Conformational Changes in Biological Molecules: New Computational Methods, and Comparison with Experiment

Thesis submitted for the degree of “Doctor of Philosophy”

By

Elad Segev

Submitted to the Senate of the Hebrew University

November 2007
This work was carried out under the supervision of Professor Robert Benny Gerber
Acknowledgements

Prof. Robert Benny Gerber,

I'd like to convey my appreciation and thanks for returning me my former love and enjoyment to science studies. You taught me scientific thinking. Your endless support and trust in me along the way enabled me to discover and find my own personal route way.

Yael my dear wife,

Thank you for your long unwearied await to your studying husband. Your endless support and understanding in the long evening and weekends when I was spending my time studying, sitting in front of the computer instead of being with my family.

My dear parents,

For always being there for me, helping me and my family at any time.

Idan and Tal my sons,

For respecting my work. For being patient enough and growing up to be independent, thus enabling me working peacefully.
1 Abstract

The study of biological molecules in the gas phase encompassing mass spectrometric, spectroscopic and theoretical approaches is a novel and thriving field of research. The determination of structure and structural changes of the biomolecules is a central challenge.

A study of structural changes in biological molecules in the gas phase is the focus of this thesis. Two newly-developed computational approaches for simulating short and long timescale dynamics of biological molecules are presented: the Gradual Molecular Dynamics simulation, which was applied to the study of structural changes in ubiquitin +13 ions in mass spectrometric experiments, and the Hybrid Molecular Dynamics / RRK algorithm, which was applied to the study of conformational changes in small biological molecules.

The Gradual Molecular Dynamics simulation approach was developed to overcome computational timescale limitation in describing the structural changes in protein mass spectrometric experiments. The approach involves a sequence of Molecular Dynamics simulations at gradually increasing temperatures: this leads to identification of major intermediate states, and to determination of unfolding pathways. The unfolding rate at any temperature can then be calculated by the RRK approach. Evidence from cross-section and electron capture dissociation data indicates that ubiquitin +13 ions lose their secondary and tertiary structure in gas-phase mass spectrometric experiments. For ubiquitin +13, three interesting intermediate states were found and the final near-linear
geometry was computed. The several energy barriers calculated for the process are in the range of $7 \rightarrow 15$ kcal/mol, and the unfolding timescale at 300K was computed to be 2 milliseconds. A simple model for estimating the ECD fragmentation pattern was applied and found to give a good agreement with experimental data which support the result for the final unfolding structure of the ion. The final structure of the simulation is also in accordance with experimental cross-section data. Moreover, cross-section calculations of the different intermediate structures of +13 ions indicate a possibility that +12 ions adopt these structures as stable conformation in the gas phase.

These results show that the combination of the new method with cross section measurements provides a powerful framework for studying protein unfolding in mass spectrometric experiments. It seems reasonable to expect that structures and structural changes of other proteins in the gas phase can be explored by this approach.

The *Hybrid* Molecular Dynamics / RRK algorithm describes the dynamics of long timescale evolution of conformational changes in small biological molecules. The approach employs classical trajectories for transitions between adjacent structures separated by a low barrier, and the classical statistical RRK approximation when the barrier involved is high. For determination of the long-time dynamics from an initial structure to a final structure of interest, an algorithm is introduced to find the most efficient pathways (sequence of the intermediate conformers). This method uses the Dijkstra algorithm for finding optimal paths on networks. Three applications of the method using an AMBER force field are presented: a detailed study of conformational
transitions in a blocked valine dipeptide; a multiple-reaction path study of the blocked valine tripeptide; and the evolution in time from the $\beta$ hairpin to $\alpha$ helix structure of a blocked alanine hexapeptide. Advantages and limitations of the method are discussed in the light of the results.

The results are very encouraging at to the use of the method to explore the evolution of structure in time for small biological molecules. Extending of the method for larger systems is an attractive challenge for the further.
## 2 Table of contents

Acknowledgements.......................................................................................................................... 3
1 Abstract........................................................................................................................................... 4
2 Table of contents.............................................................................................................................. 7
List of acronym:.................................................................................................................................. 8
3 Background ..................................................................................................................................... 10
   3.1 Introduction................................................................................................................................ 10
   3.2 Why Study Protein in the Gas Phase?......................................................................................... 13
   3.3 The Use of Force Fields............................................................................................................... 14
   3.4 Classical MD............................................................................................................................... 16
   3.5 Methods based on MD simulations............................................................................................ 17
      3.5.1 High temperature MD simulation......................................................................................... 17
      3.5.2 Massively distributed computational MD ............................................................................ 18
      3.5.3 Hyperdynamics scheme ...................................................................................................... 19
   3.6 'Double-ended' problem............................................................................................................. 20
   3.7 Search over PES.......................................................................................................................... 21
   3.8 Objective of the thesis................................................................................................................ 22
   3.9 Thesis organization..................................................................................................................... 23
4 Methodology .................................................................................................................................. 24
   4.1 Rate constant calculations.......................................................................................................... 24
   4.2 Identifying the transition state structures (SPW)........................................................................ 26
   4.3 Fitting Force Field Parameters to ab initio Calculations ........................................................... 29
5 Protein ions in the gas phase: theory and experiment................................................................. 30
   5.1 The Ubiquitin Experimental System........................................................................................... 33
   5.2 Gradual temperature MD simulations....................................................................................... 34
   5.3 Initial and intermediate structures of ubiquitin +13 ions in electrospray experiments........................................................................................................................................... 36
   5.4 Energy Barriers and Timescale Calculation ................................................................................. 40
   5.5 High Temperature MD Results versus ECD Experiments ....................................................... 44
   5.6 Comparison with cross-section experiments .............................................................................. 46
5.6.1 Cross-section calculations for ion drift experiments ................................ 47
5.6.2 Comparison of the final structure with experimental data ........................ 48
5.6.3 Comparison of intermediate states with experimental data ...................... 48
5.7 Ubiquitin summary ...................................................................................... 49
6 The HMR method results................................................................................. 50
  6.1 HMR method overview.............................................................................. 51
  6.1.1 Construction of the network of minima ............................................... 52
  6.1.2 Describing the dynamics on the network.............................................. 54
  6.1.3 Finding the optimal reaction path ....................................................... 55
  6.2 Valine dipeptide example ........................................................................ 57
  6.2.1 Valine dipeptide MD path................................................................. 58
  6.2.2 Validity of the RRK Approach ........................................................... 62
  6.2.3 HMR method versus MD reaction pathway ....................................... 63
  6.2.4 Run-time comparison ....................................................................... 64
  6.2.5 Summary of blocked valine dipeptide results: accuracy of path, timescale,
      reduced run-time .................................................................................. 65
  6.3 The blocked valine tripeptide - finding multiple paths............................... 65
  6.3.1 Valine dipeptide MD and HMR paths ................................................. 66
  6.3.2 Run-time comparison ....................................................................... 71
  6.3.3 Summary blocked valine tripeptide – multiple path run-time .............. 71
  6.4 Blocked alanine hexapeptide conformational change ................................ 72
  6.5 Phenylalanine conformational changes - HMR method and experiment ...... 74
  6.6 The HMR method summary .................................................................. 77
7 Discussion and Conclusion............................................................................ 80
8 References.................................................................................................... 82
List of acronym:

ECD - Electro Capture Dissociation
ESI - Electrospray ionization
fs – femtoseconds
Hartree-Fock (HF)
HMR Hybrid Molecular Dynamics / RRK
IMS - Ion mobility spectroscopy.
MD - Molecular Dynamics.
MS - Mass Spectrometric.
ns – nanoseconds.
PES – Potential Energy Surface
ps – picoseconds.
RRK - Rice Ramsperger and Kassel method.
SPW - Self Penalty Walk
TST - transition state theory
3 Background

3.1 Introduction

The proper functioning of cells and organisms is ensured by the correct activity of a network of thousands of proteins. Proteins and peptides are complex molecular systems with distinct stable three-dimensional structures in which their activity is usually performed by transition between different structures and conformations. The conformational transitions are the key to understanding the function of these biological molecules. The significance of the contribution of conformational transition to the function of the protein manifests itself in a number of diseases caused by improper conformational transition or misfolding of native proteins. Some such examples are Alzheimer’s, Parkinson’s and Huntington’s diseases[1-3], and prion proteins that can cause Creutzfeldt-Jakob’s and Gerstmann-Straussler-Scheinker disease [4].

In the field of conformational changes the experimental methods are usually very limited and can identify in most cases only the most stable conformation along the reaction path. Computational chemistry on the other hand enables following and simulating the conformational changes, but is limited because of its short timescale, which means that only the integration of experimental and computational methods can contribute to a better understanding of the nature of biological molecules. One of the most fascinating examples in the field is the study of the relation between the structure and the function of bacterial chaperon protein GroEL. By means of experimental methods it was found that during its functional cycle GroEL undergoes great conformational changes in a timescale
of 15 sec [5, 6]. During this cycle the protein can adopt two different conformations: the 'closed' structure in the absence of ATP and co-factors; and 'open' conformation when bound to these co-factors [7]. Even state-of-the-art computers cannot simulate classical Molecular Dynamics simulations (MD) of the protein in the latter’s timescale, and owing to this limitation the target MD, which is detailed below, and based on the two well-defined protein structures was used to reveal the whole pathway connecting the open and the closed conformations [8]. The simulations in the atomic resolution exposed a new dominant intermediate which plays a key role in the conformational transition of GroEL. The simulations demonstrated that steric interactions and salt bridges between different sub-units are the source of observed positive cooperativity of ATP binding and hydrolysis within one ring and the negative cooperativity between the two rings. Later cryo-electron microscopy results supported this prediction [9].

The three-dimensional structure and the mechanism of all proteins and peptides are determined by the nature of the underlying potential energy surface (PES), also known as 'energy landscape' [10]. The energy hypersurface of biological molecules is a very rugged surface, with energy basins and mountains of a wide range of depths or heights and spatial extent. Experimental evidence for the complexity of PES in a system with many degrees of freedom is the multiplicity of relaxation times [11] observed, for example, in the rebinding kinetics of CO to myoglobin [12, 13]. Moreover, classical MD simulations of myoglobin in the crystalline structure demonstrated that the potential energy surface of proteins is characterized by a large number of thermally-accessible minima in the
neighborhood of the native structure. In this simulation approximately 2000 minima were sampled in a 300-picosecond trajectory at room temperature.

Biomolecular systems generally are characterized by a very large number of degrees of freedom \(10^4 - 10^6\) or more and local minima [14-17]. The motions along these degrees of freedom cover a wide range of time and spatial scales [15-19], from femtoseconds and tenths of nanometers to milliseconds and micrometers. This complexity of the PES makes the search for optimal reaction paths connecting two different minima on the PES a daunting task. Part of the difficulty in finding the optimal reaction path is the necessity for high energy barriers transitions. The dynamics of the molecule is in these regions are characterized by long time periods in which the system remains trapped at certain minima. This complicates the study of evolutionary conformational changes when classical MD is used. Owing, moreover, to computational restrictions, there are only a few examples in which calculation of the microsecond timescale simulation was successfully performed on proteins and peptides [20, 21]. In most cases the MD simulation timescale is limited to a few nanoseconds[22], and thus falls short of the interesting experimentally-observed protein motion time-scales, which, for most proteins, lie somewhere between a few milliseconds to a few minutes. Several methods have been developed to enable the study of conformational changes of biological molecules, some of which, with the biological molecule studies produced in consequence, are the High temperature MD simulation, the use of massively distributed computational MD, the Hyperdynamics scheme and some other methods which are presented in details below.
Despite the considerable number of different methods proposed, the problem is still largely open. Additional theoretical tools for this problem are desirable.

The objective of this study is to investigate and learn the nature of conformational changes of biological molecules at the short (fs to ps) and long (microsecond to seconds) time scales. To study conformational transitions of the short timescale classical MD simulations were carried out. For the study of transitions on the long timescales new theoretical MD-based approaches were developed to overcome the present computational limitation of classical MD simulation. The integration of classical MD simulation with other methods enables us to give detailed trajectories in atomic resolution for further interpretation of the experimental results.

