

# Long Term Dynamics of Proteins and Peptides

K. Anton Feenstra

2002





**Long Term Dynamics  
of  
Proteins  
and  
Peptides**

**K. Anton Feenstra**

The work described in this thesis was made possible by the support of the Netherlands Foundation for Life Sciences (SLW) with financial aid from the Netherlands Organization for Scientific Research (NWO).

*cover:* Top left and right: exploded views of peptide fragments identified in two proteins (1hdn and 1ksr respectively) (See chapters 4 and 5). Bottom: conformations of the three largest clusters of a peptide (see chapter 6). Background: three-dimensional essential space sampling density of peptide trajectory ('bubble' size represents density) and conformational clusters ('ivory ball' size represents cluster size) (see chapter 6).

ISBN 90-6464-992-8

Lay-out by K. Anton Feenstra

Copyright © Anton Feenstra, 2002

Printed by Ponsen & Looijen, Wageningen

Rijksuniversiteit Groningen

**Long Term Dynamics of Proteins and Peptides**

Proefschrift

ter verkrijging van het doctoraat in de  
Wiskunde en Natuurwetenschappen  
aan de Rijksuniversiteit Groningen  
op gezag van de  
Rector Magnificus, dr. D. F. J. Bosscher,  
in het openbaar te verdedigen op  
vrijdag 5 juli 2002  
om 14.15 uur

door

**Klaas Antoni Feenstra**

geboren op 10 augustus 1971  
te Danderyd (Zweden)

Promotores : Prof. dr. H. J. C. Berendsen  
Prof. dr. A. E. Mark

Beoordelingscommissie : Prof. dr. R. Kaptein (Rijks Univ. Utrecht)  
Prof. dr. G. Robillard (Rijks Univ. Groningen)  
Prof. dr. A. J. W. G. Visser (Landb. Univ. Wageningen)

*Voor mijn familie en de toekomst ...*

*‘‘Life is like a box of chocolates –  
you never know what you get.’’*

*— Forrest Gump*



---

# Contents

<b>Contents</b>	<b>7</b>
List of Tables . . . . .	12
List of Figures . . . . .	13
<b>Abbreviations</b>	<b>15</b>
<b>Preface</b>	<b>16</b>
<b>Sources</b>	<b>18</b>
Chapters and Publications . . . . .	18
Preparation and other Software . . . . .	19
<b>Abstract</b>	<b>21</b>
Introduction . . . . .	21
Simulation Efficiency . . . . .	21
Protein Folding . . . . .	22
Cofactor Dynamics and Fluorescence . . . . .	22
Keywords . . . . .	23
<b>Samenvatting</b>	<b>25</b>
Inleiding . . . . .	25
Efficiënt Simuleren . . . . .	25
Eiwitvouwing . . . . .	26
Cofactor Dynamica en Fluorescentie . . . . .	27
Sleutelwoorden . . . . .	27
<b>1 Introduction</b>	<b>29</b>

1.1	The Cosmic View . . . . .	31
1.2	The Organismic View . . . . .	32
1.3	The Molecular View . . . . .	33
1.4	The Local View . . . . .	34
<b>2</b>	<b>Inleiding</b>	<b>37</b>
2.1	Het Kosmische Perspectief . . . . .	37
2.2	Het Organismische Perspectief . . . . .	39
2.3	Het Moleculaire Perspectief . . . . .	39
2.4	Het Locale Perspectief . . . . .	40
<b>3</b>	<b>Efficient Simulation of Large Time-scale Molecular Dynamics</b>	<b>43</b>
3.1	Introduction . . . . .	45
3.1.1	Outline of the Method . . . . .	46
3.2	Methods . . . . .	47
3.2.1	MD Simulations . . . . .	47
3.2.2	System Topology . . . . .	50
	Construction of Dummy Atoms . . . . .	50
	Out-of-plane Vibrations in Aromatic Groups . . . . .	52
3.2.3	Determination of System Properties . . . . .	53
	Motional Periods . . . . .	53
	Energy Drift . . . . .	53
	Diffusion Constant . . . . .	54
	Viscosity . . . . .	54
	Lifetime of Hydrogen Bonds . . . . .	55
3.3	Results . . . . .	55
3.3.1	Water . . . . .	55
	Energy Drift . . . . .	55
	Diffusion . . . . .	56
	Viscosity . . . . .	56
	Hydrogen Bond Lifetime . . . . .	56
3.3.2	Protein . . . . .	57
	Energy Drift . . . . .	57

---

Long-term Properties . . . . .	57
3.4 Discussion and Conclusion . . . . .	59
3.5 Acknowledgments . . . . .	61
<b>4 Hierarchical Motions in Small Proteins: Implications for Protein Folding</b>	<b>63</b>
4.1 Introduction . . . . .	65
4.1.1 Outline of the Approach . . . . .	65
4.2 Methods . . . . .	66
4.2.1 Simulations . . . . .	66
Proteins . . . . .	66
Parameters . . . . .	67
Equilibration . . . . .	67
4.2.2 Analysis . . . . .	67
Essential Dynamics . . . . .	68
Motionally Coherent Elements . . . . .	68
4.3 Results . . . . .	69
4.3.1 Simulations . . . . .	69
4.3.2 Essential Dynamics . . . . .	70
4.3.3 Consensus Dynamics . . . . .	70
Fyn-Sh3 . . . . .	70
T4 Lysozyme . . . . .	72
HPr . . . . .	76
SpoIIAA (1auz) . . . . .	78
Abp-120 (1ksr) . . . . .	78
4.3.4 General Properties . . . . .	78
4.4 Discussion . . . . .	79
4.4.1 Comparison with Experimental Studies . . . . .	80
Fyn-Sh3 . . . . .	81
T4 Lysozyme . . . . .	81
4.5 Conclusion . . . . .	84
4.6 Acknowledgments . . . . .	84
<b>5 Stability of Putative Folding Units</b>	<b>87</b>

5.1	Introduction . . . . .	89
5.2	Methods . . . . .	89
5.2.1	Simulations . . . . .	89
	Parameters . . . . .	89
	Equilibration . . . . .	90
	Production Simulations . . . . .	90
5.2.2	Peptide Stability . . . . .	90
5.3	Results . . . . .	91
5.3.1	Simulations . . . . .	91
5.3.2	Fragment Stability . . . . .	92
5.3.3	Relevance for Folding . . . . .	93
5.4	Discussion . . . . .	93
5.5	Conclusion . . . . .	95
<b>6</b>	<b>Theoretical and Experimental NMR of a Small Peptide</b>	<b>97</b>
6.1	Introduction . . . . .	99
6.1.1	Outline of the Method . . . . .	99
6.2	Methods . . . . .	100
6.2.1	Simulation Set-up . . . . .	100
	Systems . . . . .	100
	Parameters . . . . .	100
	Equilibration . . . . .	101
	Production Simulations . . . . .	101
6.2.2	Analysis of Conformational Space . . . . .	101
	Covariance Analysis . . . . .	101
	Cluster Analysis . . . . .	102
6.2.3	Analysis of Physical Properties . . . . .	102
	Diffusion and Rotational Properties . . . . .	102
	Hydrogen Bond Analysis . . . . .	102
6.2.4	NMR Experiments . . . . .	102
	Sample . . . . .	102
	Spectra . . . . .	103
6.2.5	Calculation of NMR Cross Relaxation Rates . . . . .	103

---

6.3	Results . . . . .	104
6.3.1	Sampling and Convergence . . . . .	104
6.3.2	Conformational analysis . . . . .	105
6.3.3	NMR Experiments . . . . .	107
	Assignments . . . . .	107
	ROESY Intensities . . . . .	107
6.3.4	Comparisons using Theoretical Intensities . . . . .	108
	Calculation . . . . .	108
	Sampling Effects . . . . .	109
	Theoretical Models . . . . .	109
	Complete Ensembles and Single Conformations . . . . .	113
6.3.5	Comparison between Theoretical and Experimental Intensities . . . . .	113
6.4	Discussion . . . . .	114
<b>7</b>	<b>Fluorescence and Dynamics of Flavin Adenine Dinucleotide</b>	<b>119</b>
7.1	Introduction . . . . .	121
7.1.1	Outline of the Method . . . . .	123
7.2	Methods . . . . .	124
7.2.1	Fluorescence measurements . . . . .	124
	Reagents and Sample . . . . .	124
	(Sub)nanosecond Polarized Fluorescence . . . . .	124
7.2.2	Simulations . . . . .	125
	Force Field . . . . .	125
	Initialization and Equilibration . . . . .	126
	Production Simulations . . . . .	127
	Analysis . . . . .	128
7.3	Results . . . . .	129
7.3.1	Fluorescence Dynamics . . . . .	129
7.3.2	Simulations . . . . .	132
	Stacking . . . . .	132
	Conformational Analysis . . . . .	135
	Physical Properties . . . . .	138
7.4	Discussion . . . . .	139

7.5	Acknowledgments . . . . .	142
<b>8</b>	<b>Conclusion</b>	<b>145</b>
8.1	The Future view . . . . .	146
8.2	The Ultimate View . . . . .	146
<b>9</b>	<b>Conclusie</b>	<b>148</b>
9.1	Het Toekomstperspectief . . . . .	148
9.2	Het Ultieme Perspectief . . . . .	149
<b>A</b>	<b>Redistribution of Forces on Dummy Atoms</b>	<b>152</b>
<b>B</b>	<b>FAD Force Field Parameters</b>	<b>154</b>
B.1	Atoms . . . . .	155
B.2	Bonds . . . . .	156
B.3	1-4 Lennard-Jones . . . . .	157
B.4	Angles . . . . .	158
B.5	Proper Dihedrals . . . . .	160
B.6	Improper Dihedrals . . . . .	160
	<b>Bibliography</b>	<b>163</b>
	<b>Index</b>	<b>172</b>

## List of Tables

3.1	Characteristic oscillation periods of atomic motions. . . . .	45
3.2	Atomic masses in water and physical and dynamical properties. . . . .	47
3.3	Summary of simulations of the protein in water. . . . .	49
3.4	Maximum time steps $\Delta t_{max}$ . . . . .	57
3.5	Results of simulations of the protein in water with various topologies. . . . .	57
4.1	Proteins, system sizes and overall simulation properties. . . . .	68
4.2	Essential dynamics and semi-rigid element assignments. . . . .	70
4.3	Amino acid residue composition relative to element boundaries. . . . .	79

5.1	Fragment stability for HPr and 1ksr. . . . .	91
6.1	Simulation details for pepAc. . . . .	100
6.2	Dynamical properties from simulations for pepAc. . . . .	104
6.3	Backbone hydrogen bonds of pepAc. . . . .	107
6.4	$^1H$ chemical shift assignments ( $\sigma_{1H}$ ) for pepAc. . . . .	108
7.1	Atomic charges on FAD ground and excited states. . . . .	126
7.2	Simulations of FAD in water. . . . .	127
7.3	Fluorescence lifetimes and rotational correlation times of FAD and FMN. . . . .	130
7.4	Dynamic behavior and conformations of FAD. . . . .	137

## List of Figures

3.1	Different types of dummy atom constructions. . . . .	50
3.2	Different types of dummy atom constructions for aromatic sidechains. . . . .	52
3.3	Drift in the total energy as a function of time step. . . . .	55
3.4	Inter-protein hydrogen bond distance and angle distributions. . . . .	58
3.5	Dihedral angle distributions in Lys24-NH <sub>3</sub> <sup>+</sup> . . . . .	59
4.1	Semi-rigid element assignments for Sh3. . . . .	71
4.2	Consensus dynamics results for Sh3. . . . .	72
4.3	Motional hierarchy for Sh3 on 3D structure. . . . .	73
4.4	Consensus dynamics results for T4L. . . . .	74
4.5	Motional hierarchy for T4L on 3D structure. . . . .	75
4.6	Consensus dynamics results for HPr, 1auz and 1ksr. . . . .	76
4.7	Motional hierarchy of HPr, 1auz and 1ksr on 3D structures. . . . .	77
4.8	Experimental data for Sh3. . . . .	82
4.9	Experimental data for T4L. . . . .	83
4.10	Tentative folding pathways for Sh3 and T4L. . . . .	85
5.1	Fragments from HPr and 1ksr in “exploded” view. . . . .	92
5.2	Tentative folding pathways for HPr and 1ksr. . . . .	94

---

6.1	Projection of simulations and clusters of pepAc on eigenvectors. . . . .	105
6.2	Middle structures of two largest clusters of pepAc. . . . .	106
6.3	Calculated intensities for different starting structures of pepAc. . . . .	110
6.4	Calculated intensities for different time steps of pepAc. . . . .	111
6.5	Correlation between methods for calculating intensities for pepAc. . . . .	112
6.6	Calculated <i>vs.</i> experimental intensities for pepAc. . . . .	115
7.1	Chemical structure and atom labeling of FAD. . . . .	122
7.2	Fluorescence decays of FMN and FAD. . . . .	131
7.3	Stacking and H-bonds in simulations of FAD. . . . .	132
7.4	Stacked and open conformations of FAD. . . . .	135
7.5	Relative free energy of stacking in the ground and excited state of FAD. . .	136
7.6	Hypothetical theoretical decay curve for FAD. . . . .	141

## Abbreviations

<b>AMP</b>	Adenosine Mono-Phosphate
<b>ATP</b>	Adenosine Tri-Phosphate
<b>ED</b>	Essential Dynamics
<b>FAD</b>	Flavin Adenine Dinucleotide
<b>FFT</b>	Fast Fourier Transform
<b>FMN</b>	Flavin MonoNucleotide
<b>FNR</b>	Ferredoxin NADPH-Reductase
<b>FWHM</b>	Full Width at Half Maximum
<b>GR</b>	Glutathione Reductase
<b>MD</b>	Molecular Dynamics
<b>MEM</b>	Maximum Entropy Method
<b>MSD</b>	Mean Square Displacement
<b>MSF</b>	Mean Square Fluctuation
<b>NMR</b>	Nuclear Magnetic Resonance
<b>NOESY</b>	NOE Spectroscopy
<b>NOE</b>	Nuclear Overhauser Effect
<b>PDB</b>	Protein Data Bank
<b>RMSD</b>	Root-Mean-Square Difference
<b>ROESY</b>	Rotating frame NOESY
<b>SPC</b>	Simple Point Charge
<b>TCSPC</b>	Time-Correlated Single Photon Counting
<b>TOCSY</b>	Total Correlation Spectroscopy
<b>TrxR</b>	Thioredoxin Reductase

# Preface

*In your hands you hold five years of my life. This is not an absolute value. Measured in calendar time it is close to five years, but in hindsight one cannot escape the thought that the same work might have been completed in less time. On the other hand, if I think back on what I learned and experienced in that time, it might be worth something like ten “average” years. One should allow for the fact that gaining knowledge and experience inevitably takes time.*

*For me, the journey that took me through those years have been challenging, enjoyable, instructive and interesting. I will mention some of the highlights along the road and some of the people that made it all possible.*

*The kick-off was at my first, still un-official, Lunteren meeting where my then supervisor Marcus Hemminga introduced me to David van der Spoel and to Gert Vriend. Between them, one can say that find much there is to know about biomolecular modeling and simulation, but much more importantly I found them inspiring and willing to share their knowledge with me. Although my interactions with David were much more extensive, both should be held equally responsible for dragging me into the treacherous expanses of computational biophysics.*

*If the kick-off was in Lunteren, the first gravity assisted slingshot was in Groningen using the space-time discontinuity that accompanies the highly condensed state of knowledge known to most people as Herman Berendsen. A note of caution is at place here. As is the case in celestial mechanics, navigating around Herman is counter intuitive and one can easily find oneself accelerated at relativistic levels of inspiration into quite unexpected directions. This brings to mind a ‘Star Trek – The Next Generation’ episode where Capt. Jean-Luc Picard demands a position report and Helmsman Lt. Geordi answers: “This is unbelievable, sir! According to my readings we are three days away from our previous position! At warp nine!”. That “previous position” was from mere minutes ago. . . But seriously, Herman has remained the main catalyst that drove my project over these years.*

*These first steps already took place before my work on this thesis began. In between I meddled in quite interesting ways with nearly completely insoluble protein aggregates (known as “gluten”) that needed continuous mixing. This eventually drove home the point that my future should be in the computational and molecular point of view.*

---

When I was faced with the choice between doing simulation work on proteins in Wageningen with Jacques Vervoort, or in Groningen with Herman, I was lucky enough to obtain advice from a mutual acquaintance of them both (who will remain unnamed), for which I am still very grateful. The choice was not between a good or a bad location (which I obviously could have made unaided), but between different styles of supervision. That proved to be the corner stone of the conclusion that led me to accept the position in Groningen, and to boldly go where no-one had gone before. . .

For the indispensable entertainment during the long, dark and cold interstellar jumps of the journey (hours of playing Quake), I am indebted to the richly varied (and varying) population of our group: Marc, Bert, Pieter, Peter, Maarten, Frans, Natasha and Bernard, to name a few that spring to mind.

Delicate thruster manouvring around tight turns and crowded inner star-systems was performed by and/or assisted by Ruud Scheek, Steve Hayward (now at the University of East Anglia in Norwich), Petra van den Berg (from Wageningen) and Christine Peter (from ETH Zürich). The final re-fueling stops at San Diego in early spring and Zürich in early summer 2001 were enervating and brought new and much needed motivation. The interest in my work expressed by Prof. Dave Case and Prof. Charles Brooks (at Scripps in S.D.) and Prof. Andrew McCammon (at U.C.S.D.) and members of their groups, certainly pushed me back to escape velocity, when I found myself trapped in low orbit. The same (and more) goes for the warm welcome Corianne and I received in Zürich. . . Thanks Wilfred, Christine, Roland, Phil and Franka for working with me and entertaining us. Also here I want to mention the stimulating interest Prof. Nico Vermeulen (from the V.U. in Amsterdam) showed for my work and this thesis. I am looking forward – in fact, have already started – to put my interests and experience into a different context under his supervision. Not least important was the professional and sensitive ground control during the final, and by far most taxing stage of re-entry and landing. Alan was responsible for most of that, and he compensated generously for lack of discipline and motivation from my side by simply and patiently talking me through the many moves and procedures that if done wrong could still, in sight of safe ground, crash me.

Always, whether within reach or at the other end of electronic transmissions, there has been a host of family and friends who never found it necessary to fake understanding – for they have little or none – nor to fake interest – for they have ample – for my work. Finally, I never ceased to be amazed by the virtually unlimited patience of Corianne and the uncompromised support of Laika and Sita. I sincerely hope I have not overstretched their supplies too much or too often. . .

Groningen – January 18, 2002

# Sources

## Chapters and Publications

### Chapter 3

has been published as “*Improving Efficiency of Large Time-scale Molecular Dynamics Simulations of Hydrogen-rich Systems*” by K. Anton Feenstra, Berk Hess and Herman J. C. Berendsen in the *Journal of Computational Chemistry*, Vol. 20, No. 8, 786-798 (1999) © copyright (1999) John Wiley & Sons, Inc. [www.interscience.wiley.com/jpages/0192-8651/](http://www.interscience.wiley.com/jpages/0192-8651/): Issue Contents.<sup>1</sup> Sec. 3.2.2 **Out-of-plane vibrations in aromatic groups** was submitted as an appendix to chapter 4. The remarks in sec. 3.4 “*Discussion and Conclusions*” on planar aromatic groups are an extension of the original article.

### Chapter 4

was submitted as “*Hierarchical Motion in Small Proteins: Implications for Protein Folding*” by K. Anton Feenstra, Herman J. C. Berendsen and Alan E. Mark to *PROTEINS: Structure, Function & Genetics*, first in February 2001 and in its final form in October 2001.<sup>2</sup> The appendix to the article is included in chapter 3 as sec. 3.2.2 **Out-of-plane vibrations in aromatic groups**.

### Chapter 5

is an original text (which was included in the first submission of chapter 4).

### Chapter 6

was submitted as “*Methods for Calculating NMR Relaxation Parameters of a Small Peptide and comparison with Experiment*” by K. Anton Feenstra, Christine Peter, Ruud M. Scheek, Wilfred F. van Gunsteren and Alan E. Mark in January 2002 to the *Journal of Biomolecular NMR*.

### Chapter 7

was submitted as “*Dynamic conformations of flavin adenine dinucleotide: simulated molecular dynamics of the flavin cofactor related to the time-resolved fluorescence characteristics*” by K. Anton Feenstra, Petra A. W. van den Berg, Alan E. Mark, Herman J. C. Berendsen and Antonie J. W. G. Visser in January 2002 to the *Journal of Physical Chemistry*. It is the result of a collaboration. This author was responsible for all simulation, modeling and trajectory analysis work.

### Appendix A

has been submitted as an appendix to the article of chapter 3.

### Appendix B

has been provided as supplementary material to the article of chapter 7.

## Preparation and other Software

For the production of this manuscript, the authors of the various programs (many available in the public domain as freeware and/or under the GNU general public license) used are gratefully acknowledged:

- The images for the chapter title-page of Chapter 1 were generated by (from top to bottom): me using `galaxy` and `glplanet` (from the XScreenSaver 3.33 distribution: [www.jwz.org/xscreensaver](http://www.jwz.org/xscreensaver)), Leonardo da Vinci in a study on human proportions, Bert de Groot ([bgroot@gwdg.de](mailto:bgroot@gwdg.de)) using `Bobscript` and `Raster3D` and used with permission.
- All simulations as well as much of the analysis were performed using (progressing versions) of the GROMACS software package ([www.gromacs.org](http://www.gromacs.org)), developed in-house in Groningen.
- Cartoon and ball-and-stick renderings of proteins and FAD, were produced using `Molscript` (Per Kraulis [pjk@ciclid.csb.ki.se](mailto:pjk@ciclid.csb.ki.se)<sup>3</sup>) or a modified version of `Molscript`, referred to as ‘Bob-Script’ (Robert ‘Bob’ Esnouf<sup>4</sup>). Ray-traced renderings of `Molscript` and `Bobscript` images were produced using `Raster3D` (E. A. Merritt [merritt@u.washington.edu](mailto:merritt@u.washington.edu)<sup>5,6</sup>).
- Post-processing and conversions of (rendered) bit-map images was done using `xv` (John Bradley, <ftp.cis.upenn.edu/pub/xv>) and `ImageMagick`’s `convert` and `display` (John Christy, [cristy@mystic.es.dupont.com](mailto:cristy@mystic.es.dupont.com)).
- Schemes and drawings were produced using `xfig` ([www.epb.lbl.gov/xfig](http://www.epb.lbl.gov/xfig)).
- Graphs and plots were produced using `xmgr` or its successor `xmgrace` ([plasma-gate.weizmann.ac.il/Grace](http://plasma-gate.weizmann.ac.il/Grace)).
- Text editing and post-processing of `.xpm` and PostScript figures was done using GNU Emacs ([www.gnu.org/software/emacs](http://www.gnu.org/software/emacs)).
- Document preparation and production was done using L<sup>A</sup>T<sub>E</sub>X (Donald E. Knuth).
  - Here is how much of T<sub>E</sub>X’s memory I used:
    - 49568 string characters out of 196870 (= 25% )
    - 161074 words of memory out of 350001 (= 46% )
    - 6885 multiletter control sequences out of 10000+15000 (= 69% )
    - 32317 words of font info for 114 fonts, out of 400000 for 1000 (= 8%, 11% )
    - 34 hyphenation exceptions out of 10000 (= 0% )
    - 34i,23n,33p,273b,657s stack positions out of 3000i,100n,1500p,50000b,4000s (= 1%, 23%, 2%, 1%, 16% )
  - Here is what the resulting files look like:
    - 786k May 7 2002 `thesis.dvi`
    - 3.7M May 7 09:37 `thesis.pdf`
    - 8.9M May 6 12:59 `thesis.ps`
- Citations and references were managed by BibT<sub>E</sub>X (Oren Patashnik, Stanford University).
- overall consistency control and up-to-date checking of figures, graphs and text was done using GNU Make ([www.gnu.org/software/make](http://www.gnu.org/software/make)).
- communication with co-authors of the individual papers was done using the mail readers `elm`, Netscape Mail and Mozilla Mail.



# Abstract

## Introduction

Protein folding is one of the major scientific issues of today. It is under investigation at a great many laboratories around the world, using at least as many different approaches, both experimental and theoretical. In the last few years the prevailing gap between experimental and theoretical results has progressively shrunk, but a lot remains to be done. In this thesis I have aimed to take some steps along that route, in three main areas: *i*) increasing simulation efficiency; *ii*) analysis of native protein dynamics aimed at a simplified description; and *iii*) rigorous comparison between experimental and theoretical results on a small (model) system. In a broader context I have investigated the dynamical behavior of the fluorescent cofactor FAD (see below), which is of great interest also for studying properties of the enzymes.

## Simulation Efficiency

Simulation efficiency is mainly determined by the maximum integration time-step possible for a certain simulated system. This in turn is determined by the fastest degree of freedom present in the system. For protein systems in water this is (mainly) the hydrogen bond-angle vibrations which have a characteristic motional period of around 20 fs. This limits the time step to 3 fs. The biologically relevant behavior, on the other hand, takes place on time scales of nanoseconds or much larger. This large separation of time scales allows the elimination of these fast degrees of freedom without significantly affecting the relevant behavior of the protein system. The use of angle constraints would eliminate this restriction, but would in fact impose secondary constraints on the movement of the heavy atoms (in particular it would lead to a very stiff protein backbone). The alternative is to treat all hydrogens as *dummy atoms*, i.e. the positions of the hydrogens is not integrated in time but is constructed from a combination of positions of (nearby) heavy atoms. This effectively eliminates all degrees of freedom for the hydrogen atoms and thus the restriction on maximum time-step imposed by the fast motions. A similar construction can be implemented for large planar groups, like aromatic amino acid side chains, which can additionally improve simulation stability and reliability. For some hydrogen atoms, e.g. those in water and hydroxyl groups, a rotational degree of freedom must remain. For these atoms, the motion can be retarded by modifying the mass distribution between the hy-

drogen atom and the connected heavy atom (oxygen or nitrogen) to increase the moment of inertia, and thus decrease the motional frequencies involved. Integration time-steps of up to 7 fs were used successfully, with a time-step of around 5 fs being regarded as the optimum trade-off between accuracy and efficiency.

## Protein Folding

Protein native dynamics can be described in terms of the motion of a set of (semi-)rigid bodies or motionally coherent elements. A consistent definition of these bodies or elements is crucial. The obvious assumption that secondary structure elements behave as (semi-)rigid bodies or motionally coherent elements in the native dynamics, need not be correct. Using a novel method to identify said elements from the native dynamics of a protein, based on a consensus description of rigid body motions, it was shown that they often did not coincide with major secondary structure elements. The elements identified show a hierarchical organization which may be related to a hierarchy of folding. Those elements that also show relative stability outside the protein environment, can be considered as possible protein folding nucleation sites. The relative stability of protein elements can be estimated from MD simulations. Two simulations of such an element, one started from the protein conformation of the element and another from an extended conformation, allow to assess the tendency of the fragment towards the protein conformation. This tendency can be considered as a measure for the stability of the fragment in the protein conformation. Using the hierarchical organization of the motionally coherent elements observed and the relative stability of the corresponding elements, hypothetical folding pathways were constructed. These pathways could subsequently be investigated with additional simulations and folding models, and by experiments.

From extensive simulations of one of the protein fragments (which was the best candidate folding nucleus), and NMR experiments on the corresponding synthetic peptide, comparisons were made between theoretical predictions and measured values for ROESY cross-peak intensities. It appears that these values are not very sensitive to details of the conformational ensemble of the molecule, or to details of its dynamics. Therefore, ROESY cross-peaks are of very limited value for solving the solution structure or determining the structural preference of small peptide systems. Using the theoretical intensities only, comparisons were made between different simulations. This allowed the investigation of possible artifacts arising from the enhanced efficiency protocols developed (as described above) by comparing simulations performed at different time-steps. No difference could be detected between simulations performed with time steps of up to 4 fs, and some small differences were detected with simulations performed at 7 fs.

## Cofactor Dynamics and Fluorescence

Molecular Dynamics (MD) simulations and polarized (sub)nanosecond time-resolved flavin fluorescence spectroscopy have been used to study the conformational dynamics of the flavin adenine dinucleotide (FAD) cofactor in aqueous solution. FAD displays a highly

heterogeneous fluorescence intensity decay, resulting in lifetime spectra with two major components; a dominant 7 ps contribution characteristic for ultrafast fluorescence quenching, and a 2.7 ns contribution resulting from moderate quenching. The simulations showed transitions from an 'open' to a 'closed' state, in which the flavin and adenine ring systems stack in a coplanar manner. Stacking generally occurred within the lifetime of the flavin excited state (4.7 ns in water). The fluorescence lifetime calculated from the MD simulations are in excellent agreement with the fluorescence data. Hydrogen-bonds in the ribityl-pyrophosphate-ribofuranosyl chain connecting both ring systems form highly stable cooperative networks and dominate the conformational transitions of the molecule. This work suggests that fluorescence quenching in FAD is mainly determined by the coplanar stacking of the flavin and adenine ring systems, most likely through a mechanism of photo-induced electron transfer.

## Keywords

molecular dynamics	water simulation
protein simulation	time step optimization
accuracy of integration	constraints
dynamical domains	motionally coherent elements
essential dynamics	long time-scale dynamics
conformational dynamics	peptide NMR
flavin	time-resolved fluorescence
flavoprotein	



---

# Samenvatting

## Inleiding

Eiwitvouwing is een van de grootste wetenschappelijke vraagstukken van vandaag. In vele laboratoria over de hele wereld wordt er onderzoek naar gedaan, gebruik makend van uiteenlopende aanpakken, zowel experimenteel als theoretisch. In de afgelopen paar jaar is de bestaande kloof tussen experimentele en theoretische resultaten aanzienlijk geslonken, maar er blijft nog veel te doen. In dit proefschrift heb ik beoogd om een aantal stappen in die richting te nemen, op drie gebieden: *i*) vergroten van de efficiëntie van de simulatie; *ii*) analyseren van natieve eiwitdynamica gericht op een vereenvoudigde beschrijving; en *iii*) rigoureuze vergelijking tussen experimentele en theoretische resultaten van een klein (model)systeem. In een bredere context heb ik onderzocht wat het dynamisch gedrag is van de fluorescente cofactor FAD (zie onder), welke ook van groot belang is bij het bestuderen van de eigenschappen van enzymen.

## Efficiënt Simuleren

De efficiëntie van een simulatie wordt voornamelijk bepaald door de maximum integratie tijdstap die mogelijk is voor een bepaald gesimuleerd systeem. Dit wordt weer bepaald door de snelste vrijheidsgraad die in het systeem aanwezig is. Voor eiwitssystemen in water zijn dit (met name) de waterstofatoom bindingshoek vibraties die een karakteristieke periode hebben van ongeveer 20 fs. Dit limiteert de tijdstap tot 3 fs. Het biologisch relevant gedrag, aan de andere kant, speelt zich af op tijdschalen van nanoseconden of veel langer. Deze grote scheiding in tijdschalen maakt het mogelijk om de snelle vrijheidsgraden uit het systeem te elimineren, zonder daarbij significant invloed uit te oefenen op het relevante gedrag van het eiwitstelsel. Het gebruik van hoekfixaties zou deze beperking elimineren, maar zou in feite secundaire fixaties opleggen aan de beweging van de zware atomen (in het bijzonder zou het leiden tot een zeer stijve ruggesgraat van het eiwit). Het alternatief is om alle waterstofatomen te behandelen als *dummy atomen*, dat wil zeggen dat de positie van de waterstofatomen niet in de tijd geïntegreerd wordt, maar geconstrueerd uit een combinatie van de posities van (nabije) zware atomen. Dit elimineert in feite alle vrijheidsgraden van deze waterstofatomen en daarmee de beperking op de maximum tijdstap die door deze snelle bewegingen wordt opgelegd. Een vergelijkbare constructie kan geïmplementeerd worden voor grote vlakke groepen, zoals aromatische

aminozuur zijketens, wat kan leiden tot extra verhoogde stabiliteit en betrouwbaarheid van de simulatie. Voor sommige waterstofatomen, bijvoorbeeld die in water en in hydroxylgroepen, moet een draaiingsvrijheidsgraad intact blijven. Voor deze atomen kan de beweging vertraagd worden door de massaverdeling tussen het waterstofatoom en het daaraan gebonden zware atoom (zuurstof of stikstof) aan te passen zodat het traagheidsmoment vergroot wordt, wat een verlaging van de karakteristieke bewegingsfrequenties inhoudt. Een integratietijdstap tot 7 fs werd met succes gebruikt, waarbij een tijdstap van rond de 5 fs wordt beschouwd als de optimale balans tussen nauwkeurigheid en efficiëntie.

## Eiwitvouwing

Natieve eiwit dynamica kan beschreven worden in termen van de beweging van een aantal (gedeeltelijk) starre lichamen of coherent bewegende elementen. Een consistente definitie van deze lichamen of elementen is daarbij van groot belang. De meest voor de hand liggende aanname dat secundaire structuur elementen zich gedragen als (gedeeltelijk) starre lichamen of coherent bewegende elementen, is niet noodzakelijkerwijs waar. Gebruik makend van een nieuwe methode om dit soort elementen op grond van de natieve dynamica van een eiwit te identificeren, werd aangetoond dat deze vaak niet overeenkomen met de voornaamste secundaire structuur elementen. De geïdentificeerde elementen zijn gerangschikt in een hiërarchische organisatie, welke gerelateerd zou kunnen zijn aan een hiërarchie van vouwing. Die elementen die ook een relatieve stabiliteit tonen buiten de eiwitomgeving, kunnen aangemerkt worden als mogelijke nucleatieplaatsen voor eiwitvouwing. De relatieve stabiliteit van eiwit-elementen kan geschat worden aan de hand van MD simulaties. Twee simulaties van zo'n element, de een gestart vanuit de conformatie van het element in het eiwit en de ander vanuit een gestrekte conformatie, maken het mogelijk om de neiging van het fragment richting de eiwitconformatie in te schatten. Deze neiging kan beschouwd worden als een maat voor de stabiliteit van het fragment in de eiwitconformatie. Gebruik makend van de hiërarchische organisatie van de waargenomen coherent bewegende elementen en de relatieve stabiliteit van de overeenkomstige fragmenten, werden hypothetische vouwingspaden opgesteld. Deze paden zouden vervolgens onderzocht kunnen worden met behulp van nog meer simulaties en vouwingsmodellen, en met behulp van experimenten.

Op grond van uitgebreide simulaties van een van de eiwitfragmenten (de beste kandidaat vouwingskern), en NMR experimenten op het overeenkomstige synthetische peptide, werden vergelijkingen uitgevoerd tussen theoretische voorspellingen en gemeten waarden voor ROESY kruis-pek intensiteiten. Het lijkt erop dat deze waarden niet erg gevoelig zijn voor details van het conformationele ensemble van het molecuul, of voor de details van de dynamica ervan. Daarom zijn ROESY kruis-pek intensiteiten van beperkte waarde bij het oplossen van de structuur in oplossing of het bepalen van de structuurvoorkeur van kleine peptide systemen. Gebruik makend enkel van de theoretische intensiteiten werd een vergelijking gemaakt tussen verschillende simulaties. Hierdoor konden mogelijke artefacten voortkomend uit de ontwikkelde protocollen voor verbeterde efficiëntie (zoals hierboven beschreven) onderzocht worden door simulaties, uitgevoerd met verschillende tijdsstap, te vergelijken. Er kon geen verschil waargenomen worden tussen simulaties uitgevoerd met

tijdstappen tot 4 fs, en kleine verschillen waren aanwezig met simulaties uitgevoerd met 7 fs.

## Cofactor Dynamica en Fluorescentie

Moleculaire dynamica simulaties en gepolariseerde (sub)nanoseconde tijdsopgeloste flavine fluorescentie spectroscopie zijn gebruikt om de conformationele dynamica van de flavine adenine dinucleotide (FAD) cofactor in waterige oplossing te onderzoeken. FAD heeft een sterk heterogeen fluorescentie intensiteitsverval, resulterende in een levensduurspectrum met twee belangrijke componenten: een dominante 7 ps bijdrage karakteristiek voor ultrasnelle fluorescentiedoving en een 2.7 ns bijdrage die het resultaat is van middelmatige doving. De simulaties lieten overgangen zien van een 'open' naar een 'gesloten' toestand, waarin de flavine en adenine ringsystemen 'stapelen'. Stapeling vond meestal plaats binnen de levensduur van de aangeslagen toestand van het flavine (4.7 ns). De fluorescentielevensduur bepaald op basis van de MD simulaties komen goed overeen met de fluorescentie data. Waterstofbruggen in de ribityl-pyrofosfaat-ribofuranosyl keten, die beide ringsystemen verbindt, vormen zeer stabiele coöperatieve netwerken en overheersen de conformationele overgangen van het molecuul. Dit werk suggereert dat fluorescentiedoving in FAD voornamelijk bepaald wordt door de coplanaire stapeling van de flavine en adenine ringsystemen, zeer waarschijnlijk door middel van een mechanisme van foto-geïnduceerde electronoverdracht.

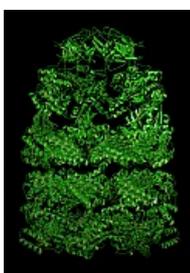
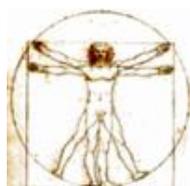
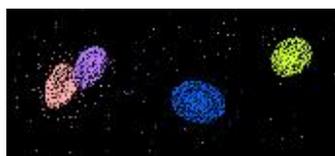
## Sleutelwoorden

moleculaire dynamica	water simulatie
eiwit simulatie	tijdstap optimalisatie
integratienuwkeurigheid	fixaties
dynamische domeinen	coherent bewegende elementen
essentiële dynamica	lange tijdschaal dynamica
conformationele dynamica	peptide NMR
flavine	tijdsopgeloste fluorescentie
flavoeiwit	



# Chapter 1

## Introduction





## 1.1 The Cosmic View

The whole of contemporary science appears to be headed towards a collective complexity barrier, in which the wide variety of systems under study share one common characteristic of being of such complexity, that a conventional description of well-defined constituent parts with well-defined interactions, no longer seems able to capture the nature of the phenomenon under observation. Still worse, in some areas even the distinction between observer and observed cannot clearly be made. Areas in which these problems appear particularly obvious, are:

- *quantum-physics*, where an “elementary particle” is never isolated but always accompanied by a shroud of “virtual particles”, and interactions are thought to be transmitted by yet more particles (worse still, the distinction between virtual and real particles, and between particles and interactions, is actually meaningless in that context);
- *nano-technology* where matter must sometimes be described as collections of atoms, or as quantum waves, or in terms of averages and densities and often new, hybrid and still poorly understood models are necessary to describe properties that emerge from the interplay of these descriptions;
- *biochemistry*, where the whole cell is seen to be a continually changing and cycling machinery of molecules, of which probably most have an influence on each other;
- *biology* or *ecology*, where it has become clear that flora, fauna, soil, water and air are inextricably linked into an eco-system that we once thought could be divided into independent parts, and we are gradually learning (the hard way) that we ourselves are an equally inextricably linked part of it;
- *sociology* and *economics* (which can be viewed as a branch of sociology with a different set of boundary conditions), in which often even the act of performing the experiment, and afterwards publishing the result, cannot be separated from the experiment itself;
- *cosmology*, which is linked to particle and quantum-physics, bringing the circle back round to where we began.

This might mean that the historical approach of science of bringing down the complexity of the world around us into a limited set of simple, well-defined rules, is coming to an end and that the remaining big questions can only be answered by tackling them in all their immense complexity and detail. Fortunately, we have recently also begun developing tools which are very well equipped for handling the enormous tasks of analyzing systems of high complexity in a highly detailed fashion. The most obvious of these is the computer which allows the storage, retrieval and processing of the raw data associated with such systems. Equally important are the mathematical methods and models that can minimize the complexity of the raw data describing the system and thus uncover the intrinsic complexity of the system itself. The compression algorithms used for image and audio data are a nice example.

This shift towards developing complex models that can describe complex systems can already be observed in a number of fields of research, like economics, various engineering practices, meteorology, quantum chemistry and protein folding, in which apparently the fruitful approaches involve large-scale simulations of the system in question. Often the level of simplification is dictated by the available computing power (e.g. in meteorology), or, conversely the maximum level of simplification required for a certain level of reliability sets a limit on the size of the systems that can successfully be taken under consideration (e.g. in quantum chemistry).

Continuing along this line of thought, it seems that major breakthroughs in years to come will most likely be linked to sophisticated and clever methods for handling and analyzing huge datasets of a very complex nature. A broadly used but still relatively simple example of this is the mathematical method of principal components analysis, which can extract general trends and correlations from a (large) set of variables. It also appears likely that cross-overs from one field of research to another will be more frequent and will have high chances of success, since these types of methods will probably have more or less general applicability, and will therefore not be restricted to a particular field. In other words, we should see an increasing emphasis on mathematical methods and models specifically equipped to deal with these large complex data sets.

## 1.2 The Organismic View

There is one area of research that might be described as being “meta-complex” as it involves aspects of several of the complex areas listed above. It can be captured under the broad label of “life sciences”, involving the interaction between the molecular basis of life (quantum-physics and biochemistry), its evolutionary history (molecular biology and biology), the environment and eco-system (ecology and sociology) and its possible applications in engineering (nano-technology). This makes it a very challenging field of research, and considering our own interest as *H. sapiens* the amount of interest it receives should not be a surprise.

Much discussion is possible about which of the complex systems associated with living organisms is most fundamental to our existence, or that of terrestrial organisms in general. Mostly, it will be a matter of taste. However, there appears to be one particular thin spot in our general understanding of life. From one side, we know how the process of evolution causes the prevalence of certain genes, and we know how genes give rise to proteins. From the other side, we know how organisms are organized and how they are “constructed” out of organs, and how those are constructed out of cells. We also know what a cell is made of, and in which way many of its processes are regulated. Many of these areas have their own problems in understanding certain aspects, but there is one thing that ties everything together of which the basic mechanism remains unknown. While we know how the genetic code is translated into a sequence of amino acid residues, and we know of many proteins how they function, the understanding of how the sequence translates itself into the functional folded protein remains elusive.

## 1.3 The Molecular View

All beings (as we know them) consist of molecules. The most interesting ones are the so-called “biological macromolecules”, namely proteins (which make up the cell’s machinery), DNA and RNA (which respectively store and transport the genetic information) and lipids (which compartmentalize the processes in the cell, and the cell itself from the outside). Proteins are the molecules in a cell that accelerate and regulate all processes involved in making a living cell alive. That, in a sense, makes them central to the understanding of how a cell functions. At the same time, some of the key features of protein molecules are poorly understood.

