

# **Role of electrostatics explored with molecular dynamics simulations for protein stability and folding**

Dissertation zur Erlangung des akademischen Grades des  
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie  
der Freien Universität Berlin

vorgelegt von  
DIPL. PHYS. GIULIA MORRA

*Berlin, Juli 2005*

1. Gutachter: Prof. Dr. E. W. Knapp, Freie Universität Berlin

2. Gutachter: Prof. Dr. C. Frömmel, Charité Berlin

Disputation am 11. November 2005

## **Acknowledgments**

I arrived in Berlin on one of those typical grey and cold February days, after leaving the early italian spring. Despite the thermal shock, I found a warm and pleasant working atmosphere in the AG Knapp. My deep thanks go to my supervisor Prof. Ernst-Walter Knapp, who has always been present with useful discussions, careful in addressing my issues and questions and, last but not least, patient and kind when I used to take my little daughter along and put her in a cradle on my desk. All group members friendly helped me in the daily fight against CHARMM and Linux. Therefore, I want to thank Dr. Peter Vagedes, Dr. Dragan Popovic and Dr. Björn Rabenstein for introducing me to the world of protein computations, Daniel Winkelmann and Björn Kolbeck for saving me many times from computer troubles. Dr. Milan Hodoscek boosted my research with his precious help. A special thank goes to my officemate Hiroshi Ishikita for many interesting discussions, challenging collaboration and nice coffee breaks. Also I would like to thank Gernot Kieseritzky for the pleasant and fruitful collaboration. I am grateful to my italian coworkers, headed by Dr. Luciano Milanesi, who gave me the opportunity to finish my thesis.

Many people outside made this work possible: my parents, who are at my side wherever I am; my husband, who always supports me and helps me to see beyond bad days; my daughter, whose love gives sense to all of it.

# Contents

<b>Abstract</b>	<b>1</b>
<b>Zusammenfassung</b>	<b>3</b>
<b>1 Protein folding and stability</b>	<b>5</b>
1.1 The protein folding problem . . . . .	5
1.2 Native and denatured state . . . . .	6
1.3 Thermodynamics of protein folding . . . . .	7
1.3.1 The unfolding transition . . . . .	8
1.4 Folding kinetics . . . . .	10
1.5 Transition state and $\Phi$ -value analysis . . . . .	11
1.6 Folding mechanisms . . . . .	15
1.7 Energy landscape and funnel theory . . . . .	16
<b>2 Molecular dynamics of proteins</b>	<b>18</b>
2.1 Introduction . . . . .	18
2.2 Ensemble and time averages . . . . .	19
2.3 Molecular dynamics of biomolecules . . . . .	20
2.3.1 Force field . . . . .	20
2.3.2 Treatment of long range interactions . . . . .	22
2.3.3 Boundary conditions . . . . .	23
2.4 Integrators: Verlet, leap-frog, velocity Verlet . . . . .	24
2.5 Set up of a (N, V, E) protein simulation . . . . .	25
2.6 Temperature control . . . . .	27
2.7 Stochastic molecular dynamics . . . . .	29
2.8 The origins of CHARMM . . . . .	30
<b>3 Implicit solvent models</b>	<b>32</b>
3.1 Introduction . . . . .	32
3.2 Poisson-Boltzmann equation . . . . .	33
3.3 Numerical solution of the LPBE and applications . . . . .	35

3.4	Generalized Born approximation . . . . .	37
3.5	EEF1 energy function . . . . .	41
3.6	Statistical mechanics of a solvated protein . . . . .	43
3.6.1	Native state and thermodynamic stability . . . . .	45
<b>4</b>	<b>Unfolding CspB by means of biased molecular dynamics</b>	<b>47</b>
4.1	Introduction . . . . .	47
4.2	Methods . . . . .	48
4.2.1	Biased unfolding . . . . .	49
4.2.2	Contact analysis . . . . .	50
4.3	Results and discussion . . . . .	51
4.3.1	Wild type protein: biased unfolding at room temperature . . . . .	52
4.3.1.1	Early phase of unfolding . . . . .	52
4.3.1.2	Later phase of unfolding . . . . .	53
4.3.2	Simulation of unfolding at elevated temperature . . . . .	55
4.3.3	Biased unfolding of mutants . . . . .	56
4.3.4	Energetics of unfolding for the wild type and mutant proteins . . . . .	58
4.4	Analysis of unfolding kinetics and energy barriers . . . . .	60
4.5	Conclusions . . . . .	60
<b>5</b>	<b>Study of the stability of an artificial ion channel</b>	<b>64</b>
5.1	Introduction . . . . .	64
5.2	Materials and Methods . . . . .	65
5.2.1	Structures . . . . .	65
5.2.2	General set-up of molecular dynamics . . . . .	67
5.2.3	Generating Cs <sup>+</sup> ion coordinates in the channel . . . . .	67
5.2.4	MD simulations with implicit solvent . . . . .	69
5.2.5	Analysis of MD simulation data . . . . .	70
5.3	Results . . . . .	71
5.3.1	Monomer and dimer structure in absence of ions . . . . .	71
5.3.2	Stabilizing effect of two bound Cs <sup>+</sup> ions . . . . .	72
5.3.3	Simulation of monomer structure in presence of explicit ions: Cs <sup>+</sup> ion binding . . . . .	75
5.4	Discussion . . . . .	76
5.4.1	Comparison between monomer and dimer structure . . . . .	76
5.4.2	Cs <sup>+</sup> ion binding reaction . . . . .	78
5.5	Conclusions . . . . .	79