3.2 Why Study Protein in the Gas Phase?

To understand the structure and dynamics of biological molecules in their native environment, it is often imperative to understand first these properties in isolated systems in the gas phase. The native conformation and dynamics of the proteins are defined by both the intramolecular interactions within the protein itself and the interactions between the protein and its solvent. The protein-solvent interactions are large and make a large contribution to the protein structure, that of thousands of kilojoules per mole for a small protein[23, 24]. By the removal of a peptide or protein from the solution to the vapor phase, it is possible to separate its hydration and intramolecular interactions and examine them independently. Studies of proteins in the gas phase therefore provide insight into intermolecular interactions and enable better understanding of stabilization interaction of tertiary and secondary structures in peptides and proteins [25]. Moreover, the intensive
current interest and experimental data in conformational transitions of biological molecules in the gas phase, including the context of structural evolution in mass spectrometric experiments, enable good interaction with interpretation of the experimental data.

3.3 The Use of Force Fields

Large molecular systems are far too complex to be treated by quantum mechanics. In force field methods, however, this difficulty is bypassed when $V$ (the potential energy) is noted as a function of nuclear coordinates. The parameters considered in the function are fitted to experimental or higher level computed data. Force field methods ignore electronic motions and calculate the energy of a system as a function of the nuclear positions alone (assuming the Born-Oppenheimer approximation). All the interactions within the molecule can be expressed by a relatively simple analytical function, in which stretch, bend, torsion and improper energies of the molecule are expressed, as well as the van der Waals and Coulomb interaction.

The form of the potential energy ($V$) function used in this thesis is taken directly from AMBER force formalism [26], and is given by the following equation (3.1):

$$V = \sum_{\text{bonds}} k_b (b - b_{eq})^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_{eq})^2$$

$$+ \sum_{\text{dihedrals}} k_\phi [1 - \cos(n\phi - \delta)]$$

$$+ \sum_{\text{impropers}} k_\omega (\omega - \omega_{eq})^2 +$$

$$\sum_{\text{nonbonded}} \left( \frac{A_{ik}}{r_{ik}^12} - \frac{C_{ik}}{r_{ik}^6} \right) + \frac{q_i q_k}{\varepsilon r_{ij}}$$

(3.1)
Here $k_b$ is the bond force constant and $b-b_{eq}$ is the atomic displacement from equilibrium. This potential model is limited by the use of harmonic forces for the representation of the bond stretches. Obviously processes involving bond cleavage events cannot be represented in this way, but conformational changes are properly represented. Bond angles are accounted for in the second term of the function, where $k_{\theta}$ is the angle force constant and $\theta-\theta_{eq}$ is the deviation of the angle from equilibrium for the 3 relevant atoms. The dihedrals (torsion angles) are presented in the third function term, where $k_{\phi}$ is the dihedral force constant, $n$ is the multiplicity of the function, $\phi$ is the dihedral angle and $\delta$ is the phase shift. The impropers are noted in the fourth term. This is the plane bending, where $k_{\omega}$ is the force constant and $\omega-\omega_{eq}$ is the out-of-plane angle. Non-bonded interactions between pairs of atoms $(i,j)$ are represented by the final two terms. By definition the non-bonded forces are only applied to atom pairs that are separated by at least three bonds. The van der Waals energy is calculated with a standard 12-6 Lennard-Jones potential and the electrostatic energy is calculated with a Coulomb potential (where $q_i$ and $q_j$ are the partial charges). In the above-mentioned Lennard-Jones potential, the $A_{ik}$ and $C_{ik}$ constants are atom-type dependent parameters. As a result of the force field formalism, not only the functional form but also the parameters, i.e. the various constants such as $k_b$, $b_{eq}$, $k_{\phi}$, $\theta_{eq}$ in equation (3.1), must be specified. To ensure correct molecular geometries, calculations of the necessary parameters must be carried out. Such parameters can be revealed through quantum mechanics methods of calculating small molecules and experimental data parameterization techniques[27]. Parameters for the most common amino acids can be found in standard force field parameter files such as CHARMM[28] and AMBER[29]. In this work special amino acid parameters, which
are not represented, or are not represented with the appropriate precision, are applied to ab initio calculations.

3.4 Classical MD

Molecular Dynamics (MD) simulations have been used as a standard tool in studies of biological molecules to provide information complementary to that obtained from experiments[30, 31]. MD simulation methods are usually based on the Verlet algorithm that solves Newton’s equation of motion[32]:

\[
X_{i+1} = X_i + V_i \Delta t + M^{-1} \cdot (\Delta t^2 / 2) \cdot F_i \\
V_{i+1} = V_i + M^{-1} \cdot (\Delta t^2 / 2) \cdot (F_i + F_{i+1})
\]  

(3.2)

where \(X_i\) and \(V_i\) are the coordinate and velocity vectors at time slice \(i\), respectively. The Mass matrix is \(M\). The integration time step is \(\Delta t\) and the force vector \(F_i\). The solution of equation (3.2) is propagated by specification of initial conditions, e.g. \(X_o\) as the initial coordinate and \(V_o\) by Boltzmann distribution. To maintain algorithm stability the integration time step \(\Delta t\) must be small (usually a few fs), and thus a limitation is fundamentally applied to long timescale simulations. Owing to this limitation all-atom MD algorithms can simulate events in the range of \(10^{-9}\) to \(10^{-8}\) seconds for typical proteins and \(10^{-6}\) seconds for very small proteins while investing outstanding efforts [33-37]. These timescales are at least one order of magnitude smaller than even the secondary structure motifs formation, such as \(\alpha\)-helices and \(\beta\)-hairpins. [38]. Therefore classical MD is not yet capable of directly simulating long timescale biological conformational changes.
3.5 Methods based on MD simulations

Despite improvements in recent years in hardware and software computational capabilities, and advances in the understanding of the properties of integrators [39], the basic-time MD simulation step-integration is still too small to overcome the timescale limitations of the simulations. It is thus essential to produce alternative methods. As a result of the simplicity and accuracy that MD simulation can provide a large number of alternative MD simulation-based methods were developed. Three of the most commonly-used MD-based methods are represented below: the high temperature MD simulation, the massively distributed computational MD and the hyperdynamics scheme.

3.5.1 High temperature MD simulation

One of the most widespread MD-based approaches is the pursuit of MD simulation at high temperatures. In the high temperature MD simulation, a standard MD simulation algorithm is used and the velocities are sampled by the use of an artificially high-temperature Boltzmann distribution. At sufficiently high temperatures high energetic barriers can be easily and rapidly surmounted. The great advantage of the high temperature MD simulation approach is its simplicity. Peptide simulations from van Gunsteren's group research gave good support for the approach [37]. It was found that during an MD simulation peptides repeatedly unfolded and refolded, and the unfolding and refolding pathways were independent of temperature. By means of the high temperature MD simulation, over a dozen folding paths of proteins with different architecture were compared successfully with those of the experiments [40-45].
There are, however, certain difficulties. The sequence of conformational changes in high and low temperatures cannot be expected to be similar in all cases. Moreover, since the temperature of the simulation is higher than that in experimental procedures, the time-scales cannot be directly inferred.

### 3.5.2 Massively distributed computational MD

Waiting for thermal fluctuations that should bring the examined system over the energetic barriers is the largest expenditure of time in MD simulations: even rare conformational change events (e.g. passing over high energy barriers), statistically may occur with very rapid trajectories. The use of massively distributed computer network methods for accelerating the simulation was suggested by Pande et al [46, 47]. There, in the first free energy minimum, a large number \( N \) of trajectories was generated, (figure 3.3a). The number of trajectories that passed over a high barrier up to a time \( t \) \( n(t) \) was then determined by the method ,as represented at figure 3.3b. The fraction of folded trajectories can be very small. Since, however, \( N \) is large, \( n(t) \) can still be statistically significant. If an exponential rate law is assumed, as was found in a study of small proteins and peptides [47], the barrier-surmounting rate can be extracted from the short-time behavior of the population of trajectories. Presently, all other simulations are placed at the same configuration space as the simulations that have crossed that barrier, as depicted in figure 3.3c. This process is then repeated until additional free energy barriers are crossed. In this way timescales and long time MD simulations can be calculated. As the molecules increase in size, however, the complexity of the “folding” process increases as well. Intermediates can be found along the folding pathways, which can make the process non-exponential and cause timescale error calculations [48-50].
Another drawback of the method is that it is not feasible for many research laboratories: over 40,000 PCs spread over the globe and managed via the interface found on the folding@home website were used by Pande.

Figure 3.3: A schematic illustration of the method used by Pande et al. (a), Starting with a large number of trajectories in the first free energy minimum (b), after several trajectories have passed over the energy barrier (c) all other simulations are placed at the same configuration space [47].

3.5.3 Hyperdynamics scheme

In the hyperdynamics scheme [51, 52], classical MD simulation is performed by means of a repulsive bias potential. The scheme modifies the PES specifically to reduce the CPU time spent evolving the system when it is trapped in high-frequency oscillations in potential minima. This is done by raising the PES in the region where molecules fluctuate. Such an increase in the potential energy of the minimum reduces the activation energy of the transition, and therefore increases the transition rate and boosts dynamical processes. As a result, this method speeds up classical MD. In hyperdynamics, transition state theory is used to relate the timescale to the simulation. This scheme studies the long-term description of the system’s evolution. The cost of the scheme is that detailed
vibrational information is lost, as well as accurate information related to the fast transitions trajectories.

### 3.6 'Double-ended' problem

As commonly formulated, problems in classical dynamics are typically 'initial value' in character. That is, they involve the determination of the evolution of a set of coordinates and momenta given their values at a specified starting-time, assuming the relevant equations of motion are known. Although not without technical challenge, such initial value problems are well posed and amenable to study with established methods [53]. In certain conditions 'double-ended' problems can be more efficient, especially for the surmounting of high energy barriers. Several methods were developed to compute the trajectory between two well-known configurations, such as the initial and the final minima. Some of these methods are:

1. The stochastic difference equation in time and length [54] is based on the formalism of searching for a trajectory that makes the action ($S$) stationary, where the action is expressed as: $E$ is the total energy, $U$ is the potential energy and $dl$ is a length element. $R_i$ and $R_f$ are the initial and the final conformations respectively. A discretized version of the action is optimized and a minimum of the functional of the trajectory (action) provides an approximation of the true classical path. Protein folding and sugar puckering paths were successfully computed by means of this method [55, 56].
2. Steered and targeted MD implemented in NAMD and CHARMM programs [57-59], which take advantage of the MD accuracy and add a time-dependent external force along certain degrees of freedom or along a local free energy gradient - to drive the reaction from the reactant to the product [60-62].

3. Gaussian chain approach method [63] and the self-penalty walk (SPW) used in my work and described in detail in section 4.3 [64].

### 3.7 Search over PES

The PES includes minima and transition states of conformations. In order to define and analyze multiple reaction paths a search over the configuration space is carried out which identifies the different minima and transition state conformations. The different rates constants between two neighboring minima are calculated with the transition state theory (TST)[65]. For example, tetra-alanine and octa-alanine conformational changes were analyzed successfully by such methods [18, 66, 67]. As a result of the enormous number of minima on an individual PES, one of the main problems of these methods is to achieve an efficient PES search. Becker et al. solved this problem by using high-temperature MD simulations followed by gradual annealing [68]. The conformational samples were pruned by removal of conformations that are highly similar to other conformations in the sample, in order to enable more efficient potential energy landscape analysis. In all of the polyalanine analysis cases the timescales were related to the simulation by use of TST.
The TST is partially quantum version of the classical statistical RRK developed by Rice and Ramsperger in 1927 and Kassel in 1928[69]. This method is challenging in particular for larger biological systems, which have a large number of local minima and saddle points. For larger biological molecules it is unfeasible to search the whole PES in order to locate the minima and to compute the rate constants [70]. Moreover, the TST is valid only for the ratio \( E_a/k_BT \gg 1 \) (where \( E_a \) is the energy barrier; \( k_B \) is the Boltzmann constant and \( T \) the temperature). TST can reproduce molecular MD timescales, with the exception of short timescales (low energy barriers)[67, 71, 72]. Since a broad distribution of the energy barriers ranging from one-tenth to tens of kcal/mol is generally characteristic of biomolecules [73, 74], even for small biological molecules in which the PES can be explored, the method cannot be applied to all of the transitions for these systems.