What is true for the life sciences as a whole, also holds for proteins. For most of the research areas involving highly complex systems, as mentioned above, one can identify areas of scientific research directed towards understanding of the working of proteins, or to the relations of a protein with its surroundings. The function of a protein as an enzyme, that accelerates a specific chemical process, can only be described and understood using a proper quantum-chemical description. The existence of a certain protein with a certain appearance and construction must be described in its evolutionary context, which brings it also in relation to its present and past functions in the cell and via that to possible roles that might have in the broader context of the organism and society. The specific shape and function a protein has is inextricably linked to its construction, which is derived from the descriptions in the genetic code, which can only be understood, again in the evolutionary context and its function in the cell, but also against the background of the basic physical knowledge of the behavior of individual atoms and molecules. The behavior of large protein complexes, like the keratin fibers that hairs are made of, or muscle fibers which are made out of countless units of actin and myosin, or the aggregates of proteins that form the amyloid plaques that can ravage a person’s (or animal’s) brains, might only be comprehensibly described on a mesoscopic level, beyond the level of individual molecules.

Previously, the so-called “protein-folding problem” was already briefly mentioned: how does a sequence of chemically linked amino acids, in total several thousands of atoms, fold itself into a specific structure? This structure often is defined to within a tenth of a nanometer, much less than half the diameter of the typical atom in proteins, a carbon atom. This cannot happen purely by chance, if one works out the odds it is impossible and this is known as the “Levinthal paradox”; protein folding appears impossible but yet it happens. Therefore some mechanism must govern the process that leads from an unstructured, freshly produced sequence of amino acids, to this fully folded, functional, ‘native’ protein.

Broadly speaking, three approaches are being pursued in the field of protein folding research: biochemical, structural and computational. In the first, one attempts to learn all there is to learn in a laboratory about the functions that can be performed by a protein, modified versions of the protein, and more or less related ‘cousins’ of the protein from different species. The second approach, also experimental, elucidates the appearance of a protein by determining the three dimensional position of all its atoms, from which an extremely detailed ‘picture’ of the protein emerges, known as the “protein structure”. The third approach is theoretical and takes a known three dimensional structure of a protein,

the known properties of its atoms (like radius, mass, charge and connections or bonds), Newtons' laws of motion and a powerful computer to predict the possible motions of all the atoms that could occur within the protein, i.e. the dynamics of a molecule are simulated.

One important point has been made clear in the last couple of years, however, that neither simulations nor experiments alone will be able to solve the issues at hand. Generally speaking, experiments lack sufficient detail and discriminatory power at the atomic level and at short timescales, and conversely, simulations lack sufficient statistics on multiple (or large) molecules and at long timescales. The way forward therefore must lie in a fruitful synthesis of simulation and experiment.

There is a two-fold hurdle in combining simulation and experiment derived data for biological molecular systems. This lies in the difference in timescale accessible to simulation and experiment, and in the comparison of the raw data of both. The timescale gap can be bridged by (ever increasing) more sophisticated experimental set-ups and by performing simulations by more efficient means on, also ever increasing, faster computers. Comparison of data can best be done by calculating from the simulations the actual raw experimental data as directly as possible.

## 1.4 The Local View

The hypothesis that started off the work described in this thesis, was about an aspect of protein folding. Early engineered four helix bundle "proteins" lacked the well-defined interior structure characteristic of "natural proteins", instead giving rise to a state often referred to as a "molten globule", indicating the presence of secondary structure (like helices) and the absence of a packed structure for the amino acid side chains in the interior. Introduction of a covalently bound heme group in the interior of the helix bundles, yielded an anchoring point around which a "natural" protein structure could form. From these results, the thought arose that also in naturally occurring proteins such an anchoring point must be present, which would induce structure in the native state structure. This anchor could consist of a salt-bridge, a sulfur bridge, a buried hydrogen bond, a stacking interaction between aromatic side chains, or an interdigitation between a number of aliphatic side chains. Due to this diversity in the nature of possible anchoring points, it seemed not to be feasible to identify them directly by searching for the constituent components. Further thinking centered around ideas of possible modes of motion of the protein around such an anchoring point, in terms of single- or double-hinged, swiveling or sliding motions. Identifying these types of motion in the protein dynamics could be a way of finding the anchoring points.

One approach for identifying anchoring points was based on the influence they should have on the correlation of motions of the surrounding protein. Therefore, a thorough analysis was performed of motional correlations in a protein in its native state as expressed during molecular dynamics simulations. An anchor would reveal itself by correlation of spatially adjacent but sequentially distant parts of the protein. Instead it was found that all significant correlation in a protein is sequential. This, added to other negative results from the search for anchors, effectively negated the initial hypothesis. It was replaced by

the notion of “motionally coherent elements” that form the elementary building blocks of a protein in a dynamical sense. Some of the preliminary work was written up in a conference proceedings as “The Domain Decomposition of a Single-Domain Protein”.<sup>7</sup> The results and reasonings that led to these (and other) conclusions are presented in chapter 4. The findings make it possible to generate a simplified description of the dynamic nature of a protein.

In the context of protein folding, dynamics obviously plays a significant role. Therefore, the “motionally coherent elements” mentioned above, might in addition be related to dynamical elements that behave more or less independently during folding, thus they might be referred to as “folding units”. Some of the possible implications of this hypothesis are investigated in chapter 5. The elements identified from the protein are treated as isolated peptides in solution. A tendency of such a peptide to adopt the protein conformation can indicate a possible role of the element in the early stages of the protein folding process. In chapter 6 one of these fragments that showed most promising as possible folding nucleation site, is examined in detail using NMR experiments and MD simulations.

A large portion of this thesis is devoted to making steps towards overcoming the hurdle of combining simulation and experiment. In simulations, time scales of 1  $\mu$ s are considered huge, whereas for many experiments such times lie well below the lower limit of temporal resolution or dead-time. Likewise, simulations of ensembles of protein molecules are rare (to say the least), but any experiment will yield results on ensembles of at least  $10^{10}$  molecules. Fortunately, recent experimental developments are heading towards the single molecule level. The most spectacular is pulling atomic force “microscopy” (AFM), where a single molecule is caught with the tip of an AFM and subsequently pulled apart. But also single molecule spectroscopic measurements are being developed, and single molecule imaging can be done using electron microscopes. Still, in general these techniques yield results on timescales that are far beyond the accessible time scales of MD simulations, so they do not yet enable direct comparisons. On the side of comparison, the main bottleneck is in the interpretations and assumptions that must often be made while processing experimental data. In chapter 3 aspects of improving simulation efficiency are addressed which extend the time scales accessible in the simulation. In chapter 6 comparisons are made between NMR data and extensive simulations of a nine-residue peptide. The calculated NMR observables will also be used to assess possible artifacts resulting from the different simulation set-ups devised in chapter 3 for increased efficiency.

A different example of dynamical behavior of complex biological molecules and of comparing results from experiments and simulations, is presented in chapter 7 where the fluorescent co-factor flavin adenine dinucleotide (FAD) is studied. The fluorescent behavior of FAD is directly linked to the dynamical behavior, which makes the comparison of MD simulations with fluorescence experiments interesting. In addition, FAD fluorescence measurements can be used as a tool for studying protein to which it is bound, but in many cases also traces of free FAD will be present. For a proper understanding of these experiments, a detailed understanding of the dynamical and fluorescent properties of FAD is important.



---

# Hoofdstuk 2

## Inleiding

### 2.1 Het Kosmische Perspectief

Het geheel van de hedendaagse wetenschap lijkt zich te bewegen richting een collectieve complexiteitsbarrière, waarbij een grote variatie aan bestudeerde systemen gemeen hebben dat ze van een dermate grote complexiteit zijn dat een conventionele beschrijving van duidelijk gedefinieerde samenstellende delen met welomschreven wisselwerkingen, niet langer in staat lijkt het wezen van het bestudeerde fenomeen te vangen. Bovendien kan in sommige gebieden zelfs het onderscheid tussen waarnemer en waargenomene niet duidelijk gemaakt worden. Gebieden waarin deze problemen bijzonder duidelijk lijken, zijn:

- *quantumfysica*, waar een elementair deeltje nooit geïsoleerd is, maar altijd vergezeld gaat van een wolk virtuele deeltjes en waar wisselwerkingen gedacht worden overgedragen te worden door nog meer deeltjes (nog sterker, het onderscheid tussen virtuele en echte deeltjes, en tussen deeltjes en interacties is feitelijk betekenisloos in die context);
- *nano-technologie* waar materie soms beschreven moet worden als een verzameling atomen, soms als quantumgolven, of in termen van gemiddeldes en dichtheden en waar vaak nieuwe, gecombineerde en nog slecht begrepen modellen nodig zijn om de eigenschappen te beschrijven die voortkomen uit het samenspel van deze beschrijvingen;
- *biochemie*, waar de gehele cel gezien wordt als een voortdurend veranderende en vernieuwende machinerie van moleculen, waarvan waarschijnlijk de meeste wisselwerkingen hebben met elkaar;
- *biologie* of *ecologie*, waar het duidelijk wordt dat de flora, fauna, bodem, water en lucht onontwarbaar verbonden zijn in een ecosysteem waarvan we ooit dachten dat het onderverdeeld kon worden in onafhankelijke delen, en we langzamerhand (door schade en schande) leren dat wij zelf er een even onontwarbaar verbonden onderdeel van zijn;

- *sociologie* en *economie* (wat gezien kan worden als een tak van sociologie met een andere verzameling randvoorwaarden), waarin vaak zelfs het ondernemen van een experiment, en achteraf het publiceren van de resultaten, niet los gezien kan worden van het experiment zelf;
- *kosmologie*, dat verbonden is met deeltjes- en quantumfysica, waarmee de cirkel rond is tot het punt waar we begonnen.

Dit kan betekenen dat de traditionele aanpak van de wetenschap om de complexiteit van de wereld om ons heen terug te brengen tot een beperkte verzameling van eenvoudige, goed gedefinieerde regels, aan zijn einde komt en dat de overgebleven grote vragen alleen beantwoord kunnen worden door ze in hun overweldigende complexiteit aan te pakken. Gelukkig we recentelijk ook begonnen met het ontwikkelen van gereedschappen die zeer goed uitgerust zijn om de enorme taak van de analyse van systemen van grote complexiteit zeer gedetailleerd te kunnen uitvoeren. Het meest voor de hand liggende voorbeeld hiervan is de computer die de opslag, opvraag en verwerking van de ruwe gegevens die bij dit soort systemen hoort, mogelijk maakt. Even belangrijk zijn de wiskundige methodes en modellen die de complexiteit van de ruwe gegevens die het systeem beschrijven, kunnen minimaliseren en zodoende de intrinsieke complexiteit van het systeem zelf kunnen blootleggen. Compressie algoritmes die gebruikt worden voor beelden en geluid zijn hiervan een mooi voorbeeld.

Deze verschuiving richting de ontwikkeling van complexe modellen die complexe systemen kunnen beschrijven, is al zichtbaar in een aantal onderzoeksvelden, zoals de economie, verschillende takken ingenieurswerk, meteorologie, quantumchemie en eiwitvouwing, waarin klaarblijkelijk de vruchtbare benaderingen gebruik maken van grootschalige simulaties van het onderhavige systeem. Vaak wordt de mate van vereenvoudiging opgelegd door de beschikbare rekenkracht (bijv. in de meteorologie), of, omgekeerd de maximale vereenvoudiging die vereist is voor een bepaald nivo van betrouwbaarheid, een limiet zet aan de afmetingen van de systemen die met succes onder beschouwing genomen kunnen worden (bijv. in de quantumchemie).

Voortbordurend op dit idee, lijkt het dat grote doorbraken in de komende jaren zeer waarschijnlijk verbonden zullen zijn met geavanceerde en slimme methodes voor het behandelen en analyseren van grote, complexe verzamelingen gegevens. Een veelgebruikte, maar nog verhoudingsgewijs eenvoudig voorbeeld hiervan is de wiskundige methode van principale componenten analyse, welke algemene tendensen en correlaties uit een (grote) verzameling variabelen kan halen. Het lijkt ook aannemelijk dat kruisbestuivingen vanuit het ene onderzoeksgebied naar een ander meer frequent zullen zijn en meer kans op slagen zullen hebben, omdat dit soort methodes waarschijnlijk een min of meer algemene toepasbaarheid zullen hebben en daardoor niet beperkt zullen zijn tot één bepaald gebied. Met andere woorden, we zouden een verhoogde nadruk op wiskundige methodes en modellen moeten zien, die specifiek uitgerust zijn om deze grote gegevensverzamelingen te behandelen.

## 2.2 Het Organismische Perspectief

Er is een onderzoeksgebied beschreven zou kunnen worden als metacomplex, omdat er onderdelen van de verschillende hierboven beschreven complexe gebieden bij betrokken zijn. Het kan gevangen worden onder de brede term “levenswetenschappen”, waarbij de wisselwerkingen tussen de moleculaire basis van het leven (quantumchemie en biochemie), de evolutionaire geschiedenis (moleculaire biologie en biologie), de omgeving en het ecosysteem (ecologie en sociologie) en mogelijke toepassingen in de technologie (nanotechnologie) een rol spelen. Hierdoor is het een uitdagend onderzoeksgebied en in het licht van ons eigen belang als *H. sapiens*, zal de omvang van de aandacht die eraan geschonken wordt, geen verrassing moeten zijn.

Er is veel discussie mogelijk over welke van de complexe systemen die samenhangen met levende wezens, het meest fundamenteel is voor ons bestaan, of dat van aardse organismen in het algemeen. Voornamelijk zal dat een kwestie van smaak zijn. Echter, er lijkt een bijzonder dun plekje te zitten in ons algemene begrip van het leven. Aan de ene kant weten we hoe het proces van evolutie het in verhoogde mate vóórkomen van bepaalde genen veroorzaakt, en we weten hoe organismes georganiseerd zijn, hoe ze “opgebouwd” zijn uit organen en hoe deze weer opgebouwd zijn uit cellen. We weten ook waar een cel van gemaakt is en hoe veel van de processen in de cel geregeld zijn. In veel van deze gebieden spelen problemen bij het begrijpen van bepaalde onderdelen, maar er is één ding waarmee alles aan elkaar hangt waarvan het basismechanisme onbekend blijft. Terwijl we weten hoe de genetische code vertaald wordt naar een bepaalde volgorde van aminozuurresiduen, en we van veel eiwitten weten hoe ze werken, begrip van hoe deze aminozuurvolgorde zichzelf vertaald in een werkend, gevouwen eiwit blijft ons ontglippen.

## 2.3 Het Moleculaire Perspectief

Alle wezens (zoals wij ze kennen) bestaan uit moleculen. De meest interessante daarvan zijn de zogenaamde “biologische macromoleculen”, namelijk eiwitten (waaruit de machinerie van een cel bestaat), DNA en RNA (respectievelijk voor de opslag en transport van de genetische informatie) en lipiden (waarmee de processen in de cel, en de cel zelf, afgebakend worden) Eiwitten zijn de moleculen in een cel die alle processen die betrokken zijn bij het levend maken van een levende cel, versnellen en besturen. Dat maakt ze, in zekere zin, een centraal element bij het begrijpen van hoe een cel functioneert. Tegelijkertijd zijn bepaalde sleuteleigenschappen van eiwitmoleculen slecht begrepen.

Net zoals voor de levenswetenschappen in het algemeen, kan men ook voor eiwitten voor de meeste van de onderzoeksgebieden die te maken hebben met sterk complexe systemen, zoals eerder genoemd, wetenschappelijke onderzoeksgebieden aanduiden die gericht zijn op het begrijpen van de werking van eiwitten, of op de verhoudingen van een eiwit met zijn omgeving.

## 2.4 Het Locale Perspectief

De hypothese waarmee de start gemaakt werd voor het werk dat in dit proefschrift beschreven wordt, ging over een aspect van eiwitvouwing. De eerste ontworpen vier helix bundel “eiwitten” misten de goed gedefinieerde binnenste structuur die karakteristiek is voor “natuurlijke eiwitten”. In plaats daarvan bevonden deze ontworpen eiwitten zich in een toestand die vaak beschreven wordt als een “gesmolten bol” (“molten globule”), waarmee de aanwezigheid van secundaire structuur (zoals helices) en de afwezigheid van een vervlochten structuur van de aminozuur zijketens binnenin aangegeven wordt. Toevoeging van een covalent gebonden heemgroep in het binnenste van de helix bundels, gaf een ankerpunt waaromheen een “natuurlijke” eiwitstructuur zich kon vormen. Uit deze resultaten ontstond de gedachte dat ook in natuurlijk voorkomende eiwitten zo’n ankerpunt aanwezig zou moeten zijn, waarmee de structuur in de natieve toestand van het eiwit geïnduceerd zou worden. Het anker zou kunnen bestaan uit een zout- of zwavelbrug, een inwendige waterstofbrug, een stapelings-wisselwerking tussen twee aromatische zijketens of een vervlechting tussen een aantal alifatische zijketens. Deze diversiteit van mogelijke ankerpunten, maakte dat het onhaalbaar leek om ze rechtstreeks te identificeren door te zoeken naar de samenstellende delen. De volgende stappen richtten zich dan ook op ideeën over mogelijke bewegingstoestanden van het eiwit rondom zo’n ankerpunt, in termen van enkel of dubbel scharnierende, draaiende of schuivende bewegingen. Identificatie van deze soorten beweging in de eiwitdynamica, zou een manier kunnen zijn om de ankerpunten te vinden.

Een aanpak voor het vinden van ankerpunten was gebaseerd op de invloed die ze zouden moeten hebben op de correlatie van bewegingen van het omringende eiwit. Daarom werd een grondige analyse uitgevoerd van bewegingscorrelaties van een eiwit in zijn natieve toestand, zoals dat tot uitdrukking komt tijdens moleculaire dynamica simulaties. Een anker zal naar voren komen door een correlatie tussen ruimtelijk nabij maar in de eiwitketen ver weg gelegen delen van het eiwit. In plaats daarvan werd gevonden dat alle betekenisvolle correlatie in een eiwit sequentieel is. Gevoegd bij alle andere negatieve resultaten van de zoektocht naar ankers, werd hierdoor de oorspronkelijke hypothese duidelijk ontkracht. In plaats daarvan kwam het idee van “coherent bewegende elementen” die, wat de dynamische eigenschappen betreft, de elementaire bouwstenen vormen voor een eiwit. Een deel van het verkennende werk is beschreven onder de titel “The Domain Decomposition of a Single-Domain Protein”.<sup>7</sup> De resultaten en redeneringen die leidden tot deze (en andere) conclusies worden gepresenteerd in hoofdstuk 4. Door deze vondsten is het mogelijk om een vereenvoudigde beschrijving te maken van het dynamische karakter van een eiwit.

Tegen de achtergrond van eiwitvouwing speelt dynamica vanzelfsprekend een belangrijke rol. Daarom zouden de “coherent bewegende elementen” die eerder genoemd werden, ook verwant kunnen zijn aan dynamische elementen die zich min of meer onafhankelijk gedragen gedurende het proces van eiwitvouwing, oftewel ze zouden ook bestempeld kunnen worden als “vouwingseenheden”. Een aantal van de mogelijke implicaties van deze hypothese wordt onderzocht in hoofdstuk 5. De elementen zoals ze in het eiwit gevonden werden, worden behandeld als onafhankelijke peptiden in oplossing. Een tendens van zo’n peptide om de eiwitconformatie aan te nemen, is een indicatie voor een mogelijke rol

van dat element in de eerste stappen van het eiwitvouwingsproces. In hoofdstuk 6 wordt een van deze fragmenten, die het meest veelbelovend leek als mogelijke vouwingskern, gedetailleerd onderzocht met behulp van NMR experimenten en MD simulaties.

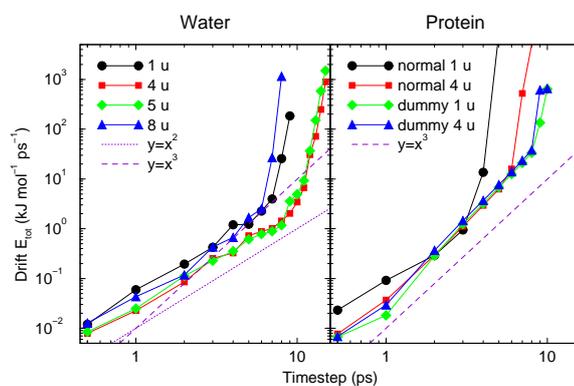
Een groot deel van dit proefschrift is gewijd aan het maken van stappen om de horde te nemen van het combineren van simulatie en experiment. In simulaties worden tijdschalen van  $1 \mu\text{s}$  als enorm beschouwd, terwijl voor veel experimenten deze tijden ver onder de ondergrens van het tijdsoplossend vermogen of de dode tijd liggen. Ook zijn simulaties van ensembles van eiwitmoleculen (op zijn minst) erg zeldzaam, terwijl een willekeurig experiment resultaten oplevert van ensembles van ten minste  $10^{10}$  moleculen. Gelukkig zijn recente ontwikkelingen op het experimentele gebied in de richting van het enkele molecuul in gang gezet. Het meest spectaculaire van deze is de trekkende atoomkracht “microscop” (pulling atomic force microscope, pulling AFM), waar een enkel eiwitmolecuul vast gemaakt wordt aan de tip van een AFM en vervolgens uit elkaar getrokken wordt. Maar ook spectroscopische metingen op enkele moleculen worden ontwikkeld, en afbeeldingen van een enkel eiwit molecuul kunnen gemaakt worden met een elektronen microscoop. Toch leveren deze technieken nog resultaten op op tijdschalen die ver voorbij de bereikbare tijdschalen van MD simulaties liggen, waardoor een rechtstreekse vergelijking nog niet mogelijk is. Op het vlak van het maken van de vergelijking, ligt de grootste drempel bij de interpretatie en de aannames die vaak gemaakt moeten worden bij het verwerken van de experimentele gegevens. In hoofdstuk 3 worden aspecten van het verbeteren van de efficiëntie van simulaties aan de orde gebracht, waardoor een verruiming van de haalbare tijdschalen in simulaties bereikt wordt. In hoofdstuk 6 worden vergelijkingen gemaakt tussen NMR gegevens en uitgebreide simulaties van een negen residu peptide. De berekende NMR gegevens zullen ook gebruikt worden om mogelijke afwijkingen te beoordelen die het gevolg kunnen van de verschillende simulatie instellingen die in hoofdstuk 3 bedacht zijn voor het verhogen van de efficiëntie van de simulatie.

Een ander voorbeeld van het dynamische gedrag van complexe biologische moleculen en van het vergelijken van resultaten uit experiment en simulatie, wordt gepresenteerd in hoofdstuk 7 waar de fluorescente co-factor flavine adenine dinucleotide (FAD) wordt bestudeerd. Het fluorescente gedrag van FAD is rechtstreeks verbonden met het dynamisch gedrag. Dit maakt een vergelijking van MD simulaties met fluorescentie experimenten interessant. Bovendien kunnen FAD fluorescentiemetingen gebruikt worden als methode om het eiwit te bestuderen waar het aan gebonden is, maar in veel gevallen zullen ook sporen van vrij FAD aanwezig zijn. Voor een gedegen begrip van deze experimenten is een gedetailleerd inzicht in de dynamische en fluorescente eigenschappen van FAD belangrijk.



## Chapter 3

# Efficient Simulation of Large Time-scale Molecular Dynamics





### 3.1 Introduction

The maximum time step in Molecular Dynamics (MD) simulations is limited by the smallest oscillation period that can be found in the simulated system. Bond-stretching vibrations are the fastest atomic motions in a molecule, typically in the order of 10 fs. A classical treatment of these motions is not correct because such vibrations are in their quantum-mechanical ground state. For a proper treatment quantum-mechanical calculations should be included. It stands to reason, therefore, that for normal (classical) MD these motions are ignored altogether, i.e. they can be better represented by a constraint.

For the remaining degrees of freedom, the shortest oscillation period as measured from a simulation is 13 fs for bond-angle vibrations involving hydrogen atoms, see Table 3.1. Taking as a guideline that with a Verlet (leap-frog) integration scheme a minimum of 5 numerical integration steps should be performed per period of a harmonic oscillation in order to integrate it with reasonable accuracy,<sup>8</sup> the maximum time step will be about 3 fs. This is slightly larger than the 2 fs which is routinely used in MD simulations of biomolecules in water. Disregarding these very fast oscillations of period 13 fs (which are also in the quantum-mechanical ground state) the next shortest periods are around 20 fs, which will allow a maximum time step of about 4 fs. In simulations with constrained bond lengths and angles, it has recently been shown that hydrogen atom dihedral angle motions (e.g. rotation of hydroxyl groups) impose a 5 fs limit on the time step, while non-hydrogen atomic collisions (Lennard-Jones “rattling”) restrict the time step to a maximum of 10 fs.<sup>9</sup>

The fastest motions in a simulation will invariably involve hydrogen atoms, since these are by far the lightest atoms present in all biological systems. The biologically relevant behavior of these systems mostly takes place on large time-scales; at least several nanoseconds but often seconds or even beyond. On these time scales hardly any influence can be expected from the tens of femtoseconds long oscillations of hydrogen atoms.

motion:	$f_c$	$I$	Period (fs)		<i>Table 3.1:</i> <i>Characteristic oscillation periods of atomic motions in MD simulations. <math>f_c</math>: force constant; <math>I</math>: moment of inertia, or atomic mass for bond stretching; calc.: calculated from eqn. 3.1; sim.: highest frequency significant peak in spectrum of angle respectively dihedral motion from simulation. An entry of “_” means not applicable, or not determinable.</i>
	(kJ mol <sup>-1</sup> )	(u nm <sup>2</sup> )	calc.	sim.	
bond stretch, H	400 000	m = 1 u	10	10	
bond stretch, heavy atoms	500 000	m = 12 u	30	20	
water libration	—	0.0059	—	28	
water rotation	—	0.0059	—	1300	
angle, H	375	0.010	32	20	
angle, heavy atoms	450	0.27	154	45	
angle -NH <sub>3</sub> <sup>+</sup> group, C-N-H	375	0.010	32	22	
angle -NH <sub>3</sub> <sup>+</sup> group, H-N-H	750	0.010	23	13	
improper, planar	167	—	—	28	
improper, tetrahedral	335	—	—	27	
dihedral, peptide-bond	33	0.20	489	28	
dihedral, -NH <sub>3</sub> <sup>+</sup> group	3.8	0.023	489	89	
dihedral, OH group	1.3	0.0094	53	43	

The obvious solution would be to constrain all bond angles involving hydrogen atoms in all molecules, in addition to all bond lengths. The generally more robust and faster LINCS constraint algorithm<sup>10</sup> is now preferred above SHAKE,<sup>11</sup> but cannot handle the highly connected constraints that arise from constraining both bonds and angles.<sup>10</sup> The cleanest solution would be to eliminate these high-frequency degrees of freedom from the system altogether.

### 3.1.1 Outline of the Method

For hydrogen atoms in large molecules (e.g. proteins) this can be implemented in a rather straightforward manner. Instead of connecting a hydrogen atom with bonds, angles and dihedrals to the molecule, the position of the hydrogen will be generated every MD step based on the position of three nearby heavy atoms. All forces acting on the hydrogen atom will be redistributed over these heavy atoms. A particle that is treated in this manner is referred to as a *dummy atom*.<sup>12,13</sup> To keep the total mass in the system constant, the mass of each hydrogen atom that is treated in this way, should be added to the bonded heavy atom. Care should be taken that for groups with internal rotation (e.g. hydroxyl- or amine-groups) only the other internal degrees of freedom of the group should be fixed, but the rotational freedom should remain.

A special case is the movement of water molecules. The internal geometries of the popular water models are already rigid, so the high-frequency motions are in this case librational motions of the whole water molecule. The only way these motions can be slowed down is by increasing the moments of inertia of the water molecule. This can be done best by increasing the mass of the hydrogen atoms while decreasing the mass of the oxygen atoms, such that the total mass will remain unchanged. A similar modification can be made for groups with internal rotational freedom, as hydroxyl or amine groups, which will display motional frequencies close to those of water. All this, of course, constitutes a non-trivial deviation from “physical reality” and requires justification.

The equilibrium distribution of a system of particles that behaves according to classical statistical mechanics (as in MD), is not dependent on the masses of the individual particles. This is not strictly true if constraints are present because of the mass dependence of the metric tensor correction, but these corrections are usually zero or negligible. Therefore we are at liberty to choose appropriate masses without affecting thermodynamic properties.

This idea was originally proposed by Jacucci & Rahman at a C.E.C.A.M. workshop in 1974<sup>14</sup> and has been investigated in some detail since<sup>12,15,16</sup> and used in our laboratory before.<sup>17</sup> None of these studies has been fine grained in the increase of the hydrogen mass. Moreover, those that had a systematical approach, have not kept the total mass of the system constant and have failed to notice that this in fact scales system time. This is most pronounced in Mao *et al.*<sup>18</sup> who report simulations at 600 K and normal time steps with all masses uniformly increased by a factor of 10, but have kept the velocities appropriate for normal masses. So actually they are simulating normal masses with a  $\sqrt{10}$  times smaller time step at 6000 K. This increase in system temperature accounts for all improved conformational sampling that was reported. Also the approach as outlined

above has not been applied to simulations of biological (macro)molecules, which require additional modifications.

The issue is, again, the time scales that are of interest for the study of biological systems. On these time scales, at least several nanoseconds, the properties of the water will average and the influence of water on the dynamics of the system is exerted through its bulk properties, like diffusion constant, viscosity and dielectric relaxation time. Hydrogen motions in proteins are almost completely uncoupled from the main chain vibrations and will therefore hardly influence the behavior of the system on these time scales.

## 3.2 Methods

### 3.2.1 MD Simulations

All MD simulations were performed using the following parameters and methods, unless stated otherwise. The Verlet integration scheme (leapfrog) was used.<sup>19</sup> The GROMOS-87 forcefield<sup>20</sup> was used, with increased repulsion between water oxygen and carbon atoms.<sup>21</sup> Explicit hydrogens were defined for the aromatic rings, the resulting parameter set is the one referred to as SW by Daura *et al.*<sup>22</sup> Periodic boundary conditions with a rectangular box were applied. LINCS<sup>10</sup> was used to constrain all covalent bonds in non-water molecules. The SETTLE algorithm was used to constrain bond lengths and angles in the water molecules.<sup>23</sup> The temperature was controlled using weak coupling to a bath<sup>24</sup> of 300 K with a time constant of 0.1 ps. Protein and water were independently coupled to the heat bath. Initial velocities were randomly generated from a Maxwell distribution at 300 K, in accordance with the masses that were assigned to the atoms. The pressure was also controlled using weak coupling with a time constant of 1.0 ps.

mass (u)		$I$	$\eta$	$D$	$\tau_{\text{H-bond}}$	drift $E_{\text{tot}}$	$\Delta t_{\text{max}}$
H	O	( $\text{u nm}^2$ )	( $10^{-4} \text{ kg m}^{-1} \text{ s}^{-1}$ )	( $10^{-9} \text{ m}^2 \text{ s}^{-1}$ )	(ps)	$\text{kJ mol}^{-1} \text{ ps}^{-1}$	(fs)
1	16	0.0059	4.3	4.08	0.67	1.04	6.6
2	14	0.0104	4.7	3.89	0.74	0.86	8.9
3	12	0.0133	4.9	3.79	0.89	0.42	10.0
4	10	0.0148	4.9	3.34	0.79	0.36	10.3
5	8	0.0148	5.1	3.50	0.84	0.47	10.4
6	6	0.0133	5.3	3.35	0.84	0.59	8.6
7	4	0.0104	5.2	3.34	0.88	0.43	7.5
8	2	0.0059	5.1	3.60	0.95	0.61	5.6
real H <sub>2</sub> O		—	8.0	2.3	0.59	—	—
real D <sub>2</sub> O		—	—	2.0	—	—	—

Table 3.2: Atomic masses in water; corresponding smallest moments of inertia  $I$ ; resulting dynamical properties: viscosity  $\eta$ ; diffusion constant  $D$ , values of H<sub>2</sub>O and D<sub>2</sub>O from Lide *et al.*<sup>25</sup> and hydrogen-bond lifetime  $\tau_{\text{H-bond}}$ , value of H<sub>2</sub>O from Montrose;<sup>26</sup> RMS drift of the total energy over 12 runs at a time step of 4 fs; maximum time step ( $\Delta t_{\text{max}}$ ) at a maximum order of 10 for the drift as a function of time step.

The starting conformation for the water simulations was generated by equilibrating 820 SPC (Simple Point Charge) water molecules<sup>27</sup> in a 3.2 nm cubic box for 100 ps at 300 K and 1 bar, using a time step of 2 fs and other parameters as stated above. Subsequently, the masses were reassigned (see Table 3.2) and an additional equilibration was done for 10 ps using the same parameters. The same procedure was applied to a smaller (1.9 nm) cubic box containing 216 SPC water molecules and an elongated box of  $1.9 \times 1.9 \times 5.6$  nm containing 648 SPC water molecules. The resulting conformations for each hydrogen mass and box shapes and sizes were used as a starting conformation for all simulations of water.

Short simulations of 1 ps starting from the equilibrated water structures of the small box of 216 SPC waters, for each set of atomic masses were performed to determine the total energy drift as function of atomic mass and time step. Time steps of 0.5 to 15 fs were used; simulations with larger time steps were not stable. A shift function<sup>28</sup> was applied for Coulomb and Lennard-Jones interactions which decreases the potential over the whole region and lets potential and force decay smoothly to zero between 0.5 and 0.75 nm. This introduces some artifacts into the simulation,<sup>29,28</sup> but it effectively removes noise from cut-off effects, enabling an accurate assessment of the simulation accuracy as determined by the time step. Neighbor list generation was performed every time step to exclude possible errors from diffusion of particles in between neighbor list updates. No pressure coupling was applied, instead the system was equilibrated to the right pressure and density and subsequently simulated at constant volume. No temperature coupling was applied.

Longer simulations of 100 ps starting from the equilibrated water structures for each set of atomic masses were performed to determine the dynamical properties of water, i.e. diffusion constant, lifetime of hydrogen bonds and viscosity. For determination of the viscosity an elongated box of 648 SPC waters was used. For the other determinations a cubic box of 820 SPC waters was used. A time step of 2 fs was used. A twin-range cut-off for non-bonded interactions was employed with a short-range cut-off for Lennard-Jones and Coulomb interactions of 0.9 nm which were calculated every simulation step and a long-range cut-off of 1.1 nm for Coulomb interactions which were calculated during neighbor-list generation, every 10 steps (20 fs). Neighbor searching was done based on the centers of geometry of the water molecules.<sup>30</sup> This is the same parameter set as was used by Van der Spoel *et al.*<sup>31</sup>

The simulations of a protein were performed with the small protein HPr (NMR PDB entry 1HDN<sup>32</sup>). This 85 residue  $\alpha/\beta$  protein consists of a four-stranded anti-parallel  $\beta$ -sheet flanked on one side by three anti-parallel  $\alpha$ -helices. The protein was solvated by generating a cubic box of SPC water molecules, such that the minimum distance between the protein and the edge of the periodic box would be 0.6 nm, resulting in a cubic box of 4.7 nm. All water molecules from the generated box of water that were within 0.23 nm of a protein atom were removed, leaving 2985 water molecules around the protein. The resulting conformation was energy minimized with harmonic constraints on the atomic coordinates of the protein. Subsequently a round of 10 ps of MD was performed, also with harmonic constraints on the atomic coordinates of the protein to relax the water orientation near the protein. The final conformation was used as starting conformation for simulations of the protein in water. Additionally, another 10 ps of MD was performed starting from this

conformation. From these two final conformations the water was removed and the protein was allowed to relax for 1 ps in a vacuum environment. The final conformations from both 1 ps vacuum simulations were used as starting conformations for simulations of the protein in vacuum.

Short simulations of 1 ps of the protein in vacuum were performed to determine the total energy drift as function of time step. Four different topology types were used: the normal topology (“normal 1  $u$ ”), with the hydrogen atoms fourfold increased in mass (“normal 4  $u$ ”), with dummy hydrogens (“dummy 1  $u$ ”) and with dummy hydrogens and remaining hydrogens fourfold increased in mass (“dummy 4  $u$ ”). Time steps of 0.5 to 7 fs were used; simulations with larger time steps were not stable. No cut-off for Lennard-Jones or Coulomb interactions was applied and no periodic boundary conditions were used. No temperature coupling was applied. Although simulations of a protein in vacuum are generally not relevant to the majority of applications, these simulations do allow for a relatively accurate estimate of the energy drift, which is not possible for a simulation of the protein in water. Distortions of the shape of the protein by the vacuum environment might influence energy drift, but are unlikely to occur within the 1 ps duration of the simulations.

Long simulations of 1 ns of the protein in water were performed to determine the long-term properties of a protein using the “normal 1  $u$ ”, “normal 4  $u$ ”, “dummy 1  $u$ ” and “dummy 4  $u$ ” topologies, with time steps ranging from 1 to 7 fs. Long-range Coulomb interactions were calculated using PPPM<sup>33,34</sup> with a grid spacing of 0.09 nm. Neighbor list generation was performed every 20 fs. No pressure coupling was applied, instead the system was equilibrated to the right pressure and density and subsequently simulated at constant volume, resulting in average system pressures ranging from 14 to 80 bar with an average of 44 bar. Within this range of pressures, no influence is likely to exist on the properties of the protein.

All MD simulations were carried out using the GROMACS molecular dynamics package<sup>35,36</sup> on a Silicon Graphics (SGI) Power Challenge with MIPS R 10 000 processors and on SGI O2 Workstations with MIPS R 5 000 processors. CPU times for the long (1 ns) runs of the protein in water on the SGI Power Challenge machine are summarized in Table 3.2.1, for a total of 130 days of CPU time.

$\Delta t$ (fs)	$N_{steps}$ ( $\times 10^3$ )	CPU time
1	1000	341
2	500	171
3	333	114
4	250	86
5	200	68
6	167	57
7	143	49

*Table 3.3: Summary simulation parameters of long (1 ns) simulations of the protein in water. From left to right: time step ( $\Delta t$ ); number of integration steps ( $N_{steps}$ ); total runtime on an SGI Power Challenge with MIPS R 10 000 processors (CPU time).*

### 3.2.2 System Topology

Normal topologies, with constraints on all bonds and no constraints on angles, were generated using standard GROMACS topology building tools. These tools were modified to include an option to produce the modified topologies containing the dummy atoms and remove all bond, angle and dihedral definitions that have become obsolete due to the introduction of the dummy atoms. Also optionally, masses of all remaining normal hydrogen atoms can be increased by a factor of 4, while subtracting this increase from the bonded heavy atom. More details are included in the following section.

#### Construction of Dummy Atoms

The goal of defining hydrogen atoms as dummy atoms is to remove all high-frequency degrees of freedom from them. In some cases not all degrees of freedom of a hydrogen atom should be removed, e.g. in the case of hydroxyl or amine groups the rotational freedom of the hydrogen atom(s) should be preserved. Care should be taken that no unwanted correlations are introduced by the construction of dummy atoms, e.g. bond-angle vibration between the constructing atoms could translate into hydrogen bond-length vibration. Additionally, since dummy atoms are by definition mass-less, in order to preserve total system mass, the mass of each hydrogen atom that is treated as dummy atom should be added to the bonded heavy atom.

All forces acting on the dummy atoms must be redistributed over the atoms used in the construction of the dummy atom.<sup>13</sup> Dummy atom positions can be easily calculated from any three atoms that have a fixed orientation with respect to each other. If the constructing atoms move significantly with respect to each other, normalized vectors will have to be used to ensure the right position for the dummy atom. This results in complicated derivatives in the force redistribution, which are described in detail in appendix A.

Taking into account these considerations, the hydrogen atoms in a protein naturally fall into several categories, each requiring a different approach, see also Fig. 3.1:

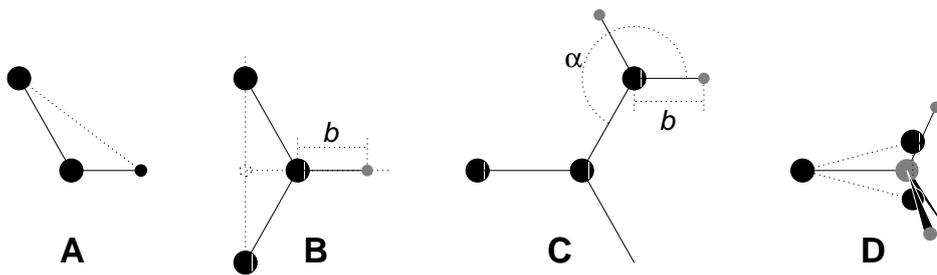


Figure 3.1: Schematic view of the different types of dummy atom constructions used. The atoms used in the construction of the dummy atom(s) are depicted as black circles, dummy atoms as grey ones. Hydrogens are smaller than heavy atoms. A: fixed bond angle, note that here the hydrogen is not a dummy atom; B: in the plane of three atoms, with fixed distance; C: in the plane of three atoms, with fixed angle and distance; D: construction for amine groups ( $-\text{NH}_2$  or  $-\text{NH}_3^+$ ), see text for details.

- *hydroxyl (-OH) or sulfhydryl (-SH) hydrogen*: The only internal degree of freedom in a hydroxyl group that can be constrained is the bending of the C-O-H angle. This angle is fixed by defining an additional bond of appropriate length, see Fig. 3.1A. This removes the high frequency angle bending, but leaves the dihedral rotational freedom. The same goes for a sulfhydryl group. Note that in these cases the hydrogen is not treated as a dummy atom.
- *single amine or amide (-NH-) and aromatic hydrogens (-CH-)*: The position of these hydrogens cannot be constructed from a linear combination of bond vectors, because of the flexibility of the angle between the heavy atoms. Instead, the hydrogen atom is positioned at a fixed distance from the bonded heavy atom on a line going through the bonded heavy atom and a point on the line through both second bonded atoms, see Fig. 3.1B.
- *planar amine (-NH<sub>2</sub>) hydrogens*: The method used for the single amide hydrogen is not well suited for planar amine groups, because no suitable two heavy atoms can be found to define the direction of the hydrogen atoms. Instead, the hydrogen is constructed at a fixed distance from the nitrogen atom, with a fixed angle to the carbon atom, in the plane defined by one of the other heavy atoms, see Fig. 3.1C.
- *amine group (umbrella -NH<sub>2</sub> or -NH<sub>3</sub><sup>+</sup>) hydrogens*: Amine hydrogens with rotational freedom cannot be constructed as dummy atoms from the heavy atoms they are connected to, since this would result in loss of the rotational freedom of the amine group. To preserve the rotational freedom while removing the hydrogen bond-angle degrees of freedom, two “dummy masses” are constructed with the same total mass, moment of inertia (for rotation around the C-N bond) and center of mass as the amine group. These dummy masses have no interaction with any other atom, except for the fact that they are connected to the carbon and to each other, resulting in a rigid triangle. From these three particles the positions of the nitrogen and hydrogen atoms are constructed as linear combinations of the two carbon-mass vectors and their outer product, resulting in an amine group with rotational freedom intact, but without other internal degrees of freedom. See Fig. 3.1D.

Additionally, all bonds, angles and dihedrals that are defined on one of the degrees of freedom that were removed, are also removed. This boils down to removing all bonds to dummy atoms, all angles that involve two or three dummy atoms and all dihedrals that involve at least one dummy atom and for which all other atoms are used in constructing the dummy atom(s). Note that this leaves the whole force field unchanged.

