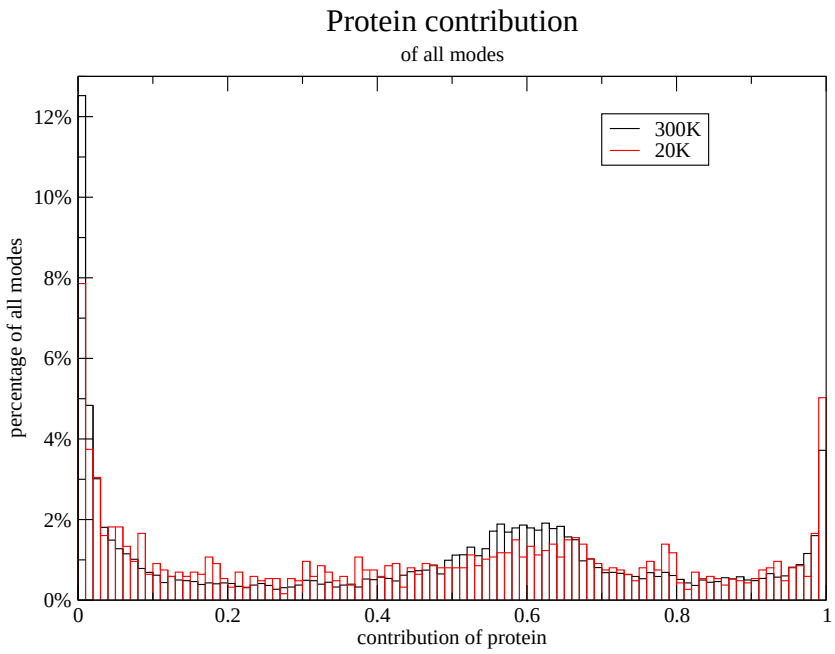

Figure 4.16: Contribution of the protein for the different escape modes

bution is wide-spread, thus most modes have some contribution from both the water and the protein.

The minimization based filtering methods were also tested. They are computed by first minimizing the potential energy or the squared gradient with CHARMM and then computing and diagonalizing the Hessian matrix after this minimization. In fig. 4.17 and 4.18 the results for the two methods are shown. As for the DW filtering, there is a slightly positive slope change in both graphs, but not a good agreement with the diffusion constant computed from the MSD is achieved.

4.3.3 Characteristics of escape filtering

Because the escape filtering gave the best results it was further analyzed. For one sample configuration at 300K we applied the escape filtering method to all modes up to $1000\,\mathrm{cm}^{-1}$. Instead of displacing along the mode to the one-dimensional minima, we displaced until the energy is 0.6kcal/mol larger than that of the starting configuration or by a fixed distance, which was

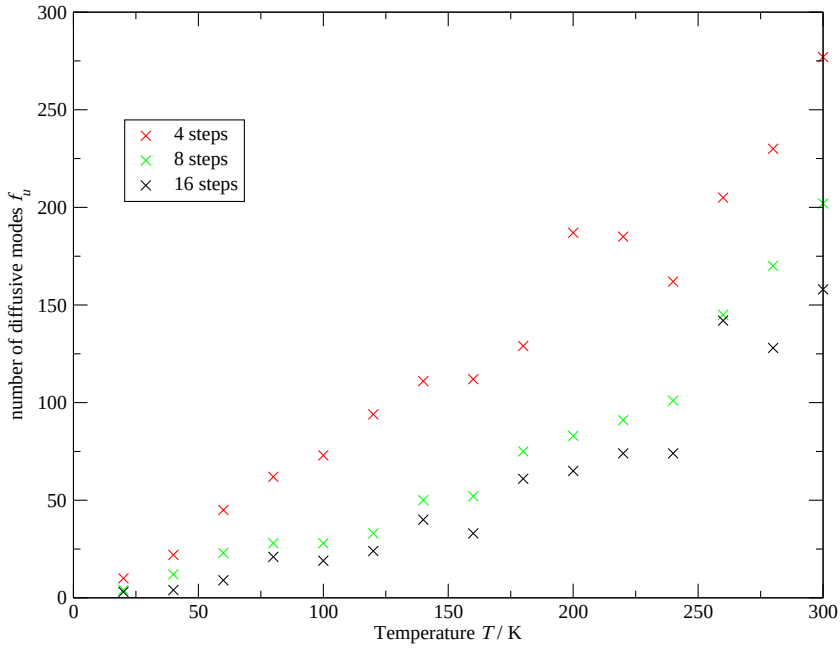

Figure 4.17: Number of diffusive modes for different number of partial minimization steps