<b>6</b>	<b>Hydrogen bonding pattern of cytochrome c</b>	<b>80</b>
6.1	Introduction . . . . .	80
6.2	Materials and methods . . . . .	81
6.2.1	Structures . . . . .	81
6.2.2	Molecular modeling and dynamics . . . . .	82
6.3	Results . . . . .	84
6.3.1	RMS fluctuations of the Bcytc backbone . . . . .	85
6.3.2	Hydrogen bond formation and disruption . . . . .	86
6.3.3	Hydrogen bond distance fluctuations . . . . .	88
6.4	Discussion . . . . .	89
6.4.1	Fluctuations and stability of secondary structure . . . . .	89
6.4.2	Heme neighborhood . . . . .	91
6.4.3	Comparison with backbone NH exchange experiments . . . . .	93
6.5	Conclusions . . . . .	95
	<b>Bibliography</b>	<b>96</b>

# Abstract

The study of stability and folding properties of proteins has been for twenty years one of the major fields of research in biological sciences and has been recently enhanced by the increasing amount of genomic data. Understanding the mechanisms which ensure stability and function of a natural or a designed protein is the keypoint for the comprehension of basic life processes and development of new drugs. Nowadays, a theory, which predicts from any given coding sequence how the corresponding protein reaches and maintains its native structure, is still lacking. Nevertheless, for a considerable number of proteins a reasonably complete picture has become available, thanks to the development of novel experimental techniques and theoretical analysis tools (chapter 1).

The molecular dynamics approach to the study of protein structure and function plays an important role in biocomputing (chapter 2). Given the current computing resources, molecular systems with many atoms, like protein complexes in water, can be simulated, providing a useful tool for verification and in some cases prediction of experimental results. However, the most evident limit of molecular dynamics simulations, besides the dependence on the accuracy of force fields and the purely classical nature, is the size, in terms of computing time and memory load, of the simulations. A nanosecond to microsecond time span is usually reached in current simulations, which is however still too short to observe major structural changes like folding or unfolding of proteins.

A significant contribution to the limitation of molecular dynamics simulations is due to the environment in which the protein under study is simulated, which may be water, an organic solvent, or a lipid bilayer. In order to reduce the computational load, implicit solvent models based on a continuum description of the surrounding medium have been developed. They vary from the rigorous treatment based on Poisson-Boltzmann equation to different empirical approaches (chapter 3). A continuum solvent is of clear advantage in terms of time and sampling power. However, the use of approximations may introduce artifacts on electrostatic interactions within the protein. Therefore, while molecular dynamics is generally established as a powerful predicting tool, the use of implicit solvents is sometimes considered as not reliable. This PhD work aims at investigating the role played by electrostatic interactions in molecular dynamics simulations of biomolecular systems under different conditions, from folding and unfolding in presence of external forces to stability and fluctuations under native conditions. In this respect, the assessment of the appropriate solvent model is crucial to the success of middle to long term

simulations on the time scale of several nanoseconds.

In chapter 4 the unfolding of the cold shock protein was simulated using two different procedures which lead to the same results, thus confirming a well defined unfolding scheme. One procedure is thermal unfolding, the other employs a bias force selecting thermal fluctuations along a suitable unfolding reaction coordinate thus favoring unfolding. The use of a rather simple implicit solvent model (EEF1 presented in chapter 3) permits an efficient sampling of phase space, including major structural changes. Introducing point mutations at charged residues, which alter the overall charge pattern of the system, results in no change of the unfolding scheme, but has an effect on energy barriers. Both findings are in agreement with experiments.