As a second option, all remaining hydrogen atoms (i.e. in hydroxyl, sulfhydryl, amine groups and water) can be increased in mass, with the increase subtracted from the bonded heavy atom. This leaves the total mass constant, but increases the moment of inertia of the group, effectively slowing down the motions. For the dummy mass and dummy atom construction of the amine group (type D as described above) this will have the net result of the dummy masses being placed further apart in accordance with the desired increase in moment of inertia.

It should be noted that in the GROMOS-87 forcefield aliphatic hydrogens are implicit, i.e. they are represented as united carbon-hydrogen atoms. For an all-atom forcefield

with all hydrogen atoms explicit an additional number of hydrogen dummy atoms will have to be constructed every time step, e.g. for the 85 residue protein, 165 hydrogen atoms are present in the GROMOS-87 forcefield but an additional 488 aliphatic hydrogens are implicit. Constructing the 165 hydrogen dummy atoms takes far less than 0.1% of the total computer time, so constructing 653 hydrogen dummy atoms will still have no noticeable effect on the total cost of simulating.

### Out-of-plane Vibrations in Aromatic Groups

When using dummy-atom constructions for non-rotating explicit hydrogens together with a time step of 7 fs at 300 K,<sup>1</sup> occasional instability of the aromatic groups occurs as a result of out-of-plane vibrations. These can be eliminated by using additional dummy-atom constructions.<sup>2</sup> The total mass of the whole group is redistributed to three of the heavy atoms. This is done such that the center of mass and moments of inertia are preserved as closely as possible. Constraints between the three atoms fix the geometry and all other atoms are constructed as dummy atoms from them.

For phenylalanine and histidine this is trivial, the heavy atoms at  $\gamma$  and  $\epsilon$  positions are kept as normal atoms. For tyrosine  $O_\eta$  and  $H_\eta$  also remain as normal atoms, with constraints between  $C_{\epsilon 1}$ ,  $C_{\epsilon 2}$  and  $O_\eta$ .  $C_\zeta$  is constructed from these three atoms. The bond angle in the hydroxyl group ( $C_\zeta-O_\eta-H_\eta$ ) is constrained by a constraint between  $C_\gamma$  and  $H_\eta$  (constraints involving a dummy atom are not implemented). The planarity of the whole group (which could still fold around  $C_{\epsilon 1}-C_{\epsilon 2}$ ) is preserved by the original improper dihedral angles, but without the instability of the separately moving atoms. For tryptophan two interaction-less masses are created at the center of mass of each of the rings and with constraints between them and with  $C_\beta$ . In Fig. 3.2 a schematic representation of these constructions is shown.

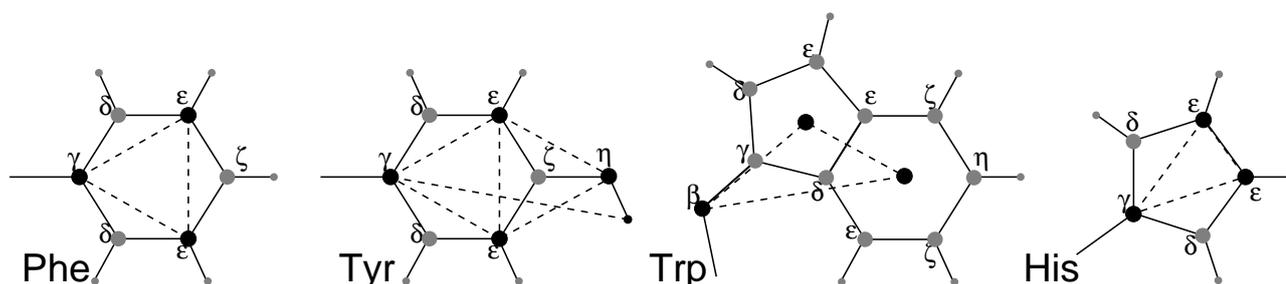


Figure 3.2: Schematic view of the different types of dummy atom constructions used for aromatic sidechains. Legend: ● atoms and masses used in the construction of the dummy atom(s); ● dummy atoms; — chemical bonds; - - constraints. Hydrogens are smaller than heavy atoms. Note, the hydroxyl hydrogen is not a dummy atom in tyrosine, and the constraint between  $C_{\epsilon 1}$  and  $C_{\epsilon 2}$  in histidine.

### 3.2.3 Determination of System Properties

#### Motional Periods

Periods of oscillation were measured from a simulation of the protein in water with a time step of 0.5 fs, by taking the highest frequency significant peak from the spectrum of the motion. Periods are also calculated from the forcefield parameters using:

$$T = 2\pi\sqrt{\frac{I}{f_c}} \quad (3.1)$$

where  $I$  is the moment of inertia, determined by bond lengths and masses, and  $f_c$  the force constant of the corresponding angle or dihedral potential, results are listed in Table 3.1. Eqn. 3.1 neglects effects of coupling with the surroundings.

#### Energy Drift

As outlined above, the simulations used to determine the drift in the total energy of the simulated system were performed with neither temperature nor pressure coupling. For the water box a shifted potential for electrostatic and Lennard-Jones interactions was used to eliminate cut-off effects and for the protein no cut-off for interactions was used. This was done to minimize as much as possible all sources of integration errors (notably cut-off effects) except for those caused by the ratio between time step and fastest motional periods. Also, double precision (8 bytes) floating point calculations were used during the simulation. The limited accuracy of summing up millions of interactions in single precision gives rise to additional drift that obscures the effects we want to investigate, especially in the smaller time-step regime. As a measure of the accuracy, the root mean square (RMS) averaged drift in the total energy obtained from a least squares linear fit over the last 0.9 ps of 1 ps simulations is used.<sup>8</sup>

Since the drift in the total energy of a well-integrated system is diffusive in nature, an appreciable number of independent simulations needs to be performed in order to get an accurate estimate of this drift.<sup>8</sup> The root-mean-square drift provides a reliable measure for the accuracy of the simulation. For the water box, 12 simulations were performed for each time-step/mass combination, each with a different random seed to generate initial velocities. For the protein, two different starting conformations were used, for which 6 simulations with different random seeds were performed for each time-step/topology-type combination. This adds up to 1440 one-ps simulations of the water box, for a total of 12.4 hours CPU time on a MIPS R 10 000 processor, and 560 one-ps simulations for the protein, for a total of 5.4 hours of CPU time.

The fluctuation of the total energy, which can be determined easily from a single simulation, is an inappropriate measure to assess the simulation accuracy of a Verlet-type (leapfrog) integration scheme, since it is of second order in time step, while the drift is of second to third order.<sup>8</sup>

### Diffusion Constant

Diffusion constants ( $D$ ) for water were calculated from the mean square displacement (MSD) of the water molecules using the Einstein relation for diffusion in three dimensions:<sup>37</sup>

$$\langle \|\mathbf{r}(t) - \mathbf{r}(0)\|^2 \rangle = 6Dt \quad (3.2)$$

$D$  was determined by a linear fit to the plot of the MSD *vs.* time.

### Viscosity

The procedure for determination of the viscosity was modified after Berendsen.<sup>38</sup> Viscosity was determined in a non-equilibrium simulation setup where an external shear-stress acceleration field was applied:

$$a_{i,x} = A \cos\left(\frac{2\pi z_i}{l_z}\right) \quad (3.3)$$

with  $a_{i,x}$  being the acceleration in the  $x$  direction,  $A$  the acceleration amplitude,  $z_i$  the  $z$ -coordinate of the particle,  $l_z$  the length of the box in the  $z$ -direction. Application of this shear-stress acceleration gradient induces a velocity gradient of the same shape. For a Newtonian fluid the dynamic viscosity  $\eta$  is given as the ratio between the applied acceleration amplitude  $A$  and the resulting velocity amplitude  $V$ :<sup>38</sup>

$$\eta = \frac{A}{V} \rho \left(\frac{l_z}{2\pi}\right)^2 \quad (3.4)$$

where  $\rho$  is the density and  $l_z$  the box length in the  $z$ -direction of the system.

The scaling procedure used in temperature coupling was modified to exclude the induced velocity gradient while applying temperature scaling.

Care was taken to choose the acceleration amplitude low enough to prevent the appearance of ordering in the water and high enough to get a velocity gradient that is discernible over the thermal velocities. An amplitude of  $0.07 \text{ nm ps}^{-2}$  was found to perform the best (E. Apol, personal communications); this results in a velocity amplitude of the order of  $0.1 \text{ nm ps}^{-1}$  which corresponds to roughly 10% of the root-mean-square thermal velocity at 300 K, which is  $1.1 \text{ nm ps}^{-1}$ . For the same reason, the acceleration field was applied along the longest edge (3 times the length of the other edges) of the rectangular simulation box.

The velocity amplitude  $V$  was calculated using a spatial Fourier component:<sup>38</sup>

$$V = \frac{2}{N} \sum_i v_{i,x} \cos\left(\frac{2\pi z_i}{l_z}\right) \quad (3.5)$$

which was stored at every time step. The viscosity was calculated from eqn. 3.4, using the average velocity amplitude over the last 90 ps of the simulations to exclude the equilibrational part in which buildup of the velocity gradient still occurs.

### Lifetime of Hydrogen Bonds

Hydrogen bonds between water molecules were defined using a simple angle and distance criterion, i.e. angle hydrogen-donor-acceptor  $\leq 60^\circ$  and distance donor-acceptor  $\leq 0.35$  nm, yielding a switch function which is 1 when a hydrogen bond is present and 0 otherwise. The hydrogen-bond lifetime is determined as the half-life time of the autocorrelation of the switch function.

## 3.3 Results

### 3.3.1 Water

#### Energy Drift

In Fig. 3.3A the energy drift is plotted as a function of hydrogen mass and time step. Note that for clarity the graphs for a hydrogen mass of 2, 3, 6 and 7  $u$  are not shown, they all lie in between the plotted graphs. The drift as a function of time step is of second order, which lies within the expected range of second to third order.<sup>8</sup> At time steps above 7 fs for the 1 and 8  $u$  simulations and above 10 fs for the 4 and 5  $u$  simulations a transition occurs from second to higher order.

Taking a rather arbitrary value of 10 as a maximum for the order of the drift as a function of the time step, a maximum time step can be determined for each hydrogen mass,

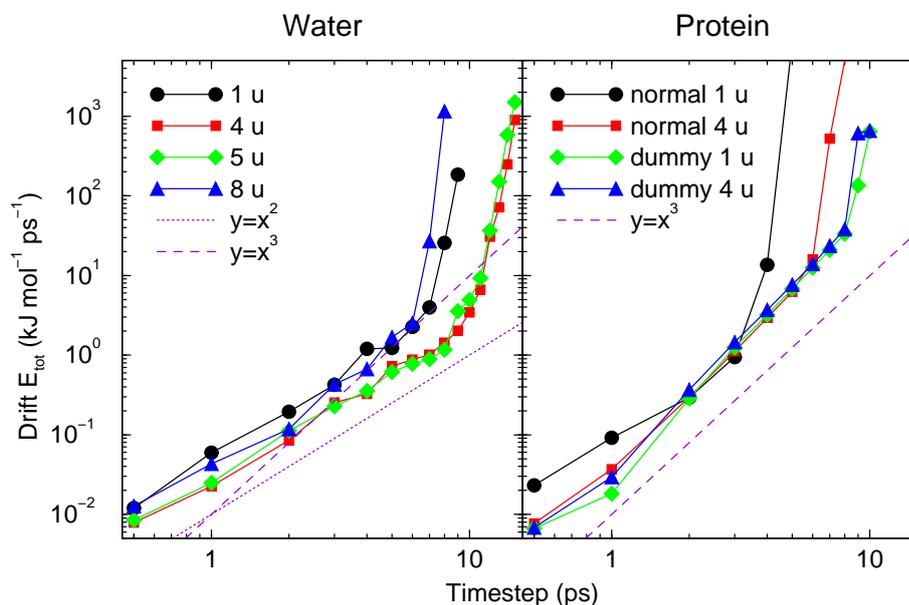


Figure 3.3: RMS averaged drift in the total energy as a function of time step of A: a box of SPC water (648 atoms) with different hydrogen masses (graphs for 2, 3, 6 and 7  $u$  are omitted for clarity). B: a small protein (805 atoms) with different topology types (“normal 1  $u$ ”, “normal 4  $u$ ”, “dummy 1  $u$ ” and “dummy 4  $u$ ”).

as is summarized in Table 3.2, these time steps correspond to the sharp increase in the slope of the plots in Fig. 3.3A. For water with the normal mass distribution, the maximum time step is 6.6 fs. The largest attainable time step of 10.4 fs occurs with water with a hydrogen mass of 5  $u$ , an increase of a factor of 1.6. According to eqn. 3.1 this is consistent with the increase in moment of inertia of a factor of 2.5. This correspondence is a clear indication that the librational frequency of (SPC) water is the major factor determining the maximum possible time step for accurate integration. Alternatively, a maximum for the magnitude of the drift could be chosen to determine maximum time-steps, however, these will be rather inconsistent due to the large fluctuation that is present in the magnitude of the drift. The very sharp increase in the order of the drift allows for a much more accurate determination of maximum time-steps.

### Diffusion

The diffusion constant as determined from the simulations for different atomic masses ranges from  $4.1 \cdot 10^{-9} \text{ m}^2 \text{ s}^{-1}$  at a hydrogen mass of 1  $u$  to  $3.3 \cdot 10^{-9} \text{ m}^2 \text{ s}^{-1}$  at a mass of 4  $u$ , or a decrease of a factor of 1.2 (see Table 3.2). Compared to the difference between diffusion constants of SPC water ( $4.1 \cdot 10^{-9} \text{ m}^2 \text{ s}^{-1}$ ) and real water ( $2.3 \cdot 10^{-9} \text{ m}^2 \text{ s}^{-1}$ ), a difference of a factor of 1.8, the variation caused by changing the hydrogen mass is relatively small.

### Viscosity

Viscosity for different atomic masses ranges from  $4.3 \cdot 10^{-4} \text{ kg m}^{-1} \text{ s}^{-1}$  at a hydrogen mass of 1  $u$  to  $5.3 \cdot 10^{-4} \text{ kg m}^{-1} \text{ s}^{-1}$  at a mass of 6  $u$ . At a mass of 4  $u$ , which yields the highest maximum time step, the viscosity is  $4.9 \cdot 10^{-4} \text{ kg m}^{-1} \text{ s}^{-1}$ , an increase of a factor of 1.1 (see Table 3.2). Compared to the difference between the viscosity of SPC water ( $4.3 \cdot 10^{-4} \text{ kg m}^{-1} \text{ s}^{-1}$ ) and real water ( $8.0 \cdot 10^{-4} \text{ kg m}^{-1} \text{ s}^{-1}$ ), a factor of 1.9, the variation caused by changing the hydrogen mass is small. Still the result can be significant in the sense that large-scale consorted motions in the simulation, e.g. domain motions of proteins, will be limited by the viscosity of water, which means that a higher viscosity of the water will result in slower protein motion. It can be expected that the maximum slowing-down will be of similar order as the increase in the viscosity of the water, i.e. about 14%.

### Hydrogen Bond Lifetime

The hydrogen bond lifetime increases monotonically with the hydrogen mass, from 0.67 ps at a mass of 1  $u$  to 0.95 ps at 8  $u$  (see Table 3.2), when the single deviating value at 3  $u$  is ignored. At a hydrogen mass of 4  $u$  the lifetime is 0.79 ps, which is an increase of a factor of 1.2 with respect to the value of normal SPC water. The lifetime of hydrogen bonds in normal SPC water (0.67 ps) may be compared with an experimental estimate of 0.59 ps,<sup>26</sup> on the basis of fluctuations in the anisotropy of molecular polarizability as determined from depolarized Rayleigh scattering measurements.

### 3.3.2 Protein

#### Energy Drift

For time steps of 1 fs and below, the energy drift of the protein (in vacuum) is diffusive in nature, as was the case for the water box. In Fig. 3.3B the magnitude of the drift is plotted against the time step used. The diffusive nature gives rise to relatively large variations in the determined drift. For larger time steps the drift becomes systematic and is always positive. The order of the drift as a function of time step also lies within the theoretically expected range of second to third order, see Fig. 3.3B. Surprisingly, the magnitude of the drift is virtually identical for all topology types, the only difference between them is the time step at which a transition from third to higher order occurs. These transition time-steps are summarized in Table 3.4, column A.

#### Long-term Properties

From Table 3.5 it is immediately obvious that the average RMS difference from the starting structure in the last 100 ps of each run with respect to the starting structure, the secondary structure content in the last 900 ps and the average total number of hydrogen bonds in the last 900 ps are not noticeably influenced by the introduction of dummy atoms, heavy

topology type	$\Delta t_{max}$ (fs)	
	A	B
normal 1 u	3	3
normal 4 u	6	4
dummy 1 u	8	7
dummy 4 u	8	7

Table 3.4: Summary of maximum time steps  $\Delta t_{max}$  for which A: the drift of the total energy as a function of the time step in short (1 ps) simulations (in vacuum) is still of third order; B: long (1 ns) simulations (in water) can be performed without errors.

$\Delta t$ (fs)	RMS difference (nm)				sec. struct. (%)				# H-bonds			
	normal		dummy		normal		dummy		normal		dummy	
	1 u	4 u	1 u	4 u	1 u	4 u	1 u	4 u	1 u	4 u	1 u	4 u
1	0.11	0.12	0.18	0.11	87	86	85	84	114	112	117	114
2	0.15	0.15	0.17	0.13	86	85	87	85	114	111	116	118
3	0.15	0.09	0.17	0.13	87	87	82	86	114	115	114	116
4	–	0.16	0.15	0.14	–	87	85	86	–	114	116	114
5	–	–	0.12	0.16	–	–	87	85	–	–	116	114
6	–	–	0.13	0.15	–	–	88	85	–	–	117	118
7	–	–	0.11	0.18	–	–	89	84	–	–	113	116

Table 3.5: Summary of results from long (1 ns) simulations of the protein in water performed at different time steps ( $\Delta t$ ) and with “normal 1 u”, “normal 4 u”, “dummy 1 u” and “dummy 4 u” topologies. Long term average properties: RMS difference of all backbone atoms with respect to the starting structure, averaged over last 100 ps; secondary structure content (% of residues not in random-coil conformation, according to the DSSP program<sup>39</sup>) averaged over 100 to 1000 ps; number of inter-protein hydrogen bonds averaged over 100 to 1000 ps. Entries of “–” indicate failure of the simulation to run without errors, this was also the case for time steps larger than 7 fs.

hydrogen atoms or large time steps.

In Table 3.4, column B, a summary is given of the maximum time steps for which a 1 ns simulation could be performed without errors. These time steps are somewhat smaller than those obtained from monitoring the energy drift, as summarized in column A. Since the energy drift was determined from simulations in vacuum, it could be expected that this difference is due to the interactions between protein and water (data not shown). However, tests of long simulations of the protein in vacuum yield the same maximum time steps as those found for the protein in water. This means that the time-step limiting factors arise from the protein and not from the water, as can also be seen from comparison of the maximum time steps found for the water box (see Table 3.2) and for the protein in vacuum (see Table 3.4).

In Fig. 3.4 inter-protein hydrogen-bond distance and angle distributions are shown. It is clear that the introduction of dummy atoms gives rise to slightly broader distributions. This is to be expected since for normal hydrogens the bond angle can adjust itself to accommodate an optimal hydrogen-bonding conformation. Removing this freedom by constructing the hydrogen as a dummy atom, makes adjustment impossible, giving rise to more sub-optimal hydrogen-bonding conformations and hence a slightly broader distribution. In contrast, no effect on the distributions is noticeable from changing hydrogen masses or time step (data not shown).

In Fig. 3.5 dihedral angle distributions are plotted for the  $\text{C-NH}_3^+$  dihedral angle, averaged over time steps (Fig. 3.5A) and topology types (Fig. 3.5B). It is clear that neither introduction of dummy atoms in the  $-\text{NH}_3^+$  group, nor increase of mass of the hydrogen atoms nor taking larger time steps has a noticeable effect on the dihedral angle distributions.

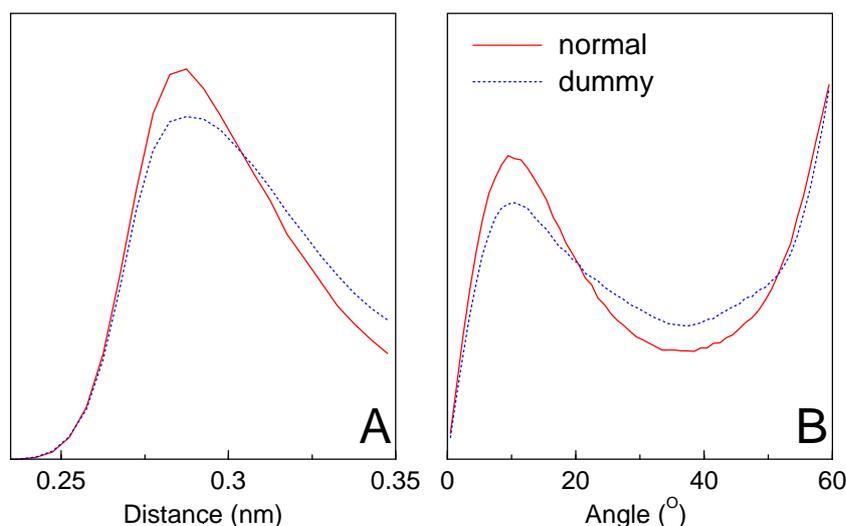


Figure 3.4:  
Inter-protein  
hydrogen-bonds:  
A donor-acceptor  
distance distribution  
and B hydrogen-donor-  
acceptor angle  
distribution, averaged  
over all simulations  
without dummy atoms  
("normal") and with  
dummy atoms  
("dummy"). All  
distributions appear to  
be insensitive to  
changes in time step  
and hydrogen masses.

### 3.4 Discussion and Conclusion

In MD simulations of proteins, in which bond lengths are constrained, the usual time step is 2 fs. This is only slightly below the absolute maximum, which is shown to be 3 fs, the 2 fs maximum being a practical limit imposed by the use of the SHAKE algorithm. In order to perform simulations at larger time steps, the hydrogen degrees of freedom should be further restricted. This can be done either by defining additional constraints, or by treating hydrogen atoms as “dummy” atoms which are constructed from neighboring “real” atoms. We have chosen the latter approach because it (*i*) avoids problems with constraints in planar groups, (*ii*) allows the combination of two constrained and one flexible angles in a plane (as for the backbone amide proton) and (*iii*) enables the use of the more stable LINCS constraint algorithm instead of SHAKE to satisfy constraints.

The removal of hydrogen degrees of freedom is not expected to cause a noticeable disturbance in the physical behavior of the system on longer time scales because the hydrogen motions are almost uncoupled from the main chain vibrations. This stands in contrast to the strongly coupled heavy-atom bond-angle vibrations that influence the accessible configurational space and can therefore not be treated as constraints.<sup>40</sup> Treating all hydrogens as dummy atoms (“dummy 1 u”) allows the time step to be increased by a factor of 2.3 (see Table 3.4, column B).

The bottleneck is now the internal rotation or libration of hydrogen-containing groups and of water molecules (see Table 3.1). The frequencies related to such motions will scale inversely proportional to the square root of the moments of inertia and can thus be decreased by modification of the atomic masses. For classical simulations the thermodynamic properties do not depend on the (distribution of) atomic masses. Dynamic properties of a protein on longer time scales will only weakly depend on the mass of hydrogen atoms in the protein and depend on the properties of water through its bulk transport properties.

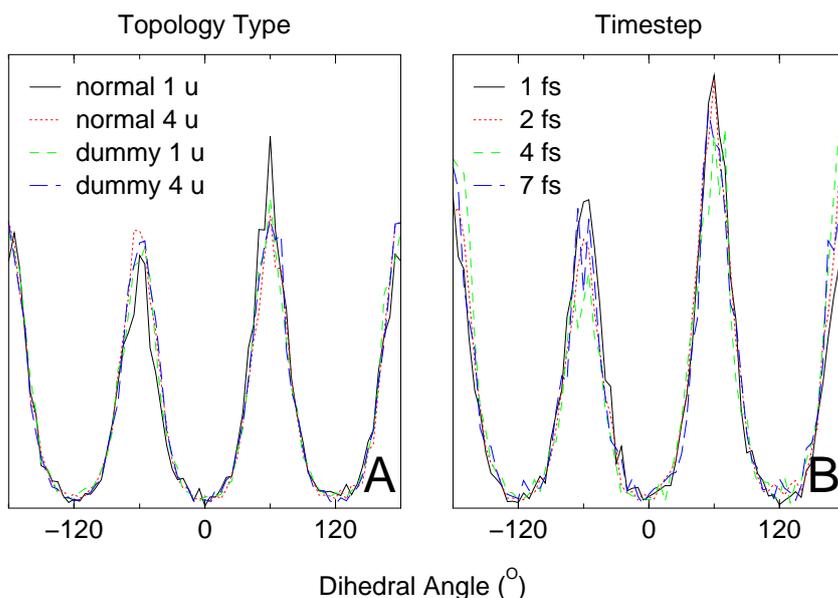


Figure 3.5: Dihedral angle distributions in  $\text{Lys24-NH}_3^+$  of the protein. A: averaged over all time steps for each topology type; B: averaged over topology types for time steps of 1, 2, 4 and 7 fs (see Table 3.5 for an overview of simulations at different time steps and topology type).

Increase of hydrogen mass by a factor of 4 with simultaneous decrease of the mass of the bonded heavy atom to preserve the total mass of the group (“normal 4  $u$ ”), only allows for a modest increase of a factor of 1.3 (see Table 3.4, column B). Combining the use of dummies with mass modification (“dummy 4  $u$ ”) allows for an increase in time step of 2.3, which is identical to that observed for “dummy 1  $u$ ”. It appears that no additional gain comes from increasing hydrogen masses in a system where most hydrogen atoms in the protein are already treated as dummy atoms, however, a gain in simulation stability is to be expected.

The viscosity of water increases and diffusion coefficient decreases by roughly 15%; therefore the net gain in simulation efficiency can be up to 15% lower than expected based on the increased time step. An additional factor can also result in an efficiency gain that will be slightly less; when a neighbor list is used, which is to be updated for example every 20 fs, counting in integration steps it must be updated more frequently.

Using both dummy atoms and modified masses, the next bottleneck is likely to be formed by the improper dihedrals (which are used to preserve planarity or chirality of molecular groups) and the peptide dihedrals. Although the improper dihedrals could be replaced by dummy atom constructions or their potential function be modified to reduce the resonance frequencies, the peptide dihedral cannot be changed without affecting the physical behavior of the protein. Thus we have approached the limit of what can be achieved without affecting the physical behavior.

Additional gain in simulation stability can be achieved by replacing the remaining improper dihedral constructions on the planar aromatic amino acid sidechains. Coupling between the several improper dihedrals in a aromatic group results in harmonics with higher characteristic oscillation periods than those found for the peptide bond, thereby posing a separate obstacle to an increased timestep. From simulations described in chapter 4 is found that this increases the simulation stability.

We would like to conclude from this discussion that measuring the drift in total energy of a simulation allows one to determine the maximum time-step given a maximum order of the drift as a function of the time step. Table 3.4, column A, shows that this criterion would allow for a time step of 3 fs for normal simulation, and an increase of a factor of 2 for simulating with hydrogen atoms of 4  $u$ . Introducing dummy hydrogens atoms will allow for a gain in maximum time-step of factor of 2.7, irrespective of the mass of remaining hydrogen atoms. As is evident when comparing columns A and B in Table 3.4, in real-life examples the maximum time-steps will be somewhat lower. It appears that monitoring the total energy drift fails to capture some important features of the simulated system which determine the integration accuracy and stability. This is most pronounced for the “normal 4  $u$ ” case; based on energy drift a maximum time step of 6 fs would be expected, but an actual simulation of a protein in water remains stable only up to time steps of 4 fs.

It seems therefore best to choose an important property (or a number of properties) of the system for which a reference value or distribution is known (either from experiment or from an accurately performed reference simulation), and monitor this property during simulation. This, however, gives rise to a new problem because for many systems, such a property might be hard to find.

The most practical approach to determine the maximum time-step is simply to determine for which time steps the simulation itself will remain stable. From Table 3.4 it can be seen that increasing hydrogen atom mass will allow for a modest increase in time step from 3 to 4 fs, introducing dummy hydrogen atoms, however, allows the time step to be increased to 7 fs and the combination of increased hydrogen atom mass and dummy hydrogen atoms will give the same time step of 7 fs, but with a slightly less fluctuations in various simulation parameters, and presumably a better long-term stability.

In a concluding summary we can say that an increase in time step from 2 to 7 fs, a factor of 3.5, for routine MD simulations of proteins in water can be achieved by constructing hydrogen atoms in the protein as dummy atoms, leading to a gain in simulation efficiency of a factor of 3 to 3.5. Additional simulation stability can be gained by increasing the mass of all hydrogen atoms with remaining degrees of freedom from 1 to 4  $u$  and by using dummy constructions for the planar aromatic groups.

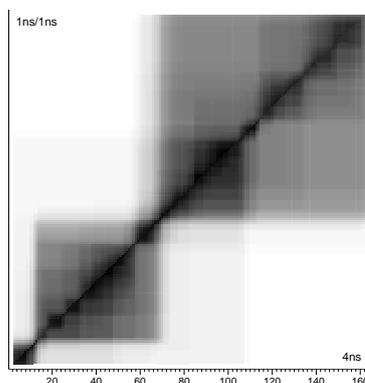
### 3.5 Acknowledgments

We would like to thank Dr. M. L. C. E. Kouwijzer for stimulating discussions and to Dr. A. K. Mazur for some very helpful suggestions. Thanks also go to our undergraduate students for conducting a number of preliminary studies: M. Bergsma, R. M. de Blouw, S. J. Boot, A. Duursma, M. van Faassen, S. Harkema, D. R. Hekstra, M. Mol and C. Smit.



## Chapter 4

# Hierarchical Motions in Small Proteins: Implications for Protein Folding





## 4.1 Introduction

Experimental estimates of the lower bound of the folding time of small proteins range from 1 to 70 microseconds.<sup>41,42</sup> Using atomic models and explicit solvent, the maximum length of time that a protein in solution can be simulated is currently on the order of hundreds of nanoseconds,<sup>43</sup> or, by supreme effort, a few microseconds.<sup>44</sup> This puts equilibrium simulations of protein folding well out of range for the foreseeable future. For this reason there is much interest in the development of simplified models that capture the essential features of a protein and the folding process, while reducing the computational cost.

One possible approach to this problem is based on the nucleation model of protein folding.<sup>45,46,47,48</sup> This proposes that proteins are constructed in a modular fashion from independent folding units or “foldons”<sup>49</sup> with a hierarchical organization. If this model is a realistic representation of protein folding, the identification of such folding elements would be an important first step in the development of simplified models to simulate protein folding. Several theoretical methods have been proposed to identify such modules, which are usually assumed to be rigid elements in a protein. Most of these methods are based on ad-hoc definitions of what will be a rigid element. Often rigid elements are simply defined as secondary structure elements.<sup>41,50,51,52,53,54</sup>

Here, we propose and test a novel method to analyze the dynamics of a protein in its native state, as obtained from Molecular Dynamics (MD) simulations, in terms of motionally coherent elements. Preliminary aspects of this work have been described previously.<sup>7</sup> In this paper we apply the approach to a series of five proteins in order to validate the approach and correlate the elements identified with folding nucleation sites proposed by others.

### 4.1.1 Outline of the Approach

Central to our approach is the identification of groups of residues that show large-scale correlated motion in a number of separate simulations or in separate parts of a single long simulation. The aim is to obtain a consensus regarding the dynamics of particular regions of the proteins. From this one can then extract motionally coherent elements which are groups of residues that, on average, show correlated motional properties. The fact that such groups or elements show distinct motional properties also suggests that they are independently stable and thus possible candidates for foldons or folding nucleation sites.

To identify groups of residues that show large-scale correlated motion in a set of independent simulations a two step procedure was used. First, in order to focus only on large scale collective modes, uncorrelated local fluctuations motions were filtered out by performing a form of Principal Components Analysis and retaining only the dominant modes. Based on the predominant collective modes obtained from the Principal Components Analysis motional elements were then identified by clustering groups of residues that undergo rotation around a similar axis and also show little internal motion. That is, the residues must not display significant motion relative to other members of the cluster. This leads to a definition of such elements for each collective mode analyzed. From the global set of motional elements, a matrix can be constructed by counting the number of

times a given pair of residues is found within the same element. Statistics on the element definitions can be improved by performing analysis of additional simulations or by analysis of separate parts of a large simulation. From this “consensus dynamics matrix” a set of consensus groups of residues with large scale correlated motion can be identified.

## 4.2 Methods

### 4.2.1 Simulations

#### Proteins

Five proteins were selected as test-cases (the PDB<sup>55</sup> entry labels are given in parentheses):

**Sh3** (**1shf**<sup>56</sup>/**1nyf**<sup>57</sup>):  $\beta$  barrel, 59 residues (Fyn-Sh3), Sh3 domain from human fyn proto-oncogene tyrosine kinase, crystal structure (**1shf**, 2 molecules in the asymmetric unit) and NMR structure (**1nyf**);

**T4L** (**2lzm**<sup>58</sup>):  $\alpha/\beta$  protein, 164 residues (T4-Lys), bacteriophage T4 lysozyme, crystal structure (**2lzm**);

**HPr** (**1hdn**<sup>32</sup>/**1poh**<sup>59</sup>):  $\alpha$ /anti-parallel- $\beta$  sandwich, 85 residues (HPr), a histidine containing phosphocarrier protein from the phosphoenolpyruvate dependent phosphotransferase system of *E. coli*, NMR ensemble (**1hdn**) and crystal structure (**1poh**);

**1auz**:<sup>60</sup>  $\alpha$ /anti-parallel- $\beta$  sandwich, 116 residues (SpoIIAA), a phosphorylatable component of the system that regulates transcription factor  $\sigma^F$  of *B. subtilis*, NMR ensemble (**1auz**);

**1ksr**:<sup>61</sup> anti-parallel- $\beta$  sandwich, immunoglobulin fold, 100 residues (abp-120), an actin-binding protein from the repeating segments (rod 4) of F-actin cross-linking gelation factor of *D. discoideum*, NMR ensemble (**1ksr**);

The proteins will be referred to by the labels as indicated above.

For the Fyn-Sh3 domain both a NMR and an X-ray structure is available. For T4 lysozyme an X-ray structure is available and for the other three proteins, which were chosen because they are small and globular, highly refined NMR structures are available. In the cases of **1hdn**, **1auz** and **1ksr**, simulations were started from each of the individual structures from the NMR derived ensemble of structures given in the respective PDB files. For the Sh3 domain (**1shf** and **1nyf**), long contiguous simulations were performed starting from the two conformations in the PDB file **1shf**, and from the single NMR-derived conformation in the PDB file **1nyf**. In addition, two long simulations of HPr (starting from the x-ray conformation given in the PDB file **1poh**) and two simulations of T4 Lysozyme (starting from the coordinates given in the PDB file **2lzm**) were provided by B. Hess. These simulations were performed in a manner similar to those described below, except for the use of a more conservative time step of 4 fs. Full details are given elsewhere.<sup>62</sup>

## Parameters

All MD simulations were performed using the GROMOS-96 forcefield<sup>63</sup> and the GROMACS molecular dynamics package.<sup>35,64</sup> The LINCS algorithm<sup>10</sup> was used to constrain covalent bonds within the proteins. Fast degrees of freedom involving hydrogen atoms were eliminated from the system as described in chapter 3.<sup>1</sup> In addition out-of-plane motion in the aromatic rings of the sidechains of tyrosine, phenylalanine and tryptophan was removed as described in sec. 3.2.2. For the hydrogen atoms with remaining degrees of freedom (rotatable groups like hydroxyl and amine, and water), the mass was increased from 1  $u$  to 4  $u$  while simultaneously decreasing the mass of the bonded heavy atom by the same amount, as described previously.<sup>1</sup> This allowed a time step of 7 fs to be used.

A twin-range cut-off for non-bonded interactions was employed with a short-range cut-off of 1.0 nm for the Van der Waals and Coulomb interactions, which were calculated every simulation step, and a long-range cut-off of 1.7 nm for Coulomb interactions which were calculated during neighbor-list update, every 21 fs (3 time steps). The temperature was maintained by weak coupling to an external bath<sup>24</sup> at 300 K with a time constant of 0.1 ps. Protein and water were independently coupled to the heat bath. The pressure was maintained by weak coupling with a time constant of 1.0 ps to an external bath of 1 bar. A relative dielectric constant ( $\epsilon_r$ ) of 1.0 was used.

## Equilibration

Individual structures taken from the respective PDB files were first energy minimized using a steepest descent algorithm. The resulting structures were each solvated in a cubic box of simple point charge (SPC) water,<sup>27</sup> with a minimum distance of 0.85 nm between the protein and the box wall. The water box was constructed by replicating a cubic box containing 216 equilibrated SPC water molecules. All water molecules with the oxygen atom closer to any protein atom than the sum of their respective Van der Waals radii were removed. Energy minimization followed by 0.2 ps of unrestrained MD using a 2 fs time step was performed to relax the systems.

Production simulations were performed starting from each of the structures in the ensemble of NMR solution structures in the respective PDB files for simulation times of 2 or 5 ns, or starting from single X-ray or NMR structures for simulation times of 600 ns. A total of 2  $\mu$ s of protein trajectories was generated. A summary of the simulation parameters is given in Table 4.1.

### 4.2.2 Analysis

The identification of motionally coherent elements consisted of a filtering step based on a Principal Components Analysis of the fluctuations of the coordinates of the backbone (N, C and  $C_\alpha$  atoms) and an assignment step based on the domain identification procedure DYNDOM.<sup>65,66</sup>

## Essential Dynamics

To identify collective motions, Principal Components Analysis (PCA) of the fluctuations in the atomic coordinates of the backbone (N, C and  $C_\alpha$  atoms), also referred to as essential dynamics (ED) analysis,<sup>67</sup> was performed on each of the protein trajectories generated. For the multiple shorter trajectories generated for **1hdn**, **1ksr** and **1auz** separate eigenvectors were calculated from each of the trajectories. The first 0.2 ns of each trajectory was excluded from the analysis to minimize possible artifacts arising from the choice of starting structure. The longer trajectories generated in the other simulations were first fragmented before the PCA analysis was performed. This was done in several ways. For the two 40 ns simulations of **1poh** and **2lzm**, 20 contiguous non-overlapping segments of 4 ns each were generated. In addition, from the two 40 ns trajectories of **2lzm** a set of 40 segments of 1 ns each separated by 1 ns was generated. The three 600 ns trajectories of Sh3 domain were divided into a total of 180 segments of 10 ns each.

For each trajectory or fragment thereof the first six eigenvectors, corresponding to the first six generalized degrees of freedom with the largest fluctuations, were identified and used for further analysis. In effect this filters the motion in the trajectory eliminating small-scale motion while retaining the larger more collective modes.

## Motionally Coherent Elements

The identification of semi-rigid elements from the filtered trajectories was achieved by clustering residues based on the direction of their rotation axes as implemented in the program DYNDOM.<sup>65,66</sup> DYNDOM is a model-free approach that defines dynamical domains based on two key criteria: i) residues within a domain must rotate around an axis with a similar orientation and ii) the “amount of motion” (measured by the mean square fluctuation normalized to the size of the domain) within a domain must be smaller than the “amount of motion” between domains, i.e. domains are more rigid than the whole. The rotation of each residue is determined from the two extremes of the projection of a

protein label	PDB entry	# Str.	# Res.	# atoms		# Sim.	length (ns)	total (ns)	box (nm)	$r_g$ (nm)
Sh3	1shf	2	59	627	7 968	2	600	1 200	4.53	1.05
	1nyf	1	58	618	7 983	1	600	600	4.53	1.03
T4L	2lzm	1	164	1 731	21 484	2	40	80	6.33	1.63
HPr	1hdn	30	85	821	7 957	30	2	60	4.84	1.17
	1poh	1	85	821	12 225	2	40	80	5.04	1.15
1auz	1auz	24	116	1 146	14 152	24	2	48	5.51	1.31
1ksr	1ksr	20	100	931	14 801	20	5	100	5.53	1.37

Table 4.1: Proteins, system sizes and overall simulation properties. From left to right: label by which the system is referred to; PDB entry label; number of structures in the PDB file; number of residues in the protein; number of atoms in the protein and in the water; number, length and total length of simulations; simulation box size and radius of gyration ( $r_g$ ) of the protein. The number of water atoms, box size and radius of gyration are averaged over the simulations.

MD trajectory onto a given ED eigenvector. In the case of multi-domain proteins clusters of residues identified by DYNDOM will correspond to the separate domains. In the case of small single-domain proteins the clusters of residues identified by DYNDOM are more appropriately thought of as semi-rigid motionally independent elements.

The input parameters of the DYNDOM program were set as follows: the maximum number of clusters, 20; the fitting segment window length for determination of residue rotation, 5 residues; the minimum element size, 5 residues; the minimum ratio of external to internal motion, 1.0. The results are, however, insensitive to the specific choice of parameters. This set of parameters was chosen because it is similar to that previously used in the analysis of multi-domain proteins.<sup>65,66,68</sup> The maximum number of clusters, the maximum number of iterations and the minimum element size set the boundary conditions for the clustering algorithm that groups residues with similar rotation. The ratio of internal to external motion determines the acceptance criterion for a given cluster. Note, DYNDOM does not require that any or all residues are members of a cluster (domain). Frequently, only part of the molecule can be assigned to a domain and in some cases no domain assignment is possible at all. Only the length of the fitting window and the ratio of the internal to external motion influence the number of elements assigned.

By counting the number of times residue pairs are found together inside each of the semi-rigid elements, a consensus matrix can be constructed which contains information on the relation between residues in different trajectories. From this, motionally coherent elements can be extracted. In order to determine a possible sequence dependence of the positions of the element boundaries, the number of occurrences  $N_{X,\Delta}$  of residue type  $X$  at various positions  $\Delta$  relative to the boundaries of the motionally coherent elements defined, and the total number of occurrences of  $N_X$  in the total sequence, were determined. From this, the relative probability  $\alpha_{X,\Delta}$  of residue type  $X$  occurring at position  $\Delta$  from a boundary was calculated using

$$\alpha_{X,\Delta} = \frac{N_{X,\Delta}}{p N_X} \quad \text{where} \quad p = \frac{N_\Delta}{N} \quad (4.1)$$

and  $N_\Delta$  is the total number of residues at positions  $\Delta$  and  $N$  in the whole sequence. Values observed for the relative occurrence  $\alpha_{X,\Delta}$  for  $|\Delta| \leq 2$  were compared with the relative probability  $\alpha_{X,random}$  expected for a random distribution of residues and its standard deviation based on a binomial distribution:

$$\alpha_{X,random} = 1 \pm \sqrt{\frac{1-p}{pN}} \quad (4.2)$$

## 4.3 Results

### 4.3.1 Simulations

In all simulations the native fold of the protein and the overall secondary structure distribution as assigned by DSSP<sup>39</sup> remained intact. The root-mean-square difference (RMSD) value of the atomic positions of the backbone atoms from the starting structures, averaged over the trajectory from 0.5 ns was calculated for each simulation. These values varied for

the separate simulations of, **1hdn** from 0.14 to 0.26 nm (average 0.18 nm), **1ksr** from 0.16 to 0.48 nm (average 0.35 nm) and **1auz** from 0.15 to 0.47 nm (average 0.34 nm). For the longer simulations the values, averaged starting from 4 ns, were as follows: for **1poh** 0.22 and 0.28 nm, for **2lzm** 0.28 and 0.31 nm, for **1shf** 0.32 and 0.51 nm and for **1nyf** 0.42 nm.