### 3.8 Objective of the thesis

Despite the large number of different theoretical methods which were developed to identify the reaction path in biological molecules there is a necessity for a method which can calculate long time trajectories as well as timescales. In this thesis two newly-developed methods are presented to calculate the conformational changes of proteins and peptides. The first method is a gradual temperature MD simulation applied to ubiquitin ions in the gas phase simulating the mass spectrometry experiments done on this ion. The second method presented is a hybrid MD/RRK method developed by us, in order to enable the study of short and long timescales conformational changes, and the multiple reaction paths of biological molecules.
3.9 Thesis organization

The thesis is structured as follows. In Chapter 4 the main methodologies used in the thesis are described: Chapters 5 and 6 describe the results; Chapter 5 describes how the ubiquitin ions’ conformational changes in mass spectrometry experiments are simulated by gradual temperature MD simulation and the simulation results are compared with electron-capture dissociation, and ion-mobility spectrometry experiments. In Chapter 6 the newly-developed hybrid MD/RRK method for analyzing reaction paths in biological molecules is presented. In this chapter we compare single and multiple-reaction paths revealed by MD simulation with the one revealed by the hybrid MD/RRK method, examine a blocked alanine hexapeptide conformational change from a $\beta$ hairpin to a $\alpha$ helix, and finally compare the hybrid MD/RRK method results of phenylalanine simulation in the gas phase with spectroscopic experiments and high level ab initio calculations. Chapter 7 presents the conclusions.
4 Methodology

In this chapter the different theoretical tools widely employed in this work for the study of conformational changes in biological molecules are presented in detail.

4.1 Rate constant calculations

RRK calculations enable rate constant calculations. The RRK theory supplies a statistical method for computing rate constants for a process at constant energy $E$ (such as that determined by MD). The starting-point, using the classical RRK theory, is to identify the reactant and the involved transition state on the PES. The reactant is a minimum of the located conformation, and the transition state structure can be found by several computational methods, including the SPW method which is described below [64]. The RRK rate is given by equation (4.1) [75, 76],

$$k(E) = A \left( \frac{E - E_0}{E} \right)^{s-1}$$

(4.1)

Where $E$ is the total energy, $E_0$ is the energy barrier, and $s$ is the number of internal degrees of freedom, namely $3n-6$ (where $n$ is the number of atoms in the non-linear molecule). $A$ can be calculated as:
\[
A = \left( \prod_{j=1}^{s} \nu_j \right)^{s-1} \left( \prod_{i=1}^{s-1} \nu_i' \right)
\]  

(4.2)

Where \( \nu_j \) is the vibrational frequency from the initial minimum geometry and \( \nu_i' \) is the vibrational frequency from the transition state conformation. The vibrational frequencies are calculated with the harmonic normal mode approximation (equation 4.3), where \( \lambda \) is the eigenvalue of the mass-weighted second derivative matrix of the potential, and \( \nu \) is given by:

\[
\nu = \frac{\sqrt{\lambda}}{2\pi}
\]

(4.3)

In principle the RRK equation parameters \( s, A, \) and \( E_0 \) can all have different values for different intermediate barriers: in the gradual MD simulations carried out for ubiquitin in Chapter 5, it was found that effectively all barriers can be realistically treated with essentially the same \( A \) and \( s \). In such cases, and for thermal temperature ensembles, the rate of barrier crossing can be represented by the Arrhenius equation [69],

\[
\ln(k) = \ln(A) - \frac{E_0}{RT}
\]

(4.4)

Basically RRK is used because of the strong evidence for its validity, at least on the semiquantitative level in unimolecular reactions theory[69]. In the present context, it
seems reasonable to employ RRK theory, which is a classical statistical theory, since
classical dynamics are used in other parts of the treatment. It is also possible to use the
more complete and rigorous RRKM theory [77], which is quantum-mechanical in part
and of which RRK is the classical limit. The RRKM rate is given by:

\[ k_{ij} = \frac{k_B T}{h} \frac{Q_{ij}^\dagger}{Q_i} e^{-\frac{E_o}{k_B T}} \]  \hspace{1cm} (4.5)

Where \( k_{ij} \) is the gas phase rate constant for the transition between minima \( i \) and \( j \), \( k_B \) is the
Boltzmann constant, \( h \) is Planck’s constant, \( Q_i \) the partition function of the ‘reactant’
state, \( Q_{ij}^\dagger \) is the partition function of the transition state, and \( E_o \) is the barrier height
measured relative to the reactant structure. In cases where some of the transitions require
a more accurate calculation, RRKM can indeed be used.

4.2 Identifying the transition state structures (SPW)

Generally, in order to find the path between two adjacent minima, we apply Ron
Elber’s[64] Self Penalty Walk (SPW). The method treats an "elastic band" that consists
of a number of replicas of the original molecule connecting the reactant to the product
[64]. The replicas are interpolated between the two minima of interest, and are connected
by harmonic weight to prevent collapse of the replica onto nearby local minima.

The SPW method minimizes the weight \( (W) \) of the path by minimizing the quantity:

\[ W = \sum_i V(R_i) + \sum_j \gamma [d_{i,i+1} - \langle d \rangle]^2 \]  \hspace{1cm} (4.6)
The first term minimizes the potential energy, where $V(R_i)$ is the potential energy of conformation $i$ (replica $i$). The second term gives weight to distances between neighboring structures that deviate from the mean value, $\gamma$ is a parameter selected on the base of experience and $d_{i,i+1}$ is the root mean square distance between structures $i$ and $i+1$. Finally, $<d>$ is the average distance between each of two neighboring structures.

Originally the SPW method was designed to find the optimal path between two known intermediates and not to identify transition state structures. This method is, however, a very computationally efficient method which doesn’t require the calculations of second derivatives of the potential energy (Hessian), which is the method used for small molecules [78-80]. To measure the accuracy of the SPW in identifying correctly the saddle points, a comparison was made of the SPW extremum point along different reaction paths in valine dipeptide PES with the calculated saddle point, using the MATLAB program, as shown in Figure 4.7a. This comparison showed that the SPW method finds the transition state structure and energy with satisfactory accuracy. The energy and spatial deviation of the transition state, calculated with the SPW method from the saddle point calculated from Hessian matrices, are in the range of several degrees in the spatial psi and phi dihedral angles and less than 3% in the energy: these results are presented in Figure 4.7c. Moreover, the extremum along the SPW pathway was found to be consistently slightly lower in energy than the exact saddle point energy. The transition state energy was not scaled, but took into consideration that the rate constants, which are derived from the RRK approach using this energy barrier, will be higher than in the classical MD simulation.
Figure 4.7: (a) A contour plot of the PES of valine dipeptide as a function of psi and phi dihedral angles. A-E are the different saddle points along the PES, in circle is the calculated saddle point using a second derivative calculation of the PES, in star the extremum along SPW path. (b) A schematic representation of the valine dipeptide and the relevant psi and phi dihedral angles. (c) The energy and spatial deviation of the transition state calculated using the SPW method compared with Hessian matrices method results.
4.3 Fitting Force Field Parameters to ab initio Calculations

In some of the studies presented in this work the AMBER parameters were either missing or weren’t sufficiently precise. Missing parameters were therefore assigned by means of Hartree-Fock (HF) calculations.

The HF model is a standard tool for computing the ground electronic state structure of relatively small molecular systems (up to 20 atoms)[81]. To calculate and minimize the energy, the electronic wave function is represented as a Slater determinant, and the optimization of the determinant is solved by use of the variation method. From the geometrical optimized structure the equilibrium distance and angles for the different bonds and angles are revealed, and implemented in the force field. In order to obtain the force constant parameters for the bonds and angles, the Hessian matrix (the second derivative of the potential energy with respect to the coordinates) of the minimized structure is computed, and the vibrational frequencies of the system are observed. The harmonic forces are computed according to the following equation:

\[ k = 4\pi^2 \nu^2 \mu \]  

(4.8)

where \( k \) is the harmonic force constant, \( \nu \) is the relevant vibrational mode frequency and \( \mu \) is the reduced mass of that specific mode. The charge distribution is taken directly from GAMESS ESP charges calculations.
5 Protein ions in the gas phase: theory and experiment

The study of conformational transitions in proteins has grown extensively in recent years, both experimentally and theoretically [82]. A major challenge in understanding the protein folding landscape is the effect of solvents. Proteins are routinely modeled theoretically using a variety of solvent models [83]. Mass Spectrometric (MS) experiments carried out on isolated, solvent-free proteins offer hope for the exploration of protein unfolding without the presence of water [84-86]. This may shed light on the differences between such a process and the corresponding unfolding in liquid solution [87]. Thus, solvent-free simulations are interesting from a fundamental point of view, as they provide a means of separating solvation effects from intrinsic properties of proteins. Calculations of protein unfolding in the gas phase seem desirable at this stage for supporting the growing volume of MS experiments. Such calculations should generate a framework for describing important aspects of the potential energy landscapes that are intrinsic to proteins.

In Electrospray ionization MS (ESI-MS) the protein, solvated in a volatile solution, is pumped through a narrow capillary and dispersed into an aerosol of highly charged droplets [88]. The charged droplets decline in size through solvent evaporation, a process that takes place in the presence of a warm flow of nitrogen gas which passes across the front of the ionization source. Eventually charged protein ions, free from solvent, are released from the droplets [88].
Within a framework of ESI-MS, a range of methods has been developed to explore structural and other properties of the protein ion produced in such experiments. Central to the field are cross-section measurements of ion mobility [89, 90] and also Electro Capture Dissociation (ECD) [91-94], both of which provide information on equilibrium and the intermediate structures of biomolecular ions [95, 96].

Ion mobility spectroscopy (IMS) experiments and the information they provide are at the focus of the present study. In these experiments protein ions traverse an environment of inert neutral gas under the influence of a weak electric field [84-86]. In general, the mobility of a protein ion in the gas phase depends on its average collision cross-section with the buffer gas. Compact conformers have smaller cross-sections and thus higher motilities than elongated ones [97-100]. Transit times for ions under these conditions can therefore be used to determine the shape and size of the gas phase proteins [101]. More sophisticated experimental methods including ion trap-IMS and IMS-IMS combinations are able to follow structural changes occurring over the millisecond to seconds timescale [102-104]. It is thus possible to determine the collision cross-section via drift time and ion mobility [105, 106]. A given model structure for a biomolecular ion can then be tested by comparing a computed cross-section to experimental data. A number of models have been proposed to calculate theoretical cross-sections. One approach is to calculate an orientation averaged projection cross-section of the model using the hard spheres scattering model [107].
To interpolate the unfolding of the protein in MS conditions it is desirable to compute the final and intermediate structures in the process, the reaction pathway and even the timescales involved. A major difficulty in addressing these objectives is the very long timescale (typically >1 ms) of these processes. One aspect of this is the presence of multiple minima separated by relatively high barriers [18, 19]; in such cases, the dynamics are often characterized by long periods during which the system remains trapped in certain minima. These challenges have led to a growing interest in developing novel methods for computing conformational changes in peptides and in proteins [47, 108, 109]. Standard MD simulations of extended molecules are typically limited to the nanosecond (ns) timescale [22]. Simplifications and approximate methods for longer timescales are of great interest, as this makes it possible to do calculations with much more modest computational resources. One such approach is the use of high-temperature simulations to overcome large barriers on a shorter timescale. Obviously it is necessary to link the high temperature results with prediction of behavior of experimental temperature. This method has proved successful in a substantial number of specific folding cases [45]. The problem is however, that the sequence of conformational changes at high and low temperatures may not be the same in all cases [110].