A very different system is presented in chapter 5, namely an artificial ion channel designed by engineering the gramicidin A peptide. Recent NMR studies show that the presence of cesium iodide triggers a structural transition from the inactive to the active form of the channel. The use of implicit solvent for this system (generalized Born approximation, see chapter 3) is challenging, due to the presence of explicit salt ions in solution. The simulations provide insight into the binding reaction of ions and suggest that these might be responsible for the stabilization of the active form. In spite of some artifacts introduced by the implicit solvent approximation, the method yet delivers solid results in qualitative agreement with experiments. The geometrical details of solvent-solute interactions might be neglected when considering large scale structural changes in proteins, which go together with fast relaxation of the solvent molecules. However, there are cases in which a detailed description of these interactions is most important, for instance the analysis of hydrogen bonding pattern under native conditions or in presence of ligands. The example discussed in chapter 6 goes along this line. The MD simulation of a bacterial cytochrome c immersed in a water box at room temperature for 3 ns is analysed in terms of fluctuations of atomic positions and of intra-protein hydrogen bonds. The presence of explicit water molecules is crucial in order to obtain a realistic dynamics of the hydrogen bonds. This work aims at correlating simulation data on fluctuations with hydrogen exchange experiments under native conditions. Although it is not possible to simulate the exchange events with short term molecular dynamics, a good qualitative correlation is reached between stability of bonds as obtained from exchange data and fluctuations as calculated from MD simulations.

To summarize, this PhD work shows with various examples that electrostatic interactions play a major role in molecular dynamics simulations of proteins. Depending on the task, it is required to find a balance between computational efficiency and geometrical details. Implicit solvent models, especially derived from Poisson-Boltzmann formalism, can be successfully used in a wide range of applications.



# Zusammenfassung

Die Studie der Stabilität und des Faltungsprozesses von Proteinen stellt seit zwanzig Jahren eine der wichtigsten Forschungsrichtungen in Biologie dar und spielt eine noch grössere Rolle seitdem neulich riesige Mengen Genomdaten verfügbar wurden. Die Frage, wie ein Protein sich faltet, damit es stabil und funktionsfähig ist, ist extrem interessant, sowohl für die biologische Forschung, als auch für die pharmakologischen Anwendungen. Heutzutage fehlt immer noch eine Theorie für Proteinfaltung und Strukturvorhersage. Dennoch ist ein ganz vollständiges Bild für eine Anzahl von Proteinen dank der Entwicklung von neuartigen experimentellen Techniken und theoretischen Analysemethoden (Kapitel 1) verfügbar.

Molekulardynamik (MD) spielt eine wichtige Rolle in Biocomputing, nämlich in der Studie über Proteinstruktur und Funktion (Kapitel 2). Es ist mittlerweile möglich, molekulare Systeme mit vielen Atomen wie Proteinkomplexe in Wasser mit MD zu simulieren. Das bietet ein hilfreiches Mittel für die Prüfung und sogar Voraussage von experimentellen Ergebnissen. Jedoch sind die schweren Bedürfnisse an Rechenzeit und Speicherplatz der Simulationen die offensichtliche Grenze von MD Simulationen, zusammen mit der Abhängigkeit von der Genauigkeit von Kraftfeldern und der klassischen Natur. Eine Nanosekunden- zu Mikrosekundenzeit wird normalerweise in gegenwärtigen Simulationen erreicht, welches immer noch zu kurz ist, um größere strukturelle Änderungen, wie Faltung und Entfaltung von Proteinen, zu beobachten. Ein bedeutsamer Beitrag zur Beschränkung von MD Simulationen ist durch die Umgebung verursacht, in der das Protein unter Studium simuliert ist. Diese Umgebung kann aus Wasser, aus einem organischen Lösungsmittel oder einer Zellmembrane aus Lipiden bestehen. Um die Computerlast zu reduzieren, sind implizite Modelle für das Lösungsmittel entwickelt worden, die von einer Kontinuumsbeschreibung des umliegenden Mittels ausgehen. Diese entweder stammen aus der Poisson-Boltzmann-Gleichung oder sind empirische Modelle (Kapitel 3). Ein kontinuierliches Lösungsmittel bietet einen Vorteil wegen geringerer Rechenzeit und effizientes Samplings. Jedoch kann der Gebrauch von Näherungswerten Artefakte in den elektrostatischen Wechselwirkungen innerhalb des Proteins verursachen. Während Molekulardynamik im Allgemeinen als leistungsfähige Methode gilt, wird manchmal der Gebrauch der impliziten Lösungsmittel als unzuverlässig betrachtet.