### 4.3.2 Essential Dynamics

For each trajectory (for **1hdn**, **1auz** and **1ksr**) and all trajectory fragments (for **Sh3**, **T4L** and **1poh**), the first six ED eigenvectors were determined. The resulting total number of eigenvectors for each of the proteins is listed in the second column of Table 4.2. The percentage of the total fluctuation contained in the subspace spanned by the first six ED eigenvectors, averaged over the trajectories of each protein is given in the third column of Table 4.2. On average the first six ED eigenvectors contained 67% of the total atomic positional mean square fluctuation (MSF) present in each trajectory. This is sufficient for the subsequent analysis to provide information on large scale motions within the proteins.

### 4.3.3 Consensus Dynamics

#### Fyn-Sh3

Figs. 4.1 and 4.2 outline the results of the analysis for **Sh3**. In Fig. 4.1 a selection of the individual semi-rigid element assignments as made by the DYNDOM method for each of the essential modes of the **Sh3** trajectories, is represented by shaded bars. The three 600 ns trajectories were split into 60 fragments of 10 ns. For each of these trajectory fragments six eigenvectors were analyzed, yielding in total  $3 \times 60 \times 6 = 1080$  eigenvectors. Element assignments for the first sixty of these eigenvectors are shown in Fig. 4.1. The separate semi-rigid elements are shaded differently. Blank rows correspond to eigenvectors for which no semi-rigid elements could be identified. The percentage of eigenvectors for which an element could be assigned is listed in the fourth column of Table 4.2. Blank areas which are present in most rows correspond to residues that were not assigned to any

Protein		total # of eigenvectors	% of MSF in eigenvectors	% of eigenv. with s.-rigid elements	% of residues in s.-rigid elements
Sh3	<b>1shf/1nyf</b> *	1 080	67	58	78
T4L	<b>2lzm</b> †	240	63	45	80
	<b>2lzm</b> ‡	120	67	52	83
HPr	<b>1hdn</b>	180	65	56	63
	<b>1poh</b> ‡	120	58	51	73
<b>1auz</b>	<b>1auz</b>	144	76	44	61
<b>1ksr</b>	<b>1ksr</b>	120	73	29	66

Table 4.2: Essential dynamics (ED) and semi-rigid element assignments. From left to right: total number of eigenvectors analyzed; percentage of the total mean square fluctuation (MSF) in the atomic positions contained in the first six eigenvectors, percentage of ED eigenvectors with a semi-rigid element assignment and percentage of residues assigned to a semi-rigid element. Long trajectories ( $\geq 40$  ns, see Table 4.1) were fragmented for analysis into adjacent 10 ns parts (\*), 1 ns parts with a 1 ns separation (†) or adjacent 4 ns parts (‡).

semi-rigid element for that particular eigenvector. The percentage of residues assigned to a semi-rigid element for those eigenvectors for which elements were identified, is listed in the fifth column of Table 4.2.

Fig. 4.2a shows the “consensus dynamics matrix” for **Sh3**. The intensity corresponds to the number of times a residue pair was observed together in the same semi-rigid element. The most prominent feature of this plot is the hierarchical block structure. Blocks in the matrix correlate with the locations of the individual semi-rigid elements as depicted in Fig. 4.1. The average intensity of a block in the consensus dynamics matrix gives a qualitative measure of the apparent rigidity of the element.

Using the matrix as a guide, subdivisions of the protein can be made as shown in Fig. 4.2b. On the most detailed level six elements can be identified in **Sh3** that contain more than 5 residues (the minimum element size used when determining the semi-rigid elements). In Fig. 4.2c a histogram of the occurrence of the beginning or end of an element at a particular point along the sequence is plotted. Peaks correspond to locations where boundaries of motionally coherent elements occur frequently and are correlated with the element boundaries as defined in the hierarchy of Fig. 4.2b.

In Fig. 4.3 the motional hierarchy outlined in Fig. 4.2b is color coded onto the three-dimensional structure of **Sh3**. At the highest level three large partially overlapping motionally coherent elements can be recognized. In the structure, the element **Sh3:1-29** (red & dark pink in Fig. 4.3a) corresponds to the first two sets of  $\beta$ -strands at the N-terminus. Element **Sh3:1-40** (red, dark pink & green) also includes the following loop-and-strand. Element **Sh3:20-58** (dark pink, green & blue) contains the complete three-stranded  $\beta$ -sheet (to the rear in Fig. 4.3), with the  $3_{10}$ -helix and C-terminal  $\beta$ -strand on one side

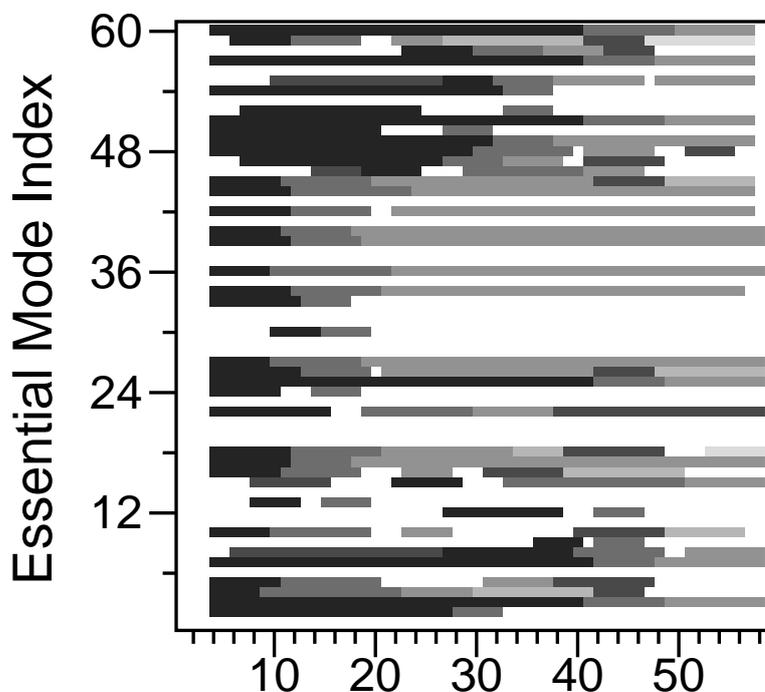


Figure 4.1: Schematic representation of a selection from all semi-rigid element assignments for **Sh3** (60 out of a total of 1080 are shown). On the horizontal axis the residues are plotted. On the vertical axis a cumulative index is plotted of all essential modes for the trajectories that were selected for display. Different shades correspond to different semi-rigid elements assigned by DYNDOM for that particular essential mode, shades do not correlate between different essential modes.

and an additional  $\beta$ -strand on the other side. The next, most detailed, level contains six non-overlapping elements, which are indicated in Fig. 4.3b. Elements 2 (residues 10-19, loop, yellow in Fig. 4.3b), 4 (residues 29-40, strand-loop-strand, green), 5 (40-48, strand-turn-strand, blue) and 6 (48-58, strand-helix-strand, light blue) form relatively compact structures within the protein.

## T4 Lysozyme

The consensus dynamics matrix for T4L is shown in Fig. 4.4a. The upper half of the matrix was derived from the analysis of 1 ns fragments taken from the long trajectories with 1 ns intervals inbetween and the lower-right half from consecutive fragments of 4 ns

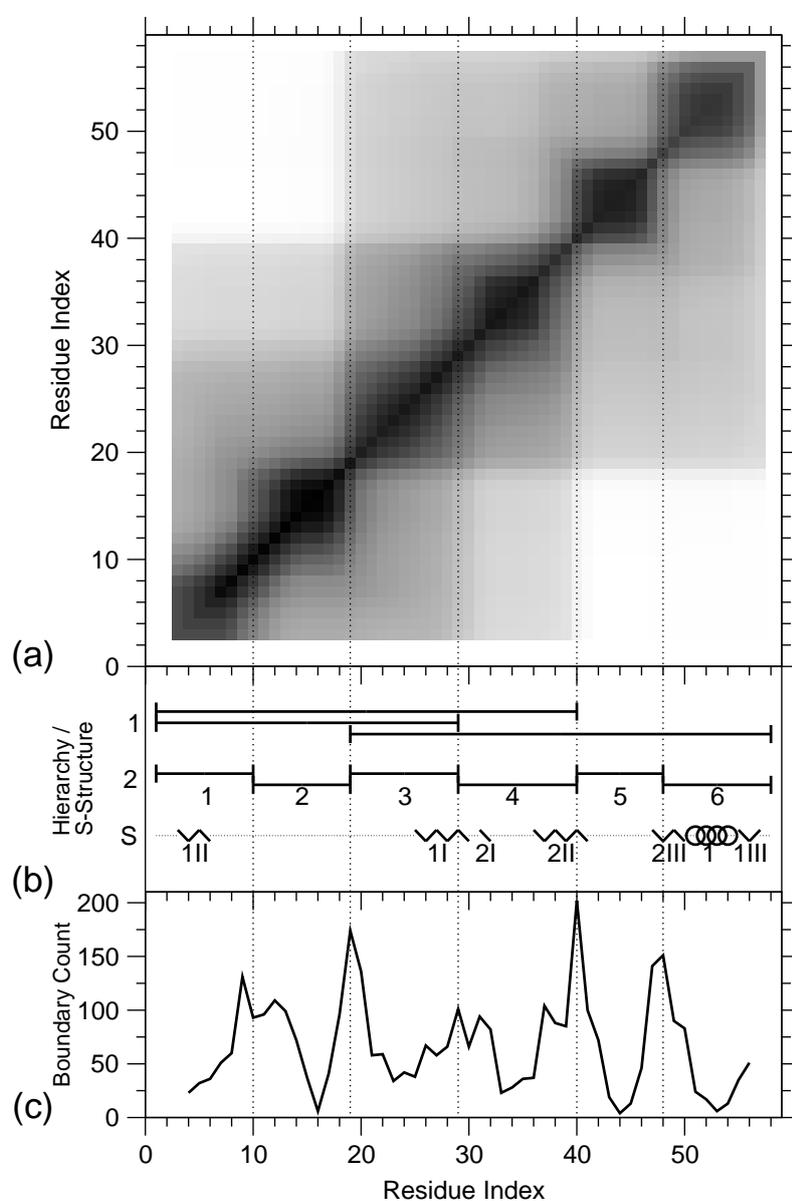


Figure 4.2: Outline of consensus dynamics results for Sh3. (a) Consensus dynamics matrix as derived from the individual element assignments (see Fig. 4.1a). Residue number is indicated on both axes. Intensities correspond to the number of times both residues of a pair can be found together in the same semi-rigid element. (b) Schematic view of the hierarchy of motionally coherent elements as observed in the matrix of a, organized into hierarchical levels (numbered from 1) and secondary structure elements (labeled 'S',  $\odot\odot\odot$ = $\alpha$ -helix,  $\wedge$ = $\beta$ -strand, labeled according to the PDB file). Horizontal lines indicate the elements. The vertical dotted lines indicate the boundaries and the correlation between hierarchical levels. Elements in the lowest hierarchical level are numbered along the sequence. (c) Count of semi-rigid element boundaries. Vertical dotted lines indicate significant peaks and correspond to element demarcations as defined in b.

each. A more detailed description of the procedure can be found in the Methods section (sec. 4.2.2). The difference between the two halves of the matrix is insignificant.

An hierarchical breakdown gives rise to three the levels depicted in Fig. 4.4b. The first (highest) level contains three motionally coherent elements, indicated by different colors in Fig. 4.5a. The three elements correspond to the N- and C-terminal domains of T4 Lysozyme (blue and green respectively in Fig. 4.5a) and the N-terminal  $\alpha$ -helix that structurally is part of the C-terminal domain (element 1, red in Fig. 4.5a). Analysis of crystal structures, however, indicate that this fragment can be associated with either of the two domains.<sup>69</sup> The large helix that extends from one domain into the other,  $\alpha$ -helix 3, is divided in the middle, with each end being assigned to a different element.

On the next level, shown in Fig. 4.5b, in the N-terminal domain two elements can be assigned and in the C-terminal domain three elements can be assigned. The protein as a whole is divided into three large parts of roughly equal size which each form relatively compact structures, and three small  $\alpha$ -helical regions, which lie inbetween the larger sections. The twelve elements identified on the most detailed level are indicated in Fig. 4.5c.

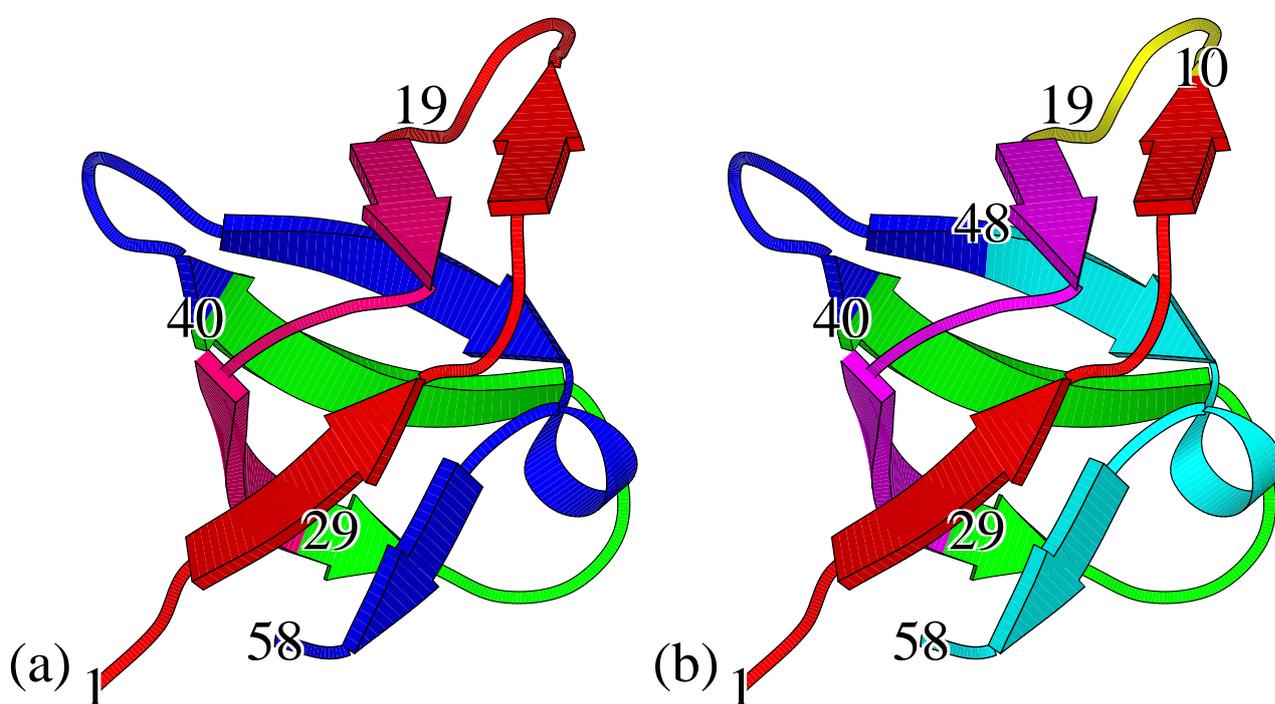


Figure 4.3: Representation of the two levels of motional hierarchy for Sh3 depicted schematically in Fig. 4.2b, color coded onto a cartoon representation of the crystal structure of the protein. Colors indicate different elements. The last residue of each element, as well as residue 1, are labeled. Plots were generated using a modified version of MolScript.<sup>3,4</sup>

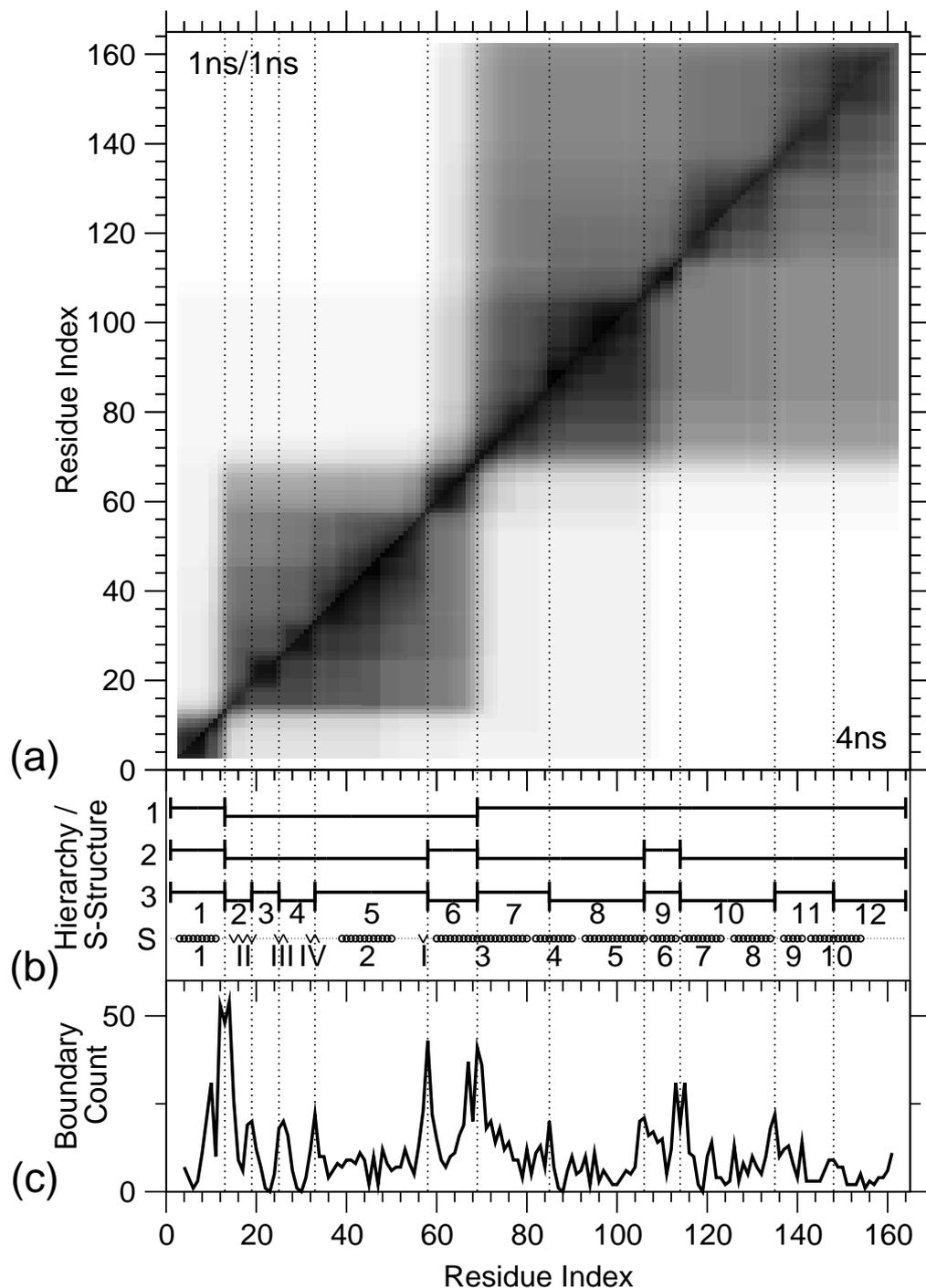


Figure 4.4: Outline of results for T4L. (a) consensus dynamics matrix (cf. Fig. 4.2a). Results for fragmenting the long trajectories into 1 ns parts with a 1 ns separation or into adjacent 4 ns parts are plotted in the upper left and the lower right half of the matrix respectively. (b) Hierarchy of elements (cf. Fig. 4.2b). (c) Demarcation counts (cf. Fig. 4.2c) summed over the results corresponding to both halves of the matrix.

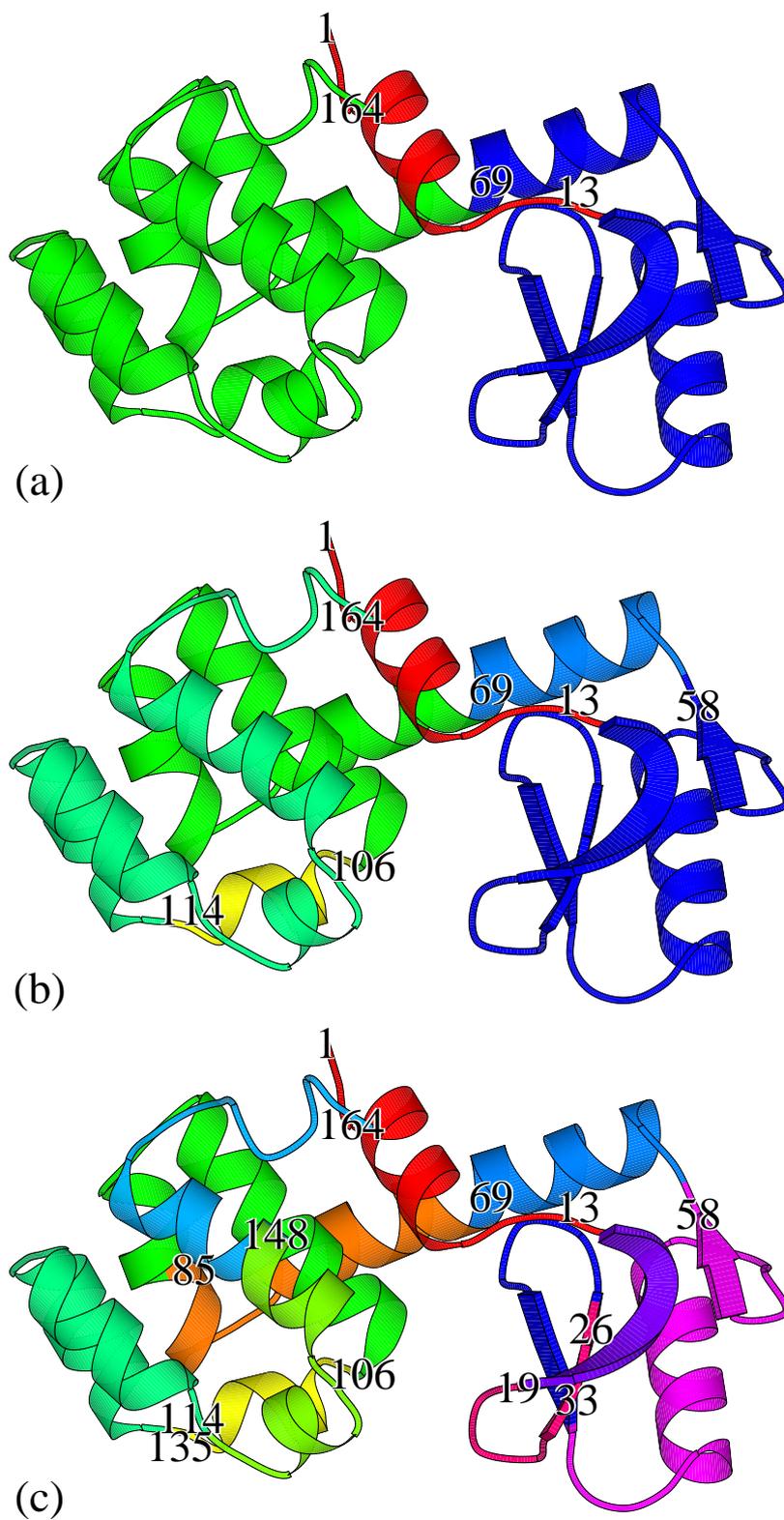


Figure 4.5:  
Representation of the  
three levels of motional  
hierarchy for T4L as  
depicted schematically in  
Fig. 4.4b (cf. Fig. 4.3).

## HPr

The consensus dynamics matrix for HPr is shown in Fig. 4.6a. The upper half of the matrix was derived from a series of simulations started from NMR structures given in PDB file 1hdn, the lower half from a series of fragments taken from two simulations started from an X-ray structure (1poh), see Methods section sec. 4.2.2. The difference between the two halves of the matrix is marginal. On the most detailed level nine separate elements could be defined (Fig. 4.6b) most of which correspond roughly to secondary structure elements. Two notable exceptions are the  $\beta$ -turn between strands III and IV which forms element 5 (residues 36-41) and the half-strand, loop, half-helix that forms element 8 (residues 62-76). Both these elements form compact structures. The elements are indicated by color coding in Fig. 4.7.

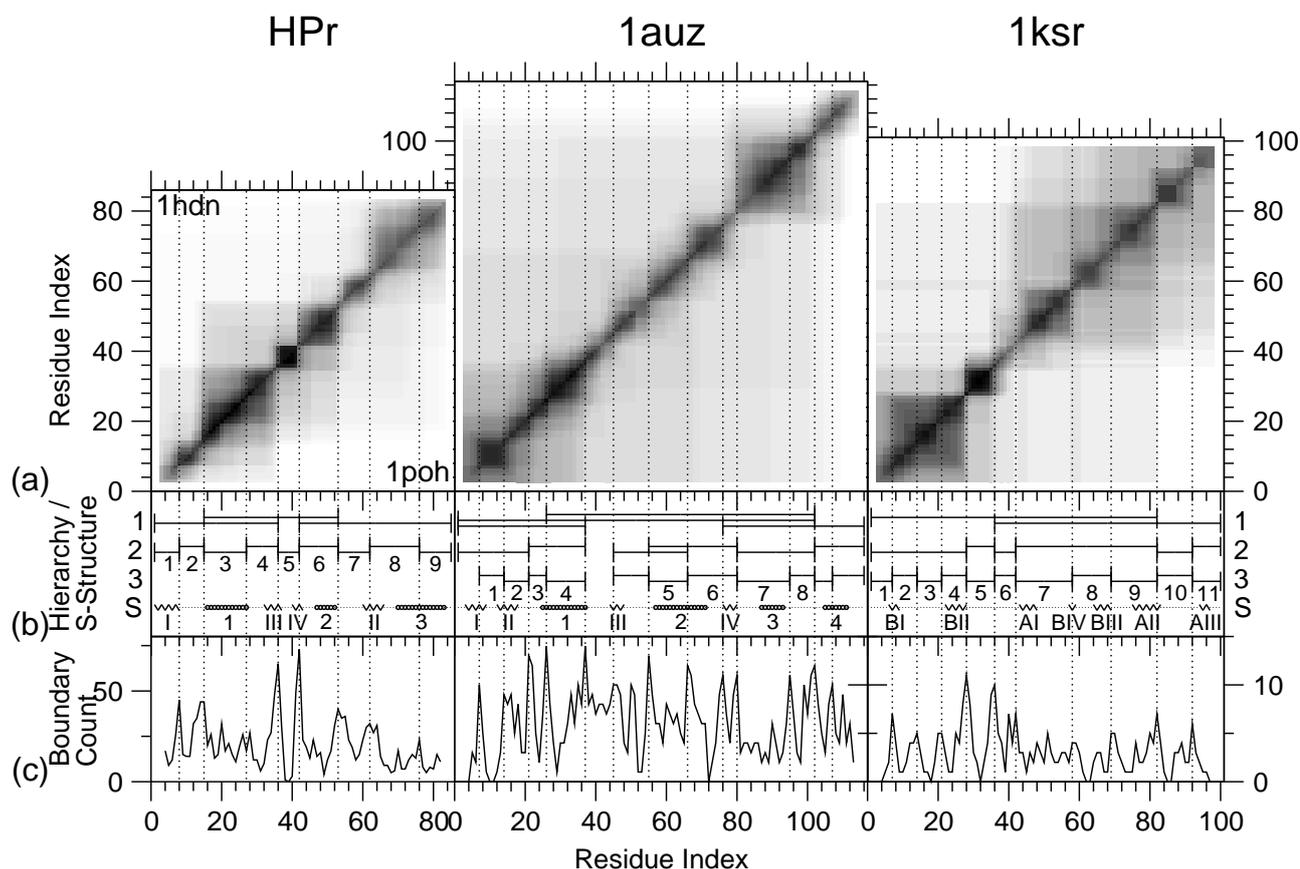
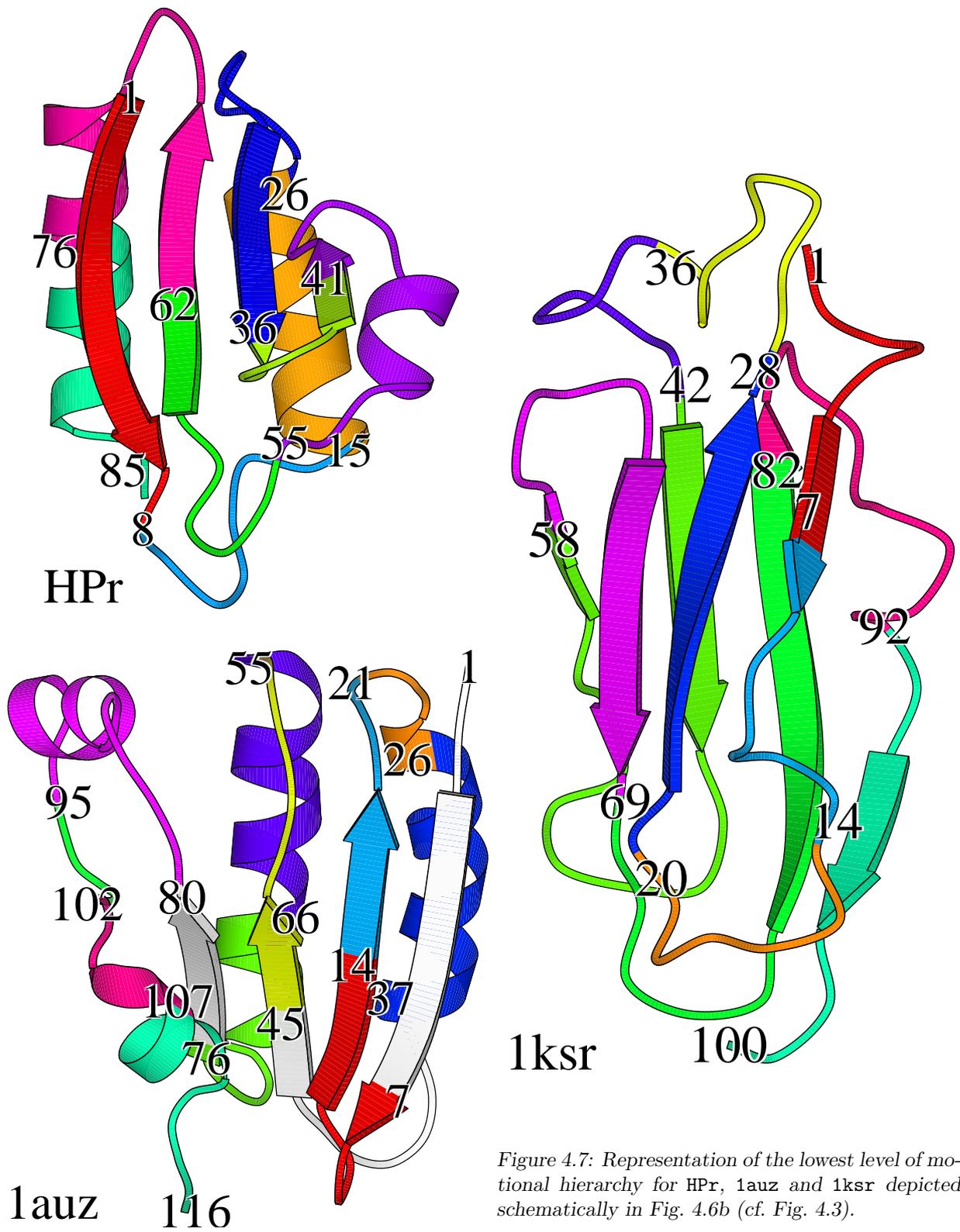


Figure 4.6: Results for HPr, 1auz and 1ksr. (a) consensus dynamics matrix (cf. Fig. 4.2a). For HPr, results for the simulations started from 1hdn and 1poh are plotted in the upper left and the lower right half of the matrix respectively. (b) Hierarchy of elements (cf. Fig. 4.2b). (c) Demarcation counts (cf. Fig. 4.2c). For HPr only the total demarcation counts corresponding to both halves of the matrix are plotted.



### SpolIAA (1auz)

For 1auz the blocked structure of the consensus dynamics matrix is much less pronounced (Fig. 4.6a). On the first level of the motional hierarchy the protein can be split into roughly three parts (Fig. 4.6b). The middle section does not show a pronounced blocked structure in the matrix. In the lowest level 8 elements can be defined as indicated in Fig. 4.7.

### Abp-120 (1ksr)

On the most detailed hierarchical level, 1ksr may be divided into 11 elements, see Figs. 4.6a and b. Some correspond to individual  $\beta$ -sheet regions. Most of the elements are not very compact in the protein. Fig. 4.7 gives a graphical representation of these elements.

## 4.3.4 General Properties

Certain elements contain very few or no boundaries of semi-rigid elements and correspond to minima of the demarcation counts in Figs. 4.2c, 4.4c and 4.6c. Most notably,  $\beta$ -turns are almost always correlated with such minimum (Sh3:30-40, Sh3:41-48, T4L:20-25, T4L:26-33, 1hdn:37-41, 1auz:8-14, 1ksr:59-69 and 1ksr:70-81). The only exception is the  $\beta$ -turn at 1ksr:50-55. Other minima in the demarcation counts are found inside certain loop regions (Sh3:11-19, Sh3:20-29, 1auz:67-76, 1ksr:15-21, 1ksr:29-36 and 1ksr:83-92) as well as inside helical regions (Sh3:49-58, T4L:34-58, T4L:85-105, T4L:106-113, 1hdn:42-52 and 1auz:27-37). This suggests a relatively high (dynamical) coherence inside some loops and for certain positions within  $\alpha$ -helices.

To a first approximation, the motionally coherent elements observed correspond to secondary structure elements, such as  $\beta$ -turns and  $\alpha$ -helices (as noted above) and to a lesser extent  $\beta$ -strands. A number of element boundaries were, however, positioned inside secondary structure elements, dividing them into parts. Most notably, longer helices (e.g. HPr helix 3 and T4L helix 3) and  $\beta$ -strands are divided approximately in the middle (e.g. HPr strand II or 1ksr strand AII) or near  $\beta$ -turns. This suggests a loss of (dynamical) coherence near the ends of secondary structure elements and a reduced coherence for some extended secondary structure elements.

Table 4.3 shows the relative abundance of residue types ( $\alpha_X$ ) at positions up to two residues from the element boundaries, calculated using eqn. 4.1. The expected standard deviation of the relative abundance based on a random distribution of residue types ( $\alpha_{X,random}$ ), calculated using eqn. 4.2, is also shown. For serine and histidine, a significant deviation from the expected occurrence is found with a statistical confidence level of more than 95 %.

No off-diagonal blocks appear in any of the consensus dynamics matrices of the proteins investigated. This is surprising as many of the motionally coherent elements identified are in close contact in the three-dimensional structure, while being sequentially distant. Possible reasons for this observation are discussed below.

## 4.4 Discussion

This work clearly demonstrates that the use of multiple starting structures for a number of relatively short protein simulations can be an efficient way to sample the accessible configurational space within a limited amount of computational time. Results of the analysis of the series of shorter simulations are essentially identical to those obtained from long contiguous simulations. Likewise, no significant differences were detected between simulations performed using a time step of 7 fs and simulations performed at 4 fs.

The analysis of the atomic motions in this study is based on the first six principal components of the motion which comprises 60 – 70 % of the total structural fluctuations. Extensive tests on HPr (data not shown) indicate that using only the first six eigenvectors results in an optimal filtering of the information available in the trajectories. Including more (higher) eigenvectors increases noise whereas using less eigenvectors diminishes the available information. From the semi-rigid elements obtained, it is possible to infer a consensus description of the motionally coherent elements of a protein. Approximately half of the total fluctuation of the native state of the protein can be attributed to the motion of a relatively small number of structural elements. In this respect, the combination of principal components analysis on atomic coordinates and semi-rigid element analysis acts as a powerful filter for the analysis of dynamical trajectories of proteins.

Motionally coherent elements appear as blocks in the consensus dynamics matrix.

res. type	$N_X$	$N_{X,\Delta}$ $ \Delta  \leq 2$	$\alpha_X$	$\alpha_{X,random}$	confidence level
Gly	43	18	0.91	1.00±0.17	
Ala	36	12	<b>0.72</b>	1.00±0.18	
Val	40	17	0.92	1.00±0.17	
Ile	27	13	1.05	1.00±0.29	
Leu	44	22	1.09	1.00±0.16	
Pro	16	7	0.95	1.00±0.27	
Phe	20	10	1.09	1.00±0.24	
Trp	5	1	<b>0.43</b>	1.00±0.48	
Met	10	5	1.09	1.00±0.34	
Cys	5	2	0.87	1.00±0.48	
Ser	27	18	<b>1.45</b>	1.00±0.29	> 95%
Thr	36	22	<b>1.33</b>	1.00±0.18	
Asn	20	7	<b>0.76</b>	1.00±0.24	
Gln	19	5	<b>0.57</b>	1.00±0.25	
Tyr	14	7	1.09	1.00±0.29	
His	9	8	<b>1.93</b>	1.00±0.36	> 95%
Asp	34	13	0.83	1.00±0.19	
Glu	37	15	0.88	1.00±0.18	
Lys	36	19	1.15	1.00±0.18	
Arg	22	9	0.89	1.00±0.23	
total	500	230			

Table 4.3: Statistics of amino acid residue composition relative to element boundaries in absolute and relative numbers. From left to right: residue type; total number of that type in all sequences ( $N_X$ ); and close to boundaries ( $N_{X,\Delta}$  for  $|\Delta| \leq 2$ ); relative occurrence of that type observed ( $\alpha_X$ ) and expected for random ( $\alpha_{X,random}$ ) and statistical confidence level for deviation of observed value from random. Relative occurrences ( $\alpha_X$ ) deviating more than one  $\sigma$  from random are indicated in bold.

Larger blocks often contain smaller ones, which suggests the motions in a protein are hierarchical in nature. The occurrence of distinct peaks and minima in the number of element boundaries lends confidence to the method. Motionally coherent elements that contain non-contiguous parts of the sequence will show up as off-diagonal blocks in the consensus dynamics matrices. Surprisingly such elements only occurred rarely in the cases tested. Nearly all motionally coherent elements consist of a single contiguous sequence. This appears to be a non-trivial property of the proteins studied. It suggests a description of the dynamics of the proteins in terms of a hierarchy of structurally and motionally coherent, sequential elements. This is consistent with models of protein folding that assume the formation of local structural elements which then assemble sequentially into the folded protein.

Little preference could be detected for the occurrence of specific amino acids at the boundaries between motionally coherent elements. Although the statistics are insufficient to draw general conclusions, the hydrogen-bonding amino acids His, Ser and Thr seem to be preferred at the boundary locations. From the available data it is not possible to extract statistics about combinations of residues or about sequence patterns.

There also does not appear to be a strong relation between the location of the motionally coherent elements and the location of secondary structure elements in the protein. Many  $\alpha$ -helices and  $\beta$ -sheets contain element boundaries and belong to more than one element. On the other hand, many elements consist of parts of two secondary structure elements together with the connecting turn or loop region, e.g. strand-turn-strand or helix-loop-strand. This suggests that loop regions can undergo concerted motion. It also suggests a much more dynamic and flexible character of secondary structure elements than is usually assumed. In part, this difference might simply reflect ambiguities resulting from the use of a discrete algorithm (such as DSSP) for the classification of secondary structure elements when conformations fall near the classification boundaries. Comparison with a continuous classification scheme might lead to more consistent results.<sup>70</sup>

A similar conclusion on the nature of secondary structure elements can be drawn from the comparison between the consensus dynamics matrices and the overall secondary structure of the proteins. T4L, Sh3 and HPr consist mainly of relatively large secondary structure elements and feature sharply demarcated blocks in the matrix. 1auz has a similarly well-ordered secondary structure, but a more blurred appearance in the consensus dynamics matrices. 1ksr in contrast has fewer large secondary structure elements (many loop and coil regions), but does show sharply demarcated blocks in the matrix. No definitive relationship between the nature of the secondary structure of a protein and its dynamical behavior is evident.

#### 4.4.1 Comparison with Experimental Studies

Despite the wealth of protein folding studies, there is little experimental information at an atomic level that can be directly compared to the results of the work we have presented. The most direct data on the mechanisms of protein folding, that is data involving the least number of assumptions, is NMR data. This includes relaxation data (NOESY/ROESY intensities or build-up curves and T1 relaxation times) and exchange data (hydrogen/deuterium

exchange protection factors). The other main source of information is mutational studies, primarily the so-called  $\Phi_f$  analysis which attempts to relate the effects of single amino acid substitutions on the rates of folding and unfolding of proteins to the nature of the folding transition state, and studies involving the stability or dynamics of isolated peptide fragments taken from different parts of the protein.

### Fyn-Sh3

The folding and stability of Fyn-Sh3 has been extensively studied by Riddle *et al.*<sup>71</sup> who have performed  $\Phi_f$  analysis and applied an *ab initio* folding method. The effects of single alanine and glycine mutations at many positions along the sequence on the rates of folding and unfolding ( $\Phi_f$  analysis) were analyzed.<sup>71</sup> The results of this work are schematically depicted in Fig. 4.8b. Blocks correspond to regions of the protein considered to be important in the transition state of folding and hence are formed early in the folding process. The size of the blocks corresponds to their relative importance. Note that the least important (number five) region spans two non-consecutive regions, residues 3-6 and 50-57. The positions of these regions correlate fairly well with the elements identified from the consensus dynamics matrix (Fig. 4.8a). Not surprisingly, the two most important folding regions are  $\beta$ -turns.

The *ab initio* folding method ROSETTA was used to predict a progression of folding events<sup>71</sup> (Fig. 4.8c). ROSETTA ranks the elements according to the probability they will adopt a particular conformation based on an analysis of protein structural databases. The order in which separate regions of the protein fold according to this algorithm, roughly corresponds to the order observed for the five fragments found in the  $\Phi_f$  analysis (Fig. 4.8b). Regions of the protein that display distinct behavior in the ROSETTA analysis correlate well with the motionally coherent elements identified from our analysis. It is interesting that the most important and early folding segment comprising residues 40-48 is also the element with the clearest demarcation in our analysis (see Fig. 4.2c, element 5). The early folding of element 5, followed by folding of segments 3-6 (residues 19-56) in the next hierarchical level, is compatible with all available data.