In this chapter we propose a pragmatic, practical approach to overcoming high-energy barriers and simulating long timescale trajectories of proteins in the gas phase. Our approach borrows ideas from well-established methods for folding, whether in the gas phase or in solution, and adapts them to the specific problem addressed. The method is
mostly based on Voter’s [51, 52] and the Replica Exchange method [21, 111] in which calculations at a range of temperatures are employed to overcome energy barriers.

The following chapter presents the study of the unfolding of ubiquitin +13 ions in ES-MS conditions. The intermediate and final structures in the process, the unfolding pathway, and the timescale involved are the targets of the study. The approach taken borrows ideas from previous high temperature MD methods and adapts these to provide a practical tool requiring modest computational means. The results are tested against experimental data, which are admittedly limited of the final unfolded structure for which cross section data is available. As will be seen, the results are quite encouraging.

5.1 The Ubiquitin Experimental System

Ubiquitin has been the topic of extensive theoretical and experimental studies [112, 113] and thus was chosen as a model system for this work. Ubiquitin is a small globular protein of 76 amino acids (figure 5.1a) found in all eukaryotes, and is involved in post-translational protein modification, cell metabolism, transcriptional regulation, signal transduction and gene regulation [114-116]. Ubiquitin is partially folded under destabilized conditions such as low pH and water/alcohol mixtures. The structure of ubiquitin at pH≈2 and 60% CH₃OH was characterized using NMR, and was named the “A state” (figure 1b) [117]. While the focus of this study is the characterization of protein conformation transitions by means of theoretical tools, ESI-MS [84] emerges as an important experimental tool for the characterization of these features. With this conceptually-simple approach, charged folded proteins and non-covalent interactions in the gas phase can be studied directly [86]. To enable study of the conformational changes
in atomic resolution of gas-phase ubiquitin +13 ions, the protein was modeled in accordance with its \textit{A state} structure, and was charged according to Williams et al.’s calculations [118].

![Figure 5.1](image)

Figure 5.1: a) The native state globular structure of ubiquitin [119]. b) The \textit{A state} conformation, stable at low pH methanol water solution as was generated according to Ernst et. al [117].

### 5.2 Gradual temperature MD simulations

The approach taken here is an adaptation of the widely used high temperature MD approach, where structural evolution is accelerated through the use of elevated temperatures. Obviously, sufficiently high barriers can be relatively easily surmounted [34]. A major advantage of this approach is its simplicity. The main problem is, however, that the sequence of conformational transitions may change at more elevated
temperatures, at least in certain cases [110]. To ensure that the simulation at high
temperature followed the same sequence of conformational transitions as at the
experimental temperature, MD simulations were carried out using a set of gradually
increasing temperatures. This made it possible to check at each stage that the process did
not switch to a different reaction pathway. As the temperature is increased by a
sufficiently small increment, the system overcomes the next barrier within the simulation
time window. In order to implement the approach, the Moil MD simulation package was
employed [120]. A set of 15 different simulations of 5 ns duration were initiated at the
initial state of ubiquitin +13 at the assumed experimental temperature [117]. Following
this simulation step the temperature was raised by 50K and a set of 15 new trajectories
were computed to investigate the structural changes, again for the timescale of 5 ns. The
magnitude of the temperature increase is determined by experience. Thus: the trajectories
were inspected for continuity of behavior from one temperature to the other. That is, the
high temperature result was accepted provided that the initial pathway of structural
changes corresponded closely with the one found at the lower temperature simulations. In
the case of ubiquitin +13 ions, the different conformations were defined by a set of
distances between noncovalently bonded atoms. The specific characterization and
identification of the important intermediates in this process is discussed in section 3.2.
In cases where jumps of 50K are inadequate and do not correspond to the same early
unfolding path, one has the option of using a smaller temperature jump. At each
successive temperature the unfolding process may proceed across a barrier to a new
maximum, but the path up to that point was required to agree with the lower temperature
trajectories. This process was repeated until a conformation was obtained beyond which higher temperature simulations did not predict new structures.

To be able to calculate the rate of conformational changes at a low (experimental) temperature for which MD is too slow, we employed the approximate RRK method for rate calculations (equation 4.1). These RRK rates were tested against full MD simulation in small peptides and were found to provide a satisfactory approximation [108]. Although the parameters $\omega$ and $s$ can in principle have different values for different barriers, we found here that they are essentially the same for all barriers encountered. In this case and for thermal temperature ensembles, the rate of barrier crossing can be reduced to Arrhenius behavior (equation 4.4) [75], which allows convenient calculation of rate constants at any temperature.

### 5.3 Initial and intermediate structures of ubiquitin +13 ions in electrospray experiments

The ubiquitin +13 ions are formed by ESI experiments in ~2:3 water/methanol solutions at low pH. The N-terminal half of the $A$ state was modeled by an antiparallel $\beta$-sheet and a central $\alpha$-helix using SYBYL and BALLView programs [121, 122]. Experimentally, acidic amino acids are found to be protonated in the low pH solution (pH~2) [118]. The protein was charged according to the calculations of Williams et al. [117]. The determined parameters were integrated into the force field. A realistic description of the initial state of the ES process requires simulations of the protein in a bulk solvent. However, there is evidence from recent MD simulations that the protein structure in
solution remains almost intact until solvation is nearly completed [123]. In the simulations in the present study, we therefore used the protein structure in solution as the starting point for the structural changes in MS. Thus the model here assumes that the charging of the protein and the removal of the water are the driving forces for the structural changes in the MS experiments. The timescales for conformational changes provided here do not include the contribution of the evaporation process and may therefore be lower than the actual experimental timescales.

The trajectory calculations began with 15 trajectories initiated at the $A$ state with no water hydration layer. These calculations were carried out at the assumed experimental temperature of 400K, resulting in the formation of a structure which we refer to as Intermediate-1 after a mean time of 1860 ps. All of the 15 trajectories led to this structure within the timescale of the simulation at 400K (5 ns). No other Intermediate state with a significant lifetime was identified in the MD simulations at 400K. The Intermediate-1 (figure 5.3b) is characterized by the disappearance of the helical structure from residues 41-76 that is present in the initial state. This region carries a charge of +7 located at protons on the side chains. In the absence of a solvation layer the repulsion between the charges overcomes the classical H-bonds [123] that hold the helix in this region together in the $A$ state. This, then, is the mechanism for the disappearance of the helical element in the $A$ state.

The transition between the $A$ state to Intermediate-1 is characterized in terms of the distance between the $C\alpha$ of Ala46 and the C terminal: at distances larger than 72 Å the
Intermediate-1 structure is considered to have been formed. The Intermediate-1 structure persists throughout the MD trajectories at 400K for 5 ns. An accelerated transition to the next Intermediate structure is obtained when the temperature is raised to 500K. Again this was simulated by computing 15 MD trajectories starting from the A state structure at the new temperature. Continuity of qualitative behavior is observed between 400K and 500K throughout the reaction pathway from the A state to Intermediate-1. This continuity was also observed in simulations at the intermediate temperature of 450K (the results are not shown here). The trajectories at 500K lead, within the computational time window of 5 ns, to the formation of a new structure – Intermediate-2 (figure 5.3c). To examine the structural change, first let us consider Intermediate-1. In this structure the hydrogen of the first N’ methionine and the oxygen of glutamate 16 are in tight contact. This interaction represents an energy barrier that functions as the gatekeeper of the two beta sheet strands. At this point, the two β sheets, which are structured like an interlacing strand (similar to a zipper), disengage. This transition is characterized by an increase of more than 20 Å in the distance between the Cα of Glu17 and the N terminal. The transition from Intermediate-1 to Intermediate-2 seems to correspond to a pattern often observed in the RRK type of dissociations of polyatomic molecules and complexes [75], and at the increased temperature of 500K sufficient thermal energy reaches the relevant bonds within 5 ns or less. In this case these are the hydrogen bonds that hold the β-sheets together. It can be noted that this transition occurs over a fairly narrow, sharply defined time interval, much more so then the transition from A state to Intermediate-1, as presented in figure 5.4b and discussed in chapter 5.4.
The Intermediate-2 structure remains stable over 5 ns throughout the simulations at 500K. The next transition is observed when temperature is increased to 600K. This transition leads to the formation of an elongated structure (see figure 5.3d). The transition takes place following the break of the hydrogen becomes possible after the bond between the positively-charged lysine 11 and the neutral glutamate 34 residues.

The near linear structure consists of a short $\alpha$ helical element continued by an elongated backbone. Shortly afterwards (within picoseconds), the helical structure disappears almost entirely in high temperature simulations. The structural transition from Intermediate-2 to the near linear species is characterized by the increase in the distance between the C$\alpha$ of Lys11 and the C$\alpha$ of Lys33. When this distance increases beyond 21 Å, we consider the transition to the near linear structure to have taken place.

The results of this subsection can be summarized as follows. The simulations show that in MS conditions the initial, desolvated A structure evolves into a near linear structure. Two important intermediate states are formed, and at each conformational transition a significant element of the secondary structure is destroyed. The final elongated structure has virtually no elements of secondary structure.
5.4 Energy Barriers and Timescale Calculation

The computed trajectories yield timescales for the observed structural transitions, but these are not generally computed at the experimental temperatures. As pointed out in the previous section transformations such as Intermediate-1 to Intermediate-2 require very long trajectory simulations at the experimental temperature. We employed the RRK approximation as described in the methodology section, to compute the transition timescales. Using the MD trajectories at high energies (or temperatures) we computed the transition rate at these temperatures. In these calculations we used only the
trajectories that result in a transition over the specified timescale. Thus at 400K all 15 trajectories passed from the \textit{A state} to the Intermediate-1 structure within a half-life time of 1.7 ns. At 500K 3 out of 15 trajectories continued from the \textit{A-state} through Intermediate-1 to reach Intermediate-2. In this case the half-life time of the transition to Intermediate-1 was 251 ps. At 600K 9 trajectories out of the 15 computed ones went through Intermediate-1 and Intermediate-2 to reach the final elongated conformation. The formation lifetime of the elongated structure was 4.2 ns. Out of the 15 trajectories, 14 reached Intermediate-2 within 800 ps and all reached the first Intermediate within 76 ps. At each temperature level, the rate constant was calculated according to the first order reaction equation (equation 5.3),

\begin{equation}
k = \frac{\ln 2}{t_{1/2}} \tag{5.3}
\end{equation}

where \( t_{1/2} \) is the time in which half the simulations reached the next intermediate or final conformation. The temperature dependence of the results corresponds quite well to the Arrhenius equation. Figure 5.4a shows the plots of \( \ln(k) \) versus \( 1/RT \). Here \( k \) is the rate for any of the three structural transitions calculated, and the values plotted were obtained from the high temperature MD simulations. The energy barriers extracted from the slopes from figure 5.4a are as follows: the barrier for the transition from \textit{A state} to Intermediate-1 is 7.2 kcal/mol., the barrier for the transition from Intermediate-1 to Intermediate-2 is 11.3 kcal/mol., and the barrier for the transition from Intermediate-2 to the near linear structure is 15.7 kcal/mol. These barriers correspond to the simultaneous breaking of one
to three hydrogen bonds. Qualitatively, this is in agreement with the mechanism of the transitions as discussed in the previous subsection. It is useful to illustrate the transitions as revealed in the trajectories by considering the results presented in figure 5.4b. This figure shows an example of a trajectory computed at 600K for 5 ns. Curve A describes the distance between Ala46 and the C terminal. The transition between the A state and Intermediate-1 occurs after about 300 ps. Curve B describes the distance between Glu17 and the N terminal; here the transition between Intermediate-1 and Intermediate-2 is sharply defined and takes place around 800 ps. Another sharp transition is observed between Intermediate-2 and the final near linear structure. This transition occurs after approximately 4 ns. Obviously the transition times differ for the different trajectories but the values in figure 5.4b are typical. The total unfolding timescale in different temperatures are presented in figure 5.4c.
Figure 5.4: a) ln(k) versus $\frac{1}{RT}$ plot. The blue line is the plot of ln(k1) the transition rate from the initial to intermediate-1, the green line ln(k2) the transition rate towards intermediate-2, and the red line ln(k3) the transition to the final elongated conformation. b) The different distances which characterize the different intermediates along a 600K 5 ns trajectory. Curve A represents the distance between Ala46 to the C’, Curve B represents the distance between Glu17 to the N’, and curve C represents the distance between Lys11 and Lys33. c) The calculated timescale of the transition from the A state to the elongated unfolded state.