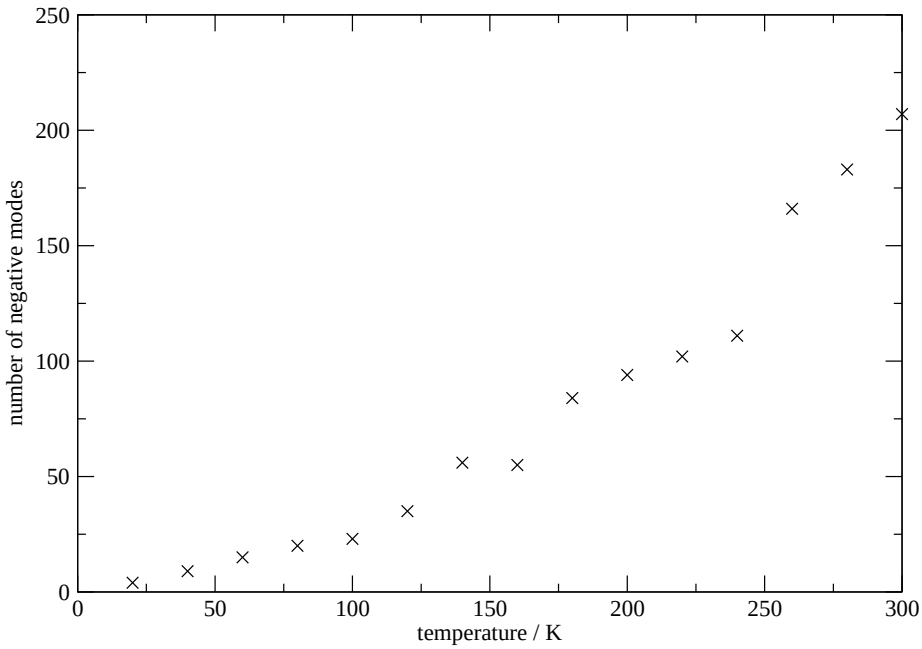

Figure 4.18: Number of diffusive modes after gradient minimization

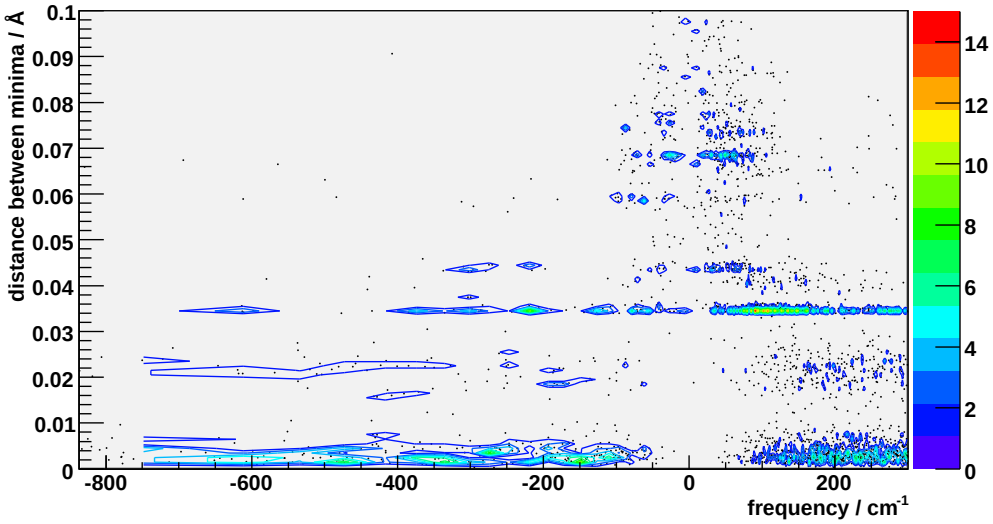

Figure 4.19: Number of modes as a function of frequency and distance between minima

explained in the last subsection because it was also used for the decision making regarding the cut-off. With both the distance and the energy criteria the single and double wells can be analyzed alike. The result of this is shown in fig. 4.11 and 4.19.