Overall, a reasonable correspondence is observed between the positions of the motionally coherent elements and the various regions identified in the analyses performed by Riddle *et al.*<sup>71</sup>

### T4 Lysozyme

The folding and stability of T4 Lysozyme has been extensively studied. Hilser & Freire used  $\Phi_f$  analysis.<sup>72</sup> Lu & Dahlquist used pulsed hydrogen exchange experiments.<sup>73</sup> Najbar *et al.* analyzed the stability of isolated peptide fragments taken from T4 Lysozyme.<sup>74</sup>

The effect of mutations at many positions along the sequence on the rate of folding was measured and expressed as the logarithm of the folding rate ( $\ln \kappa_f$ )<sup>72</sup> is shown in Fig. 4.9b. High values of  $\ln \kappa_f$  correspond to more stable regions of the protein. Qualitatively, regions with higher  $\ln \kappa_f$  correlate with regions with lower demarcation counts (Fig. 4.4c), most notably for the peaks of  $\ln \kappa_f$  inside helices 1, 5, 7 and 10.

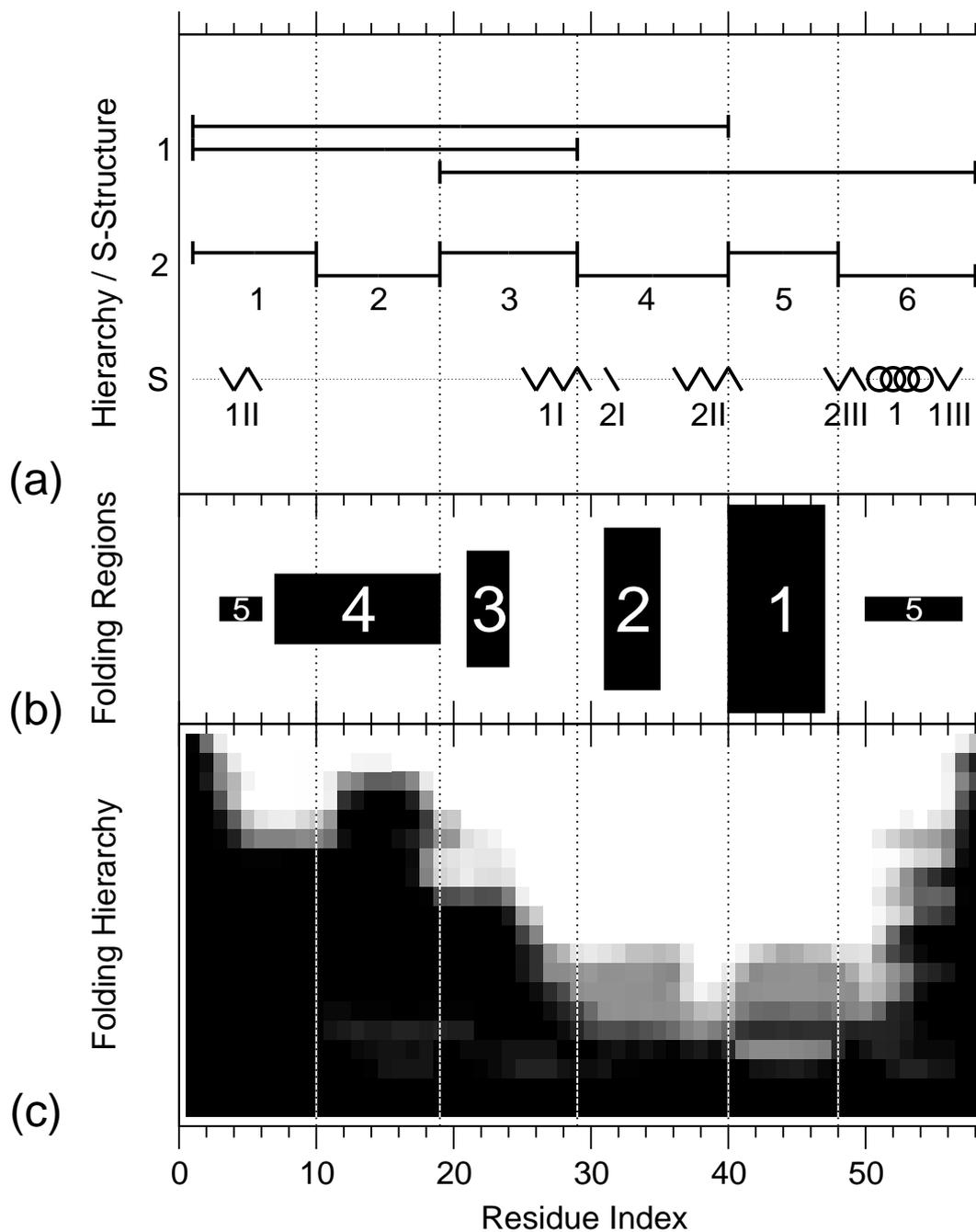


Figure 4.8: Comparison with experimental data for Sh3 from Riddle *et al.*<sup>71</sup> (a) Hierarchy of motionally coherent elements (see Fig. 4.2b). (b) Five structural regions in order of importance in the folding transition state.<sup>71</sup> Block height corresponds to relative importance. (c) Sequence of folding events as determined by the *ab initio* folding method ROSETTA<sup>71</sup> (plot modified from original paper with permission), folding progresses from the bottom upwards. Black corresponds to residues without native contacts and white to residues with all native contacts formed.



From pulsed hydrogen exchange experiments, the protection factors for 84 individual residues in T4L were determined by Lu & Dahlquist<sup>73</sup> (see Fig. 4.9c). This procedure can identify sites in the protein that are protected early in folding (around 80 ms). Several regions show significant protection (more than a factor of 20):<sup>73</sup> two parts in the N-terminal  $\beta$ -sheet region,  $\alpha$ -helix 1 and 5 as well as some residues in helix 3. These regions correspond to relatively low demarcation counts in Fig. 4.4c.

For several peptide fragments from T4L the importance to the folding was determined by Najbar *et al.* from analysis of stability and other properties of the isolated peptides in solution.<sup>74</sup> Three peptide fragments were identified as potential folding initiation sites. Two of these, residues 1-13 and 92-107, correspond roughly to elements identified in our analysis. The third potential folding initiation site identified by Najbar *et al.* corresponds to  $\alpha$ -helix 3, which bridges the two domains of T4L and corresponds to two consecutive elements in our analysis.

## 4.5 Conclusion

It is generally accepted that the separate domains in multi-domain proteins fold independently. Furthermore, for some proteins (e.g. lysozyme<sup>75</sup>) it has been suggested that one of the domains must fold before another can complete its folding. Other studies suggest local patterns, such as  $\beta$ -turns or one turn of an  $\alpha$ -helix initiate folding. This leads to a concept of a hierarchy of folding, with the assembly of protein structural domains at the top level, and a series of independent folding units of several residues in length at the most basic level. Our hypothesis was that such a folding hierarchy may be reflected in a hierarchy of motions in the protein in its native state. What we have found is that there is a hierarchical organization of motionally coherent elements within proteins. The motional hierarchy observed most likely reflects the folding hierarchy and as such is a guide to construct a hierarchy of folding. Within this framework it is still possible to devise *multiple* folding pathways that describe all elements observed as is illustrated in Fig. 4.10. Nevertheless the procedure described in this paper provides an objective means to identify boundaries between potential folding elements.

## 4.6 Acknowledgments

I would like to thank B. Hess (RuG, Groningen) for providing the long simulations of HPr and T4 Lysozyme.

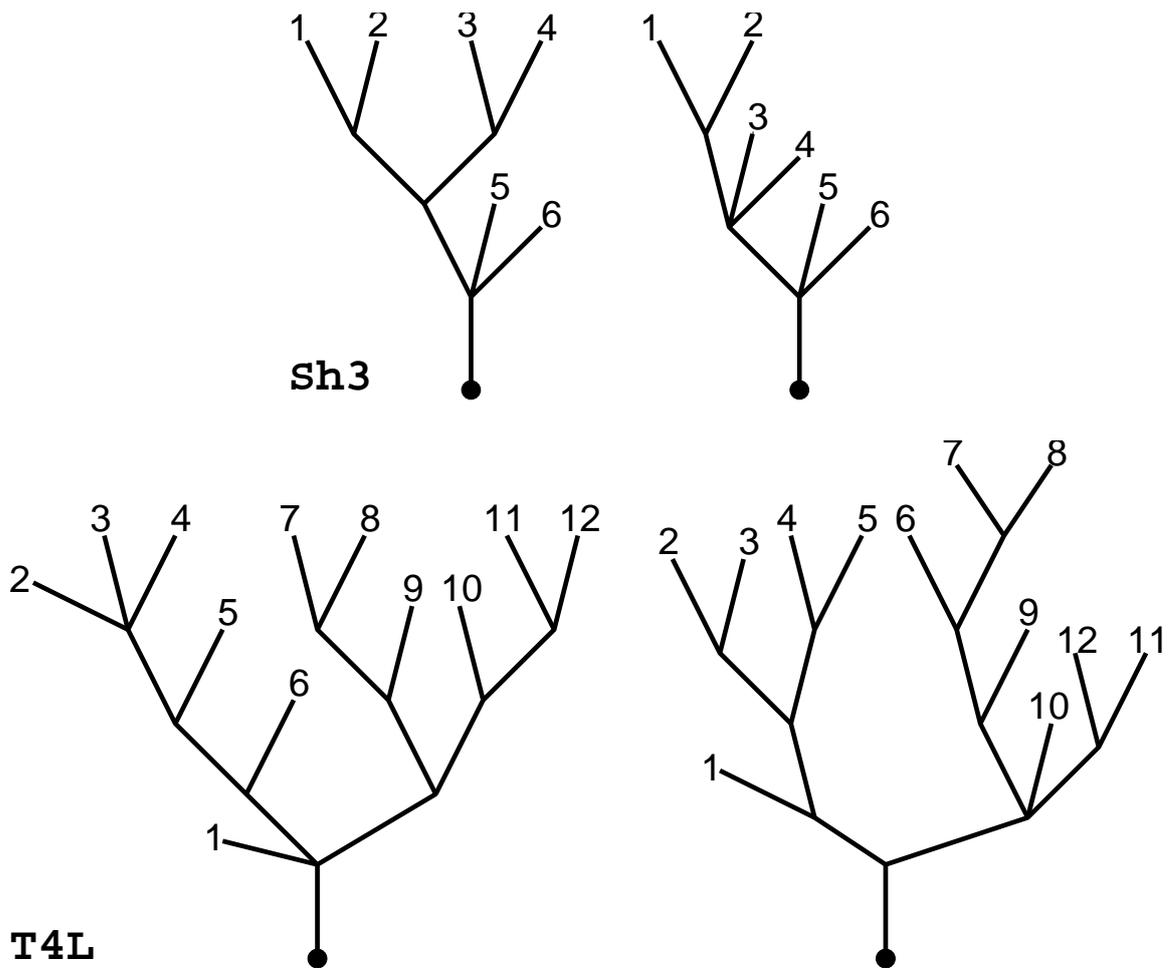
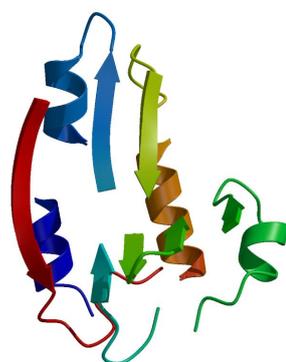


Figure 4.10: Non-unique examples of tentative folding pathways for Sh3 and T4L where folding can be thought to progress from the outer branches inwards. These pathways can account for the occurrence of all motionally coherent elements observed and correspond to the hierarchy as observed in Figs. 4.2 and 4.4.



## Chapter 5

# Stability of Putative Folding Units





## 5.1 Introduction

In chapter 4 fragments of **HPr** and **1ksr** were identified on the basis of concerted motions and rigid body analysis. This results in fragments that, on average, show concerted motion within them, and relatively independent motion between them.

These fragments of the protein are proposed as candidate foldons and/or nucleation sites. The next requirement is that such a fragment should not only appear stable in the protein environment, but also isolated in solution. That assumption will be tested by performing simulations of the isolated fragments in solution. Ideally, for a proper assessment of the stability of a certain conformation of a peptide, extensive equilibrium simulations should be obtained, from which cluster analysis can extract various conformational clusters and corresponding occupancies. If the largest, most occupied cluster then corresponds to the conformation from the protein environment, the protein conformation of the fragment can be said to be stable. That would imply that the fragment is likely to play a significant role in the early stages of folding, as a foldon or folding nucleation site.

In cases where achieving equilibrium during a simulation is not feasible, a more qualitative assessment of the peptide's stability in the protein conformation can be obtained by performing two separate simulations of adequate length for a peptide starting from the protein conformation and from an extended simulation. In that way, the tendency of the peptide to remain close to, respectively tend towards, the protein conformation can be assessed. While not a definite proof, such a tendency is a strong indication for actual stability of the fragment in the protein conformation. This is in fact the procedure that was applied to the nine fragments of **HPr** and eleven of **1ksr**. From this we were able to draw conclusions on the existence of several possible folding nucleation sites in both proteins.

## 5.2 Methods

### 5.2.1 Simulations

#### Parameters

All MD simulations were performed using the GROMACS molecular dynamics package.<sup>35,64</sup> The GROMOS-96 forcefield<sup>63</sup> was used. Fast hydrogen atom degrees of freedom as well as out-of-plane degrees of freedom in aromatic residues were eliminated from the system as described in chapter 3.<sup>1,2</sup> The combined modifications allow a time step of 7 fs to be used reliably. The LINCS<sup>10</sup> constraint algorithm was used for all covalent bonds.

A twin-range cut-off for non-bonded interactions was employed with a short-range cut-off for Van der Waals and Coulomb interactions of 1.0 nm which were calculated every simulation step, and a long-range cut-off of 1.7 nm for Coulomb interactions which were calculated during neighbor-list update, every 21 fs (3 time steps). The temperature was controlled using weak coupling to a bath<sup>24</sup> of 300 K with a time constant of 0.1 ps. Protein and solvent were coupled to the heat bath independently. The pressure was also controlled

using weak coupling with a time constant of 1.0 ps to a bath of 1 bar. A relative dielectric constant ( $\epsilon_r$ ) of 1 was used.

## Equilibration

Conformations of the peptide fragments were taken from the NMR ensembles of the proteins. For HPr, the first structure in PDB<sup>55</sup> file 1hdn<sup>32</sup> was used. For 1ksr the first structure of PDB file 1ksr<sup>61</sup> was used. A short description of both proteins is given in sec. 4.2.1. An additional conformation of each fragment was generated in a completely extended state using the BUILD option of the WHATIF package.<sup>76</sup>

Both protein-derived and extended conformations were first subjected to energy minimization using the steepest descent algorithm. A cut-off for electrostatic interactions of 1.8 nm was used in the energy minimizations. The resulting structures were solvated in cubic boxes of simple point charge (SPC) water,<sup>27</sup> with a minimum distance of 0.85 nm between the peptide and the box. The water box was built by stacking cubic boxes containing 216 equilibrated SPC water molecules. All water molecules within Van der Waals distance from the protein were removed. Another round of energy minimization followed by 0.2 ps of unrestrained MD using a 2 fs time step were performed to relax the newly built systems.

## Production Simulations

For each fragment of HPr and 1ksr as defined in Fig. 4.6 two simulations of 20 ns each were performed. For HPr fragments 1 and 2 were combined into a single one in the simulations. An overview of the fragments and the simulations performed is given in Table 5.1. Termini were uncharged, -COOH and -NH<sub>2</sub>, to minimize the difference between the fragment in isolation, and in its environment in the protein. In order to be able to make a direct comparison with experimental studies (which will be described in chapter 6), one selected peptide (HPr:37-41) was extended by two residues at either end (corresponding to HPr:35-43), the N-terminus acetylated and the C-terminus amidated, -COCH<sub>3</sub> and -CONH<sub>2</sub> respectively. Two simulations of 250 ns each were performed using this peptide.

### 5.2.2 Peptide Stability

The stability of the isolated fragments in solution was analyzed from the two separate simulations performed for each fragment, one started from the conformation of the peptide in the protein and one from an extended conformation respectively. The peptide was considered stable, if in both simulations the conformation of the peptide remained close to or tended toward the protein conformation. The size-independent difference factor  $\rho_{sc}$ <sup>77</sup> is used to compare different conformations. This parameter is similar to the well-known atomic positional root-mean-square difference (RMSD), but is normalized to exclude size and overall-shape related factors. It is also dimension-less instead of expressing a distance. Values of the  $\rho_{sc}$  parameter can therefore be directly compared between molecules of different sizes and with high internal mobility.

For the selected peptide, for which extensive simulations were performed, additional analysis of the stability was performed by conformational clustering based on the atomic positional RMSD. The cluster algorithm as described by Daura *et al.*<sup>78</sup> was used. This method yields non-overlapping clusters of structures with an RMSD with respect to the central structure within a certain cut-off. A cut-off of 0.1 nm for the RMSD between backbone atoms was used.

## 5.3 Results

### 5.3.1 Simulations

Separate fragments of HPr (9 fragments) and 1ksr (11 fragments), which could represent stable folding units, have been identified (see chapter 4). These are depicted schematically at the bottom hierarchical level (level 2 for HPr and 3 for 1ksr) in Fig. 4.6b. Fig. 5.1 shows a view of the fragments as if they have exploded from the two proteins. For HPr the first two fragments (1 and 2) have been combined into a single one. For each of the resulting 19 fragments, see Table 5.1, two simulations in water were carried out. One simulation was started from the protein conformation (taken from the first conformation

pro- tein	fragments				difference $\rho_{sc}$			
	id	start	end	length	folded	extended		
HPr	1+2	1	15	15	0.89		0.75	
	3	16	26	11	0.58		0.64	
	4	27	36	10	0.46	~	0.66	
	5	37	41	5	0.45	~	0.47	
	pepAc <sup>a</sup>	35	43	9	0.55		0.57	
		6	42	55	14	0.43	~	0.53
		7	56	61	6	0.64		0.57
		8	62	76	15	0.50	~	0.50
	9	77	85	9	0.57		0.65	
1ksr	1	1	7	7	0.61		0.50	
	2	8	14	7	0.37	~	0.39	
	3	15	21	7	0.41	~	0.49	
	4	22	28	7	0.88		0.37	
	5	29	36	8	0.25	=	0.64	
	6	37	42	6	0.40	~	0.40	
	7	43	58	16	0.71		0.76	
	8	59	69	11	0.61		0.62	
	9	70	82	13	0.65		0.86	
	10	83	92	10	0.58		0.70	
	11	93	100	8	0.86		0.86	

<sup>a</sup> derived from fragment 5 with two residues appended on either side

Table 5.1: Fragment definitions and stability estimates for HPr and 1ksr as indicated in Fig. 4.6b, for HPr fragments 1 and 2 were combined.  $\rho_{sc}$  is the size-independent difference factor<sup>77</sup> of the backbone-atoms with respect to the conformation of the fragment in the protein averaged over the simulation excluding the first 2 ns. Simulations were started from the protein conformation ('fold') or a fully extended conformation ('extd'). Significant values of  $\rho_{sc}$ <sup>77</sup> are indicated as follows; highly similar,  $\rho_{sc} \leq 0.3$  (=), similar,  $0.3 < \rho_{sc} \leq 0.5$  (~). Fragments that score similar or better in both simulations are marked in bold (~).

in the respective PDB files) and designated “folded”. The other was started from a fully extended conformation and designated “extended”.

RMSD matrices of the trajectories of the isolated fragments provide a comprehensive insight into the behavior of the protein fragments during simulation (data not shown). Most matrices show clear indications of a beginning of convergence to a consensus cluster of conformations, i.e. many separate repeating off-diagonal blocks indicating the hopping between an extensive but limited number of conformational states. For HPr, these are fragments 4, 5, 7, and to a lesser degree fragments 6 and 8. The fragments that show this repetitive block structure best are not surprisingly also the smallest: it is most prominent in fragment 5, which is only 5 residues in size (see Table 5.1 for details).

### 5.3.2 Fragment Stability

The stability of the isolated fragments in the protein conformation can be estimated from the size-independent difference factor  $\rho_{sc}$  of the “extended” and “folded” runs with respect to the conformation of the peptide in the protein, as shown in Table 5.1. Low  $\rho_{sc}$  values in both runs indicates that the peptide tends towards the conformation found in the protein, and therefore is relatively stable. High values in the “folded” run indicate the opposite, i.e. relatively low stability. High values in the “extended” run indicate either relatively

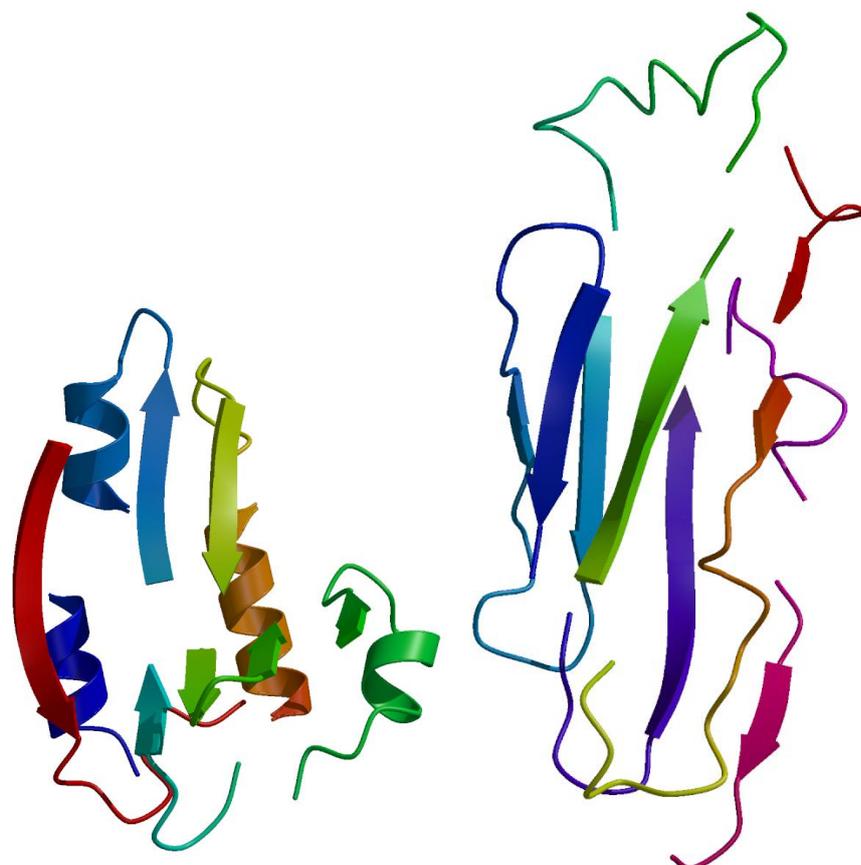


Figure 5.1:  
Overview of  
fragments from HPr  
and 1ksr  
(respectively on left  
and right-hand side)  
in an “exploded”  
view of the proteins.  
See Fig. 4.7 for a  
view of the “intact”  
proteins.

low stability, or poor convergence and therefore insufficient sampling of conformational space for a reliable estimate of the stability to be made.

The fragments that show the highest stabilities of the protein conformation and good convergence are **HPr:37-41** and **HPr:62-76** (fragments 5 and 8 respectively) and **1ksr:8-14**, **1ksr:15-21** and **1ksr:37-42**, (fragments 2, 3 and 6 respectively), see Fig. 4.6b. Fragment **HPr:37-41** is a short strand-turn-strand motif and **HPr:62-76** a strand-turn-helix motif, while the fragments from **1ksr** are all parts of loop regions. Due to the large size of fragment 8 from **HPr** and the borderline  $\rho_{sc}$  values, it is possible that this fragment is not actually stable in isolation (see Table 5.1).

We can define an “apparent poor convergence” as having low  $\rho_{sc}$  values for the “folded” and high values for the “extended” runs. Note, poor convergence is not restricted to the longer fragments. One of the shortest fragments, **1ksr:29-36** has not converged during the 20 ns of simulation, though on average better convergence would be expected for shorter fragments. Clearly much longer simulations would be required to ensure full convergence in all cases.

### 5.3.3 Relevance for Folding

The  $\beta$ -hairpin from **HPr**, fragment **HPr:37-41**, designated fragment 5 (see Table 5.1), was chosen for detailed examination with extended MD simulations and verification by NMR experiments, the results of which will be presented in chapter 6. To minimize possible effects of the ‘artificial’ termini of the fragment, it was extended on both sides by two residues. This elongates the  $\beta$ -sheet by two residues and results in fragment **HPr:35-43**, designated fragment pepAc (see Table 5.1). Several simulations of  $> 200$  ns were performed.

The similarity of the runs of fragment pepAc (**HPr:35-43**) with respect to the conformation of the peptide in the protein is less than that found in the runs of fragment 5 (**HPr:37-41**), see Table 5.1. This can be attributed in part to the much broader range of conformations that is sampled during these long runs than is possible during the relatively short simulations of fragment 5. Therefore the decreased similarity should not be taken as an indication for reduced long-term stability of fragment pepAc relative to fragment 5. A more exhaustive analysis of the clusters and dynamical behavior of pepAc is given in chapter 6.

## 5.4 Discussion

Fragment 5 (**HPr:37-41**) was the most dynamically distinct fragment identified in the consensus dynamics matrix in chapter 4. In simulations, it was comparatively stable isolated in solution with the dominant cluster corresponding to the protein native conformation populated for 26 % of the time. Therefore, it is proposed as the main candidate folding unit or nucleation site for the folding of **HPr**. Fragment 8 of **HPr** was also apparently marginally stable, but due to its relatively large size, there is doubt on the actual reliability of this assessment. For **1ksr**, three fragments were found stable from our criteria.

In sec. 4.5 tentative folding pathways were proposed for the proteins Sh3 and T4L, see Fig. 4.10, based on the motional hierarchy depicted in Figs. 4.2 and 4.4. In an analogous way, folding pathways were constructed for HPr and 1ksr from the hierarchies depicted in Fig. 4.6 and are depicted in Fig. 5.2. In these graphs, folding is thought to progress from the outer branches inwards. The fragments that comparative stability are marked with dotted circles. On the assumption that these fragments are likely to be involved in the first stages of folding, a statement can be made on the probability of the proposed folding pathways. For HPr,  $P_2$  seems unlikely due to the late incorporation of the relatively stable fragment 5. In  $P_1$  and  $P_3$  both fragments 5 and 8 are located on the outer branches of the

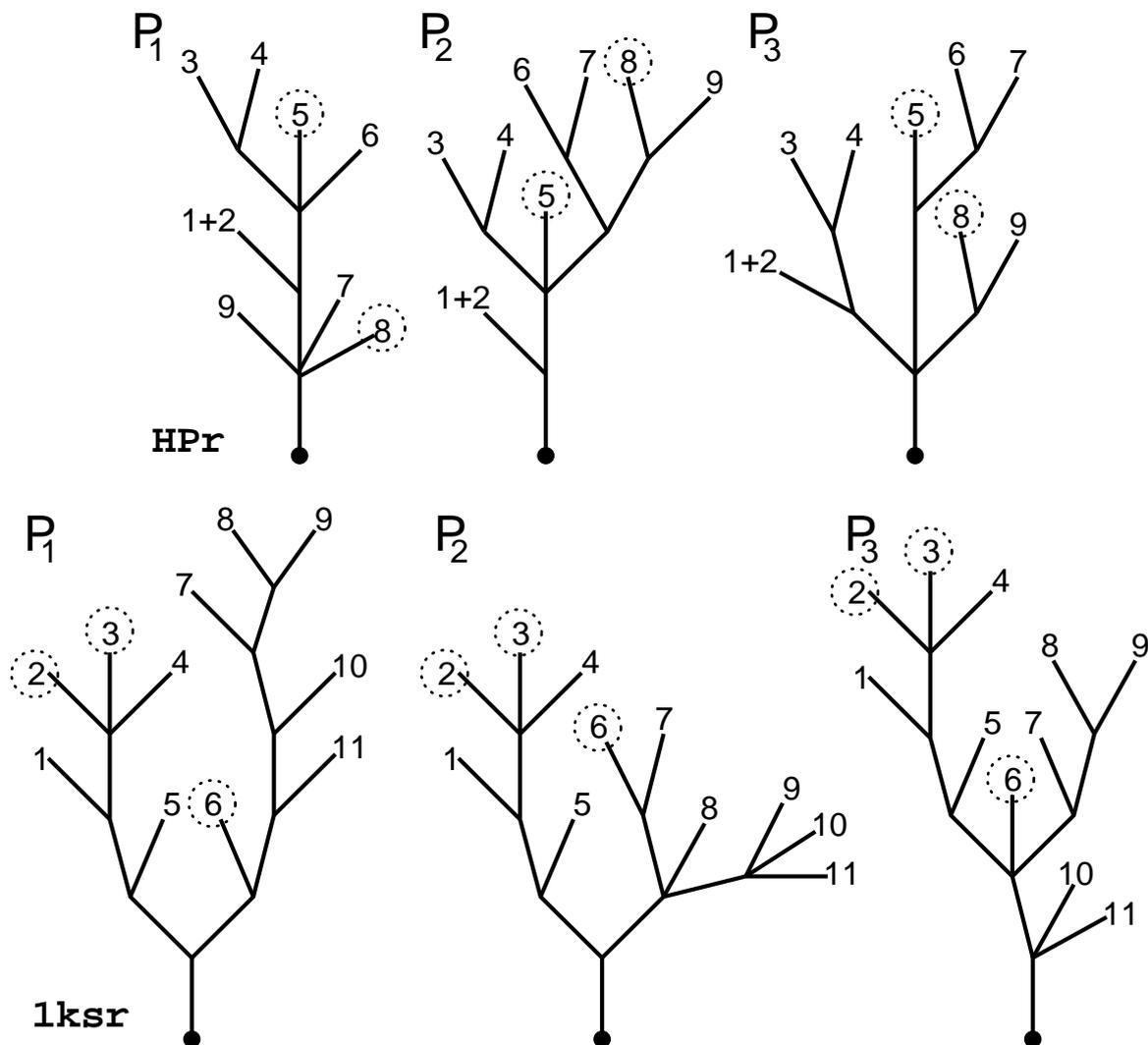


Figure 5.2: Non-unique examples of tentative folding pathways for HPr and 1ksr where folding can be thought to progress from the outer branches inwards. These pathways can account for the occurrence of all motionally coherent elements observed and correspond to the hierarchy as observed in Fig. 4.6. Fragments estimated to be stable are indicated with a dotted circle (see also Table 5.1).

folding tree, which indicates early involvement. Likewise, for 1ksr  $P_2$  seems most likely, and  $P_1$  and  $P_3$  unlikely due to the late incorporation of fragment 6.

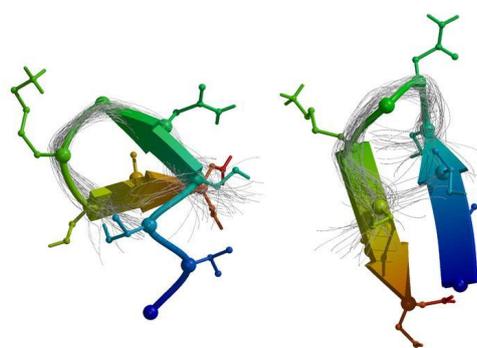
## 5.5 Conclusion

This method is the second of a series of small steps towards a comprehensive and fully functional simplified description of protein behavior. The first step was taken in the previous chapter (chapter 4), where a method for the extraction of motionally coherent elements from a protein was described. A next small step is taken in the next chapter (chapter 6), where the possibility of obtaining an experimental confirmation of the stability of one of the fragments is assessed. Many additional hurdles still remain, the major one being the determination of the fragment demarcations; in order to be able to use this approach without prior knowledge of a protein's structure, it will be necessary to reliably correlate the observed fragment demarcations with specific locations in the protein sequence, which may be recognizable by the presence of a certain type of residues or pattern of residues.



## Chapter 6

# Theoretical and Experimental NMR of a Small Peptide





## 6.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is currently the only experimental method that can yield high resolution structural information on peptides and proteins in solution. In particular, interproton dipolar cross relaxation rates derived from nuclear Overhauser effect (NOE or ROE) experiments can provide through space information on the proximity of different groups. In general the observed cross relaxation rates of NOE intensities are directly related to interatomic distances, assuming the molecule (peptide or protein) can be approximated as an isotropically tumbling rigid structure or in some cases an ensemble of isotropically tumbling structures. While this is in practice a reasonable assumption in the case of well structured larger molecules, mounting evidence suggests that for small flexible molecules such as peptides ignoring the internal or intramolecular dynamics and its correlation with the overall motion may give rise to severe artifacts in the interpretation.<sup>79, 80, 81, 82, 83</sup>

### 6.1.1 Outline of the Method

The availability of molecular dynamics (MD) simulations of small peptide systems that are much longer than the overall tumbling time of the molecules means that it is possible to follow approaches that involve much less severe assumptions. For example, from the dynamical correlations in the simulation the cross relaxation rates can be computed directly. First, spectral densities are obtained by a Fourier transform of the time-correlation functions of each of the proton pair distances. Then, a set of coupled differential equations that describe the relaxation of the coupled system of spins, can be constructed in the form of a relaxation matrix from the spectral densities and subsequently solved by diagonalization.<sup>84</sup> The incorporation of such dynamical correlations in the calculated cross-relaxation rates means in turn that the results will potentially become more sensitive to the precise details of the simulation. This will include the extent of sampling of conformational space and the fine details of the dynamics as determined by the time step scheme used.

In the current paper we consider these issues by investigating in detail the case of a nine-residue peptide of the histidine containing phosphocarrier protein HPr from *E. coli*.<sup>32</sup> This particular peptide has been proposed as a possible folding nucleation site and thus was expected to adopt a stable spatial fold, as described in chapter 4.<sup>2</sup> Molecular dynamics (MD) simulations described in chapter 5 support this notion. Here, we investigate whether the occurrence of a particular fold can be verified by comparing cross-relaxation rates calculated using different approaches to incorporate atomic motion, with cross-relaxation rates observed experimentally.

## 6.2 Methods

### 6.2.1 Simulation Set-up

#### Systems

Simulations were performed on a nine-residue peptide fragment corresponding to residues 35-43 from the protein HPr, an 85 residue  $\alpha$ /anti-parallel- $\beta$  sandwich protein.<sup>32</sup> HPr is the histidine-containing phosphocarrier protein from the phosphoenolpyruvate-dependent phosphotransferase system of *E. coli*. The amino-acid sequence of the fragment is V T S N G K S A S. The N-terminus was acetylated and the C-terminus amidated, -COCH<sub>3</sub> and -CONH<sub>2</sub> respectively, in order to minimize the difference between the isolated peptide and the fragment in the protein.

#### Parameters

All MD simulations were performed using the GROMOS-96 force field<sup>85,63</sup> and the GROMACS molecular dynamics package.<sup>35,64</sup> The LINCS algorithm<sup>10</sup> was used to constrain the length of all covalent bonds. Time steps used were 2, 4 and 7 fs. For the simulations at 4 and 7 fs, fast hydrogen atom and were eliminated from the system where possible, or high-frequency motions were reduced by modifying the mass distribution, as described in chapter 3.<sup>1,2</sup> This allows a time step up to 7 fs to be used. A twin-range cut-off was used to evaluate the non-bonded interactions. Within the short-range cut-off of 1.0 nm the Van der Waals and Coulomb interactions were calculated every simulation step. Within the long-range cut-off of 1.4 nm Coulomb interactions were only re-calculated during the neighbor-list update, about every 20 fs (10, 5 or 3 steps, depending on the time step). The temperature was maintained by weak coupling to a bath<sup>24</sup> at 300 K with a time constant of 0.1 ps. The protein and water were coupled independently. The pressure was maintained by weak coupling with a time constant of 1.0 ps to an external bath of 1 bar. A relative dielectric constant ( $\epsilon_r$ ) of 1.0 was used.

starting structure	$\Delta t$ (fs)	H/ plane	label	length (ns)	nr H <sub>2</sub> O	box (nm)	$R_{gyr}$ (nm)
folded	2	n	F <sub>2</sub>	573	744	2.91	0.58
folded	4	d/m	F <sub>4</sub>	250	852	3.00	0.57
folded	7	d/m	F <sub>7</sub>	600	852	3.00	0.58
extended	4	d/m	E <sub>4</sub>	196	1 538	3.63	0.59
extended	7	d/m	E <sub>7</sub>	691	1 538 881 <sup>a</sup>	3.63 2.93 <sup>a</sup>	0.58
total <sup>b</sup> /average <sup>c</sup>				2 310	666 1 538	2.37 3.63	0.58

Table 6.1: Overview of the five simulations of the nonapeptide in water. From left to right: starting conformation; time step ( $\Delta t$ ); treatment of hydrogen atoms and planar groups (H/plane): **n** normal or **d/m** dummified and modified masses;<sup>1</sup> identifying label; simulation length; number of water molecules; average box length; average radius of gyration  $R_{gyr}$  of the peptide in the simulation,  $R_{gyr}$  of folded starting structure is 0.55 nm, that of the extended starting structure is 0.99 nm.

<sup>a</sup> changed from a cubic box to a truncated octahedron at 303 ns  
<sup>b</sup> for simulation length    <sup>c</sup> for cubic and octahedral box respectively

## Equilibration

The “folded” structure of the nonapeptide was taken from the NMR solution structure of HPr (PDB<sup>55</sup> entry 1hdn<sup>32</sup>). This structure and a fully extended structure were first energy minimized using a steepest descent algorithm. The resulting structures were each solvated in a cubic or truncated octahedral periodic box of simple point charge (SPC) water<sup>27</sup>, with a minimum distance of 0.6 nm between the peptide and the box wall. The water box was constructed by replicating a cubic box containing 216 equilibrated SPC water molecules. All water molecules with the oxygen atom closer to any protein atom than the sum of their respective Van der Waals radii were removed. Simulation E<sub>7</sub> (Table 6.1) was changed from a cubic to a smaller truncated octahedral box after 303 ns by removing the water falling outside the octahedral box after centering the peptide. Energy minimization followed by 100 steps of unrestrained MD using a 2 fs time step was performed to relax the systems.

## Production Simulations

Five separate simulations were performed for lengths varying from 200 to 700 ns, using two different starting structures, folded and extended, and three different MD integration time steps, 2 fs, 4 fs and 7 fs. An overview of the simulations performed is given in Table 6.1.

### 6.2.2 Analysis of Conformational Space

#### Covariance Analysis

To analyze the sampling of conformational space, Principle Components Analysis (PCA) of the fluctuations in the atomic coordinates of the backbone (N, C and C<sub>α</sub>) atoms, also referred to as essential dynamics (ED) analysis,<sup>67</sup> was performed. This in turn was used as a measure of convergence between trajectories by determining the overlap in the subspaces sampled. The subspace sampled in a trajectory is reflected by the positional covariance matrix  $M$  determined during the PCA analysis. As a measure of the overlap  $\mathcal{S}$  between two trajectories, the following expression was used:<sup>62</sup>

$$\mathcal{S}(M_1, M_2) = 1 - \sqrt{\frac{\text{tr}([\sqrt{M_1} - \sqrt{M_2}]^2)}{\text{tr}(M_1) + \text{tr}(M_2)}} \quad (6.1)$$

where  $M_1$  and  $M_2$  correspond to the positional covariance matrices of the peptide in the two trajectories and  $\text{tr}(X)$  denotes the trace of matrix  $X$ . Extreme values of  $\mathcal{S}$  are 1 for identical matrices (i.e. completely overlapping subspaces) and 0 for orthogonal subspaces (i.e. no overlap).  $\mathcal{S}$  is proportional to the overlap of the square root of the fluctuations. A more detailed description of this measure and a summary of its properties is given by Hess.<sup>62</sup>

## Cluster Analysis

Clustering of structures from the simulations into conformations was performed based on the atomic positional root-mean-square difference (RMSD) of the backbone atoms of the middle five residues. A RMSD cut-off of 0.1 nm was used for determining neighboring structures belonging to one conformation. The procedure was as follows. For each structure, the number of neighbor structures was determined. The structure with the largest number of neighbors is the central structure of the first cluster or conformation, and all its neighbors are members of that cluster. These structures were removed from the pool of structures and the procedure repeated until no structures remained in the pool. A more complete description of this clustering algorithm is given by Daura *et al.*<sup>78</sup>

### 6.2.3 Analysis of Physical Properties

#### Diffusion and Rotational Properties

Diffusion constants ( $D$ ) were calculated from the mean square displacement (MSD) of the center of mass of the peptide in each separate simulation using the Einstein relation for diffusion in three dimensions and a linear fit of a plot of the MSD *vs.* time. Rotational correlation times ( $\tau_{Rot}$ ) for the peptide were calculated from a single-exponential fit to the averaged autocorrelation function for each of the components of the normalized normal vector of the plane through the  $C_\alpha$  atoms of residues 3, 5 and 7.

#### Hydrogen Bond Analysis

A hydrogen bond is considered to exist if the distance between the hydrogen and the acceptor atom is less than 0.25 nm and the angle donor-hydrogen-acceptor is larger than 120 degrees.<sup>39</sup> Only hydrogen bonds present for more than 4 % of the trajectory were considered significant. Hydrogen bond lifetimes were estimated from the integral of the autocorrelation of the hydrogen bond existence function after subtracting the asymptotic value at infinite time and normalization. This function is assigned a value of one when the hydrogen bond exists and zero otherwise. Average hydrogen bond lifetimes were calculated from the correlation function averaged over all hydrogen bonds.

### 6.2.4 NMR Experiments

#### Sample

The nonapeptide NAc-V-T-S-N-G-K-S-A-S-CONH<sub>2</sub>, was purchased from Eurosequence, Groningen as a freeze-dried powder. It was synthesized using solid-phase Fmoc chemistry from the C-terminus to the N-terminus and purified by HPLC to better than 95 %. The N-terminus was acetylated and the C-terminus amidated. Samples were prepared to the following result: 50 mM potassium phosphate buffer at pH 6.5, with 5 % (v/v) D<sub>2</sub>O, approximately 1 mg/ml azide and 8 mg/ml or 9 mM of peptide, in a total volume of 300  $\mu$ l.

## Spectra

All spectra were recorded on a Varian Unity Inova 600 MHz NMR spectrometer.  $^1\text{H}$  TOCSY,<sup>86,87</sup>  $^1\text{H}$  NOESY<sup>88</sup> and  $^1\text{H}$  ROESY<sup>89,90</sup> experiments were performed. Watergate detection<sup>91</sup> was used in the NOESY and ROESY experiments to suppress the water resonance. In addition, for the ROESY experiments the water was pre-saturated. NOESY experiments were performed with a mixing time of 200 ms. ROESY experiments were performed with mixing times of 50, 100, 150, 200 and 300 ms. A spectral width of 8 kHz was used in both dimensions and free-induction decays were acquired with 1 024 data points. The maximum evolution time was 64 ms and 81 ms for the NOESY and ROESY experiments respectively. Positive and negative frequencies were discriminated as described by States *et al.*<sup>92</sup> The sample temperature was 5 °C for all experiments. Data processing was performed using the program package SNARF.<sup>93</sup> Data points were weighted by a Lorentzian-to-Gaussian transformation in both frequency domains. Polynomial baseline corrections were applied in both domains. Cross-peak intensities were measured from peak heights, which is estimated to be accurate to within about 10 %.