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>Unfolding Timescale [sec]</th>
</tr>
</thead>
<tbody>
<tr>
<td>200K</td>
<td>1.3E+3</td>
</tr>
<tr>
<td>300K</td>
<td>2.0E-3</td>
</tr>
<tr>
<td>400K</td>
<td>3.4E-6</td>
</tr>
<tr>
<td>500K</td>
<td>6.6E-8</td>
</tr>
<tr>
<td>600K</td>
<td>4.7E-9</td>
</tr>
</tbody>
</table>
5.5 High Temperature MD Results versus ECD Experiments

Electron-capture dissociation (ECD) represents one of the most recent and significant advancements in tandem mass spectroscopy for identification and characterization of post-translational modifications of polypeptides. In comparison with the conventional fragmentation techniques, such as collisionally-induced dissociation and infrared multi-photon dissociation[124], ECD provides more extensive sequence fragments, while allowing the labile folding modifications to remain intact during backbone fragmentation. This unique attribute qualifies ECD as an efficient tool for the detection and localization of secondary structures [125, 126]. In ECD experiments, multi-charged ions produced by electrospray ionization are trapped within the confines of a combination of magnetic and electrostatic fields of a mass spectrometer. Afterwards, the ions are irradiated with an electron beam, generated by an electron gun [91, 93]. The positively-charged residue, which is stabilized by the carbonyl group, is known from electronic structure considerations to have a high affinity to the incoming electron [127], leading to rapid (picosecond timescale) backbone dissociation [123]. This ECD cleavage process is much faster than intra-molecular energy randomization [125, 128], so the cleavage site becomes extensive and with no specificity for cleavage sites within the backbone [94, 129, 130]. The mechanism of the cleavage reaction is represented below:

\[
\begin{array}{c}
\text{O} \quad \text{F} \quad \text{N} \quad \text{H} \quad \text{N} \quad \text{H} \\
\text{C} \quad \text{N} \quad \text{H} \quad \text{C} \quad \text{H} \quad \text{R} \\
\text{C} \quad \text{N} \quad \text{H} \quad \text{C} \quad \text{H} \quad \text{R} \\
\text{C} \quad \text{N} \quad \text{H} \quad \text{C} \quad \text{H} \\
\end{array}
\]

Both reaction mechanisms are initiated when the charged amine group is in close proximity to the carbonyl backbone group. The model we used to assign a weight to each possible fragmentation pattern is based on the following assumptions: 1. Electrons are
assumed to be uniformly available across the molecule. 2. The cleavage probability for a
given conformation is assumed to take place when the protonated residue and carbonyl
group are suitably close. To predict the cleavage site pattern, a short timescale MD
simulation was carried out for the elongated configuration. Ten trajectories were
computed each for 5 nanoseconds’ duration at 600K.
The results were used to sample the conformational structures of the protein: based on the
above assumption the cleavage site pattern is proportional to the time interval over which
the distance between the carbonyl and the relevant protonated group is equal to or less
than a critical value of two angstrom. The above model is similar in essence to the
assumptions of an approach for predicting patterns of NECD in cytochrome C, which
produced satisfactory results [123].

The comparison between the calculated fragmentation pattern and the experimental data
[123] is shown in Figure 5.6. On the whole, agreement between theory and experiment is
satisfactory, and this supports the elongated structure to which the model was applied.
However, the agreement is not complete and deviations particularly occur at the 44-59
cleavage site, and more specifically at the 44-48 cleavage site. We note that this site is
strongly hydrophobic. Such regions have little affinity with the protonated residues, but it
is possible that once the electron attacks the protonated group can follow. Such a
correlation between proton supply and the electron attack is not present in our model and
it seems likely that this may explain the discrepancy with the experimental data. Most of
Figure 5.6, however, shows a systematic agreement between experiment and theory,
which indicates that the elongated structure fits well into the ECD fragmentation pattern.

45
Much of our current understanding of protein ions in the gas phase comes from ion mobility spectrometry. In this experiment protein ions traverse an environment of inert neutral gas under the influence of a weak electric field. In general, the mobility of a protein ion in the gas phase depends on its average collision cross-section with the buffer gas. Compact conformers have smaller cross-sections and thus higher mobilities than elongated ones [97-100]. Transit times for ions under these conditions can therefore be used to determine the shape and size of the gas phase proteins [101].

5.6 Comparison with cross-section experiments

Figure 5.6: A comparison between cleavage sites in the ECD experiments (in gray), and calculated cleavage sites (in black).
5.6.1 Cross-section calculations for ion drift experiments

In order to evaluate the relevance of the intermediate and final structures obtained in our simulations we calculated cross-sections of the structures for comparison with experimental data. Experimental collision cross-sections reported in the literature were obtained by measuring the ubiquitin ion mobility in helium using an electrospray time-of-flight mass spectrometer equipped with a drift cell [102-104, 131]. In these experiments the ion drift time, given by the helium–ion collision cross-section [89, 132] is measured under thermal conditions. Kinetic theory predicts that ion mobility experiments yield an orientation-averaged momentum transfer collision cross-section, which can be evaluated theoretically for a given structure and helium interaction potential [132]. The methods employed here use a Monte Carlo approach for integrating momentum transfer over all collision geometries and the thermal energy distribution [107, 133]. The helium deflection angles entering the calculation are evaluated by trajectory calculations [107, 133]. For large ions such as ubiquitin the exact shape of the interaction potential employed in the trajectory calculations has little effect on the computed cross-section. Our work indicates that for a system the size of ubiquitin, a model with hard spheres placed at the position of each atom (exact hard sphere model [107]) yields values in agreement (within <1% on average) with more sophisticated helium-protein potential interactions including Lennard-Jones and charge–induced dipole terms [107, 133]. Uncertainties about calculated cross-sections reported here are ±1% (two standard deviations of a sample consisting of repetitive calculations) given by the convergence criterion for the numerical integrations.
5.6.2 Comparison of the final structure with experimental data

Based on the above simulations ubiquitin +13 ions traveling for hundreds of microseconds through a heated ESI capillary (5 cm long by 0.05 mm$^2$ cross-section) are expected to partially or fully unfold depending on the experimental temperature. ESI-IMS experiments carried out in the Clemmer et. al. laboratory proceed on a timescale well above a millisecond and show that ubiquitin +13 ions have a cross-section of 2115 Å$^2$ for helium [102, 131]. Using a hard sphere scattering model [107] to estimate the collision cross-section of our final model structure yields a value of 2137 Å$^2$ which is consistent with experimental data (deviation of ~1%).

5.6.3 Comparison of intermediate states with experimental data

More sophisticated experimental methods including ion trap-IMS and IMS-IMS combinations [103] are able to follow structural changes occurring on the millisecond to seconds timescale. In such experiments it was found that structural transitions of +7 ubiquitin ions formed by ESI include unfolding from a compact initial structure via a partially folded state to a final elongated structure [103]. A more thorough study on +12 ions yielded cross-sections for stable intermediates in the unfolding process of 1767, 1845, and 1998 Å$^2$ and a final structure of 2072 Å$^2$ [104]. This result is in near quantitative agreement with our simulation of the +13 ions. We found the same number
of stable or metastable species (figure 5.3a-d) with calculated cross-sections of 1686, 1829, 1986, and 2137 Å². Thus, it is very likely that the +12 charge state of ubiquitin unfolds via very similar intermediates as those found here for +13 ubiquitin (a deviation of 4.5%, 0.8%, 0.6%, 3% respectively). These results indicate that the +12 ion may adopt at least two of the intermediate structures of the +13 ions. The ion mobility time-of-flight mass spectrometer enables us to record nested drift (flight) time distributions for complex mixtures in fractions of a second. Such a method can be used to follow and identify the conformational transitions of ubiquitin ions in the gas phase, the different intermediate stages and timescales [134].

5.7 Ubiquitin summary

In summary, the experimental cross-section is in striking accord with the cross-section computed for the near linear structure of ubiquitin +13. This leads to the conclusion that in the experimental conditions of Clemmer et al., unfolding proceeds to the final near linear structure. We have shown previously that at an initial temperature of 400K, complete unfolding to the near linear structure takes place within microseconds. The temperature in the Clemmer et al. experiment is not accurately known, but if we assume it to be 400K or more, then the timescale calculations of the present paper indicate that the protein is expected to undergo complete unfolding in the experimental conditions. The cross-section computed here for the ubiquitin +13 ion can be compared with the experimental cross-section of ubiquitin +12, and the results suggest that there may be a similarity between the corresponding structures. However, this must be treated with caution, since it is based entirely on comparison of cross-section data for the two ions. No
calculations for the intermediate and final stage of ubiquitin +12 were calculated here, and this is obviously an interesting challenge for the future.

6 The HMR method results

The complexity of the landscape and the limitation of classical MD simulation have led to a growing interest in developing novel methods for computing conformational changes in peptides and in proteins. Despite the considerable number of different methods proposed, the field is still open to a large extent and additional theoretical tools for this problem are desirable. Classical MD simulation methods were used to study amino acids and gradual MD simulation was applied to the study of conformational changes of ubiquitin ions in the gas phase with long timescales. In these studies the classical MD simulation was applied to simulate the whole trajectory, the surmounting of low as well as high energy barriers, and to sample the different reaction paths. Through the objective of studying the reaction paths and dynamics of long timescale evolution of conformational changes in small biological molecules the Hybrid Molecular Dynamics / RRK (HMR) algorithm was developed. The approach employs classical trajectories for transitions between adjacent structures separated by a low energy barrier, and the classical statistical RRK approximation when the barrier involved is high. In determining the long-time dynamics from an initial structure to a final structure of interest, an algorithm is introduced for determining the most efficient pathways (sequence of the intermediate conformers). This method uses the Dijkstra algorithm for finding optimal paths on networks. Three applications of the method are presented: a detailed study of conformational transitions in a blocked valine dipeptide; a multiple reaction path study of
the blocked valine tripeptide; and the evolution in time from the β hairpin to α helix structure of a blocked alanine hexapeptide. Advantages and limitations of the method are discussed in the light of the results. The results presented in this chapter were published as a cover paper (Figure 6.1) in the Physical Chemistry Chemical Physics (PCCP) journal [108].

Figure 6.1: The cover page of PCCP (November 2006) journal presenting the HMR method.

**6.1 HMR method overview**

In HMR the problem of conformational changes is modeled as a motion between two adjacent minima within a network of minima of the system, according to computed rates. The method involves two stages: the first stage is to construct a network of minima, and the second step deals with the dynamics of motion in the network. For the first stage, the
algorithm distinguishes between two types of energy barriers, which are treated differently. The first type of barriers is of low energy, and connects two minima. The crossing of these barriers is treated by MD simulations. In the algorithm, the low energy barriers are defined as those which can be surmounted by a short timescale MD simulation of 100 picoseconds. The second type of barrier is of high energy, in which case the transitions are treated by SPW as described in Chapter 4. The second stage of the method, which describes the dynamics of the network, includes two steps: first the transition time between any two connected neighboring minima is computed and then optimal paths are found between two minima of interest in the network by means of Dijkstra's algorithm.

In general, the HMR method can be used in two variants. The first variant finds the optimal path between two given conformations of interest, and the second variant starts from an initial state and allows a free evolution in time.

### 6.1.1 Construction of the network of minima

A set of elementary hopping events is determined for each pair of neighboring minima. Depending on the timescale involved, the transition time is computed by either a short MD trajectory or by a RRK calculation. This is decided by the timescale of the MD simulation required for the elementary transition. Technically the following procedure is applied:

1. For low energy barriers, a 100 picosecond MD simulation is performed:
   
   this takes advantage of the accuracy and efficiency of this method for
crossing low energy barriers. The MD package in MOIL is applied for this purpose [120]. For high-energy barriers, the transfer between conformations does not occur in this short simulation time. The timescale of 100 picoseconds has been chosen for the following reasons: first, in our test cases, the thermal equilibrium is achieved in this time (data not shown); second, this timescale is short enough to require a low computational effort.