One sees that the histogram of the distances between the minima has large peaks for certain distances. By clustering the minima from those pairs of minima with these peak distances and measuring the potential energy barrier between members of one cluster by Conjugate Peak Refinement, it was found out that many pairs connect the same basins, thus displacement along some modes leads to the same basin. The large number of modes with distances between the minima above 0.05 Å, which can be seen in fig. 4.19, are an effect of the energy criterion. It is absent when displacing always by the same distance. The reason is that for the modes with frequencies around zero the curvature is very small and thus the displacement is very large for them with the energy criterion and thus increasing the chance of going to another basin.

The fraction of modes with a distance between the minima above the

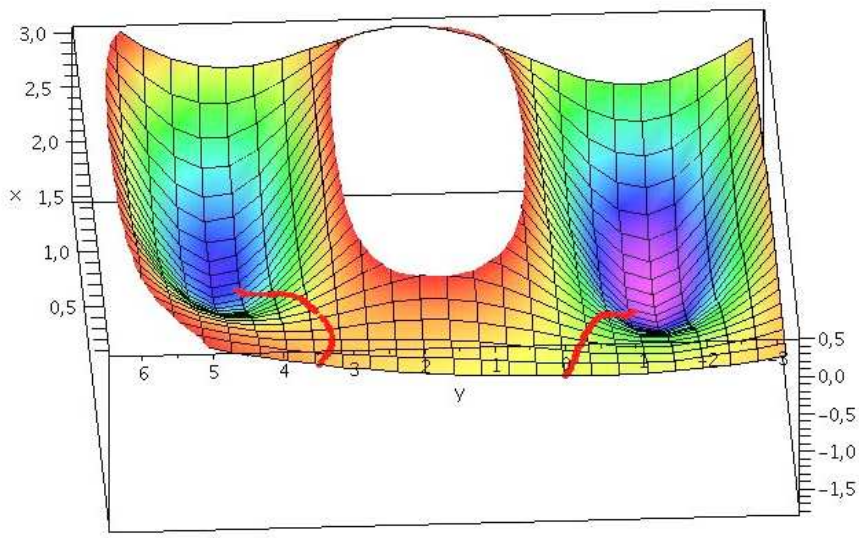

Figure 4.20: An example landscape showing an escape mode with positive curvature. Shown are the minimizations after displacing along the y axes.

cut-off is almost the same for all modes including the positive and negative ones. To illustrate why modes with positive curvature can be escape modes an example landscape is shown in Fig 4.20. Displacement from an instantaneous configuration at $x = 0.1$ and $y = \pi$ along y leads to configurations with only slightly higher energy and steepest-descent paths to two different minima.

This seems to question the approach of the escape filtering method to only analyze the DW modes. Because of the large number of minimization steps which would be required, a filtering method analyzing all modes by minimization can not be tested and compared to the escape filtering. The comparison to the other filtering methods showed that the escape filtering resulted in a much better agreement to the self-diffusion constant, thus the combination of the DW and the escape property has giving the good result. A possible explanation of this is that the system needs to be close to a saddle point, which is tested by the DW filtering as explained earlier, and the mode has to be a true DW and not a false one, which is tested by the escape filtering.

Further reducing the number of modes, which have to be tested by minimization, would be beneficial to computation time. Thus we looked for