### 6.2.5 Calculation of NMR Cross Relaxation Rates

Theoretical ROESY spectra were calculated using three different approaches. The cross relaxation rates obtained are directly related to intensities and can thus be compared to intensities in the corresponding experimental 2D-NMR spectra.

The first approach (called “flexible”) accounts for aspects of internal dynamics by computing the spectral density functions for all proton pairs from the time correlation functions of the interproton vectors and by solving the system of coupled differential equations for the intensities by diagonalizing the relaxation matrix as described by Peter *et al.*<sup>84</sup> Compared to the conventional analysis of NMR spectra and MD simulations by comparing experimentally derived distances and distance averages from simulation trajectories this approach has the advantage of including all effects of internal dynamics as well as of spin diffusion on the spectral intensities. Internal dynamics come in via fluctuations of the interproton distance, individual correlation times for the motion of each interproton vector and possible effects of coupling of internal dynamics with overall tumbling of the molecule, which influence the functional form of the time correlation function of the affected interproton vectors and thus the spectral density functions. Compared to the procedure described in the reference<sup>84</sup> small modifications were made. Specifically, in this study the spectral density functions are computed for *all* proton pairs and not only for a selected set. This was made possible by the use of fast Fourier transform (FFT) algorithms when computing the time correlation functions.

The second approach (called “rigid”) computes  $r^{-6}$ -averages,  $\langle r^{-6} \rangle$ , for all interproton vectors over all structures of a trajectory or a subset of such structures, e.g. the member structures of a cluster or conformation. Spectral densities are computed using a Lorentz function, an effective correlation time ( $\tau_{\text{eff}}$ ) which is the same for all proton pairs, and the  $r^{-6}$ -averages. From these spectral densities, the relaxation matrix is constructed and diagonalized to produce the intensities. This approach takes into account spin diffusion

effects, but not all effects of internal dynamics, because dynamics in the system are only included by their effect on the distance average, not through the angular motion of the interproton vector.

The third approach (called “naive”) computes  $r^{-6}$ -averages for all interproton vectors, and intensities are simply assumed to be proportional to these distance averages. Thus, fluctuations in the molecules are again only accounted for through the distance averaging, and, because the relaxation matrix is not diagonalized to solve the coupled system of equations for the intensities, spin-diffusion is neglected. This is close to the conventional way of analyzing NMR spectra in terms of a single averaged structure and analyzing MD simulations by comparing experimentally derived distances with simulated distance averages. For a mixing time of zero the intensities obtained by the “rigid” and the “naive” approach are identical, which corresponds to the well known fact, that spin-diffusion effects are negligible for short mixing times.

## 6.3 Results

### 6.3.1 Sampling and Convergence

Five separate simulations of the nonapeptide were performed using two different starting conformations and three different time step schemes (see Table 6.1). To estimate the degree of overlap between the regions of configurational space sampled in the different trajectories, their respective covariance matrices were used to obtain an  $\mathcal{S}$  value based on eqn. 6.1. The values obtained ranged between 0.8 and 0.9 (data not shown), which indicates a very high level of overlap.<sup>62</sup> A graphical representation of this overlap is presented in Fig. 6.1A-E which show projections of the separate trajectories onto the first two eigenvectors of the combined trajectories. The similarity in the configurations sampled was also assessed by calculating the root-mean-square difference (RMSD) of the backbone atom positions for the middle five residues of the peptide for all possible pairs of structures of the combined trajectories. The variation of the RMSD values within a given trajectory was similar to the variation between trajectories. There are frequent occurrences of high and low RMSD values between different trajectories demonstrating the regions of conformational space sampled in the different trajectories are very similar.

The dynamical properties of the nonapeptide such as the diffusion constant ( $D$ ), the

sim.	$D$ ( $10^{-9} \text{ m}^2 \text{ s}^{-1}$ )	$\tau_{rot}$ (ps)	$\tau_{HB}$ (ns)	$\tau_{eff}$ (ps)	Table 6.2: Dynamical properties of the nonapeptide from the simulations. From left to right: simulation label (sim.); diffusion constant for the peptide ( $D$ , with a standard deviation of $0.1 \cdot 10^{-9} \text{ m}^2 \text{ s}^{-1}$ ); overall rotation correlation time for the peptide ( $\tau_{rot}$ ); average correlation time for internal hydrogen bonds ( $\tau_{HB}$ ); and the effective correlation time ( $\tau_{eff}$ ) used in the “rigid” and the “naive” calculation methods of ROESY spectra as described in the text.
F <sub>7</sub>	0.55	213	0.35	65	
F <sub>4</sub>	0.50	261	0.37	80	
F <sub>2</sub>	0.45	252	0.61	80	
E <sub>7</sub>	0.60	215	0.38	65	
E <sub>4</sub>	0.59	236	0.53	75	
avg.	0.54	235	0.45	73	

rotational correlation time of the peptide ( $\tau_{rot}$ ), and the hydrogen bond lifetime ( $\tau_{HB}$ ) for each of the five simulations are summarized in Table 6.2. The differences between the trajectories for  $D$  and  $\tau_{rot}$  are small and within the uncertainty of the calculations.

### 6.3.2 Conformational analysis

Based on the backbone RMSD values described above, cluster analysis was performed on the combined trajectories, using a cut-off for the RMSD of 0.1 nm. The trajectories are dominated by two clusters with roughly equal populations. The first cluster (23% occurrence) corresponds to a  $\beta$ -turn conformation with a one-residue register shift in

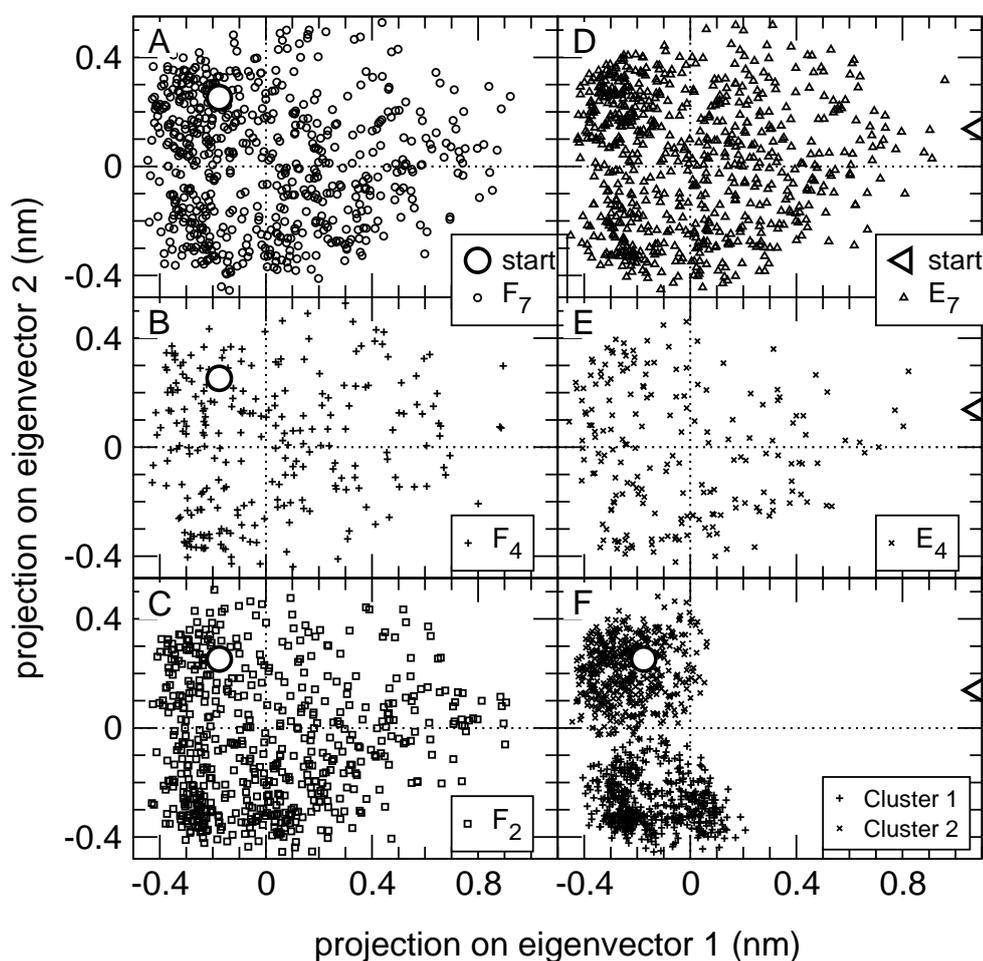


Figure 6.1: Comparison between sampled space of each of the five trajectories (panels A-E) and comparison between the two largest clusters of the trajectories combined (panel F). Shown are the projection of structures taken from each of the trajectories onto the first two eigenvectors of the five trajectories combined (small symbols), a projection of structures from the two largest clusters from the combined trajectories (small symbols in panel F), and a projection of the two starting structures: large circle for the folded structure and large triangle for the extended structure. One structure in every 1 ns of the trajectory was used.

hydrogen bonding with respect to the conformation found in the protein. The second cluster (22% occurrence) corresponds to the  $\beta$ -turn conformation found in the protein. No other cluster is populated for more than 8% of the time. These minor clusters cover a wide range of conformations. For example, the third cluster (8% occurrence) has a curled conformation with a single turn of  $\alpha$ -helix at one end. Fig. 6.2 shows the middle structures of the two largest clusters, as well as backbone traces for a selection of structures belonging to each cluster.

Fig. 6.1F shows a projection of the conformations from the two largest clusters onto the two largest eigenvectors from the ED analysis of the combined trajectories. It is clear from this projection that the clusters are well-separated in conformational space. Approximately 500 transitions in and out of each of the two largest clusters were observed during the trajectory, suggesting that the relative populations are statistically reliable.

Table 6.3 lists the occurrence of backbone-backbone hydrogen bonds in the combined trajectories of more than 4%. The maximum occurrence of any hydrogen bond in the combined trajectory is only 11%, indicating that the peptide is very dynamic. Many hydrogen bonds, however, were observed with a higher occurrence in specific clusters.

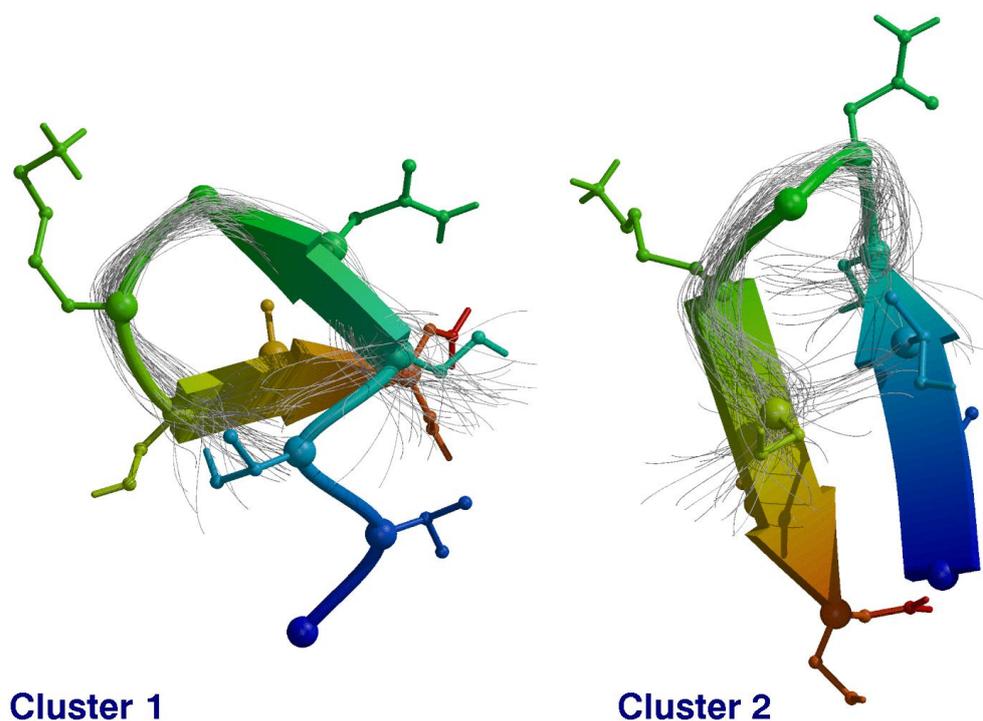


Figure 6.2: Structures from the two largest clusters from the combined trajectories. Cartoon models represent the middle structure of each cluster. Thin gray lines represent backbone traces of every tenth structure from the corresponding cluster. Plots were generated using a modified version of MolScript<sup>3,4</sup> and Raster3D.<sup>5,6</sup>

The hydrogen bonding pattern in cluster 2 corresponds closely with that observed in the ensemble of NMR solution structures of the whole protein. No significant differences in hydrogen bonding patterns were found between the five separate simulations (data not shown).

### 6.3.3 NMR Experiments

#### Assignments

Chemical shift assignments derived from TOCSY and NOESY experiments are listed in Table 6.4. Nuclei with an identical chemical environment or for which no distinction in chemical shift could be made, are grouped together. These are indicated by an asterisk ('\*') in place of a number in the atom names. The majority of resonances could be unambiguously assigned.

#### ROESY Intensities

ROESY spectra were recorded using a series of 5 mixing times. Intensities for all 5 mixing times could be measured for 113 cross peaks. Of these, 10 correspond to non-overlapping

Hydrogen-Bond		Occurrence in			
Donor	Acceptor	Protein	Comb.	Cluster nr	
N-H	O		traj's	1	2
Thr 2	Ace 1	-	4	4	5
Thr 2	Ser 9	-	4	<b>14</b>	-
Ser 3	Val 1	<b>17</b>	5	5	9
Ser 3	Lys 6	-	4	-	<b>13</b>
Ser 3	Ser 7	-	5	6	8
Asn 4	Val 1	-	6	5	-
Asn 4	Thr 2	<b>13</b>	6	<b>11</b>	5
Asn 4	Ser 7	-	8	<b>33</b>	-
Gly 5	Ser 3	-	5	5	-
Gly 5	Ala 8	-	6	-	<b>10</b>
Lys 6	Ser 3	<b>93</b>	6	-	<b>18</b>
Lys 6	Asn 4	<b>70</b>	9	-	<b>18</b>
Ser 7	Ser 3	-	5	-	9
Ser 7	Asn 4	-	<b>11</b>	<b>37</b>	-
Ser 7	Gly 5	-	5	6	5
Ala 8	Val 1	<b>93</b>	5	-	<b>12</b>
Ala 8	Ser 3	-	5	-	<b>12</b>
Ala 8	Asn 4	-	7	<b>21</b>	-
Ala 8	Gly 5	-	5	5	-
Ala 8	Lys 6	-	4	5	6
Ser 9	Thr 2	-	5	<b>17</b>	-

Table 6.3: Selected hydrogen bonds in the backbone which occur in the combined trajectories for more than 4% (21 in total). From left to right occurrence in the ensemble of protein NMR solution structures, in all five trajectories and in the two largest clusters. An occurrence larger than 10% is indicated in bold. An occurrence smaller than 4% is considered insignificant and is indicated by '-'.

cross-peaks between non-adjacent residues, i.e. are structurally most relevant. In addition, 85 non-ROE locations were identified. These are defined as positions in the spectra which correspond to the chemical shifts of a pair of protons where the experimental intensities are below the noise cut-off value, i.e. no peak is observed. A comparison between experimental and theoretical intensities was made for the 113 ROE peaks and the 85 non-ROE's. Approximately half of the peaks involve backbone hydrogens. Results are presented for a mixing time of 200 ms.

### 6.3.4 Comparisons using Theoretical Intensities

#### Calculation

ROE intensities were calculated for all possible proton pairs for each of the simulations using the three different approaches described in the Methods section, “flexible” (accounts for all aspects of internal dynamics and spin-diffusion effects), “rigid” (includes spin-diffusion effects but accounts for conformational fluctuations only through distance averaging) and “naive” (no spin-diffusion effects; only distance averaging). The CH<sub>2</sub>, NH<sub>3</sub> and NH<sub>2</sub> internal proton pairs were excluded from the comparisons as the distances between these proton pairs depend solely on the bonded interaction parameters of the force field and the method, with which hydrogens are reconstructed from united atoms. The total of 63 protons in the peptide (of which 56 were observable, see Table 6.4), translates into 53 relaxation sites (using a single site for methyl protons, and one site for all other protons)

Residue Atom	$\sigma_{1H}$	Residue Atom	$\sigma_{1H}$
Ace H <sub><math>\alpha^*</math></sub>	2.048	6 Lys H	8.255
1 Val H	8.371	H <sub><math>\alpha</math></sub>	4.381
H <sub><math>\alpha</math></sub>	4.155	H <sub><math>\beta_1</math></sub>	1.853
H <sub><math>\beta</math></sub>	2.074	H <sub><math>\beta_2</math></sub>	1.757
H <sub><math>\gamma^{**}</math></sub>	0.951	H <sub><math>\gamma^*</math></sub>	1.405
2 Thr H	8.450	H <sub><math>\delta^*</math></sub>	1.656
H <sub><math>\alpha</math></sub>	4.263	H <sub><math>\epsilon^*</math></sub>	3.026
H <sub><math>\beta</math></sub>	4.444	7 Ser H	8.542
H <sub><math>\gamma_2^*</math></sub>	1.206	H <sub><math>\alpha</math></sub>	4.450
3 Ser H	8.550	H <sub><math>\beta^*</math></sub>	3.882
H <sub><math>\alpha</math></sub>	4.447	8 Ala H	8.644
H <sub><math>\beta_1</math></sub>	3.891	H <sub><math>\alpha</math></sub>	4.359
H <sub><math>\beta_2</math></sub>	3.851	H <sub><math>\beta^*</math></sub>	1.417
4 Asn H	8.658	9 Ser H	8.362
H <sub><math>\alpha</math></sub>	4.734	H <sub><math>\alpha</math></sub>	4.361
H <sub><math>\beta^*</math></sub>	2.852	H <sub><math>\beta^*</math></sub>	3.872
H <sub><math>\delta_{21}</math></sub>	7.730	NH <sub>2</sub> H <sub>N1</sub>	7.616
H <sub><math>\delta_{22}</math></sub>	7.029	H <sub>N2</sub>	7.289
5 Gly H	8.502		
H <sub><math>\alpha^*</math></sub>	3.946		

Table 6.4: <sup>1</sup>H chemical shift assignments ( $\sigma_{1H}$  in ppm) for all observable <sup>1</sup>H nuclei, 56 in total. Multiple <sup>1</sup>H nuclei which are bonded to the same heavy atom and have identical chemical shifts are grouped together, which is indicated using a ‘\*’ instead of a number in the atom name.

for which intensities for 1378 pairs and 53 diagonal elements were calculated. These are referred to as “proton pair intensities”. The effective correlation time  $\tau_{\text{eff}}$  used in calculating the “rigid” intensities was empirically adjusted such that the calculated intensities of the backbone proton pairs match as closely as possible those from the “flexible” approach. It is much smaller than the overall rotational correlation time of the peptide ( $\tau_{\text{rot}}$ ), because it includes contributions from internal dynamics. The values of  $\tau_{\text{eff}}$  for each of the five simulations are listed in Table 6.2.

### Sampling Effects

To illustrate the effects of simulation protocol and sampling on the calculated intensities, examples of the correlation between the calculated proton pair intensities from different simulations using a “rigid” and a “flexible” treatment and a mixing time of 200 ms are shown in Figs. 6.3 and 6.4. Fig. 6.3 addresses the influence of the starting structure (E versus F) and the simulation lengths (200 ns versus 600 ns), i.e. the degree of sampling, by comparing the spectra of F<sub>7</sub> with E<sub>7</sub> (lengths:  $\sim 600$  ns) and F<sub>4</sub> with E<sub>4</sub> (lengths:  $\sim 200$  ns). In all cases a high degree of correlation was found ( $> 0.99$ ), which indicates, that the spectra are either not sensitive to differences between the simulations or that the simulations are indeed not very different. The correlations are in general higher for the longer simulations (F<sub>7</sub> *vs.* E<sub>7</sub>: 0.998 “flexible” and 0.9996 “rigid”) than for the shorter ones (F<sub>4</sub> *vs.* E<sub>4</sub>: 0.991 “flexible” and 0.999 “rigid”), which shows that the spectra converge upon better sampling to a “final” theoretical spectrum, which is independent from the starting structure. The data can also be weighted relative to the subjective structural relevance by using  $I_{\text{ROE}}^{1/6}$  instead of  $I_{\text{ROE}}$ .<sup>94</sup> Since  $I_{\text{ROE}}^{-1/6}$  is related to interatomic distance,  $I_{\text{ROE}}^{1/6}$  can be thought of as a reciprocal apparent distance. Using this weighting gives a slightly clearer separation of the correlation: F<sub>7</sub> *vs.* E<sub>7</sub> yields 0.99, whereas F<sub>4</sub> *vs.* E<sub>4</sub> yields 0.97. The correlations for the “flexible” spectra are lower than for the “rigid” ones. As the explicit calculation of the spectral density functions makes the “flexible” calculation more sensitive to dynamical details in the simulation than the “rigid” calculation method, we conclude from this observation that the major differences between the simulations, for which it takes more time until they are sufficiently sampled so that they are converged, lie in the dynamical processes. In the figure it is also indicated, which of the reported intensities belong to backbone-proton pairs. Their correlation between different simulations is very similar to the correlation found, if all pairs are considered. This shows that the similarities and differences observed are not a mere effect of sidechains (and sidechain dynamics) only.

Fig. 6.4 shows examples of the correlation between simulations with different time steps and hydrogen/planar group treatment starting from the folded structure. Apparently, the simulation protocol has little effect on the properties of the simulated ensemble that affect the NMR spectra.

### Theoretical Models

In this section we compare theoretical spectra that were computed with the different models described above in order to assess the importance of the various approximations for

spectral intensities. Fig. 6.5 shows the correlation of the theoretical intensities computed with the “flexible” approach with those computed with the “naive” approach (Fig. 6.5, panel A), the correlation between the “rigid” and the “naive” approach (panel B), and the correlation between the “flexible” and the “rigid” approach (panel C). All representations are given in logarithmic scaling, which weights the peaks according to their importance. Less intensive peaks which often correspond to structurally relevant distant interactions are emphasized. With the comparison between “rigidly” and “naively” computed spectra (panel B) the importance of spin diffusion is highlighted. The effect on single intensities (especially with rather weak signals) can easily be a factor of two or more relative to other

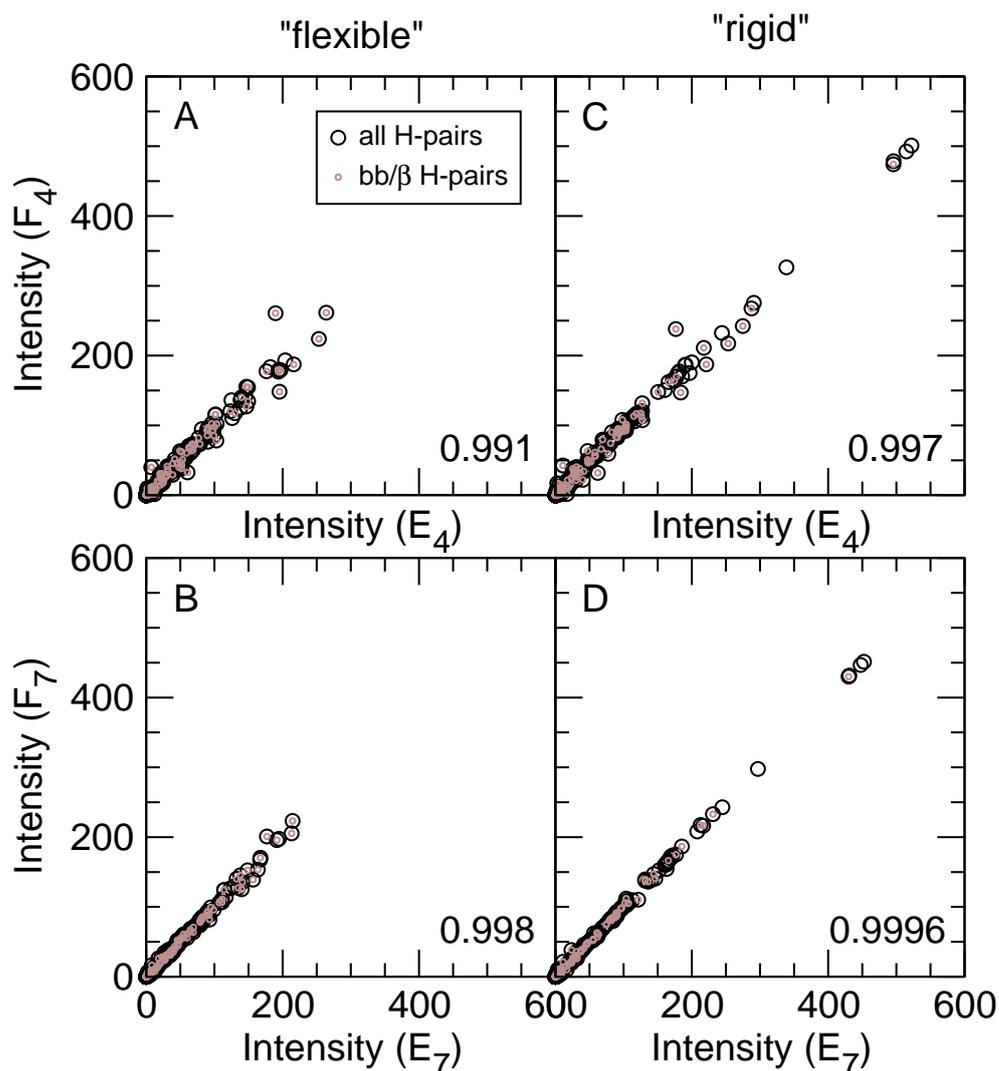


Figure 6.3: Comparison of theoretical proton pair intensities using the “flexible” (panels A/B) and “rigid” (panels C/D) methods between simulations with different starting structures (E: extended; F: folded) and of different lengths ( $E_4$ ,  $F_4$ :  $\sim 200$  ns;  $E_7$ ,  $F_7$ :  $\sim 600$  ns). Distinction is made between all proton pairs and those involving backbone or  $\beta$  protons. Numbers refer to the correlation coefficients for a linear fit to the plot of all proton pairs.

pairs. Looking at the figure, the important effect is not the deviation of the slope of the correlation curve from 1, it is well known that spin diffusion in general lowers intensities because it adds possible relaxation processes. What is relevant, is, that this effect is different for different proton pairs, which changes their relative intensities. This results in the spread of the points and the deviation of the correlation coefficient from 1. The effect is greater for longer mixing times and is significant for mixing times longer than 100 ms for this system.

The comparison of the “flexible” and the “rigid” approach (Fig. 6.5C) points out those effects of internal dynamics on the spectral intensities that go beyond simple distance averaging and may be important for small, highly flexible molecules. In such molecules,

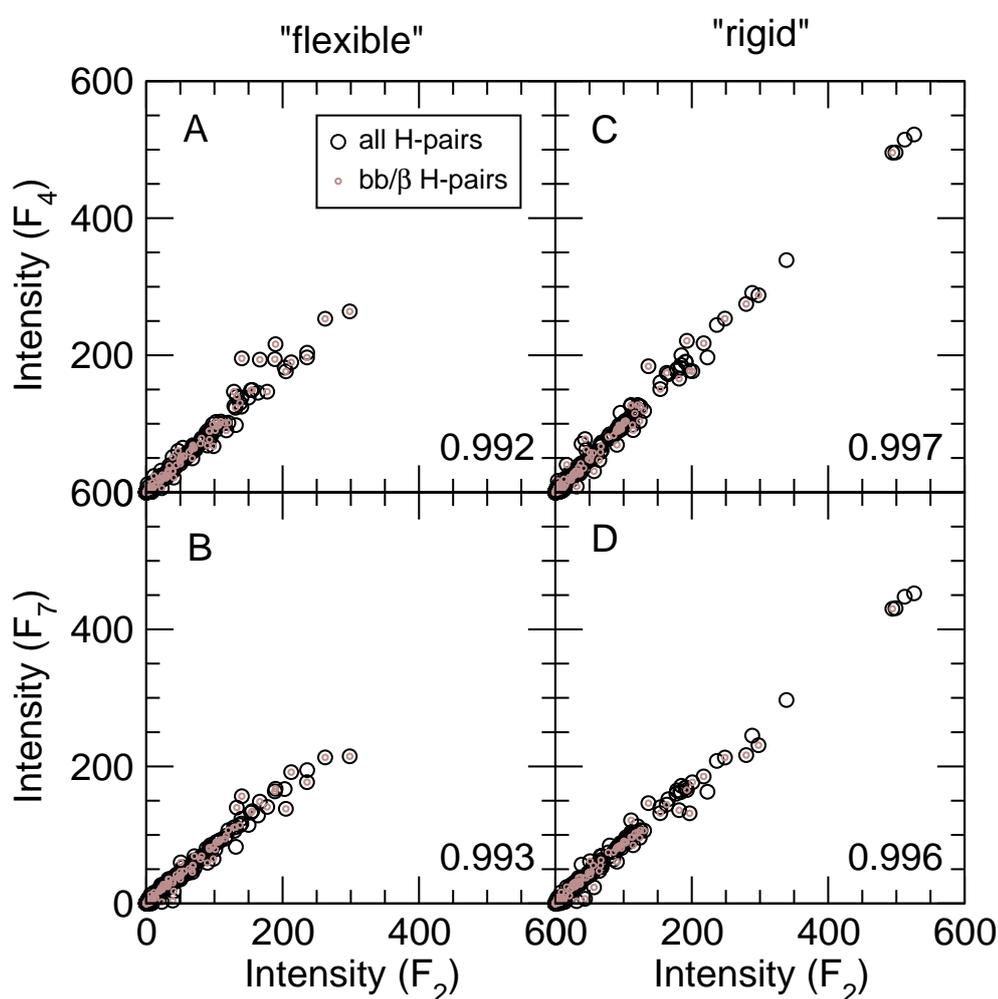


Figure 6.4: Comparison of theoretical proton pair intensities using the “flexible” (panels A/B) and “rigid” (panels C/D) methods between simulations with different time steps/hydrogen treatments ( $F_2$ : normal dynamics with  $\Delta t = 2$  fs;  $F_4$  and  $F_7$ : dumified hydrogens and modified masses with  $\Delta t = 4$  fs and 7 fs respectively; see Table 6.1). Distinction is made between all proton pairs and those involving backbone or  $\beta$  protons. Numbers refer to the correlation coefficients for a linear fit to the plot of all proton pairs.

separate regions may, for example, show individual correlation times which affect the intensities of proton pairs located in those regions. Again, it is not the slope of the correlation curve that is relevant. Additional motions will in general shorten the “effective” correlation time and induce further relaxation processes. Instead, we want to focus on the spread of the points. This spread shows that internal dynamics have a different effect on different intensities and thus on the relative lengths of the derived distances if one determines structure from NMR spectra. The comparison between the “flexible” and the “naive” approach (Fig. 6.5A) summarizes these influences on the NMR spectra.

As we see from the three panels of Fig. 6.5 discussed so far, the effects mentioned above

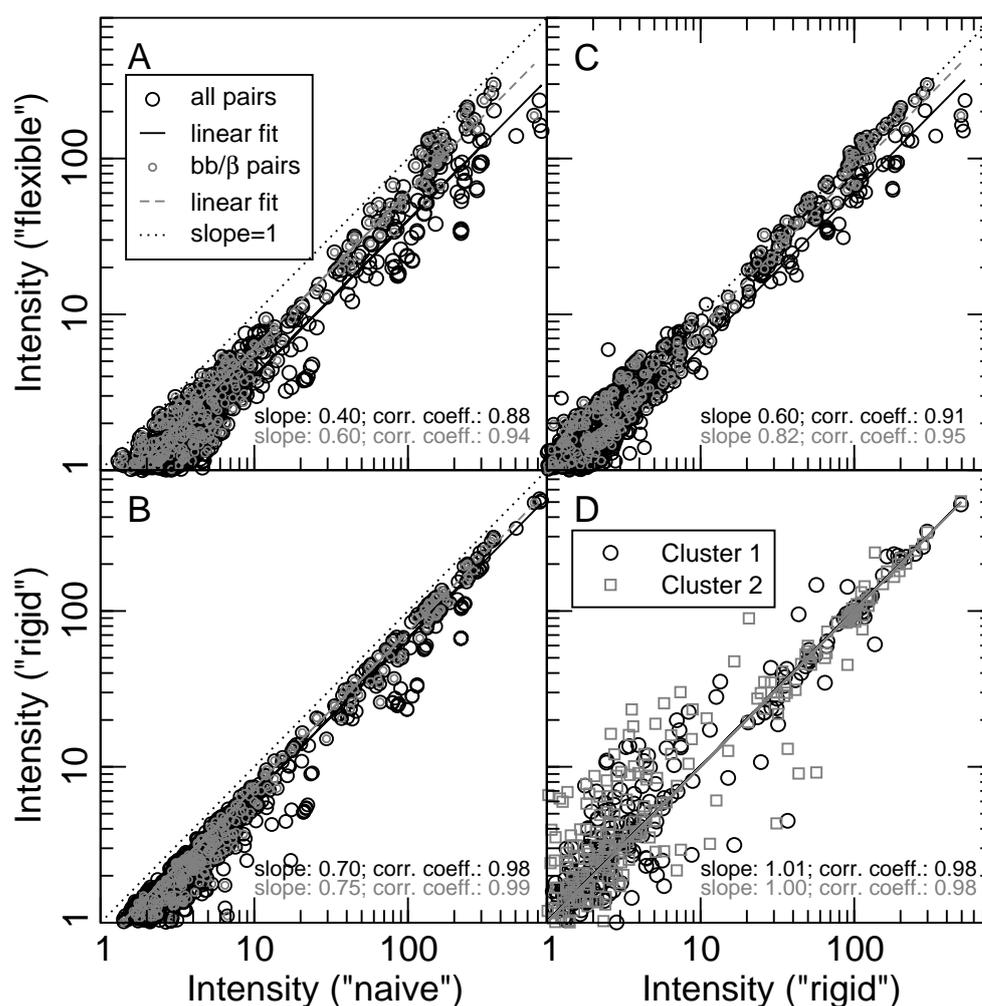


Figure 6.5: Comparison between theoretical proton pair intensities, calculated using three different approaches at a mixing time of 200 ms and for the  $F_2$  trajectory and for the two largest clusters of the combined trajectories. The intensities obtained with the three methods, “naive”, “rigid” and “flexible”, are plotted against each other in panels A-C. Distinction is made between all proton pairs and those involving backbone or  $\beta$  protons. Panel D shows intensities calculated using the “rigid” method for the two largest clusters vs. the whole trajectory (for backbone protons only). Linear fits are also shown and the corresponding slopes and correlation coefficients are reported.

influence proton pairs, but seem less relevant if only backbone protons are considered. This leads to the conclusion, that while for the nonapeptide simulated properly accounting for all possible influences on NMR spectra does indeed affect the results. If one is only interested in backbone conformations the effect is not large.

### Complete Ensembles and Single Conformations

Upon calculating theoretical spectra of selected groups of structures that represent a particular backbone conformation, i.e. all members in one cluster, and comparing those (“rigid”) spectra with the “rigid” spectrum of the complete simulation trajectory the sensitivity of these ROESY spectra to the backbone conformation can be investigated. It is also of interest, whether single conformations, for example the most important clusters, are sufficient to describe the complete simulated conformational ensemble in the sense that they yield a ROESY spectrum similar to that of the complete simulation trajectory. Alternatively, whether a particular cluster from the simulated conformational ensemble corresponds more closely with the experimental NMR spectroscopy data than the other clusters. In Fig. 6.5D the correlation of the “rigid” spectra of the first two clusters with the “rigid” spectrum of the complete simulation is displayed (for backbone protons only) in logarithmic scaling. The wide spread of the points shows, that the spectrum is highly sensitive to conformation. The correlation coefficient is nevertheless very high, and in fact higher than the correlation for the comparison between the “flexible” and the “rigid” approach; 0.95 for “flexible” *vs.* “rigid” and 0.98 for both clusters 1 and 2 *vs.* complete trajectory. This high correlation coefficient is probably due to the fact, that it was computed with a linear fit, which strongly emphasizes the peaks with high intensities. But, these proton pairs are very close in space and usually also close in sequence. This means that they are of little relevance structurally. A visual judgment, looking at the scatter of the points is thus more appropriate. Both cluster 1 and cluster 2 represent the spectrum of the complete simulation equally badly. Using  $I_{\text{ROE}}^{1/6}$  instead of a  $I_{\text{ROE}}$ ,<sup>94</sup> weighting the correlation coefficients obtained correspond better to the ‘visual’ impression gained from Fig. 6.5D: “flexible” *vs.* “rigid” gives 0.99, cluster 1 *vs.* complete trajectory gives 0.87 and cluster 2 *vs.* complete trajectory gives 0.84.)

In summary, none of the predominant conformations is sufficient to describe the spectrum that would result from the simulated ensemble. In addition, neither of the two conformations represents the ensemble significantly better than the other. It is probably not possible to represent a spectrum of such a flexible compound by just one conformation. As the spectra seem to be very sensitive to conformations, any comparison between experimental spectra and the spectra computed from MD simulations becomes meaningless if the simulated ensemble does not sample the populations of “relevant” conformations correctly.

#### 6.3.5 Comparison between Theoretical and Experimental Intensities

Intensities calculated from the simulation data were also compared to intensities obtained experimentally. The theoretical intensities were generated by summing the individual

proton pair intensities for all pairs possibly associated with a location in the experimentally obtained spectra, based on the chemical shift assignments listed in Table 6.4 and the observed peak widths. Experimental intensities for which the theoretical prediction would depend primarily on bonded interaction parameters in the force field were excluded from the comparison. In total there were 98 locations for which peak intensities could be determined experimentally. In addition, 85 locations were considered for which peak intensities could be calculated but for which no peak was present in the experimental spectrum (non-ROE's).

Fig. 6.6 shows the correlation between the calculated and experimental intensities using the three different methods. The correlation between the theoretical and experimental intensities is virtually identical for the three methods (*vis.*  $\sim 0.8$ ). The noise level observed in the experimental spectra is about 1 in the units used in Fig. 6.6. Absolute scales for experimentally derived intensities and those calculated using the various methods cannot be compared directly. Therefore, it was required that the predicted non-intensities were below 5. The average calculated intensity for the non-ROE's was 6.7 for the "naive" method, 4.1 for "rigid" and 3.2 for "flexible". This however is not a significant difference as the theoretically computed intensities from the "naive" approach are systematically higher than those from the "rigid" approach which are again higher than those from the "flexible" approach. This is discussed in section "Theoretical Models" and is also reflected in the different slopes in Fig. 6.6.

Based on the correlation between the experimental and theoretical intensities, no distinction can be made between the three calculation methods used. None of the methods reproduces the experimental data better than the others. The deviation from experiment may originate in the trajectories generated, i.e. may be due to shortcomings in the force field and/or simulation methods, since all three methods used to calculate the intensities perform equally badly. It should be noted that the comparison in Fig. 6.6 was performed using the  $F_2$  simulation, which does not have the special treatment of hydrogens and planar groups. Results obtained for the other trajectories (data not shown) are very similar to those for  $F_2$ .

Although the spectra computed from the complete simulation and from single clusters are very different, it is not possible to select a conformation or ensemble that gives significantly better agreement with the experimental data. In all the cases investigated the correlation of the theoretically computed intensities with the experimental ones is in similar ( $\sim 0.8$ , data not shown).

## 6.4 Discussion

In the simulations the nonapeptide adopts in aqueous solution a comparatively stable conformation: the dominant cluster or conformation corresponding to the native conformation of the nonapeptide in the protein HPr is populated for 22% of the time. The hydrogen bonding pattern of the second most populated conformation is similar to that of the nonapeptide in the protein HPr. NMR experiments indicated that the nonapeptide is partly structured in solution. However, the data measured is only compatible to a limited

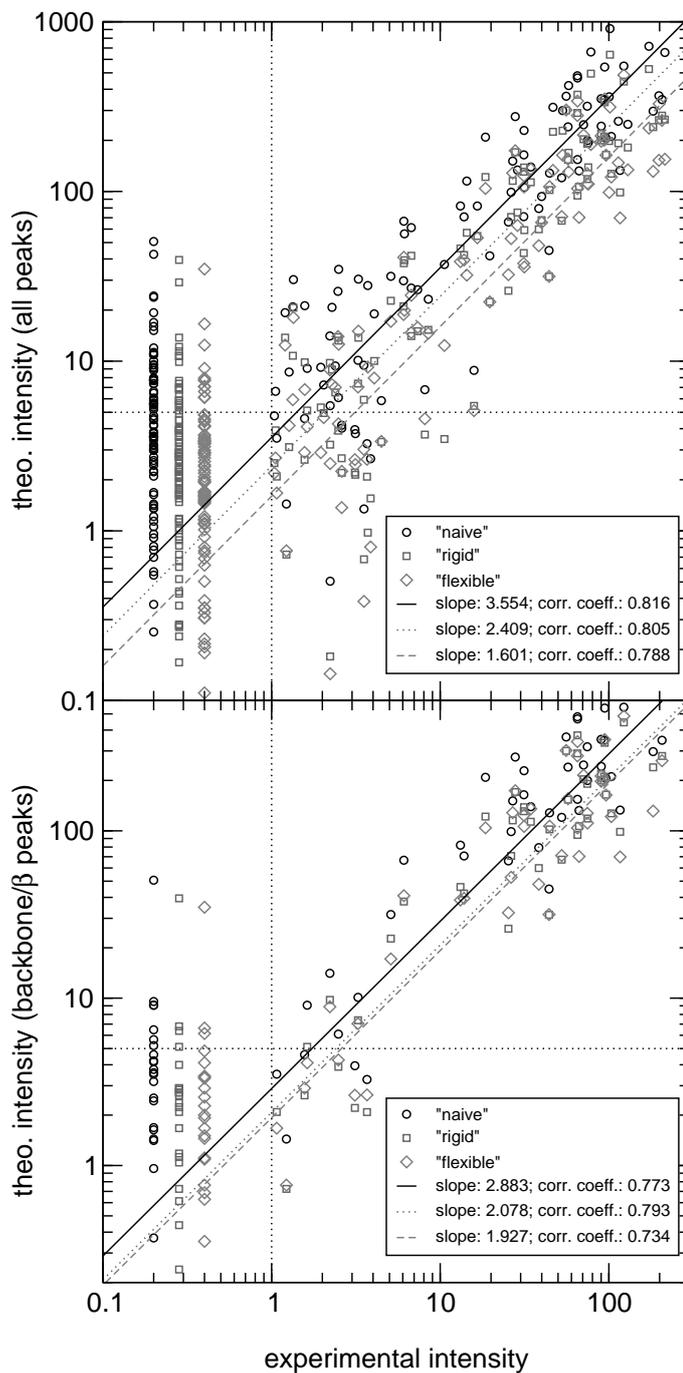


Figure 6.6: Comparison between theoretical and experimental ROE peak and non-ROE intensities. Experimental non-peak ‘intensities’ are set to arbitrary values of between 0.2 and 0.4 (on the left-hand side in the plot) with an offset added for clarity between the three data sets in each plot. Dotted lines indicate the noise level (1) as determined from the experimental spectra. Theoretical intensities are shown for three models for the molecular motion. The top panel shows all proton pairs, the bottom panel only those involving backbone or  $\beta$  protons.

extent with the conformations sampled in the simulations.