In dieser Doktorarbeit wird die Rolle elektrostatischer Wechselwirkungen in den MD Simulationen biomolekularer Systeme unter unterschiedlichen Bedingungen, wie Faltung und Auf-faltung durch externe Kräfte, Stabilität und Fluktuationen im nativen Zustand, untersucht. In

dieser Hinsicht ist die Einschätzung des passenden Lösungsmittelsmodells entscheidend zum Erfolg der Simulationen auf der Zeitskala einiger Nanosekunden.

In Kapitel 4 wurde die Auffaltung des Kälte- Schock Proteins mit zwei unterschiedlichen Verfahren simuliert, die zu den gleichen Resultaten führen und somit einen definierten Entfaltungsprozess bestätigen. Ein Verfahren ist thermische Auffaltung, das andere setzt eine externe Kraft ein, die thermische Fluktuationen entlang eine bestimmte Reaktionskoordinate vorwählt, was folglich Auffaltung verursacht. Die Anwendung von einem einfachen impliziten Modell für das Lösungsmittel (EEF1 aus Kapitel 3) erlaubt ein effizientes Sampling des Phasenraumes, wobei große Veränderungen der Proteinstruktur beobachtet werden. Mutationen an geladenen Aminosäuren, die die gesamte Ladungsverteilung des Proteins ändern, haben keinen Effekt auf den Auffaltungsprozess aber wirken auf die Energiebarrieren. Beide Entdeckungen sind in Übereinstimmung mit Experimenten.

Ein ganz anderes System ist in Kapitel 5 vorgestellt, nämlich ein künstlicher Ionenkanal, der aus dem Peptid Gramicidin A hergestellt wurde. Neue NMR Daten zeigen, dass Cäsiumjodid einen strukturellen Übergang von der unaktivierten zur aktiven Form des Kanals auslöst. Der Gebrauch des impliziten Lösungsmittels für dieses System (generalisierte Bornnäherung, siehe Kapitel 3) ist in diesem Fall aufwändig, wegen des Vorhandenseins der expliziten Salzionen im Lösungsmittel. Trotz einigen Artefakten, die durch das implizite Lösungsmittelmodell eingeführt werden, liefert die Methode dennoch Ergebnisse, die mit Experimenten übereinstimmen. Die geometrischen Details der Wechselwirkungen zwischen Lösungsmittel und Protein können vernachlässigt werden, wenn große Strukturänderungen betrachtet werden, bei denen sich die Lösungsmittelmoleküle schnell neu ordnen. Jedoch gibt es Fälle, in denen eine ausführliche Beschreibung dieser Interaktionen am wichtigsten ist, zum Beispiel die Analyse der Wasserstoffbrücken unter nativen Bedingungen oder in Präsenz von Liganden. Das Beispiel, das in Kapitel 6 besprochen wird, geht entlang diese Linie. Ein bakterielles Cytochrom c wird in Wasser bei Raumtemperatur für 3 ns simuliert. Fluktuationen von Atompositionen sowie Wasserstoffbindungen unter Proteinatomen werden analysiert. Explizite Wassermoleküle sind hier entscheidend, um eine realistische Dynamik der Wasserstoffbindungen zu erreichen. Diese Arbeit versucht, Simulationsdaten über Fluktuationen mit Wasserstoffaustauschexperimenten unter nativen Bedingungen aufeinander zu beziehen. Obwohl es nicht möglich ist, den Austauschprozess mit kurzfristiger Molekulardynamik zu simulieren, wird eine gute qualitative Übereinstimmung zwischen Stabilität der Bindungen (aus Experimenten) und Fluktuationen (aus MD Simulationen) erreicht.

Als Schlußfolgerung, zeigt diese Doktorarbeit mit verschiedenen Beispielen, dass elektrostatische Interaktionen eine Hauptrolle in den Molekulardynamik-simulationen der Proteine spielen. Abhängig von der Aufgabe wird es angefordert, eine Balance zwischen Effizienz der Berechnung und geometrischen Details zu finden. Implizite Modelle für das Lösungsmittel, besonders von der Poisson-Boltzmann-Gleichung abgeleitet, können in verschiedenen Fällen erfolgreich benutzt werden.