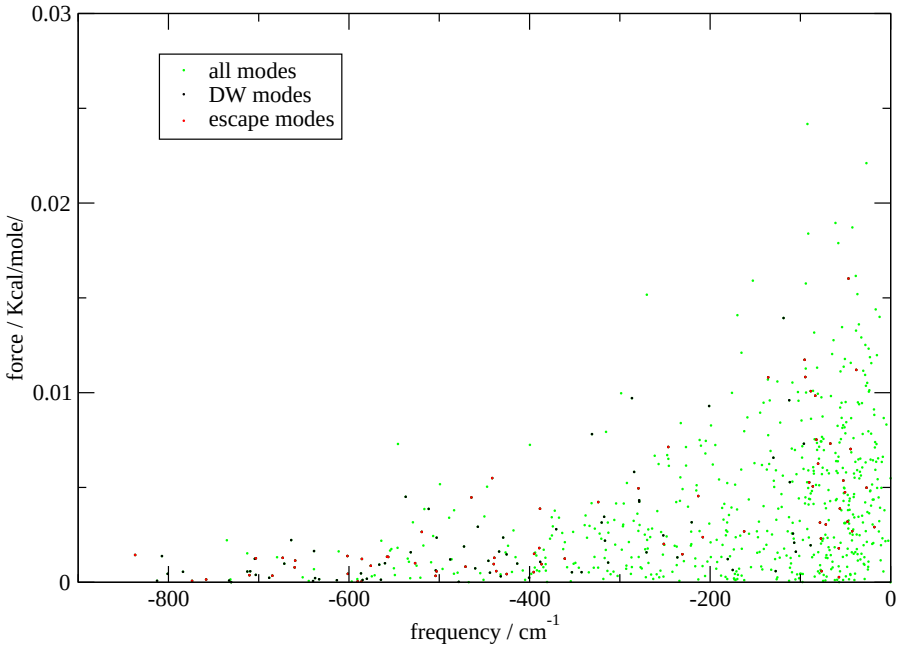

Figure 4.21: The force projected onto the mode and the frequency for all negative modes

properties all escape modes have but can be computed without minimization. In fig. 4.21 the modes for one configuration are plotted as a function of frequency and absolute force, and are colored by their type. This shows that one can not define a cut-off for either frequency or force to separate the escape modes.

Chapter 5

Conclusion

With the Instantaneous Normal Mode method the self-diffusion constant can be computed from the average number of diffusive modes of a few sample configurations. We applied this method to two sample proteins, myoglobin and a scorpion toxin. The computed myoglobin spectra have the temperature dependence predicted by INM. Spectra with reduced solvation were also computed. Out of those the one with a shell of water was very similar to the full solvation. For an accurate INM calculation a filtering method, computing the subset of the diffusive modes out of all negative modes is essential. Four such methods had been proposed and we tested them with the smaller system, the scorpion toxin. Out of those the escape mode filtering reproduced the self-diffusion constant including the slope change around the dynamical transition. However the total computation time required for this method was 3 CPU-years (Athlon-MP and Opteron CPUs). An even larger computational time required for a full minimization, only allowed a partial minimization and thus required the use of a cut-off for the distance between the minima. The exponent of the power fit of the fraction of unstable modes as a function of the self-diffusion depends on this cut-off and thus the expected linearity could not be checked. Therefore future work should focus on improving the minimization scheme, so that it is faster and the result is less dependent on the cut-off. Then the INM method with escape filtering could become useful for biological macromolecules and would allow to investigate e.g. the reasons for the strong coupling seen between the protein and the water motions.