Within the accuracy given by the correspondence between the experimental and theoretical peak intensities, no distinction can be made between the different methods used to derive the theoretical intensities. Even when intensities are naively assumed to be directly proportional to  $\langle r^{-6} \rangle$  averaged distances, a similar correspondence with the experimental data is found. All three methods show a large spread in intensities, in the order of a factor of 10. It should be noted that the corresponding spread in interproton distance would be considerably smaller, about a factor of  $10^{1/6} \approx 1.5$  or a deviation of around 50 %.

Using properties calculated from the different trajectories, the effects of performing simulations at different time steps and using different hydrogen atom and planar group treatments could be investigated. The sampling of conformational space was analyzed by calculating the overlap between different trajectories (by means of the  $\mathcal{S}$  parameter described in the *Methods* and *Results* sections) and by comparing the projection of the trajectories onto the two largest eigenvectors as determined from the trajectories combined (Fig. 6.1). In both cases, no effect of the size of the time step is apparent. Also a direct comparison of all conformations from all trajectories performed by calculating the RMSD of the atomic positions of the backbone of the middle five residues, to generate a RMSD matrix also showed that the trajectories were effectively indistinguishable.

Marginal differences in the occurrence of two out of the 21 hydrogen bonds analyzed were observed between simulations at different time steps. Likewise, dynamical parameters such as the diffusion, rotational correlation time, average hydrogen bond lifetime, average cluster lifetime all show only marginal differences between the different simulations and simulation set-ups.

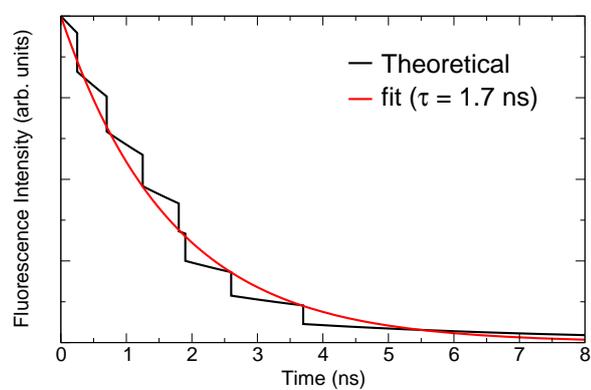
Our results demonstrate that for the case of a small flexible polypeptide in solution different approaches to the treatment of molecular motion can lead to significant variations in calculated NOE/ROE intensities and hence in the interpretation of experimental spectra. This shows once again that for small, flexible peptides in solution, 2D-NMR spectra alone provide insufficient information to determine what are the dominant conformations of the peptide. Only when simulations are of sufficient accuracy to provide very detailed information on the nature of the molecular motion and can be used to predict the experimental spectra with high accuracy will we have a true picture of the behavior of these systems at the atomic level.





## Chapter 7

# Fluorescence and Dynamics of Flavin Adenine Dinucleotide





## 7.1 Introduction

Flavoproteins are an interesting class of enzymes for studying the dynamic behavior of biomacromolecules. One reason for this is the large amount of detailed information on both catalytic mechanism and atomic structure that is available for many members of this wide-spread family of redox enzymes. The feature which makes these enzymes particularly suitable for investigating the role of conformational dynamics in catalysis is the flavin cofactor itself. This prosthetic group is not only the redox-active group situated in the heart of the active site, but it is also a naturally fluorescent group that emits green light. In the last decade, time-resolved fluorescence and fluorescence anisotropy studies have yielded information on the dynamics of the active site of a variety of flavoproteins (for an overview see Van den Berg & Visser<sup>95</sup> and Van den Berg *et al.*<sup>96</sup>).

A complicating factor in studying the fluorescence properties of these enzymes is the fact that in most flavoproteins the flavin cofactor is non-covalently bound. In principle it is thus possible to detect fluorescence from the free cofactor as well as the protein-bound one. In enzymes such as glutathione reductase (GR), lipoamide dehydrogenase and thioredoxin reductase (TrxR), the dissociation constant for the flavin cofactor is very low (pM-nM range). For these enzymes, careful sample preparation is sufficient to avoid traces of free flavin. However, in other flavoenzymes such as ferredoxin NADPH-reductase (FNR) and para-hydroxybenzoate hydroxylase, the flavin is less tightly bound. In these cases, the presence of free flavin has to be taken into account, especially when the fluorescence quantum yield of the free flavin cofactor is high compared to that of the protein-bound cofactor. For this, knowledge of the fluorescence properties of free flavin cofactors under the experimental conditions used is required.

For several decades, the physical properties of isoalloxazine ring and the two most common flavin cofactors, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD, Fig. 7.1) have been subject of investigation (for reviews on electronic and structural properties of flavins we refer to Song;<sup>97</sup> Leijonmarck;<sup>98</sup> and references therein). The remarkably low fluorescence of FAD with respect to free riboflavin was first reported by Weber.<sup>99</sup> An intramolecular ground-state complex between the isoalloxazine ring and the adenine moiety was proposed to prevail in aqueous solution, resulting in the formation of a non-fluorescent complex. In steady state fluorescence experiments, the fluorescence quantum yield of FAD was found to be 9 times lower than that of FMN. By enzymatic digestion of the diphosphate bridge of the FAD molecule the fluorescence intensity increased to equal that of free FMN. From these experiments it was proposed that FAD exists in two conformations; a so called 'closed' conformation, in which the isoalloxazine and adenine rings interact through  $\pi$ - $\pi$  interactions in a stacked conformation resulting in very efficient fluorescence quenching, and an extended, 'open' conformation that is responsible for the (remaining) fluorescence of FAD. In solution, the FAD molecule is considered to be predominantly in the stacked conformation. Under the assumption that the equilibrium of stacking and unstacking is the same in ground state and first excited state, it can be reasoned that about 80 % of the FAD molecules are in the closed conformation and that the lifetime of this intramolecular complex is 27 ns.<sup>100</sup> Dispersion forces are considered to be the principal factor favoring flavin-adenine stacking. Strong acidic conditions, non-polar solvents and even polar solvents such as formamide have been reported to prevent complex

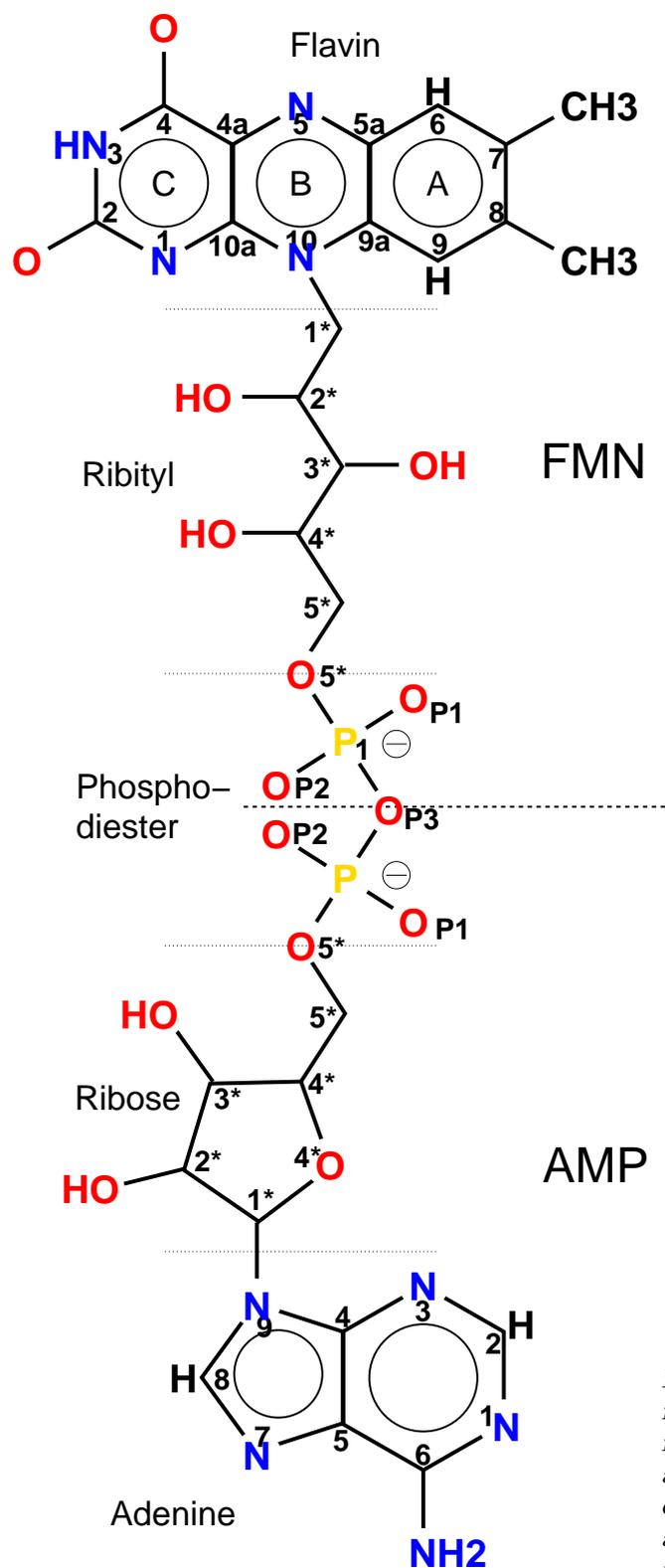


Figure 7.1: Representation of FAD including atomic labeling. The FMN and AMP moieties are indicated separately, as well as nomenclature for various other moieties. Atom labels for the AMP moiety are prepeded by an 'A' when referred to in the text.

formation in FAD<sup>101</sup> (and references therein).

Although the existence of an 'open' and 'closed' conformation of the FAD molecule is now generally accepted, little is known about the structural details and conformational dynamics in solution. Crystallographic studies on the complex between free riboflavin and adenosine<sup>102,103</sup> revealed an average structure of the  $\pi$ - $\pi$  stacked isoalloxazine and adenine ring systems under crystal packing forces. In aqueous media, support for a stacked conformation was obtained from ultraviolet resonance Raman spectroscopy, where both chromophores showed Raman hypochromism.<sup>104</sup> From NMR studies, different models were proposed for the interaction between the flavin and adenine moieties, including both intramolecular stacking,<sup>105,106</sup> and parallel intramolecular hydrogen bonding between the pyrimidine-like ring of the flavin involving N<sub>3</sub> and O<sub>4</sub>, and the adenine NH<sub>2</sub> and N<sub>7</sub>.<sup>107</sup> All NMR studies, however, were complicated by the tendency of flavins to form intermolecular complexes at millimolar concentrations.

Early time-resolved fluorescence measurements have provided a picture consistent with the dynamic model proposed by Weber and co-workers. Decay analysis of FMN revealed a fluorescence lifetime of 4.7 ns, while FAD yielded two lifetime components of resp. 2.8 ns (72 %) and 0.3 ns (28 %).<sup>108</sup> In the last decade, technical progress has provided a broader time-window to sample the dynamics of the excited state. In this paper, the flavin cofactors are therefore re-examined using sub-nanosecond resolved time-correlated single photon counting experiments. Fluorescence lifetime properties, however, relate to the ensemble of molecules, and do not yield information on the atomic level. For proper understanding of the process of stacking and the effect of dynamics on the lifetime of the excited state, a more detailed picture is needed.

### 7.1.1 Outline of the Method

In this study we have investigated the structural dynamics of FAD using a combination of time-resolved fluorescence and MD simulations. Polarized (sub)nanosecond-resolved fluorescence experiments under various temperature and solvent conditions yield experimental data on the dynamic behavior of the flavin cofactor. Nanosecond molecular dynamics simulations in water give insight into the dynamic behavior of the FAD molecule. Changes in the charge distribution were applied to mimic the electronic effect of transition from the ground state to the excited state following light absorption<sup>109</sup> and to investigate its effect on the dynamics of the molecule. Special attention was given to the interrelation between the MD data and fluorescence data in terms of fluorescence quenching and rotational behavior.

## 7.2 Methods

### 7.2.1 Fluorescence measurements

#### Reagents and Sample

FAD and FMN of the highest purity available were purchased from Sigma. A Biogel P2 column (Biorad) equilibrated with the appropriate measuring buffer, was used prior to the fluorescence experiments to remove possible traces of degradation products of FAD. Chromatography experiments with FMN using fluorescence detection showed that FMN did not contain any fluorescent impurities. All measurements were carried out in 50 mM potassium phosphate buffer *pH* 7.5, except for the control experiments performed in D<sub>2</sub>O. Buffers were made from nanopure-grade water (Millipore) and were filtered through a 0.22  $\mu\text{m}$  filter (Millipore). The samples had a maximum OD of 0.10 at the wavelength of excitation. The concentration was calculated from the molar extinction coefficient, resp.  $\epsilon_{450,\text{FAD}} = 11.3 \text{ mM}^{-1}\text{cm}^{-1}$  and  $\epsilon_{445,\text{FMN}} = 12.5 \text{ mM}^{-1}\text{cm}^{-1}$ .<sup>110</sup> Fluorescent grade glycerol was purchased from Merck. Sample preparations were performed at 277 K in the dark.

#### (Sub)nanosecond Polarized Fluorescence

Time-correlated single photon counting (TCSPC) was used to detect polarized time-resolved fluorescence on a (sub)nanosecond time-scale. Details of both the set-up and measurements procedures are described elsewhere,<sup>111,112</sup> and only a short outline will be given below. A mode-locked CW Nd:YLF laser was used to synchronously pump a cavity-dumped dye laser. Vertically polarized light of 450 nm (Stilbene 420) or 460 nm (Coumarine) was used to excite the sample with a frequency of 594 kHz and a duration of 4 ps FWHM. The excitation intensity was adjusted to yield a detection frequency of 30 kHz in the parallel component. Parallel and perpendicular polarized fluorescence were detected through resp. 557.9 nm (for excitation at 450 nm) or 526.0 nm (for excitation at 460 nm) interference filter (Schott, Mainz, Germany, half band width of 11.8 nm) in combination with KV 520 cut-off filter (Schott). FMN data and regular FAD data were collected in a multichannel analyzer with a time-window of 1024 channels at typically 15 – 20 ps/channel. For a better determination of the ultra-short fluorescence lifetime of FAD a time-window of 8000 channels at typically 3 – 4 ps/channel was used. The dynamic instrumental response of the set-up was obtained at the emission wavelength using erythrosine B in water ( $\tau = 80$  ps at 293 K) as a reference compound.<sup>113</sup> This instrumental response function is approximately 40 ps FWHM, which makes registration of 5 – 10 ps lifetime components realistic. The temperature of the samples was controlled using a liquid nitrogen flow set-up with a temperature controller (model ITC4, Oxford Instruments Inc., Oxford, United Kingdom).

Analysis of the fluorescence intensity decay  $I(t)$  and anisotropy decay  $r(t)$  was performed with a model of discrete exponentials using the 'TRFA Data Processing package' of the Scientific Software Technologies Center of *Belarussian State University*, Belarus. Global analysis of the total fluorescence decay was performed through linking the fluorescence lifetime constants for multiple data sets, global analysis of the anisotropy decay

through linking the rotational correlation time constants. In addition, data were analyzed through the maximum entropy method (software package from Maximum Entropy Solutions Ltd. Ely, United Kingdom), in terms of distributions of decay times, for which no a priori knowledge of the system is required (Brochon, 1994). The average fluorescence lifetime  $\langle\tau\rangle$  was calculated from the lifetime spectrum  $\alpha(\tau)$  according to

$$\langle\tau\rangle = \frac{\sum_{i=1}^N \alpha_i \tau_i}{\sum_{i=1}^N \alpha_i} \quad (7.1)$$

where  $N$  is the number of  $\tau_i$  values of the  $\alpha(\tau)$  spectrum. The relative fluorescence quantum yield  $Q$  was determined from the TCSPC data as:

$$\frac{Q_{FAD}}{Q_{FMN}} = \frac{\langle\tau\rangle_{FAD}}{\langle\tau\rangle_{FMN}} \quad (7.2)$$

A detailed description of the principles of the analysis of the polarized fluorescence data can be found elsewhere.<sup>111,112,114</sup>

## 7.2.2 Simulations

### Force Field

The MD simulations were performed with the GROMACS 2.0 software.<sup>35</sup> The parameter set for FAD was constructed from parameters taken from the standard building blocks of the GROMOS-96 forcefield for FMN and ATP.<sup>63</sup> Bond lengths, angle and proper and improper dihedral definitions and parameters from the FMN and ATP building blocks were used. For the adenosine moiety and the ribityl chain (from C<sub>1</sub> to C<sub>5</sub>) partial charges were taken from the GROMOS-96 force field. Hydrogen atoms on the adenine and isoalloxazine rings were defined explicitly.<sup>115</sup> Improper dihedrals were added to the bonds on N<sub>5</sub> and N<sub>10</sub> of flavin ring B (see Fig. 7.1), in accordance with the planar structure observed in crystallographic data and molecular geometry optimizations of flavin in the oxidized state.<sup>116,117</sup> Partial charges on the flavin ring in the ground state and the first excited triplet state were based on ab initio molecular orbital calculations in vacuo on isoalloxazine,<sup>118</sup> and semi-empirical MINDO-3 calculations on lumiflavin (7,8,10-trimethylisoalloxazine)<sup>119,120</sup> (see Table 7.1). The charge distribution in the first excited singlet state is similar to that of the first excited triplet state.<sup>119</sup> Partial charges for the pyrophosphate moiety were taken from ab initio molecular orbital calculations on Pi4 (Pi4).<sup>121</sup> A full description of the forcefield parameters, insofar not identical to the GROMOS-96 standard building blocks for FMN and ATP, is given in Table 7.1. A complete description of all force field parameters used is available in appendix B.

## Initialization and Equilibration

Starting structures for the unstacked conformation were taken from high-resolution crystal structures of different flavoproteins that contain non-covalently bound FAD. In *E. coli* glutathione reductase (GR)<sup>122</sup> and *E. coli* thioredoxin reductase (TrxR),<sup>123</sup> the flavin cofactor is bound in an almost completely extended conformation. However, in ferredoxin NADPH-oxidoreductase (FNR) from the *cyanobacterium Anabaena*, the flavin cofactor is bound in a distinct bent conformation.<sup>124</sup> The coordinates of the FAD in the above-mentioned crystal structures of GR, TrxR and FNR were taken to yield three different extended starting structures (FAD-ext1, FAD-ext2, FAD-ext3 respectively). For the stacked conformation of FAD, starting structures were generated from unrestrained MD simulations in vacuo of the extended conformation at 300 K using a relative dielectric constant of 10 to mimic the charge-shielding effect of solvent. These in vacuo simulations yielded stacked conformations of the FAD molecule within tens of picoseconds. Two stacked conformations were selected (FAD-sta1, FAD-sta2), in which the adenine moiety stacks on opposite sites of the flavin ring. Care was taken to select frames in which the conformation of the stacked FAD was relatively far away from its energy minimum. All five starting structures were solvated with the extended simple point charge water model (SPC/E)<sup>125</sup>, which gives good correspondence with experimental data for the dynamic behavior of water. A box size of 4.5 nm was chosen to ensure that the minimum distance between the molecule and its pe-

Isoalloxazine rings			Pyrophosphate	
Atom	Charge		Atom	Charge
	$S_0$	$S_1$		
N <sub>1</sub>	-0.43	-0.31	C <sub>5*</sub>	0.13
C <sub>2</sub>	0.51	0.48	O <sub>5*</sub>	-0.52
O <sub>2</sub>	-0.45	-0.37	P	1.85
N <sub>3</sub>	-0.35	-0.35	OP <sub>3</sub>	-0.92
H <sub>3</sub>	0.35	0.35	OP <sub>2</sub>	-1.00
C <sub>4</sub>	0.43	0.40	OP <sub>1</sub>	-1.00
O <sub>4</sub>	-0.43	-0.40	A <sub>P</sub>	1.85
N <sub>5</sub>	-0.19	-0.24	A <sub>OP</sub> <sub>1</sub>	-1.00
C <sub>6</sub>	-0.12	-0.14	A <sub>OP</sub> <sub>2</sub>	-1.00
H <sub>6</sub>	0.19	0.18	A <sub>O</sub> <sub>5*</sub>	-0.52
C <sub>7</sub>	0	0	A <sub>C</sub> <sub>5*</sub>	0.13
C <sub>7M</sub>	0	0		
C <sub>8</sub>	0	-0.02	<b>Adenine rings</b>	
C <sub>8M</sub>	0	0	Atom	Charge
C <sub>9</sub>	-0.18	-0.17	A <sub>C</sub> <sub>2</sub>	0.22
H <sub>9</sub>	0.19	0.17	A <sub>H</sub> C <sub>2</sub>	0.14
N <sub>10</sub>	-0.05	-0.07	A <sub>C</sub> <sub>8</sub>	0.22
C <sub>4a</sub>	0.06	0.09	A <sub>H</sub> C <sub>8</sub>	0.14
C <sub>5a</sub>	0.06	0.09		
C <sub>9a</sub>	0.14	0.09		
C <sub>10a</sub>	0.27	0.22		

Table 7.1: Partial charges on the isoalloxazine ring of FAD in the ground state ( $S_0$ ) and first excited singlet state ( $S_1$ ) and at other positions in FAD, insofar not identical to the GRO-MOS96 standard building blocks for FMN and ATP. A full description of the force field parameters is listed in appendix B.

periodic images was larger than the cut-off used for the Lennard-Jones interactions (1.0 nm). Depending on the starting structure, the number of water molecules varied between 2407 and 2415. In the next step, energy minimization was carried out using a steepest descent algorithm. Two  $\text{Na}^+$  ions were added to the system to compensate for the two negative charges on the phosphate groups, by replacing the two water molecules with the lowest electrostatic potential. The solvent was pre-equilibrated during 20 – 30 ps to yield a pressure of 1 bar at 300 K, using weak coupling to both an external pressure and temperature bath<sup>24</sup> with harmonic constraints on the atomic coordinates of the FAD molecule. A cut-off for Lennard-Jones interactions and short-range electrostatic interactions of 1.0 nm was applied. During neighbor-list updates, every 20 fs long-range electrostatics were calculated with a cut-off of 1.6 nm. After equilibration, the length of the box edges was in between 4.16 and 4.18 nm.

### Production Simulations

MD simulations were carried out starting from each of the 5 different starting structures – three in an extended conformation, and two in a stacked conformation – with ground state charges at 300 K. Lennard-Jones and short-range electrostatic interactions were calculated using a cut-off of 1.0 nm. Long range electrostatic interactions were calculated

Label	State	Starting structure	Length
FAD-ext1S0	$S_0$	1gre (glutathione reductase) <sup>a</sup>	0.0–2.5 ns
FAD-ext1S1	$S_1$	FAD-ext1S0 at 200 ps	0.2–3.0 ns
FAD-ext2S0	$S_0$	1tde (thioredoxin reductase) <sup>a</sup>	0.0–3.7 ns
FAD-ext2S1	$S_1$	FAD-ext2S0 at 200 ps	0.2–3.7 ns
FAD-ext3S0	$S_0$	1que (ferredoxin NADP <sup>+</sup> reductase) <sup>a</sup>	0.0–3.7 ns
FAD-ext3S1	$S_1$	FAD-ext3S0 at 200 ps	0.2–4.2 ns
FAD-ext4S0	$S_0$	FAD-ext3S0 at 200 ps	0.2–8.2 ns
FAD-ext4S1	$S_1$	FAD-ext3S0 at 200 ps	0.2–4.2 ns
FAD-ext5S1	$S_1$	FAD-ext3S0 at 200 ps	0.2–4.2 ns
FAD-sta1S0	$S_0$	In vacuo stacked FAD-1 <sup>b</sup>	0.0–2.5 ns
FAD-sta1S1	$S_1$	FAD-sta1S0 at 200 ps	0.2–2.5 ns
FAD-sta2S0	$S_0$	In vacuo stacked FAD-2 <sup>b</sup>	0.0–2.2 ns
FAD-sta2S1	$S_1$	FAD-sta2S0 at 200 ps	0.2–2.2 ns
FAD-sta3S1	$S_1$	FAD-ext1S0 at 500 ps <sup>c</sup>	0.5–2.5 ns
FAD-sta4S0	$S_0$	FAD-ext1S0 at 200 ps	0.2–8.2 ns
FAD-sta4S1	$S_1$	FAD-ext1S0 at 200 ps	0.2–8.2 ns

<sup>a</sup> Atomic coordinates for FAD in the open conformation were extracted from the indicated PDB<sup>55</sup> entries of high resolution crystal structures of three different flavoproteins. <sup>b</sup> Corresponds to stacked conformations that were formed during in vacuo simulations. <sup>c</sup> Corresponds to a stacked conformation that formed spontaneously in water.

Table 7.2: Overview of the molecular dynamics simulations of FAD in water. From left to right are shown: the label; the electronic state (ground state  $S_0$  or first excited singlet state  $S_1$ ); starting structure; and simulation length.

during the neighbor-list update, every 20 fs using the Particle-Particle Particle-Mesh approach (PPPM).<sup>126,127</sup> A time-step of 2 fs was used. After 200 ps the simulations were forked. In one fork, the charges on the isoalloxazine ring were reassigned to correspond to the first excited singlet state (Table 7.1) in order to mimic the electronic effects of light absorption. In the other, the ground state charges were maintained. A summary of the simulations performed is presented in Table 7.2. An eleventh run (FAD-sta3) was generated using the conformation of FAD from the ground-state MD run of starting structure FAD-ext1S0 that had spontaneously stacked in water. In this simulation, the process of light absorption was simulated using the time frame at 500 ps, where the molecule was in an equilibrated stacked state. In order to estimate the relative free energy difference between the open and the closed conformation in the ground and excited states, trajectories of several nanoseconds in which the molecule remained in the open or closed conformation were required. In order to obtain more sampling in the open conformation, three additional runs (FAD-ext4S0, FAD-ext4S1, FAD-ext5S1) were performed starting from FAD-ext3S0 (200 ps frame), using different starting velocities. Two additional simulations in the stacked conformation (FAD-sta4S0 and FAD-sta4S1) starting from trajectory FAD-sta3S1 (500 ps frame) were performed. Simulations were run for lengths from 2 to 8 ns, with a total simulation time of  $\sim 65$  ns. An overview of the characteristic parameters of the runs is presented in Table 7.2.

## Analysis

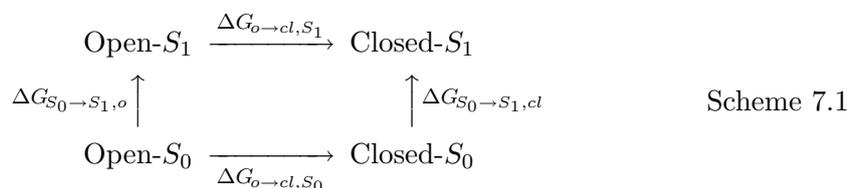
Stacked conformations were defined as having a maximum distance between the center of mass of the adenine rings to the plane of the isoalloxazine rings (see Fig. 7.1) of 0.6 nm.

For calculating the planarity of the flavin ring, the angle between the benzene-like and pyrimidine-like rings of the isoalloxazine ring (rings A and C respectively in Fig. 7.1) was taken from the  $C_8-C_{5a}$  and  $C_{10a}-N_3$  vectors. The angle between the isoalloxazine ring and adenine ring was taken from the angle between the planes defined by  $N_5$ ,  $C_{9a}$  and  $C_{10a}$  from the flavin and  $^A N_1$ ,  $^A C_2$  and  $^A C_8$  from the adenine respectively. The rotational correlation time of the isoalloxazine ring was calculated using the vector  $C_8-N_3$  which is very close to the direction of the emission dipole moment.<sup>128</sup>

Hydrogen bonds were determined using a simple distance and angle cut-off criterion (distance donor-acceptor = 0.35 nm and angle hydrogen-donor-acceptor =  $60^\circ$ ).<sup>39</sup> To refine the information on hydrogen-bonding, the trajectories were analyzed for hydrogen bonds between the different regions of the FAD (flavin ring, ribityl chain, phosphodiester bond, ribofuranosyl and adenine part) separately. Cluster analysis of the ribityl-pyrophosphate-ribofuranosyl chain was performed on the atomic positional root-mean-square differences (RMSD) of different backbone conformations from the simulations. A RMSD cut-off for the RMSD of 0.1 nm was used for determining neighbouring conformations. Cluster analysis was performed according to the following procedure: For each conformation, the number of neighbors is determined. The conformation with the largest number of neighbors is the central structure of the first cluster, and all its neighbors are members of that cluster. These conformations are removed from the pool of all conformations and the procedure is repeated to generate subsequent clusters until the pool is empty. A complete

description of this clustering algorithm has been given by Daura et al.<sup>82</sup>

The relative free energy difference between the open and closed conformation for the ground state and excited state ( $\Delta\Delta G_{o/cl}^{S_0/S_1}$ ) was calculated using a free energy perturbation approach<sup>129</sup> in conjunction with the following thermodynamic cycle:



Since the open/closed equilibrium (corresponding to  $\Delta G_{o \rightarrow cl, S_0}$  and  $\Delta G_{o \rightarrow cl, S_1}$ ) is computationally inaccessible, the free energy differences are calculated by perturbing the electronic state in the open and closed ensemble respectively ( $\Delta G_{S_0 \rightarrow S_1, o}$  and  $\Delta G_{S_0 \rightarrow S_1, cl}$ ) in a single step and using the following relation:

$$\Delta\Delta G_{o/cl}^{S_0/S_1} = \Delta G_{o \rightarrow cl, S_0} - \Delta G_{o \rightarrow cl, S_1} = \Delta G_{S_0 \rightarrow S_1, o} - \Delta G_{S_0 \rightarrow S_1, cl} \quad (7.3)$$

where  $\Delta G_{S_0 \rightarrow S_1, o}$  and  $\Delta G_{S_0 \rightarrow S_1, cl}$  are calculated from:

$$\Delta G_{S_0 \rightarrow S_1, cl} = G_{S_1, cl} - G_{S_0, cl} = -RT \ln \left\langle e^{-\frac{E_{S_1} - E_{S_0}}{RT}} \right\rangle_{S_0, cl} \quad (7.4)$$

where  $\langle \dots \rangle_{S_0, cl}$  denotes the average over the ensemble of configurations in the closed and the ground state ( $S_0$ ), and  $E_{S_0}$  and  $E_{S_1}$  stand for the energy of the conformations of the ensemble in the ground and excited state respectively. Four reference states were considered. The open and closed states, in ground and excited states respectively, were perturbed to the opposite electronic state in a manner analogous to that described above. From this, using eqn. 7.4 four free energy differences are calculated (*vis.*  $\Delta G_{S_0 \rightarrow S_1, cl}$ ,  $\Delta G_{S_0 \rightarrow S_1, o}$ ,  $\Delta G_{S_1 \rightarrow S_0, cl}$  and  $\Delta G_{S_1 \rightarrow S_0, o}$ ).  $\Delta G_{S_0 \rightarrow S_1, cl}$  and  $\Delta G_{S_1 \rightarrow S_0, cl}$  should converge to the same value, and likewise  $\Delta G_{S_0 \rightarrow S_1, o}$  and  $\Delta G_{S_1 \rightarrow S_0, o}$  should. From these converged values, the  $\Delta\Delta G_{o/cl}^{S_0/S_1}$  can be calculated and the measure of convergence is a measure for the accuracy.

## 7.3 Results

### 7.3.1 Fluorescence Dynamics

Polarized sub-nanosecond resolved TCSPC data of FAD reveal excited state dynamics over a wider time-range than earlier studies performed at a lower time-resolution<sup>108</sup> While analysis of the essentially non-quenched fluorescence decay of FMN confirms a predominant lifetime component of 4.8 ns, analysis of the extraordinarily heterogeneous fluorescence decay of FAD in terms of discrete exponentials yields a lifetime pattern with components covering the dynamic range from picoseconds up to several nanoseconds (Table 7.3). Ultrafast fluorescence quenching is reflected in the steep leading edge of the experimental

Sample	Temp (K)	$\phi$ (ps)	$\phi$ range (ps)	Fluorescence components		
				$\tau_i$ (ns)	range (ns)	$\alpha_i$
FAD	277	0.35	(0.32-0.37)	0.008	(0.004–0.008)	0.78± 4%
				0.075	(0.072–0.081)	0.11± 11%
				0.90	(0.83–1.1)	0.02± 15%
				3.6	(3.3–3.8)	0.09± 39%
FAD	293	0.20	(0.18-0.23)	0.007	(0.002–0.009)	0.83± 3%
				0.10	(0.03–)	0.01± 15%
				0.54	(0.37–)	0.02± 3%
				2.7	(2.6–3.1)	0.14± 7%
FAD	300	0.16	(0.14-0.19)	0.008	(0.003–0.012)	0.72± 12%
				0.086	(0.019–)	0.04± 63%
				0.55	(0.25–)	0.03± 17%
				2.4	(1.9–2.6)	0.20± 21%
FAD	313	0.11	(0.10-0.12)	0.010	(0.007–0.011)	0.60± 10%
				0.072	(0.052–0.079)	0.12± 32%
				1.2	(1.1–1.4)	0.13± 36%
				2.2	(1.9–2.8)	0.12± 32%
				5.1	(4.6–)	0.02± 65%
FAD in 80 % glycerol	293	15.1	(14.3-15.8)	0.012	(0.006–)	0.12±140%
				0.13	(0.090–)	0.12± 16%
				0.81	(0.69–1.0)	0.22± 44%
				4.8	(2.9–5.1)	0.48± 17%
				7.1	(5.3–)	0.06± 14%
FAD in D <sub>2</sub> O	293	0.24	(0.20-0.27)	0.005	(0.002–)	0.81± 4%
				0.049	(0.013–)	0.05± 6%
				0.44	(0.16–)	0.02± 7%
				3.1	(3.0–3.8)	0.11± 11%
				6.4	(3.2–)	0.01± 9%
FMN in D <sub>2</sub> O	293	0.21	(0.18-0.24)	$\leq 0.004$ <sup>a</sup>		0.00± 10%
				0.38	(0.077–)	0.04± 17%
				5.3	(5.3–5.4)	0.96± 7%
FMN	293	0.18	(0.16 -0.20)	1.5	(1.3–1.7)	0.12± 2%
				4.7	(4.7–4.8)	0.88± 1%

<sup>a</sup> Data suggested the possibility of a third fluorescence lifetime component of a few ps at the border of the detection limit that could not clearly be resolved.

*Table 7.3: Fluorescence lifetimes ( $\tau_i$ ) and rotational correlation times ( $\phi$ ) of FAD and FMN as a function of temperature and solvent composition. Standard experiments were performed in 50 mM potassium phosphate buffer at pH 7.5. Presented are time constants ( $\tau_i$ ) and fractional contributions ( $\alpha_i$ ) of the fluorescence lifetime components as obtained from global analysis of multiple experiments. The values in the parentheses are obtained from a rigorous error analysis at the 67 % confidence level as described by Beechem.<sup>130</sup> For components with low amplitudes upper confidence limits were not always found. For the fractional contributions  $\alpha_i$ , the standard deviation is given.*

fluorescence decay of the FAD samples (Fig. 7.2), yielding a predominant lifetime constant in the order of 5 – 10 ps, which is close to the detection limit of the set-up used. At 293 K the ultra-short component ( $\tau = 7$  ps) was found to have a contribution of 83 %, while a component with a time constant of 2.7 ns had an amplitude of about 14 %. In addition, a minor percentage of intermediate components (at 293 K: 0.5 ns for 2 % and 0.1 ns for 1 %) was needed for an optimal description of the excited state dynamics. Rigorous error analysis at the 67 % confidence interval of the global analysis results showed that the time constants of the minor intermediate components are less well defined than those of the major components (Table 7.3). Analysis in terms of fluorescence lifetime distributions with the MEM method confirmed the general requirement of four lifetime constants and yielded identical results. Calculation of the relative fluorescence quantum yield from the amplitudes and time constants of the TCSPC data of FAD and FMN using eqn. 7.2 yields  $Q_{FAD} = 10\%$  of  $Q_{FMN}$  showing rather good agreement with the value of 12 % obtained from steady state fluorescence experiments.<sup>99</sup>

Fluorescence anisotropy decay analysis revealed the overall tumbling motion of the FAD and FMN molecules free in solution (Table 7.3). Increasing temperature in the range between 277 K and 313 K results in a shortening of the rotational correlation time of FAD corresponding with the concomitant change in solvent viscosity and kinetic energy (0.20 ns at  $\sim 293$  K, 0.16 at 300 K).

In the fluorescence lifetime data, increasing temperature between 277 K and 313 K results in a shortening of the nanosecond decay time. From 293 K to 313 K a clear decrease in the amplitude of the ultra-short lifetime is found in favor of the longer lifetime components. This can be explained by a shift in equilibrium towards the open conformation.

A more complete shift towards an open, non-quenched conformation is obtained in

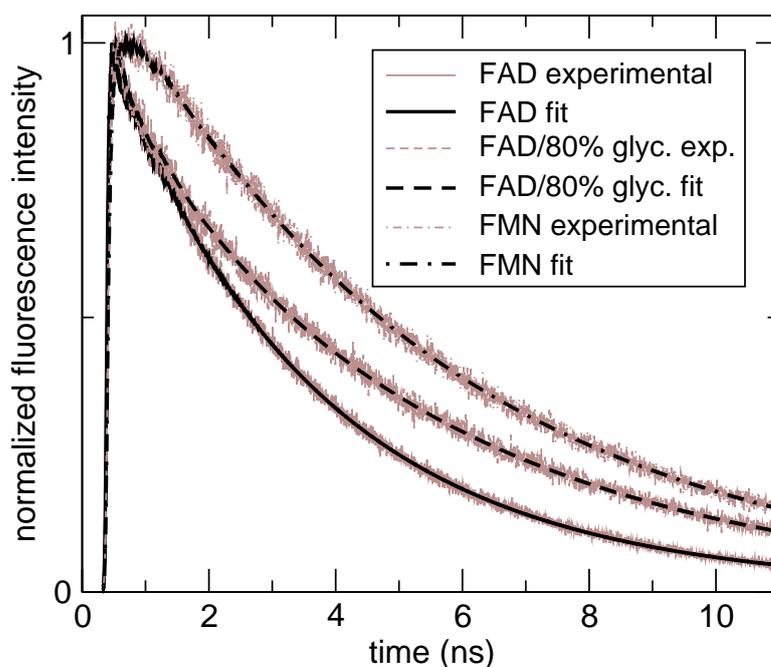


Figure 7.2: Partial experimental total fluorescence decays and corresponding theoretical data retrieved from the multi exponential fits of FAD (solid) and FMN (dashed) in  $D_2O$ , and FAD in 50 mM potassium phosphate buffer with 80 % glycerol, pH 7.5 (dot-dashed) at 293 K. Experimental data (gray) and theoretical data (black) overlap very well. In order to clearly reveal the differences between the data, only part of the normalized fluorescence intensity decays (logarithmic scale) and only the first part of the twelve ns time-window are shown.

80% glycerol (Fig. 7.2 and Table 7.3), which is consistent with earlier steady-state spectroscopic studies showing that non-polar solvents prevent stacking interactions (see reference<sup>101</sup> and references therein). Measurements of FAD in deuterated water revealed only a slight shift of the longest fluorescence lifetime component towards longer time values, leaving the amplitudes virtually unaffected (see Table 7.3). For FMN, a similar effect on the predominant fluorescence lifetime was obtained. These results show that deuterium slightly influences the intrinsic fluorescence lifetime of the isoalloxazine ring, and that significant involvement of hydrogen bonding interactions in the fluorescence quenching mechanism should not be expected. Fluorescence depolarization analysis of FMN and FAD in deuterium revealed slightly reduced rotational dynamics with respect to water (Table 7.3).

### 7.3.2 Simulations

#### Stacking

In order to obtain more insight into the relation between conformational dynamics of the FAD molecule in solution and its fluorescence characteristics, molecular dynamics simulations were performed. For this study, five different starting structures were used, three of which represent the FAD in an 'open' conformation and two which represent the flavin and adenine rings in a stacked (or 'closed') conformation. The electronic effect induced by light absorption was simulated by instantaneously adjusting the partial charges on the atoms of the flavin ring. An overview of the simulations performed is given in Table 7.2.