2. As soon as the molecule fails to make the transition in the MD simulation within 100 picoseconds, hopping to neighboring minima based on the SPW method (SPW elementary hopping events) is considered. To find these neighboring minima, a single torsional angle is varied and geometry optimization is performed. Use of the method as described above will explore the whole (or at least a very large part) of the PES. Owing to the exponential growth of conformations with the number of rotatable bonds,[135] such a process will limit the method to small peptides I order. To avoid the rapid growth in the number of minima, a reduced set of torsional angles are used for the search for minima. This reduced set includes only the torsional angles that are different from the final known product conformation. As a result, the number of minima is reduced dramatically, the search is directed toward the known product conformation and the search is computationally realistic. If only the initial structure is known, a free propagation in time is carried out with an exhaustive search over all the rotatable bonds.
The search for the relevant minima along the path is carried out here over a limited set of dihedral angles, which is motivated by geometrical intuition. Comparisons with classical MD simulation show that in this search the relevant minima and optimal pathways are found (presented at 6.2.3 and 6.3.1). A range of interesting methods have been proposed recently for generating sets of minima connected by transition states. This includes the basin-hopping method of Wales and Scheraga [136-138]. The method used here is directly tied to conformational dynamics and tested by MD simulations. It thus combines computational efficiency with a close connection to trajectory dynamics.

### 6.1.2 Describing the dynamics on the network

From the MD, or SPW elementary hopping events described above, we calculate the set of transition rates between any pair of relevant adjacent minima. Thus the set of pathways on the network between any two minima of interest can be defined. The optimal path between these minima is evaluated by Dijkstra's algorithm[139].

The details are as follows: for low energy barriers the time it takes to move between two minima using an MD simulation is taken directly from the MD simulation. In the case of a high energy barrier, however, where the molecule is forced to “jump” to the neighboring minima, the time of moving between two minima is taken from the classical RRK calculations which supply a statistical method for computing rate constants, for a process at constant energy E (such as that determined by MD). The starting-point, using the classical RRK theory, is to identify the reactant and the involved transition state on
the potential energy surface. The reactant is the minimum of the located conformation, and the transition state structure is found using the SPW method as described above [64].

The SPW method [64], is used here to locate the transition state structure. In all of our test cases regarding the valine dipeptide and the valine tripeptide, the SPW method located a correct transition state structure. The SPW method is a fast method which usually calculates good approximation for transition state structures. In cases where the SPW is not accurate enough, the Conjugate Peak Refinement method can be used[140]. Methods which use Hessian and related matrices [78-80, 141] are very useful for small molecules, but in the case of a molecule with several tens of atoms the manipulation of these matrices is computationally demanding and they cannot be used effectively.

6.1.3 Finding the optimal reaction path

Dijkstra's algorithm[139] is an algorithm that solves the single-source shortest path problem for a directed network. Before we describe the algorithm, some other details are necessary. The network consists of edges which are lines that connect vertices which are dots. A directed network is a network where the edges are directed – that is, each edge has a direction following from one vertex to another vertex. We mapped our problem as a directed graph in which the minima are the vertices and each edge is an ordered weighted pair of vertices \((u,v)\) representing a connection from a minimum conformation \(u\) to a minimum \(v\); weights of the edges are the transition times between minimum \(u\) to minimum \(v\). Therefore \(w(u,v)\) is the non-negative weight of time moving directly from minimum \(u\) to minimum \(v\). The time of a path between two minima is the sum of times of
the edges along that path. The algorithm works by keeping track of the cost \( d[v] \) of the shortest path found so far between the initial minimum conformation \( s \) and conformational minimum \( v \). Initially, this \( (d[v]) \) value is 0 for the source vertex \( s \) \( (d[s]=0) \), and infinity for all other vertices. When the algorithm finishes, \( d[v] \) will be the cost of the shortest path from \( s \) to \( v \), or infinity, if no such path exists.

The basic operation of Dijkstra's algorithm is edge relaxation: if there is an edge from \( u \) to \( v \), then the shortest known path from \( s \) to \( u \) \( (d[u]) \) can be extended to a path from \( s \) to \( v \) by adding edge \( (u,v) \) at the end. This path will have the time \( d[u]+w(u,v) \). If this is less than the current \( d[v] \), we can replace the current value of \( d[v] \) with the new value. Edge relaxation is applied until all values \( d[v] \) represent the time of the shortest path from \( s \) to \( v \). An example of Dijkstra's algorithm process is presented step by step (Figure 6.2).

Figure 6.2: The execution of Dijkstra's algorithm. The source is the leftmost vertex. The shortest path estimates for each vertex are shown within the vertices, and the weight \( w(u,v) \) is written near the edges. a) The initial graph conditions situation just before the
first iteration. b) –e) the situation after each successive iteration. f) the final values, and the optimal path to each one of the vertices[142].

6.2 Valine dipeptide example

In order to compare the HMR method and MD simulation results, the valine dipeptide was chosen as a model system. The main advantage of the system is that the conformational changes in this molecule can be described by only three dihedral angles. In Figure 6.3a, a schematic representation of the molecule is presented. Figure 6.3b shows the potential energy surface in the space of the two dihedral angles $\psi, \chi$. In this map the energy was optimized with regard to all other coordinates. Although the valine dipeptide has 8 minima, only 5 of these appear on the two-dimensional contour plot of Figure 6.3b. The highest energy conformation (at -46.035 kcal/mol) was marked as "A" and was chosen to be the initial structure for the simulation. The energy levels of conformations B and C are -48.263 kcal/mol and -48.511 kcal/mol respectively, and the dihedral angles $\psi, \chi$ are: A(155, -75), B(61,-73), C(75,-171). These conformations are represented in a ball and stick representation and are displayed via the VMD program in Figure 6.3b [143].
Figure 6.3: a) Valine dipeptide. The three soft torsional angles $\phi, \psi, \chi$ are shown. b) A contour map of the potential energy of the valine dipeptide as a function of two soft torsions ($\psi, \chi$), the dihedral angles are given in degrees, the three minima of interest A, B and C are represented using the VMD program [143].

6.2.1 Valine dipeptide MD path

In order to learn more about the natural properties of the valine dipeptide dynamics, 100 quenching MD simulations were performed, starting at the A structure with an initial
temperature of 250 K. A Boltzmann distribution was used to initialize the velocities of
the simulation, under constant energy conditions. The equations of motion were
integrated with a time step of 1.0 fs. All van der Waals and electrostatic cutoffs were
included.

The MD trajectory analysis showed a very similar spatial reaction path in all of the tr
jectories: the system passes from A to B, spends a short time in B while equilibrating
with A, and then reaches to C. One of the reaction paths is presented in Figure 6.4a. For
simplification, only the last section of the trajectory is shown and the transitions between
A and B with quasi equilibration between the two are not given. Quenching MD
simulations were used to estimate the amount of time the molecule spends in each
minimum. In a quenching MD simulation, the trajectory is studied using recursive
structural optimizations at picosecond intervals. After each time interval of 1.0
picoseconds the energy is minimized starting from the configuration reached: this gives
the conformation that is nearest to the structure at that trajectory point. In Figure 6.4b, the
output of the simulation is presented. From the results presented in Figure 6.4b, the
transition from A to B is a rapid one. The high rate of the transition is attributed to the low
energy barrier in between A and B conformations (0.6 kcal/mol). The conformational
change is mainly in the backbone chain (\(\psi\) and \(\phi\) dihedral angles) and the half-life time of
the transition path, which is the time that it takes for half of the trajectories to pass from A
to B, is \(3.5 \times 10^{-11}\) seconds. The reverse reaction, from B to A, passes over a higher energy
barrier (2.84 kcal/mol) and its half-life time is \(1.27 \times 10^{-9}\) seconds.
The transition from $B$ to $C$ passes over a higher energy barrier (3.2 kcal/mol) compared with the barriers mentioned above, and, as a consequence, the path is confined to go through a more narrow region in the vicinity of the transition state. The restricted configuration space in the transition region is shown in Figure 6.4b.
Figure 6.4: a) Potential energy map for valine dipeptide in the $\psi, \chi$ angles. In black is the pathway along a specific MD trajectory showing a conformational transition between minima $A$ to $C$. Note path passes through $B$ conformation. b) The full quenching MD simulation result, starting at the $A$ conformation through the $A \leftrightarrow B$ quasi equilibrium and to the final $C$ conformation.
6.2.2 Validity of the RRK Approach

Consider now the differences between the computed RRK and MD conformational transition paths. The half life-time for the RRK approach was determined from the RRK rate constant (Chapter 4). The half-life time is given by

\[
\frac{T_{1/2}}{2} = \frac{\ln 2}{k}
\]

(6.5)

where \(T_{1/2}\) is the half-life time and \(k\) is the rate constant. The results are represented in Table 6.6. The combination of finding the transition state with SPW and the RRK approach was not accurate in cases of low energy barriers, specifically the passage from A to B conformation with a 0.6 kcal/mol energy barrier. In this example the ratio between the rates computed from RRK is lower by a factor of 10 compared with the MD simulation result. The RRK results are much better when the energy barrier is high, for example, in the transitions from B to A the 2.8 kcal/mol energy barrier leads to a half-life time of \(1.27 \times 10^{-9}\) seconds in the MD simulation and \(8.15 \times 10^{-10}\) seconds for the RRK approach. The timescale obtained from RRK is consistently lower than that from the MD simulation. There are two main reasons for the difference in timescale results: 1. The transition structure found using the SPW is consistently lower in energy than the real transition state structure. 2. The RRK method inherently overestimates transition rates, since it essentially ignores recrossing effects.

In summary, in cases of low barriers the RRK results can be in serious error so this approximation should not be used: for high barriers the method is more reliable and much more efficient then MD.
Table 6.6: Half-life times calculated from the MD simulation and from the RRK approach

<table>
<thead>
<tr>
<th></th>
<th>Dynamics $T_{1/2}$</th>
<th>RRK $T_{1/2}$</th>
<th>Ratio (dynamics/RRK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A \rightarrow B$</td>
<td>$3.5\times10^{-11}$ sec</td>
<td>$4.96\times10^{-12}$ sec</td>
<td>10.17</td>
</tr>
<tr>
<td>$E_0=0.611$ kcal/mol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B \rightarrow A$</td>
<td>$1.27\times10^{-9}$ sec</td>
<td>$8.15\times10^{-10}$ sec</td>
<td>2.24</td>
</tr>
<tr>
<td>$E_0=2.84$ kcal/mol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B \rightarrow C$</td>
<td>$1.25\times10^{-8}$ sec</td>
<td>$4.43\times10^{-9}$ sec</td>
<td>2.84</td>
</tr>
<tr>
<td>$E_0=3.2$ kcal/mol</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.2.3 HMR method versus MD reaction pathway

In order to evaluate the accuracy of the path obtaining from the HMR method, we consider the evolution from initial state $A$ to final state $C$. The optimal reaction path found by the HMR was the one passing through conformation $B$ as shown in Figure 6.7. It should be kept in mind that the contour plot represents only two dimensions of the potential energy surface, and the entire potential energy surface exhibits eight minima, and as a consequence there are several possible reaction paths. The trajectory found by the HMR method is the same as the one computed by the MD simulation.

The HMR method computes the overall time of a conformational change in terms of transitions between adjacent minima. In applications of HMR, some contributions come from RRK while others come from MD. In the present example, it is encouraging that the total HMR rate is of reasonable accuracy, but more important is the fact that the HMR
method correctly describes the dynamical evolution in time, and it predicts the correct status of the system at any time of the process.

Figure 6.7: The optimal path calculated from the HMR method. The two parts of the path are: the MD trajectory marked as MD part, and the hopping above a high energy barrier using the SPW method marked as “jump”.