Appendix A

CHARMM script for landscape based filtering

```
!program computes escape modes
!the energy profile is computed
!then the 1-dim minima are minimized and the distance
!between them is computed

!needed variables:
!CRDFILETYPE: sample configuration file type
!CRDFILE: sample configuration file
!INMFILE: hessian matrix
!N: largest mode which should be computed
!PROFILEFILE: file name to store profile
!STEP: beginning step size for profile


set I 25

!FORMat (1PG15.10)

!Loop over modes
LABEL LOOP0

    !read sample configuration
    open read @CRDFILETYPE unit 16 name @CRDFILE
    read coor @CRDFILETYPE unit 16
```

```
close unit 16

!read eigenvector
vibran nmode 1
open read unit 8 file name @INMFILE
read norm mode @N thru @N file unit 8

IF @I LT 20 THEN CALC STEP @STEP * .75
IF @I GT 30 THEN CALC STEP @STEP * 1.5

open write unit 9 card name @PROFILEFILE@N.dat
fill comp mode 1 fact @STEP  !this is off by 2%

end


gete print unit 9
set STARTE ?ENER
set J 0
set min0 0

!move along mode and compute energy and save profile
LABEL LOOP1
        INCR J
        set I 0
        set DOWN 0
        set LAST @STARTE
        set MIN -1

        FORMat (F16.5)
        !Exploring Profile
        LABEL LOOP2
                coor add
                gete print unit 9
                calc DIFF ?ENER - @LAST
                if @i eq 0 then -
                    if ?ENER GT @starte then -
                            incr min0
                IF @DIFF GT 0 THEN -
                    IF @MIN EQ 0 THEN -
                            SET MIN @I
```

```
                IF @DIFF LT 0 THEN -
                    IF @MIN EQ -1 SET MIN 0

                incr I by 1

                set LAST ?ENER
                calc DIFF ?ENER - @STARTE
                IF @DIFF LT @STOPE GOTO LOOP2
        FORMAT

        set MIN@J @MIN
        calc MI - @I
        coor scale fact @MI comp
        coor add

        calc II 1/ @I
        coor scale fact @II comp
        IF J LT 2 THEN GOTO LOOP1

if @min0 eq 2 then if @min1 eq -1 then set min1 0
if @min0 eq 2 then if @min2 eq -1 then set min2 0

!if not a DW, then skip minimization
if @MIN1 le 0 then GOTO NOMIN
if @MIN2 le 0 then GOTO NOMIN

coor scale fact @MIN1 comp
coor add

calc OFFS - ( @MIN1 + @MIN2 ) /  @MIN1
coor scale fact @OFFS comp
coor add comp

!distance between 1-dim minima
coor rms
set 1RMSCHARMM ?RMS

!write 1-dim minima coordinates
open write card unit 15 name @MINDIM1CRD1FILE@N.pdb
write coor comp pdb unit 15
close unit 15
```

```
    open write card unit 15 name @MINDIM1CRD2FILE@N.pdb
    write coor pdb unit 15
    close unit 15

    !minimization
    mini abnr nstep 1000
    coor swap
    mini abnr nstep 1000

    !distance between minima
    coor rms

    calc 1RMS ( @MIN1 + @MIN2 ) * @STEP / sqrt( ?nato )
    calc RMS1RMS ?RMS / @1RMS
    echo mode @N RMS: ?RMS , -
        RMS between 1-dim Minima: @1RMS,-
        MINS: @MIN1, @MIN2, RMS/1RMS: @RMS1RMS, -
        STEP: @STEP, 1RMS_C: @1RMSCHARMM

    open write card unit 15 name @MINCRD1FILE@N.pdb
    write coor comp pdb unit 15
    close unit 15

    open write card unit 15 name @MINCRD2FILE@N.pdb
    write coor pdb unit 15
    close unit 15

    GOTO MIN

    LABEL NOMIN
    echo mode @N only 1 minimum

    LABEL MIN
    decr N

!end condition for Loop over modes
!(has to be changed for going upwards)
    IF @N GT 0 GOTO LOOP0
```

Appendix B

MPI program to run CHARMM distributed on a cluster

```
#include <stdio.h>
#include <mpi.h>
#include <math.h>
#include <unistd.h>
#include <string.h>
#include <error.h>
#include <stdlib.h>
#include <pthread.h>

extern int errno;
int rank,size,task0;

#define debug printf

/*
The programm starts one central thread
distributing to the nodes, which request a
single new task each time they have finished one
*/

//Request next task number per MPI
int next_task(int *task) {
  int flag=0;
  MPI_Request request;
```

```
  MPI_Send(NULL,0,MPI_CHAR,0,0,MPI_COMM_WORLD);
  MPI_Irecv(task,1,MPI_INT,0,1,MPI_COMM_WORLD,
        &request);
  while(!flag) MPI_Test(&request,&flag,
        MPI_STATUS_IGNORE);
  return *task>=0;
}

//thread function answering the task requests
//sending a finsish tag as soon as no tasks are
//left
void P(int* range) {
  MPI_Request request;
  MPI_Status status;
  int i,msg,finished,flag;
  //0: sending task, 1: sending end-token
  for (finished=0;finished<2;finished++) {
    //0: loop over tasks, 1: loop over rank
    for (i=0;i<(!finished?*range:size);i++) {
      msg=!finished?i:-1;
      MPI_Irecv(NULL,0,MPI_CHAR,MPI_ANY_SOURCE,
        0,MPI_COMM_WORLD,&request);
      MPI_Test(&request,&flag,&status);
      while(!flag)
        sleep(1);MPI_Test(&request,&flag,&status);
      MPI_Send(&msg,1,MPI_INT,status.MPI_SOURCE,
        1,MPI_COMM_WORLD);
    }
  }
}

int main (argc, argv)
     int argc;
     char *argv[];
{
  /* starts MPI */
  MPI_Init (&argc, &argv);
  /* get current process id */
  MPI_Comm_rank (MPI_COMM_WORLD, &rank);
  /* get number of processes */
  MPI_Comm_size (MPI_COMM_WORLD, &size);
```

```
  if (rank==0 && argc<5) {
    printf("Usage: %s CHARMM INPUT OUTPUT START EXCLUSIVE-END\n",
        argv[0]);
    exit(1);
  }

  int start=atoi(argv[4]);
  int end=atoi(argv[5]);
  int range=end-start;


  pthread_t thread_p;
  if (rank==0) pthread_create(&thread_p, NULL,
        (void *) P, &range); //starting central thread
  int task;

  //work loop for each node
  while (next_task(&task)) {
    task+=start;
    char buf[1000];
    int i;
    sprintf(buf,"%s task=%d <%s > %s%d.out ",
        argv[1],task,argv[2],argv[3],task);
    for (i=6;i<argc;i++) {
      sprintf(buf,"%s %s",buf,argv[i]);
    }

    if (system(buf)==-1) //starting work command
      printf(strerror(errno));
  }
  //all tasks finished
  if (rank==0) pthread_join(thread_p, NULL);

  MPI_Finalize ();
  return 0;
}
```

Bibliography

L. Angelani, R. Di Leonardo, G. Ruocco, A. Scala, and F. Sciortino. Saddles in the Energy Landscape Probed by Supercooled Liquids. *Phys. Rev. Lett.*, 85:5356, 2000.

S. Bembenek and B. Laird. Instantaneous Normal Modes and the Glass Transition. *Phys. Rev. Lett.*, 74:936–939, 1995.

H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne. The Protein Data Bank. *Nucleic Acids Res.*, 28:235–242, 2000.

H. Berman, K. Henrick, and H. Nakamura. Announcing the worldwide Protein Data Bank. *Nature Structural Biology*, 10:980–980, 2003. URL `http://www.pdb.org/`.

D. Bicout and G. Zaccai. Protein Flexibility from the Dynamical Transition: A Force Constant Analysis. *Biophysical Journal*, 80:1115–1123, 2001.

A. Bizzarri, A. Paciaroni, and S. Cannistraro. Glasslike dynamical behavior of the plastocyanin hydration water. *Physical Review E*, 62:3991–3999, 2000.

B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4:187–217, 1983. doi: 10.1002/jcc.540040211. URL `http://www3.interscience.wiley.com/cgi-bin/abstract/109583611/ABSTRACT`.

J. Bryngelson and P. Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA*, 84:7524–28, 1987.

L. Bu and J. Straub. Vibrational Frequency Shifts and Relaxation Rates for a Selected Vibrational Mode in Cytochrome c. *Biophysical Journal*, 85:1429–1439, 2003.

M. Cho. Instantaneous normal mode analysis of liquid water. *J. Chem. Phys.*, 100 (9):6672–6683, 1994.

J. Chowdhary and T. Keyes. Finding diffusive directions in supercooled liquids by partial minimization of the potential energy. *Physica A*, 314:575–582, 2002.

R. Daniel, J. Smith, M. Ferrand, S. Hery, R. Dunn, and J. Finney. Enzyme Activity below the Dynamical Transition at 220K. *Biophys. J.*, 75:2504–2507, 1998.

R. Daniel, R. Dunn, J. Finney, and J. Smith. The Role Of Dynamics In Enzyme Activity. *Annual Review of Biophysics and Biomolecular Structure*, 32:69–92, 2003.

T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

B. Derrida. Random-Energy Model: Limit of a Family of Disordered Models. *Phys. Rev. Lett.*, 45(2):79–82, 1980.

B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B*, 24(5):2613–2626, 1981.

E. Dijkstra. Go to statement considered harmful. *Communications of the ACM*, 11:147, 1968.

C. Donati, F. Sciortino, and P. Tartaglia. Role of the unstable directions in the equilibrium and aging dynamics of supercooled liquids. *Phys. Rev. Lett.*, 85: 1464, 2000.

W. Doster, S. Cusack, and W. Petry. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature*, 337:754–756, 1989.

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

M. Ferrand, A. Dianoux, W. Petry, and G. Zaccai. Thermal Motions and Function of Bacteriorhodopsin in Purple Membranes: Effects of Temperature and Hydration Studied by Neutron Scattering. *PNAS*, 90:9668–9672, 1993.