6.2.4 Run-time comparison

It should be emphasized that the full computational power of the HMR method can be seen for a complex system for which full MD is not at all feasible. In this case, we had to choose a simple system for the purpose of comparison: nevertheless, even for this simple system the time advantage of the HMR method is substantial.
When we compared the run-time of the HMR method and the MD simulation on the same computer (2.6GHz Pentium 4 processor computer), the HMR method run-time was faster by a factor of 56 (18 min for a single MD simulation compared with only 19 sec for the HMR method).

6.2.5 Summary of blocked valine dipeptide results: accuracy of path, timescale, reduced run-time

The valine dipeptide was a test model. It was relatively easy to compare the MD simulation results with the HMR method results. By this comparison it was shown that the HMR method is able to find accurate reaction paths, as well as timescales, with a reduced CPU run-time.

6.3 The blocked valine tripeptide - finding multiple paths

The valine dipeptide molecule represents a simple system for which a single significant reaction path was determined from dynamics. In this section, we consider an example for which several reaction paths were determined. The valine tripeptide has six soft degrees of freedom (six dihedral angles). The six relevant degrees of freedom are shown in Figure 6.8. In this section, the results of the HMR method are compared with the MD simulation trajectories.
In order to give us some idea of the initial and final structures with which to work, a high temperature quenching MD simulation was performed. From the MD simulation, the global minimum was found to have a potential energy of -82.360 kcal/mol and was named the “final conformation”. A higher energy conformation with a potential energy of -80.405 kcal/mol was therefore chosen to be the initial structure, and named the “initial conformation”. The two conformations differ mostly in their C terminal amino acid residue torsion angle, marked as $\chi_c$ in Figure 6.8. This torsional angle differs from 59 degrees in the initial state to -72 degrees in the final state.

### 6.3.1 Valine dipeptide MD and HMR paths

One hundred quenching MD trajectories of the blocked valine tripeptide were computed. All of the trajectories were calculated with the same parameters as the blocked valine dipeptide simulation, except that the initial temperature was doubled to 500K. The reason for the doubling of the initial temperature was to accelerate the crossing of the energy
barriers. The simulations were calculated for 20 nanoseconds, and roughly 20% of the simulations reached the final (global) conformation.

The minimizations during the quenching MD simulation were performed every pico-second. Two main paths connecting the initial and the final structures were found. The first path, named $P_1$ (path 1), is a direct path. Along this path no significant changes in the dihedral angles of the N terminal amino acid were found, and all of the changes were at the C terminal dihedral angles. A quenching MD simulation that is the result of this path is presented at Figure 6.9a. The second path, named as the $P_2$ (path 2) is an indirect path, which includes changes in the N terminus amino acid residue torsion angle, as well as in the C terminal torsion angles. The quenching MD simulation which resulted from this path is presented at Figure 6.9b. Both the direct and indirect paths are more complex than the single path of the blocked valine dipeptide. Each of the paths here includes several local minima that may play a role as intermediate states. There are also several different quasi equilibrium conformational changes, for example, as seen in the direct path quenching MD simulation, shown in figure 6.9a, the quasi equilibrium transition between the initial conformation with conformation $P_1.a$ can be seen. In the indirect path result in Figure 6.9b, the quasi equilibrium transition of $P_2.b \leftrightarrow P_2.b'$ is observed. This large number of conformations and fast transition conformational changes makes each path slightly different from the others. In order to characterize each of the paths, all of the trajectories were divided into two main groups: the $P_1$ (direct) and the $P_2$ (indirect) paths. Afterwards, the common intermediate conformations along the path of
every group of paths was found. In this way, we found the set of conformations that are essential for describing the path.

Thus the direct path (P1) follows the sequence of minima: initial conformation → P1.a → P1.b → final conformation

The indirect path (P2) follows the sequence of minima: initial conformation → P2.a → P2.b → P2.c → final conformation

The HMR was then compared with the MD simulation. After a network connecting the initial and the final conformation was found, Dijkstra's algorithm found the optimal path to be the following P3 path: initial conformation → P2.a → P2.b' → P2.c' → final conformation. This path starts in a similar manner to that of the P2 indirect path. After the second intermediate conformation, however, it seems to deviate from the indirect path. Nonetheless, a deeper look into the quenching MD simulation results (seen in Figure 6.9b) showed that P2.b' is an intermediate conformation with a fast transition to the P2.b conformation. One hundred MD simulations were performed in order to evaluate the similarity between these two conformations (P2.b' and P2.b). All MD simulations started at the P2.b' conformation, and the average time for each simulation to reach P2.b was 1.6 picoseconds. The same procedure was later performed in order to evaluate the transition time between P2.c' to P2.c. For this transition the time was less than 1.0 picoseconds. The first path found by the HMR method (P3) is therefore similar to the indirect path found by the quenching MD simulation, except that this path is passing through a higher energy intermediate conformation. This high energy path (P3) can be
easily understood, because energy barriers between the higher energetic intermediates conformations are lower in energy.

In order to calculate an additional path, the minima of $P3$ path were removed from the network, and the Dijkstra algorithm was operated once, more revealing path $P4$: initial conformation $\rightarrow P1.a \rightarrow P1.b \rightarrow$ final conformation, which is exactly the same as the $P1$ direct path found by the MD simulation.
Figure 6.9: a) A single quenching MD simulation part representing the direct ($P1$) reaction path. b) A single quenching MD simulation part representing the indirect ($P2$) reaction path.
6.3.2 Run-time comparison

In one sense the MD simulation is a bit wasteful, because 80% of the trajectories did not reach the final conformation within the time of the computation. The run-time of the HMR method, however, was only about 30 minutes, compared with a single MD of 20 nanosecond trajectory requiring about 9 hours of computer run-time.

6.3.3 Summary blocked valine tripeptide – multiple path run-time

Following section 6.2, where we found that the HMR method accurately calculates a single reaction path and transition timescales, the valine tripeptide was used to assess whether the HMR method has the capability to find multiple reaction paths. The HMR method found the two reaction paths established by an MD simulation. The first path calculated by the HMR method differed from the MD simulation path owing to its passing through higher energy intermediates conformations. It should be borne in mind that some of the MD trajectories pass through these “high energy” intermediates. Note that in the MD simulation any of the minima which do not belong to the main path are populated only for an extremely short time interval, and from there the system quickly moves to the main path configuration. The second path calculated by the HMR method was similar to the MD simulation path. I note that the method used here is actually a variant of the Dijkstra’s algorithm which made possible the determination of more of a single pathway.
6.4 Blocked alanine hexapeptide conformational change

The alanine hexapeptide example poses a serious challenge for calculations of conformational transition problems. At the outset we noted that in this example the alanine hexapeptide undergoes a transformation from the β hairpin to a α helix-like conformation. We did not carry out MD simulations for this example, since such calculations are computationally quite demanding.

On the other hand the HMR method was applied and proved to be very successful: the network constructed by the HMR method included 170 different minima. In order for the network to be built, the elementary hopping events (jumps) were strictly limited to torsional angles that differed from the helix structure by a minimum of 15 degrees. The optimal path between the conformations of interest was found to pass through three intermediate minima, as shown in Figure 6.10.

It was important to investigate whether the HMR method in this case really obtained the most important reaction path, and to ensure that the reduced set of minima was sufficient to describe it. This was tested in the following way: a search of a larger portion of the PES was performed. The search was conducted using the HMR method with no limitation on the torsional angle search, and it resulted in a network containing 2812 minima. On this network the optimal path was found to be the one similar to that found by the reduced set of torsional angles. While the test we carried out is not rigorous or complete, the similarity found between the paths supports the assumption that the HMR method, even though it searched over a reduced set of minima, was indeed able to identify and determine the most important reaction path.
Figure 6.10: The optimal reaction explored by the HMR method between the initial \(\beta\) hairpin to the \(\alpha\) helix conformation. Dashed grey lines represent hydrogen bonds.
6.5 Phenylalanine conformational changes - HMR method and experiment

In order to emphasize the advantage and the possible applications of the HMR method the phenylalanine molecule in the gas phase, which was extensively studied in spectroscopic experiments and high level ab initio tools [144, 145], was chosen. To predict properties of phenylalanine including structural transitions between different conformers, it is an advantage to construct a global PES: such a potential is obviously expected to describe correctly the structures of at least the important different conformers and the barriers that separate them. Ab initio methods clearly predict quite accurately the structures of the conformers, and also they yield fundamental vibrational frequencies in good accord with the spectroscopic data [144]. There is therefore strong evidence for the precision of the ab initio methods. On the other hand the ab initio calculations cannot be readily used for dynamics calculations, or be applied for other purposes that require a global, computational efficient PES. We therefore developed a simple AMBER-type force field that can conveniently serve to analyze properties of the PES. While the standard AMBER force field parameters are not adequate to be reproduced in the PES expected for such a system, AMBER parameters were fitted to HF-631G* calculation for the global minimum conformer of phenylalanine. This led to determination of partial charges, equilibrium bond distances and angles. The fitted AMBER-type force field produced in this way was used to predict several other conformers. Special efforts were carried out on analyzing this model in the conformational space in which intermolecular hydrogen bond O-H·····N stabilizes the molecule (as at the global minima). In this region the three conformers as named in Meijer’s article [144], I, III, IV, were identified. The geometrical
comparison between the ab initio conformations to the force field conformation is presented in Figure 6.11. In this comparison, the geometrical similarity between the ab initio and the force field models is clearly observed.

![Figure 6.11: The three different conformations of phenylalanine which are in the scope of this section: each conformation is represented in cyan colors for carbons (in the left of each square) for the force field conformation representation, and in orange colors for carbons (in the right of each square) for the ab initio conformation representation. The global minimum is masked as I, with a relative ab initio energy of 0 cm$^{-1}$, and the two other conformations are marked as III, and IV with a relative energy of 177 cm$^{-1}$ and 336 cm$^{-1}$ respectively. The experimental data and ab initio calculation demonstrate that the transition between conformation I to conformation IV passes through conformation III. To give an understanding of this transition an analytical contour plot of the PES of phenylalanine as a function of two dihedral angles was calculated (Figure 6.12). This contour plot displays the energetic cause for the transition through intermediate III.](image)
Figure 6.12: Phenylalanine PES contour plot in respect of the two dihedral angles (N-Cα-C-CH2 and H-N-Cα-CH2) and the three minima (I, III, IV) appear on it.

It has to be borne in mind that the contour plot can and does represent only two dimensions of the PES. In order to search for the real optimal reaction path the HMR method was applied. Using the HMR method the optimal reaction path was found to pass through conformation III as expected from observation of the analytical potential (Figure 6.12).

Moreover the relative energies for these three conformations were calculated and found to be in good accord with the ab initio calculations as presented in Table 6.13 (less than 1% difference in the calculated energy). A more stringent test of the quality of the fitted analytical potential and the HMR method is the comparison of heights of the conformational transitions barriers between the classical model and the ab initio
calculations. The results are shown in Table 6.13, and the comparison both for barrier heights and the different minimum energy is satisfactory.

<table>
<thead>
<tr>
<th>Conformation</th>
<th>Ab initio energy</th>
<th>Force field energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0 cm(^{-1})</td>
<td>0 cm(^{-1})</td>
</tr>
<tr>
<td>Transition state I-III</td>
<td>1725 cm(^{-1})</td>
<td>1714 cm(^{-1})</td>
</tr>
<tr>
<td>III</td>
<td>177 cm(^{-1})</td>
<td>177 cm(^{-1})</td>
</tr>
<tr>
<td>Transition state III-IV</td>
<td>386 cm(^{-1})</td>
<td>399 cm(^{-1})</td>
</tr>
<tr>
<td>IV</td>
<td>336 cm(^{-1})</td>
<td>329 cm(^{-1})</td>
</tr>
</tbody>
</table>

Table 6.13: An energetic comparison of the ab initio and classical energetic calculation based on the HMR method.

Based on the above it seems that the force field which can only be accepted to be approximate is nevertheless a faithful representation of the true PES, as reflected in ab initio calculations of equilibrium forms and transition barriers. The use of fitted potentials combined with the HMR method is of course very convenient for computational uses and extensive dynamics simulations, with a good comparison for experiment.