B. Fitch, A. Rayshubskiy, M. Eleftheriou, T. Ward, M. Giampapa, M. Pitman, and R. Germain. Molecular dynamics—blue matter: approaching the limits of concurrency for classical molecular dynamics. In *SC '06: Proceedings of the 2006 ACM/IEEE conference on Supercomputing*, 2006. ISBN 0-7695-2700-0.

N. Foloppe and A. MacKerell, Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.*, 21:86–104, 2000.

H. Frauenfelder, F. Parak, and R. Young. Conformational Substates in Proteins. *Annual Review of Biophysics and Biomolecular Structure*, 17:451–479, 1988.

J. Gezelter, E. Rabani, and B. Berne. Can imiginary instanteneous normal mode frequencies predict barriers to self-diffusion? *J. Chem. Phys.*, 107:4618, 1997.

J. Hayward and J. Smith. Temperature Dependence of Protein Dynamics: Computer Simulation Analysis of Neutron Scattering Properties. *Biophys. J.*, 82: 1216–1225, 2002.

D. Housset, C. Habersetzer-Rochat, J. Astier, and J. Fontecilla-Camps. Crystal structure of toxin II from the scorpion Androctonus australis Hector refined at 1.3 A resolution. *J. Mol. Biol.*, 238:88–103, 1994.

W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

T. Keyes. Entropy, dynamics, and instantaneous normal modes in a random energy model. *Phys. Rev. E*, 62:7905–7908, 2000.

T. Keyes. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, chapter Imaginary-frequency, Unstable Modes, and Relaxation Dynamics in Liquids. Chapman and Hall/CRC, 2006.

T. Keyes. Unstable modes in supercooled and normal liquids: Density of states, energy barriers, and self-diffusion. *J. Chem. Phys.*, 101(6):5081, 1994.

T. Keyes. Normal mode theory of diffusion in liquids for a broad temperature range. *J. Chem. Phys.*, 103:9810–9812, 1995.

T. Keyes. Instantaneous normal mode approach to liquid state dynamics. *J. Phys. Chem. A*, 101:2921, 1997.

T. Keyes, G. V. Vijayadamodar, and U. Zurcher. An instantaneous normal mode description of relaxation in supercooled liquids. *J. Chem. Phys.*, 106:4651, 1997.

T. Keyes, J. Chowdhary, and J. Kim. Random energy model for dynamics in supercooled liquids: N dependence. *Physical Review E*, 66:51110, 2002.

V. Kurkal, R. Daniel, J. Finney, M. Tehei, R. Dunn, and J. Smith. Enzyme Activity and Flexibility at Very Low Hydration. *Biophysical Journal*, 89:1282–1287, 2005.

E. La Nave, A. Scala, F. W. Starr, F. Sciortino, and H. E. Stanley. Instantaneous Normal Mode Analysis of Supercooled Water. *Phys. Rev. Lett.*, 84:4605, 2000.

E. La Nave, A. Scala, F. W. Starr, H. E. Stanley, and F. Sciortino. Dynamics of supercooled water in configuration space. *Phys. Rev. E*, 64:036102, 2001.

E. La Nave, H. E. Stanley, and F. Sciortino. Configuration Space Connectivity across the Fragile-to-Strong Transition in Silica. *Phys. Rev. Lett.*, 88:035501, 2002.

A. Lesk. *Introduction to protein science.* Oxford University Press New York, 2004.

W. Li and T. Keyes. Pure translation instantaneous normal modes: Imaginary frequency contributions vanish at the glass transition in CS_2. *J. Chem. Phys.*, 107:7275–7277, 1997.

W. Li and T. Keyes. Instantaneous normal mode theory of diffusion and the potential energy landscape: Application to supercooled liquid. *J. Chem. Phys.*, 111(12):5503–5513, 1999.

W.-X. Li, T. Keyes, and F. Sciortino. Three-flavor instantaneous normal mode formalism: Diffusion, harmonicity, and the potential energy landscape of liquid CS_2. *J. Chem. Phys.*, 108:252–260, 1998.

E. Lindahl, B. Hess, and D. van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7:306–317, 2001.

B. Madan and T. Keyes. Unstable modes in liquids: density of states, potential energy, and heat capacity. *J. Chem. Phys.*, 98:3342, 1993.

A. Moll, A. Hildebrandt, H. Lenhof, and O. Kohlbacher. BALLView: a tool for research and education in molecular modeling. *Bioinformatics*, 22:365, 2006.

P. Moore and T. Keyes. Normal mode analysis of liquid CS_2: Velocity correlation functions and self-diffusion constants. *J. Chem. Phys.*, 100(9):6709–6717, 1994.

K. Moritsugu and J. Smith. Langevin model of the temperature and hydration dependence of protein vibrational dynamics. *Journal of physical chemistry. B, Condensed matter, materials, surfaces, interfaces, & biophysical chemistry*, 109: 12182–12194, 2005.

L. Mouawad and D. Perahia. Diagonalization in a mixed basis: A method to compute low-frequency normal modes for large macromolecules. *Biopolymers*, 33:599–611, 1993.

G. Ordway and D. Garry. Myoglobin: an essential hemoprotein in striated muscle. *J Exp Biol*, 207:3441–3446, 2004.

F. Parak, E. Frolov, A. Kononenko, R. Mossbauer, V. Goldanskii, and A. Rubin. Evidence for a correlation between the photoinduced electron transfer and dynamic properties of the chromatophore membranes from Rhodospirillum rubrum. *FEBS Lett*, 117:368–72, 1980.

F. Parak, E. Frolov, R. Mossbauer, and V. Goldanskii. Dynamics of metmyoglobin crystals investigated by nuclear gamma resonance absorption. *J Mol Biol*, 145: 825–33, 1981.

J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem*, 26:1781–1802, 2005.

S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117:1–19, 1995.

W. Press. *Numerical recipes in C: the art of scientific computing.* Cambridge University Press, 1992.

V. Reat, R. Dunn, M. Ferrand, J. Finney, R. Daniel, and J. Smith. Solvent dependence of dynamic transitions in protein solutions. *PNAS*, 97:9961–9966, 2000.

M. Riberio and P. Madden. Unstable modes in ionic melts. *J. Chem. Phys.*, 108: 3256, 1998.

F. Sciortino and P. Tartaglia. Harmonic Dynamics in Supercooled Liquids: The Case of Water. *Phys. Rev. Lett.*, 78:2385–2388, 1997.

G. Seeley and T. Keyes. Normal-mode analysis of liquid-state dynamics. *J. Chem. Phys.*, 91:5581–5586, 1989.

J. Smith, K. Kuczera, and M. Karplus. Temperature-dependence of myoglobin dynamics: neutron spectra calculated from molecular dynamics simulations of myoglobin. *Proc. Natl. Acad. Sci. USA*, 87:1601–1605, 1990.

J. Smith, T. Becker, S. Fischer, F. Noe, A. Tournier, G. Ullmann, and V. Kurkal. *Soft Condensed Matter Physics in Molecular and Cell Biology*, chapter Physical and functional aspects of protein dynamics, page 225. CRC Press, 2006.

F. Stillinger and T. Weber. Hidden structure in liquids. *Phys. Rev. A*, 25(2): 978–989, 1982.

F. Stillinger and T. Weber. Packing Structures and Transitions in Liquids and Solids. *Science*, 225(4666):983, 1984.

R. Stote, A. Dejaegere, D. Kuznetsov, and L. Falquet. `http://www.ch.embnet.org/MD_tutorial/pages/MD.Part2.html`, 1999.

J. Straub and J. Choi. Extracting the Energy Barrier Distribution of a Disordered System from the Instantaneous Normal Mode Density of States: Applications to Peptides and Proteins. *The Journal of Physical Chemistry*, 98:10978–10987, 1994.

J. Straub and D. Thirumalai. Exploring the Energy Landscape in Proteins. *PNAS*, 90:809–813, 1993.

R. Tilton Jr, J. Dewan, and G. Petsko. Effects of temperature on protein structure and dynamics: x-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320K. *Biochemistry*, 31:2469–2481, 1992.

L. Verlet. Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev*, 159(1):98–103, 1967.

L. Verlet. Computer "Experiments" on Classical Fluids. II. Equilibrium Correlation Functions. *Phys. Rev.*, 165:201–214, 1968.

J. Vojtechovsky, K. Chu, J. Berendzen, R. Sweet, and I. Schlichting. Crystal Structures of Myoglobin-Ligand Complexes at Near-Atomic Resolution. *Biophysical Journal*, 77:2153–2174, 1999.

D. Wales. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses.* Cambridge University Press, 2003.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den ....

(Roland Schulz)