### 6.6 The HMR method summary

The *HMR* method for finding important pathways of conformational changes of biological molecules was introduced and tested for several examples. The method considers
elementary transitions between adjacent local minima that may participate in conformational dynamics from an initial to a final state of interest. The crossing of low barriers is treated by MD, while transitions across high barriers are described by the statistical RRK method. Information from rates of elementary hopping events between neighboring minima is used in an algorithm that treats the dynamics for important pathways of the network. The examples considered include comparisons with MD simulations for simple systems where full trajectories can easily be carried out. The results of these examples suggest that HMR is indeed able to identify the main reaction path, and therefore yields a correct description of the overall process. A further application, which is computationally more challenging, was presented in the case of the β hairpin to α helix transition in alanine hexapeptide. In this case comparison with MD simulation was not presented, as it seems that the latter is computationally quite difficult. A limited test suggests, however, that in this case also the correct path for this process was found. Finally an extensive comparison of HMR results with experimental data and ab-initio calculations presents the capability of this method in contributing to the understanding and prediction of spectroscopic results.

We find it very useful that, in the cases studied, a treatment based on a few dominant transition pathways between initial and final states is successful. Systems of different behavior probably exist but it is very useful both computationally and for physical insight efficiency to have such a treatment for systems as presented here. The HMR method has a number of advantages, in addition to computational efficiency, at least for small biological molecules: it provides a clear geometric and dynamical
description of the process. It also yields a time estimate for the overall conformational change. Many open questions remain. The method depends on determination of the local minima that may be involved, and for large biological molecules this is a difficult task. Nevertheless, it seems that for small or even intermediate biological molecules the method is quite promising, in the sense of finding the relevant intermediates, the correct energy barriers, the optimal reaction paths, and the timescale involved in the process.
7 Discussion and Conclusion

In this work two new methods for solving the conformational changes of biological molecules problems are described, a gradual MD simulation method and the Hybrid MD/RRK method. The conformational changes problem in biological molecules is a complicated problem which has been studied for decades [146]. None of the methods developed in the field, including those presented by us, can give a full solution to this complex problem. Both of our methods are, however, a step further in understanding the conformational changes in biological molecules. Owing to the importance we see in a tight comparison and collaboration with experiment, the theoretical results using both of the methods are compared successfully with experimental data.

The first method presented in this work is the Gradual Temperature MD simulation: gradual increasing temperature MD simulations were applied to ubiquitin +13 ions in the gas phase. These simulations identified the final structure which has experimental evidence support-based on data from ECD fragmentation [123] and cross-section experiments [131]. These simulations present new stable and important intermediates which are involved in the transition conformation of ubiquitin +13 ions in the gas phase. These intermediates do not have any support in ubiquitin +13 ions experiments: they do, however, have a high similarity to +12 ubiquitin ions cross-section results, which might indicate that the +12 ions adopt as stable conformations the intermediate conformations of the +13 ions.

A major challenge for experiment is to address the intermediate structures and perhaps also to determine the timescale of the unfolding process. These results were presented to
Prof. M. Bowers of the University of California, Santa Barbara, which decided to contribute to this work from the spectroscopic point of view.

The second method presented in this thesis is the Hybrid MD/RRK. Owing to the complexity and the different applications which can be used in this method, it was first essential to compare the results with computational methods. By comparing valine dipeptide HMR results with classical MD simulation results it was proven that the HMR method enables one to identify a correct optimal reaction path while calculating the correct transition timescales. The valine tripeptide example demonstrated the ability of the method in finding several accurate reaction paths, and the valine hexapeptide demonstrated the capability of the method in overcoming the classical MD simulation timescales limitation. Finally the method was successfully applied to an experimental system revealing the correct relevant conformations, reaction path and energy barriers of the different conformational changes in phenylalanine molecule in the gas phase.

The HMR method was found to be applicable for small and intermediate biological molecules, allowing a good and fast exploration of the PES conformations, identifying optimal reaction paths and the intermediates involved in the path. These calculations, in addition to the correct energy barriers calculation and timescales, present a fast and accurate tool enabling theoretical research on the characteristics of conformational changes of biological molecules, as well as a tool for predicting and giving atomic resolution insights for spectroscopic experiments.
8 References


122. SYBYL 7.0. 2004, Tripos Inc., : 1699 South Hanley Rd., St Louis, Missouri 63144, USA.


שינווי מבנה מולקולות ביולוגיות:
שיטות חישוב חדשות והשוואה בין סיווג

היבור לעevity קבלת תואר דוקטור לפילוסופיה
נואת
אלעד שבב

הונג לסיציון האוניברסיטה העברית, בירודס
אוקטובר 2007
סיןוני מבנה בмолקולות ביולוגיות: השיטות החדשות והשוואה להניכים

הброchure לפני הшение הפורמלי דוקטור לפילוסופיה
מאת
אלעד שגב

הוגש לישראל האוניברסיטת העברית, בירושלים, אוקטובר 2007
עבודה זו נועשת בהדרכת של פרופסור רוברט בני גרבר.
olt תקציר

ולמולקולות ביוולוגיות וכן בתוכא והזמחים מתקיימים מעין יחסים, הממקים יסודות ניוסי טים

שימו כך ermögית בסקוטרופקופיה וסקוטרופקופיית מסות,ornaכתביוווקייפיתתואיתבῐ

המזרחי בחר את הקביעה המולקולות עם אחראי Şיוויניו מונחים בפאזה הจ้าה.

ובобще הוא עוזב עביה בשניים מספרים של מולקולות ביוולוגיות בפ.PERMISSION

 MOZGوح של שיטות הידרוגぶり ודוחות ולאש יזוהה בinstead ציטוטים זמנים, מוגנות:ahashe,

ידגניקת פרתוח, נראו שمائית בלגיית השיוווקים המבSerializedName של כן -13 ישלחולב ביבקטי

בנוסף סקוטרופקופיית מסות, והשלכה, הטבלה ידגניקת מולקולריות וגישה

המרכזי בתחום הוא קביעה של מבנה המולקולות בפ.A וobserver והזמחים.

שבิตה והידרוגぶり המולקולריות הפרתוחים כיד להטבר על מספרים של�ımız

ששות הדגניקת המולקולריות הפרתוחים מתוחזב כיד להבבר על מספרים של הידNonQuery.

משלב סקוטרופקופיה של ידגניקת מולקולריות ברז סקוטרופקופיות מוסת. וגשה ואמסברת שאר

מגני היבניםдумалים של ניוזי לרהי התקהל התואר (unfolding) החלבון. קביעה הים של השיוווקים

המגביני בחזרה החלבון מחושב בשיטת סטטיסטים המובсадת על מושאות

 Aussie ראיות באלקטרו ספני bohbת מפלעה (cross section) ששים על כל כייבקו13 ו bags שאיבד

At המבנה ששונן והמגבירים של בניימון סקוטרופקופיות מוסת. עבד כייבקו13 ו bags מעתי

بيقיני כי מבנה המגביני לניוזי השופת. מ宮סי הארגונית עבר התוקלד נמצאת בחר של

40K מכל קילו קרחית למול, זומן של 2 מוליברנתות וtextfield עזר שני המבנה -13.36 K

עד 15 קילו קרחית למול, זומן של 2 מוליברנתות וtextfield עזר שני המבנה -13.36 K

سبقונה עבר יסודי פורק אללקטרוני (ECD ) נתן ושואנה טובח

לתסוי, והתאונה גוות תומך לזכר של מבנה הסופת של ייבקו13 ו bags שתחבל

בסקוטרופקופיה. בוחר על כי, וחישוב ח腿部 הפועלת על מבנה הסופת מנוף מבנים מפורמים על ימכ

הבולנים של ייבקו13 ו bags עם ברי ייבקו13 +13.36 K בכל המבנה נatonin המבנה -13.36 K

ותיביוס באמף ניסיון.
The integration of a new method with the existing framework for data analysis can provide insights into the understanding of the results. It is reasonable to assume that this method could be used for the study of structures and the behavior of proteins in the gas phase, as well as changes in the structure of proteins in this state.

The combination of molecular dynamics and RRK describes the dynamics of the method that integrates short paths calculated in biological molecules and classical changes in structures for moving between close structures separated by low energy barriers, as well as the use of RRK approximation for calculating optimal paths for moving between barriers.

Three applications were studied in this method, which were applied on networks and finally, the ability of the algorithm to identify the number of reaction pathways in the peptide, peptide... Results: The transition from the structure of alpha helix to the structure of beta... changes in the structure of elastin in the studied. This method is very promising in terms of the use of the method for the study of changes in structure, the adaptation of the method for use in larger molecules presents a future challenge. In biological molecules...
## תוכן עולימיים

<table>
<thead>
<tr>
<th>מספר</th>
<th>מטרות</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>תכנייר</td>
</tr>
<tr>
<td>7</td>
<td>תוכן עולימיים</td>
</tr>
<tr>
<td>10</td>
<td>רקע</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>מספר</th>
<th>מובא</th>
<th>מודע ל램וד הלכלוכים לפאזה הגזית</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>3.1</td>
<td>השמש לבש עליה</td>
</tr>
<tr>
<td>14</td>
<td>3.2</td>
<td>דרגמן מפלס קלאסי</td>
</tr>
<tr>
<td>16</td>
<td>3.3</td>
<td>שטוח מתוכנות של דרגמן מפלס קלאסי</td>
</tr>
<tr>
<td>17</td>
<td>3.4</td>
<td>דרגמן מפלס קלאסי בסטמייינות</td>
</tr>
<tr>
<td>18</td>
<td>3.5</td>
<td>שילוב בין דרגמן מפלס קלאסי</td>
</tr>
<tr>
<td>19</td>
<td>3.6</td>
<td>ביצת שטח קרבה</td>
</tr>
<tr>
<td>20</td>
<td>3.7</td>
<td>סרוכת השטח הפוטנציאלי</td>
</tr>
<tr>
<td>22</td>
<td>3.8</td>
<td>הפרת תיווה ו</td>
</tr>
<tr>
<td>23</td>
<td>3.9</td>
<td>מכוון התיהו</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>מספר</th>
<th>שיטות</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>4.1</td>
</tr>
<tr>
<td>26</td>
<td>4.2</td>
</tr>
<tr>
<td>29</td>
<td>4.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>מספר</th>
<th>תיאוריה ונסיון יונים של חלבונים בפאזה הגזית</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>5.1</td>
</tr>
<tr>
<td>34</td>
<td>5.2</td>
</tr>
<tr>
<td>36</td>
<td>5.3</td>
</tr>
<tr>
<td>40</td>
<td>5.4</td>
</tr>
<tr>
<td>44</td>
<td>5.5</td>
</tr>
<tr>
<td>46</td>
<td>5.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>מספר</th>
<th>תכנייר</th>
<th>חישובי</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>5.6.1</td>
<td>Cross section</td>
</tr>
<tr>
<td>48</td>
<td>5.6.2</td>
<td>חישובי</td>
</tr>
<tr>
<td>48</td>
<td>5.6.3</td>
<td>חישובי</td>
</tr>
</tbody>
</table>
97

50 סיכום ייבוקית

HMR - תקציר על שיטה

51 ביניית רשת המבנימים הצייבים

52 תיאור הדינמיקה על הרשת

54 מיצאת השלבים האופטימאליים על הרשת

57 ואלין דו פפטיד

58 ריכוז מטוקולורית של ואלין דו פפטיד

62 בחרת קירו

63 חישוב ובו הדינמיקה של ואלין דו פפטיד

64 חישוב ובו הרשת

65 סיכום ואלין דו פפטיד

66 ואלין טרי פפטיד

66.1 מסלול דיאמיוניק מטוקולורית

66.2.1 מתא לאולן דו פפטיד

66.3 של אולן דו פפטיד

67.3.1 של אולן דו פפטיד

67.3.3 מתא לאולן טרי פפטיד

67.4 ואלו הפפטיד

68 ואלו הפפטיד

68.1 ואלו הפפטיד

68.2 שאזור נפגעי מבנים בגלי אוולן דו פפטיד

69 ואלו הפפטיד

70 ואלו הפפטיד

80 ציון מסכנת

82 ביבליוגרפיה