All MD trajectories, both in the ground state and the excited state, were analyzed for the occurrence of flavin-adenine stacking, which is described in detail in "Methods" sec. 7.2.2. The distance between the centers of mass of the flavin and adenine ring systems was used to monitor the process of stacking and unstacking, and to characterize the FAD conformations (see Fig. 7.3). In one run, the FAD molecule remained in the open conformation throughout the simulation of 8 ns. In seven other runs started from 'open' conformations, the isoalloxazine and adenine rings became stacked in the course of the simulation. While in some cases a nearly instantaneous collapse from a completely extended

*Figure 7.3: (next page) Characteristics of the eight molecular dynamics simulations of FAD starting from the open, extended conformation. In blue the distances between the centers of mass of the flavin and adenine rings are depicted. Vertical dotted blue lines correspond to the snapshots shown in Fig. 7.4. The distance plots provide a good qualitative overview of the changes in molecular conformation during the runs. Conformations with a distance of  $\leq 0.6$  nm were regarded as stacked. Run *ext4S0* extends up to 8 ns and shows no stacking. In red (from the upper sub-panel downward) the number of hydrogen bonds between resp. the flavin ring and ribityl chain, the ribityl chain with itself, the ribityl chain and the pyrophosphate moiety, and the ribose-moiety and the pyrophosphate moiety are given for each of the trajectories. Note the presence of hydrogen-bonding networks between the ribose and pyrophosphate moieties, and between the ribityl chain and the pyrophosphate moiety in the extended conformation in most of the trajectories. In the stacked conformations, hydrogen bonds between the flavin and ribityl chain can occur (*ext1S0*, *ext1S1*, *ext4S1*).*

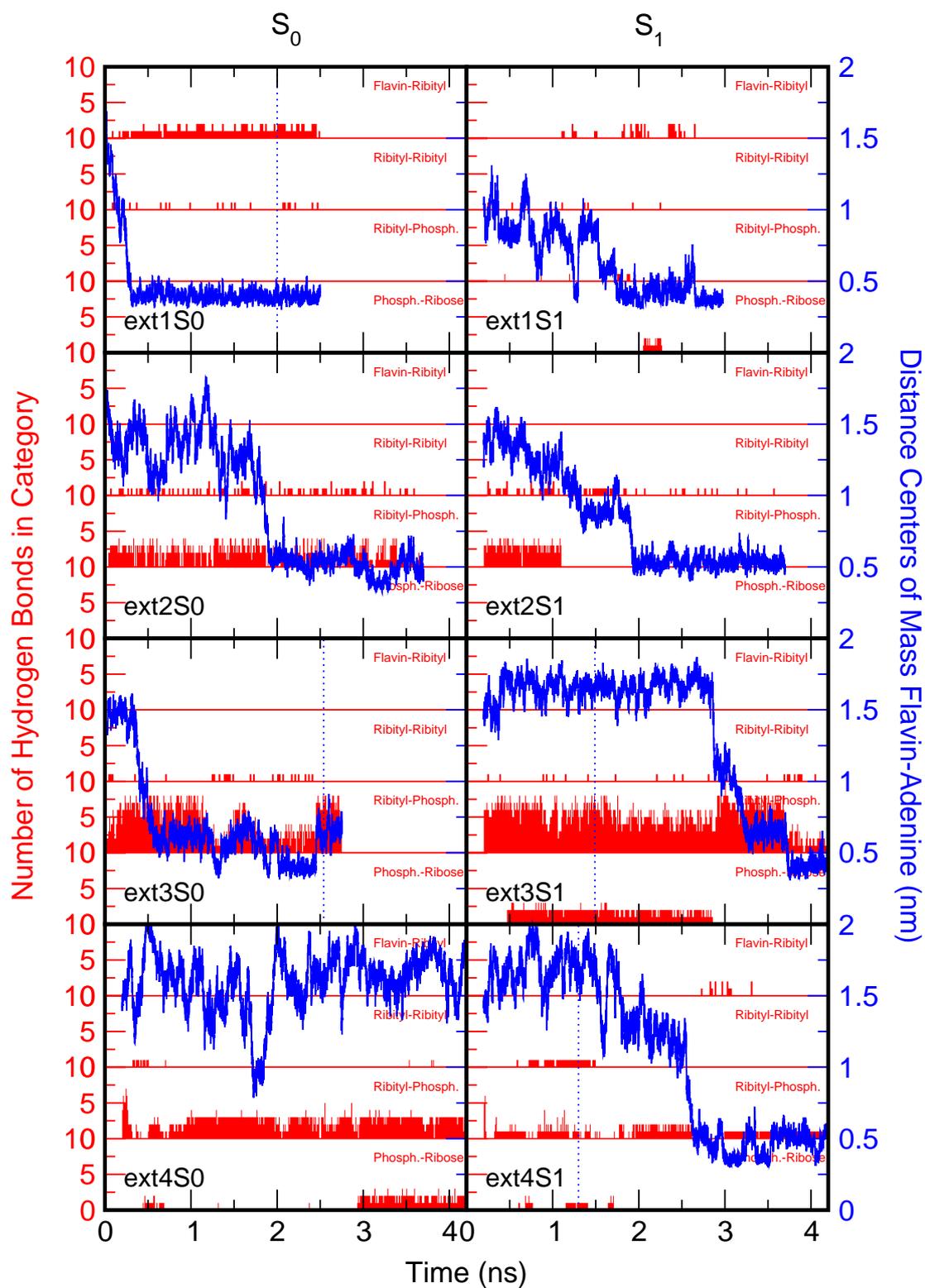


Figure 7.3 (caption on previous page)

conformation to a highly stable stacked conformation occurred (e.g. complete stacking within a few hundred ps in FAD-ext1S0), other simulations were characterized by the presence of many different intermediates some of which were relatively long-lived (up to about 1 ns, e.g. FAD-ext2S1, FAD-ext3S1). The simulations revealed that the process of stacking proceeds through the adoption of a more compact but still open conformation in which the ribityl-pyrophosphate-ribofuranosyl chain is bent. The time necessary to reach this conformation appears to be dependent on the presence of particular intramolecular hydrogen bond networks. Subsequently, the isoalloxazine and adenine rings come into contact.

The first intramolecular interactions observed between 'non-neighbouring' parts of the molecule involved in most cases the hydrophilic ring of the flavin (particularly O<sub>2</sub> and N<sub>3</sub>, see Fig. 7.1) and the six-membered ring of the adenine. This interaction is sometimes preceded by a contact between the flavin and the (hydroxyl groups of the) sugar. An alternative pathway for stacking involves the disruption of a hydrogen bonding network between the AMP-phosphate and the sugar. The adenine subsequently moves close to the flavin via the ribityl chain (FAD-ext3S1). After disruption of a second set of hydrogen bonds (between the ribityl chain and the FMN-phosphate), the six-membered ring of the adenine and the hydrophobic ring of the flavin (ring A in Fig. 7.1) stack. It should be mentioned that different routes of stacking result in different closed conformations (see the next section).

Once in a stacked conformation with considerable overlap of the ring systems, unstacking of the flavin and adenine rings occurred only occasionally and was brief (generally a few ps to 50 ps at most). Hardly any complete unstacking of the flavin and adenine rings was observed in molecules with full stacking between the rings (FAD-ext1, FAD-sta1, FAD-sta2, FAD-sta3; for the runs starting from stacked conformations, data are not shown), irrespective of the electronic state of the flavin. The less overlap between the ring systems in the stacked state, the easier was unstacking. In the run FAD-ext3S0 a closed conformation with proper stacking was hampered (see Fig. 7.4). Here, unstacking of the rings lasting up to several hundreds of picoseconds was more common. However, the FAD remained in a compact state with the adenosine and ribityl moieties in contact. In none of the simulations, 'unfolding' of the molecule to a fully extended conformation was observed.

In order to determine whether the process of light absorption influences the equilibrium between the open and closed conformations of the FAD, free energy calculations were performed to determine the  $\Delta\Delta G_{ocl}^{S_0/S_1}$  of the stacking transition in the ground state and in the excited state using eqn. 7.4 and the thermodynamic cycle as depicted in Scheme 7.1. The calculations yielded similar free energy differences between the stacked and unstacked conformation for the ground state and for the excited state, in principle suggesting that in fluorescence the process of light absorption itself indeed does not influence the conformational equilibrium (Fig. 7.5). The statistics of the calculations are limited by the relatively short lifetime of the open conformation, however, so that differences cannot be clearly distinguished. From Fig. 7.5 it can easily be estimated that the convergence between forward and backward perturbations is on the same order of magnitude of the resulting  $\Delta\Delta G_{ocl}^{S_0/S_1}$ , indicating that statistics are the limiting factor in obtaining an accurate estimate for  $\Delta\Delta G_{ocl}^{S_0/S_1}$ .

### Conformational Analysis

The range of conformations observed for the FAD molecule can be best described by the relative positions and orientations of the flavin and adenine rings, and the conformation of the connecting ribityl-pyrophosphate-ribofuranosyl backbone. This is summarized in Table 7.4. Note, the stacked starting structures FAD-sta1 and FAD-sta2 were created in vacuo while the other stacked conformations were produced in solution. It appears that in solution, the adenine ring can stack on either side of the flavin and there is no preferred orientation of the ribityl-pyrophosphate-ribofuranosyl backbone. In general, the isoalloxazine and adenine rings stack coplanar. In water, the FAD tends to adopt a fully parallel stacked conformation similar to those created in vacuo (average angle between the flavin and adenine planes of  $12.5^\circ$ ). In several simulations, (intermediate) conformations were found in which the flavin and adenine rings were slightly tilted (about  $30^\circ$ , see Table 7.4), but still had substantial contact. Occasionally, water molecules intercalated on one of the edges between the rings, and a hydrogen-bonded water bridge was formed between

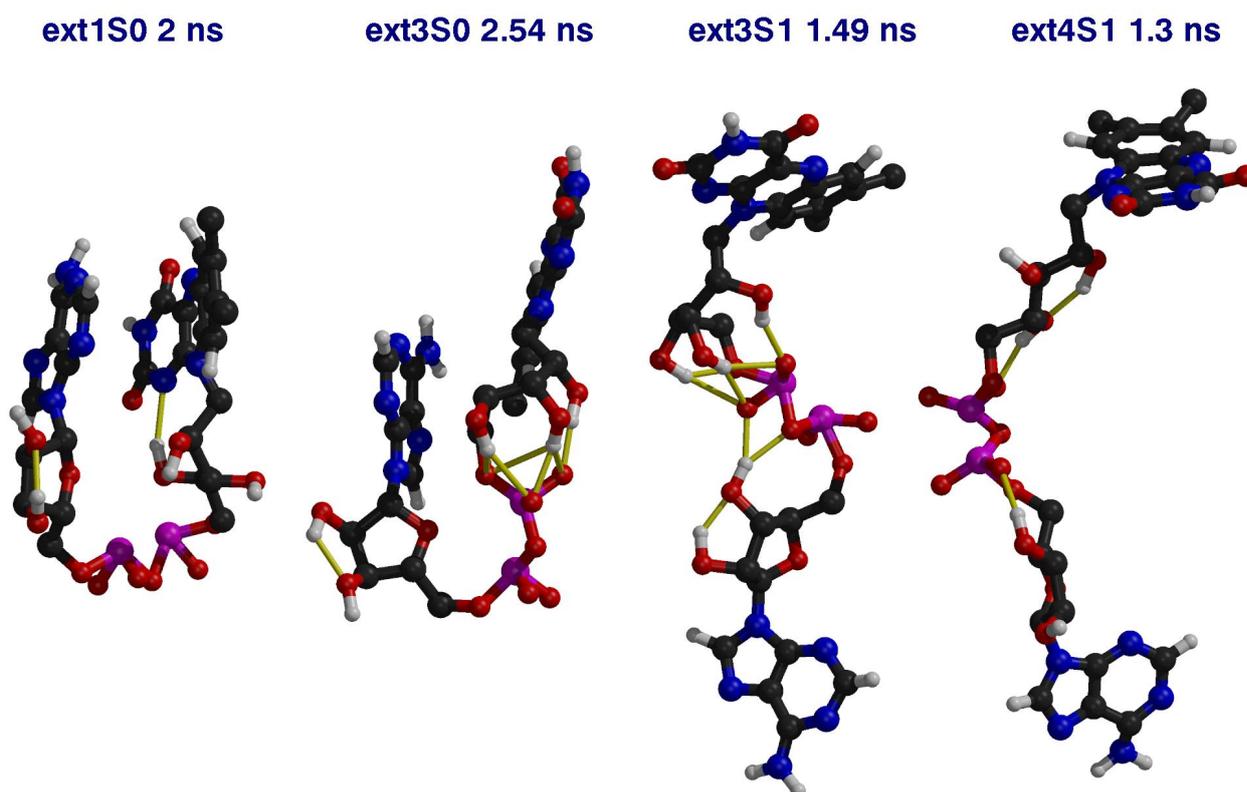


Figure 7.4: Snapshot pictures of stacked and open conformations observed in the molecular dynamics simulations. Shown are from left to right: FAD-ext1S0 at 2 ns, FAD-ext3S0 at 2.54 ns, FAD-ext3S1 at 1.49 ns, and FAD-ext4S1 at 1.3 ns. Hydrogen bonds are indicated with yellow lines. Of special interest are the extensive hydrogen bonding networks involving the phosphate and sugar moieties. Note the strain in the configuration of the phosphate groups. Plots were generated using MolScript<sup>3</sup> and Raster3D.<sup>5,6</sup>

the rings, resulting in decreased coplanarity. Stacking also influenced the planarity of the flavin ring system: While in the open conformations an average angle of  $15.1^\circ$  was found between ring A and C of the isoalloxazine (Fig. 7.1), this angle was slightly, but systematically, smaller in the stacked conformations ( $13.1^\circ$ , see Table 7.4). The largest deviations from planarity were found during conformational transitions involving intramolecular interactions with the isoalloxazine ring. After stacking, the flavin and adenine had some lateral freedom, but were restricted by the conformation of the connecting chain. In general, this mobility decreased with increasing overlap of the rings. The average position of the stacked ring systems was not influenced by the changed charge distribution from ground state to excited state. In many of the stacked conformations, a high strain occurred on one or both of the phosphate groups resulting in a significant deviation from the tetrahedral configuration. This may be due to an overestimate of charge interactions in

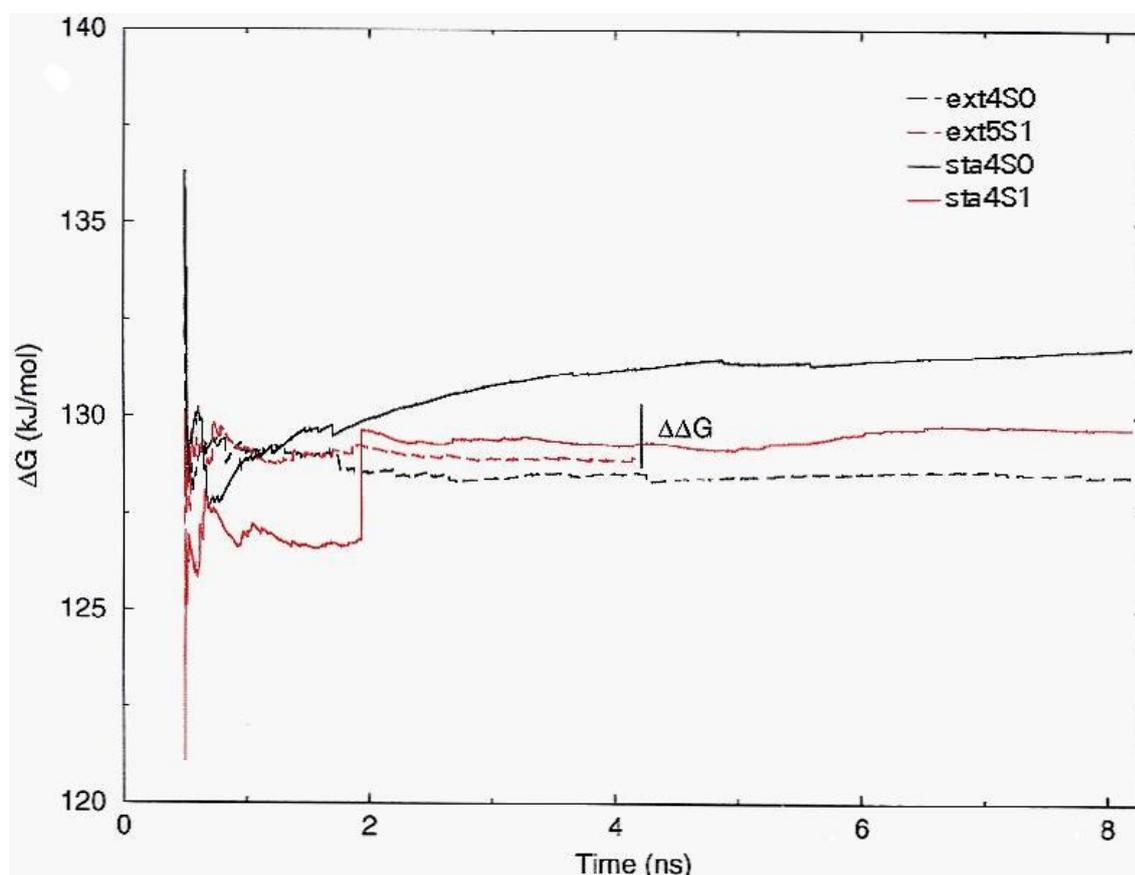


Figure 7.5: Cumulative average of the relative free energy of stacking in the ground state and excited state. For the stacked conformations, two additional runs starting from the conformation stacked in water (starting with the frame at 500 ps from trajectory FAD-ext3S0) were performed to yield 8 ns of data for both the ground state and excited state. For the open conformations, the 8 ns trajectory of FAD-ext4S0 and an additional run in the excited state yielding 4.2 ns of open conformations were used. Relative free energies for the ground state and first excited state were calculated using both charge distributions for the same trajectory. For the stacking transition, no significant free energy difference between the ground state and excited state was observed.

the force field of the phosphodiester bridge, as these charge interactions are abundant in the stacked conformation.

The conformations of the ribityl-pyrophosphate-ribofuranosyl backbone connecting the flavin and adenine moieties were analyzed through the RMSD of the atomic positions, and grouped in clusters of highly similar conformations. From the cluster analysis it can be concluded that in runs from different starting structures, different conformations of the backbone chain were sampled, and that only few transitions from one cluster to another occurred. Specific dihedral configurations of the backbone chain stabilized by hydrogen bond networks appear to have very long lifetimes, thereby restricting the rings to certain conformations. Conversely, in the stacked situation the strong interactions between the rings may restrict rearrangements of the backbone chain. Surprisingly, the

Run	Stacking Time (ns)	Stacked conformations				Average $\theta$ flavin ( $^\circ$ )		$\tau_{rot}$ (ps)	
		Position	Chain	Plane	$\theta$ Fla-Ade ( $^\circ$ )	Open	Stacked	Open	Stacked
FAD-ext1S0	0.26	C	R	Si	11.0 $\pm$ 6.5	16.0 $\pm$ 7.9	11.6 $\pm$ 6.4	–	131.1
FAD-ext1S1	1.24	C/BC	R	Si	13.2 $\pm$ 8.2	18.2 $\pm$ 8.9	14.0 $\pm$ 8.4	116.9	146.6
FAD-ext2S0	1.86	C	R	Si	31.3 $\pm$ 18.3	15.5 $\pm$ 8.0	13.7 $\pm$ 7.5	149.5	194.6
FAD-ext2S1	1.87	C	R	Si	34.6 $\pm$ 9.4	13.8 $\pm$ 7.7	13.9 $\pm$ 7.3	131.1	234.7
FAD-ext3S0	0.62	A	S	Re	26.0 $\pm$ 14.8	13.9 $\pm$ 8.0	13.6 $\pm$ 7.5	115.5	144.1
FAD-ext3S1	3.28	B	S	Re	12.3 $\pm$ 8.5 35.3 $\pm$ 13.7 <sup>a</sup>	14.4 $\pm$ 8.0	13.6 $\pm$ 7.5	118.6	146.2
FAD-ext4S0	> 8.2	–	–	–	–	14.9 $\pm$ 8.3	–	121.9	–
FAD-ext4S1	2.63	BC	S	Si	30.3 $\pm$ 23.1	14.3 $\pm$ 8.2	17.3 $\pm$ 8.4	173.3	204.0
FAD-sta1S0	–	BC	S	Si	13.0 $\pm$ 8.2 7.1 $\pm$ 11.4 <sup>b</sup>	–	12.9 $\pm$ 7.1	–	176.7
FAD-sta1S1	–	BC	S	Si	13.1 $\pm$ 7.5	–	12.3 $\pm$ 6.2	–	132.8
FAD-sta2S0	–	BC	S	Re	12.2 $\pm$ 7.2	–	10.9 $\pm$ 6.0	–	127.4
FAD-sta2S1	–	BC	S	Re	11.8 $\pm$ 7.0	–	11.6 $\pm$ 6.5	–	143.9
FAD-sta3S1	–	C	R	Si	13.0 $\pm$ 9.1	–	12.0 $\pm$ 6.7	–	116.3

<sup>a</sup> transition to a tilted angle around 3.3 – 3.7 ns    <sup>b</sup> transition to a tilted angle around 1.2 – 2.4 ns

Table 7.4: Overview of parameters characterizing the dynamic behavior and the stacked conformations in the molecular dynamics simulations. Parameters that could not be retrieved due to the absence (or limited existence) of a particular conformation or transition are indicated by ‘–’. From left to right: **Run**; **Stacking Time**: time interval to the first transition from an open to a stacked conformation. Unstacking (rapidly) followed by re-stacking was observed regularly. FAD-ext3S0 and FAD-ext1S1 (single event) showed long periods of unstacking (up to 250 ps), for further details see ‘Results’ and Figs. 7.3 and 7.4; **Position**: Average relative positions of the adenine and isoalloxazine rings. The ring(s) that is (are) closest to the center of mass of the adenine is (are) given (A = hydrophobic ring, B = middle ring, C = hydrophilic ring; see Fig. 7.1); **Chain**: Direction in which the ribityl-pyrophosphate ribofuranosyl chain is twisted, going from the flavin to the adenine (R = clockwise, S = anticlockwise); **Plane**: Side of the isoalloxazine plane on which the adenine stacks (Si = ‘on the front’, Re = ‘on the back’ of the isoalloxazine ring as depicted in Fig. 7.1);  **$\theta$  Fla-Ade**: Average angle between the adenine and flavin rings in the stacked conformation.  **$\theta$  flavin**: Average angle between ring A and C of the isoalloxazine ring in the open and stacked conformations.

starting structures stacked in vacuo have backbone conformations differing distinctly from the structures resulting from stacking in water. This is also reflected in the relative position of the stacked rings, as shown in Table 7.4.

### Physical Properties

An important conclusion from the MD simulations is that in solution the backbone of the FAD molecule behaves much more rigid than we had expected. To gain insight into the effect of hydrogen bonds between the different building blocks of the FAD, the trajectories were analyzed in terms of numbers of hydrogen bonds between two separate regions of the molecule (flavin ring, ribityl chain, pyrophosphate, ribofuranosyl and adenine ring). The high density of hydrogen-bonding donors and acceptors in the chain provides an excellent framework for intramolecular hydrogen bonding. Intramolecular hydrogen bonding networks between the sugar and the phosphodiester bond appear to have a prominent effect on the lifetime of the open conformation (see Fig. 7.3). Hydrogen bonding between the ribityl chain and the phosphodiester bond contributes to this as well. For this network, the number of hydrogen bonds may even temporarily increase just before stacking, as upon bending of the chain the distant phosphate comes within hydrogen bonding distance to the ribityl chain as well (e.g. ext3S1). Both the formation of hydrogen bonds and the concomitant breaking of the hydrogen bonds appeared to be highly cooperative. Breaking up a cooperative network involves a free-energy barrier that can 'lock' the molecule into a particular open conformation for long periods of time (up to hundreds of ps). The interplay between the energetically favorable stacking of the ring systems on the one hand, and the cooperative hydrogen bond formation between the ribityl chain and the FMN-phosphate on the other hand, even resulted in frequent stacking, unstacking and stacking again in the same conformation (FAD-ext3S0).

In some stacked conformations, especially in FAD-ext1, a hydrogen bond was formed between the hydroxyl group attached to C<sub>4</sub> or C<sub>3</sub> of the ribityl chain and the N<sub>1</sub> of the flavin. In literature, <sup>131</sup> interaction between a hydroxyl group of the ribityl chain and the N<sub>1</sub> of the flavin was proposed to be involved in the intramolecular photo-reduction of the flavin, resulting in subsequent photo-degradation. As a candidate for this, the hydroxyl group of the C<sub>2</sub> of the ribityl chain was proposed. This hydrogen bond was found in the MD simulations as well, but only in rare occasions.

In all simulations, one Na<sup>+</sup> ion was either initially positioned at 0.2 nm from the ester oxygen of the phosphodiester bond, or migrated to this position during the run (between 0.1 and 2 ns). Sometimes the second Na<sup>+</sup> ion moved towards this position as well. No interactions between Na<sup>+</sup> ions and the flavin ring were found.

The rotational mobility of the FAD molecules was analyzed through a second order Legendre polynomial of the autocorrelation function of the rotational orientation of the flavin ring; for this, the N<sub>3</sub>-C<sub>8</sub> axis was used which has a direction very close to that of the emission dipole moment.<sup>128</sup> The rotational correlation time of the flavin was 147 ± 25 ps. A small difference was found between the rotational correlation time of the flavin in the stacked and the unstacked parts of the simulations. Analysis of the eleven trajectories in which only the stacked conformations were taken into account resulted in a rotational

correlation time of  $154 \pm 33$  ps, while for the unstacked conformations a time-constant of  $129 \pm 13$  ps was found. In every single run in which the molecule was in both an unstacked and a stacked conformation for a considerable period of time, the rotational correlation time for the unstacked situation was significantly smaller than that for the stacked one (Table 7.4). This may be explained by additional mobility of the isoalloxazine ring in the plane perpendicular to the ribityl-pyrophosphate-ribofuranosyl backbone. The values for the rotational correlation time of 162 ps obtained from the simulations (Table 7.4) are in very good agreement with the time constant of 160 ps found in the time-resolved fluorescence anisotropy measurements (Table 7.3).

## 7.4 Discussion

In this study we have demonstrated the combined use of molecular dynamics simulations and (sub)nanosecond time-resolved fluorescence spectroscopy to obtain more detailed insight into the dynamic structure of complex fluorophores such as the flavin adenine dinucleotide cofactor. Polarized fluorescence experiments with high time resolution yield a broad picture of the dynamic landscape for the ensemble of molecules in solution, while MD simulations provide a detailed view on the underlying molecular structures and possible transition pathways. The limitation of the simulations, however, is whether it is possible to simulate long enough to sufficiently sample the conformational space.

The molecular dynamics simulations confirm that the FAD molecule in aqueous solution is predominantly in a compact conformation in which the flavin and adenine moieties are stacked. Although unstacking of the flavin and adenine ring systems occurs, the molecule is mainly in a conformation in which the flavin and adenine rings show (large) overlap, occasionally alternated by conformations in which the ring systems interact via direct or water-mediated hydrogen bonds.

These simulation results rule out several potential mechanisms for ultra-rapid fluorescence quenching in FAD: Interactions between the flavin N<sub>5</sub> and the amino group of the adenine ring – mostly bridged through a water molecule – were not often found in the MD simulations, and a perpendicular orientation of the flavin and adenine in which the lone electron pair of N<sub>5</sub> interacts with the  $\pi$ -system of the adenine was not found at all. Likewise, a quenching mechanism involving hydrogen bonding between the adenine and the 2-keto function of the flavin as proposed by Tsibris *et al.*,<sup>132</sup> can be excluded on the virtual absence of such interactions. Fluorescence experiments in deuterium oxide did not give any evidence for the involvement of hydrogen bonds in fluorescence quenching either. In addition, the absence of substantial hydrogen bonding between the flavin and adenine moieties renders a significant population of molecules with parallel intramolecular hydrogen bonding between the flavin and adenine, as proposed by Raszka & Kaplan<sup>107</sup> for the unstacked state, highly unlikely.

The stacked conformations of the flavin and adenine rings observed in the MD simulations, however, can well explain the ultra-rapid fluorescence quenching in FAD. The MD simulations reveal large overlap of the coplanar flavin and adenine ring systems indicating large  $\pi$ - $\pi$  overlap. The interactions between the adenine and isoalloxazine rings may

yield additional de-excitation routes through internal conversion. Most likely, however, the mechanism of flavin fluorescence quenching involves a photo-induced charge transfer interaction between the adenine (donor) and flavin (acceptor): Photo-induced electron (or hydrogen) transfer from electron-rich donors to the isoalloxazine is a well known property of flavins.<sup>133,134,111,135</sup> In flavoproteins that contain a tyrosine or tryptophan residue adjacent to the flavin, photo-induced electron transfer to the flavin has been shown to result in highly efficient flavin fluorescence quenching with time constants in the picosecond and sub-picosecond range.<sup>111,136,112,137,135</sup> Although classical charge transfer interactions between flavin and purines have been reported not to contribute significantly in the ground state, photo-induced electron transfer should be considered for ultrafast excited state-quenching in the complex. For photo-induced electron transfer, the relation between the free energy of the reaction, the reorganization energy and the distance between donor and acceptor is expressed in the Rehm-Weller equation.<sup>138,139</sup> Based on the redox potentials for the adenine reported by Seidel *et al.*<sup>140</sup> (about 1.5 V at pH 7.5), and FMN ( $-0.24$  V), photo-induced electron transfer is likely to occur at the wavelength of excitation (energy for the  $S_0 \rightarrow S_1$  transition ( $\pi - \pi^*$  character) at 450 nm is 2.76 eV). In general, typical rate constants for electron transfer derived from a variety of biological and (semi)synthetical systems lie in the order of  $1 \text{ ps}^{-1}$  to  $0.1 \text{ ps}^{-1}$  for donor-acceptor distances of 0.5 nm, while distances of 1.0 nm generally result in rates between  $10 \text{ ns}^{-1}$  and  $1 \text{ ns}^{-1}$ .<sup>141</sup> The picosecond fluorescence quenching observed for FAD corresponds to a rate constant of  $\sim 0.15 \text{ ps}^{-1}$  which lies in the expected time range for electron transfer. As the time constant is close to the detection limit of the TCSPC set-up used, the rate constant for quenching might be somewhat underestimated. Recently, femtosecond fluorescence quenching was observed in the flavoenzymes riboflavin binding protein<sup>137,135</sup> and the D-amino acid oxidase benzoate complex,<sup>137</sup> which contain a (coplanarly) stacked complex between the isoalloxazine ring and an aromatic ring (resp. tryptophan and tyrosine, and benzoate). Such ultrafast processes would fall below the detection limit of our experimental set-up. The oxidation potentials of tryptophan and tyrosine, however, are somewhat lower than that of adenine, thus accelerating an electron transfer reaction. Recently reported transient absorption spectra of FAD<sup>142</sup> revealed a time constant for excited state quenching of about 4 ps, which is in close agreement with the results presented here. Although the study on the transient absorption of FMN and FAD in water and formamide definitely confirmed the rapid excited state quenching of flavin-purine complexes in aqueous solution, the molecular mechanism of quenching was not discussed.

Besides the picosecond fluorescence lifetime component discussed above, the fluorescence lifetime patterns of FAD contained a nanosecond component and minor but definite contributions of intermediate time constants (100 ps – 1 ns). In an extreme model considering each component of the lifetime spectrum as a separate conformational state with a conformational lifetime longer than the fluorescence lifetime of that particular state ('conformational substates' model), these could be interpreted as separate conformations of the FAD molecule with a (very) small, but significant population. However, in the simulations we observed the FAD molecule to experience conformational transitions during the lifetime of the excited state, on time-scales ranging from several tens of picoseconds to nanoseconds.

Based on the fluorescence characteristics of FMN together with the MD simulations,

however, we believe that for FAD in the stacked conformation only, the flavin fluorescence will be quenched immediately. When the picosecond fluorescence lifetime constants of FAD represents the 'closed' conformation and the other lifetimes the (various) 'open' conformations of the molecule, one can obtain an estimate of the populations. Under the assumptions that only the observed ultrafast component reflects the stacked conformation, and that the equilibrium constant for the excited state is identical to that for the ground state (which seems reasonable from the calculations of the free energy differences for the stacking equilibrium in the ground state and excited state) the population of FAD molecules in the open conformations, as calculated from the weight of the amplitudes from the TCSPC data is about 17% at 293 K.

The open, unstacked conformations of the FAD are responsible for the significant 2.7 ns fluorescence lifetime component. The MD simulations have shown that the FAD molecule can exist in an extended conformation for considerable periods of time, longer than the intrinsic fluorescence lifetime of the flavin. However, if a considerable number of FAD molecules would remain in an extended open conformation with no intra-molecular interactions between the flavin and adenosine moieties during the lifetime of the excited state, a nanosecond fluorescence lifetime constant identical to that of FMN (4.7 ns) would be expected. An explanation for the 2.7 ns lifetime component of FAD that cannot be fully excluded is the existence of a particular compact conformation with a long conformational lifetime (compared to the lifetime of the excited state) in which the flavin does not stack, but interacts with other parts of the molecule resulting in less efficient flavin fluorescence quenching. The long-lived conformation in which the flavin and adenine moieties are close, but do not stack (Fig. 7.4), resembles the X-ray structure of the FAD cofactor bound to DNA photolyase,<sup>143</sup> the only flavoprotein known thus far in which the cofactor is bound in a non-extended conformation. This particular conformation, however, occurred in one run only and therefore cannot explain the magnitude of the 2.7 ns fluorescence component.

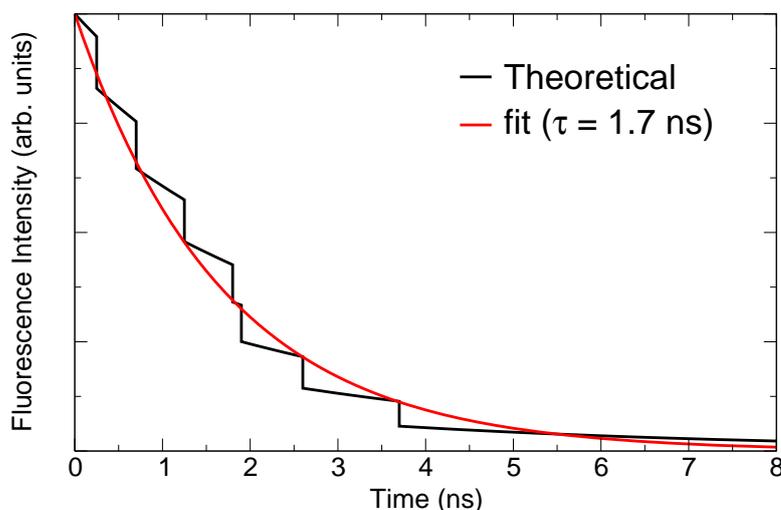


Figure 7.6: Hypothetical fluorescence decay curve calculated from the molecular dynamics simulations under the assumptions that the limited number of FAD starting structures in the extended conformation adequately represents the conformational ensemble of 'open' FAD molecules in solution, that any stacked conformation will instantaneously quench and that the intrinsic fluorescence lifetime of the flavin moiety is identical to that of FMN ( $\tau_{\text{int}} = 4.7$  ns). From a mono-exponential fit, a theoretical fluorescence lifetime of 1.7 ns is obtained.

A more likely hypothesis is therefore that the nanosecond fluorescence lifetime component of FAD corresponds to the conformational lifetime of the open conformation. Support for this is found in the MD simulations, where except for one run, all simulations starting from an extended conformation of the molecule stacked within 5 ns. Under the assumptions that the limited ensemble of FAD starting structures in the extended conformation adequately represents the conformational ensemble of 'open' FAD molecules in solution, and that as soon as a FAD molecule stacks, quenching occurs, a hypothetical fluorescence decay curve can be constructed using the following formula:

$$I_{theo}(t) = N_{open}(t) \cdot e^{t/\tau_{intr}} \quad (7.5)$$

where  $I_{theo}$  is the calculated theoretical intensity at time  $t$ ,  $N_{open}(t)$  is the number of open molecules at time  $t$ , as can be extracted from Fig. 7.3, and  $\tau_{intr}$  is the intrinsic fluorescence lifetime of the flavin moiety (4.7 ns).

The theoretical fluorescence lifetime retrieved from this hypothetical curve is 1.7 ns (Fig. 7.6), which is not far from the nanosecond lifetime observed with time-resolved fluorescence (2.4 ns at 300 K). Together, the time-resolved fluorescence data and molecular dynamics simulations support the picture that the majority of FAD molecules is in a 'closed' conformation in which the adenine and flavin rings are stacked, and that a small percentage of FAD molecules can remain in the open conformation for nanoseconds of time.

## 7.5 Acknowledgments

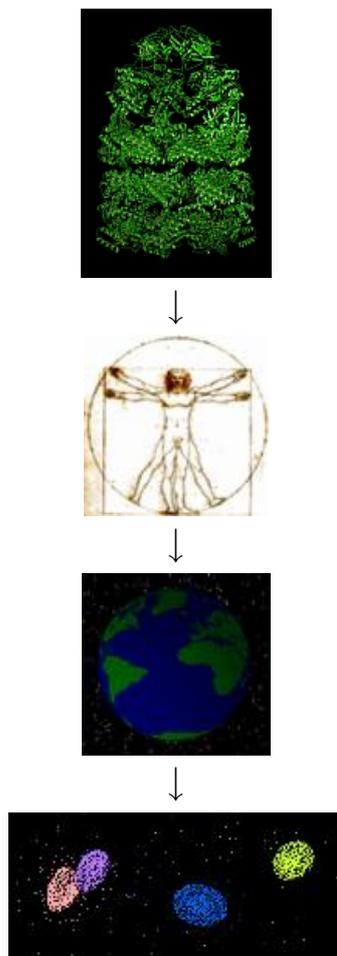
We thank B. Hess for adjustments of the software package, Dr. D. van der Spoel for assistance in early MD pilot-experiments, A. van Hoek for help with the fluorescence experiments, and Dr. W. J. H. van Berkel for valuable discussions and suggestions.





# Chapter 8

## Conclusion



## 8.1 The Future view

The sheer complexity of a protein system might make it virtually intractable for conventional scientific approaches, in the sense that the intricate relationships between all constituent parts of the protein are in themselves so complex and convoluted that any simplification or division into separate entities, as is generally done in science, must be artificial and arbitrary and worse, only valid for the level at which this simplification or subdivision was made. This means that no general rules derived from any simplified model of a protein will hold for the real protein system. This statement is bold, since it implies that only a model that describes the whole protein (and possibly a substantial part of its surroundings) in virtually exact detail, will ever be able to describe the protein behavior in such a way that any conclusions can be derived from it. It also implies that, if we happen to stumble on only part of the “correct” description, it might be difficult, if not impossible, to tell that this description was in fact correct. In other words, it becomes very difficult to gauge if a certain direction of research on proteins is “on track”, i.e. likely to provide useful results in the long run. It would be like a blind man stumbling around in the dark, not knowing where he is or where he is going, only to be able to “feel” the destination when he is right at it (the worst case being when the destination cannot even be “felt”, making it no more efficient than random diffusion is as a means of finding your way home after a wild party).

In reality it might not be so bad, but the question of which is the minimum complexity required in a model to appropriately describe the behavior of a protein, remains essentially unanswered. At least two extreme cases seem clear, though; a full quantum-dynamical description of the protein is not likely to be necessary, and a lattice model without solvent clearly lacks certain distinctive properties of proteins. All-atom MD simulations with explicit solvent are gradually becoming more established and appear to give comprehensive, reproducible and confirmed results, but a more efficient approach which lies inbetween the all-atom-and-solvent and the beads-on-a-lattice approach might still be attainable. It is not unlikely, however, that the actual computational work involved in such an approach, will not be much lower than that for an atomic model, in which case there is no gain that will render the effort worth its while.

## 8.2 The Ultimate View

The bottom line of this thesis is not the major breakthrough any PhD student hopes to achieve, but rather incremental advancements on several aspects of the protein folding process.

First, the efficiency of performing classical MD simulations of biological, hydrogen- and water-rich systems can be improved at least two-fold without significant loss in accuracy, and three-fold with only marginal loss in accuracy.

Second, sequentially correlated motions in native state protein dynamics dominate over spatial correlations. This opens the possibility of devising a simplified description of protein dynamics in terms of locally coherent sequential fragments. Building on the

notion that protein dynamics and folding are related, these sequential fragments and the observed correlations between them can be used as a guide for proposing a sequence of folding events. One can envision a method that uses this information to fold a protein *ab initio* by first finding the stable conformation of one or a few of the most stable elements, and using those conformations as a scaffold for the neighboring elements to fold onto. The question that remains is whether such an approach will yield an actual improvement in terms of simulated time scales versus CPU time usage. Irrespective of this possible gain in efficiency of simulating the folding process, actually implementing the method and applying it will be an instructive test of the assumptions described here.

Finally, as we have shown for peptide MD simulations and NMR experiments, the gap between theoretical and experimental results on many issues of protein dynamics (and folding) remains so large as to preclude definite conclusions. On the other hand, as we have shown for FAD simulations and fluorescence measurements, many problems remain for which interesting comparisons can be made. Luckily much progress is being made to overcome the remaining obstacles, both from the sides of experiments and simulations, and also in terms of methods of comparison.

Thus, we have shown that it is possible to describe a protein as a set of more or less independently moving units. Whether it will be possible to accurately describe complex processes like protein folding using the simplified model based on the independent units, remains to be seen. The complexity of the interactions between the units certainly indicates that such a reductionist approach is not guaranteed to be successful. Nevertheless a step has been made towards uncovering the intrinsic complexity of protein dynamics.

# Hoofdstuk 9

## Conclusie

### 9.1 Het Toekomstperspectief

De enorme complexiteit van eiwitssystemen maakt het wellicht zo goed als ontoegankelijk voor conventionele wetenschappelijke benaderingen, in de zin dat de verhoudingen tussen alle samenstellende onderdelen van het eiwit zélf zo complex en nauw verwoven zijn, dat elke vereenvoudiging of onderverdeling in afzonderlijke entiteiten, zoals gebruikelijk is in de wetenschap, noodzakelijkerwijs kunstmatig en willekeurig zijn en erger nog, enkel van toepassing op het nivo waarvoor deze vereenvoudiging of onderverdeling gemaakt was. Dit betekent dat uit een vereenvoudigd model van een eiwit geen algemene regels afgeleid kunnen worden die geldig zijn voor het echte eiwitstelsel. Dit is een boude uitspraak, omdat het inhoudt dat alleen een model dat het hele eiwit (en mogelijk een belangrijk deel van de omgeving ervan) precies en gedetailleerd beschrijft, ooit in staat zal zijn het gedrag van het eiwit zo te beschrijven dat het mogelijk is om conclusies te baseren op dat model. Het houdt ook in dat, als we per ongeluk slechts een deel van de “juiste” beschrijving ontdekken, het moeilijk, zo niet onmogelijk, zal zijn om te bepalen of die beschrijving feitelijk correct was. Met ander woorden, het wordt dan zeer lastig in te schatten of een bepaalde onderzoeksrichting op het goede spoor zit, in de zin dat het op lange termijn bruikbare resultaten zal opleveren. De situatie is dan vergelijkbaar met die van een blinde man die in het donker ronddwaalt, zonder enig idee van waar hij is of waar hij heengaat, en slechts in staat het eindpunt te “voelen” als hij er al pal tegenaan staat.

In werkelijkheid is het wellicht niet zo erg, maar de vraag wat de minimaal benodigde complexiteit in een model is om het gedrag van een eiwit zinnig te beschrijven, blijft grotendeels onbeantwoord. In ieder geval zijn twee uitersten duidelijk; een volledig quantummechanische beschrijving van het eiwit is waarschijnlijk niet nodig, en een roostermodel zonder oplosmiddel mist duidelijk bepaalde karakteristieke eigenschappen van eiwitten. MD simulaties waarbij alle atomen meegenomen worden, inclusief het oplosmiddel, worden langzamerhand meer en meer tot de gevestigde orde gerekend en lijken begrijpelijke, reproduceerbare en bevestigde resultaten op te leveren, maar een efficiëntere aanpak die tussen de alle-atomen-met-oplosmiddel en balletjes-op-een-rooster aanpak in ligt hoort nog

tot de mogelijkheden. Echter, het is niet onwaarschijnlijk dat het feitelijke rekenwerk dat zo'n aanpak met zich meebrengt niet veel minder zal zijn dan dat voor een atomistisch model. In dat geval is er geen winst die de investering de moeite waard zal maken.

## 9.2 Het Ultieme Perspectief

De slotuitkomst van dit proefschrift is niet de grote doorbraak die elke promovendus hoopt te bereiken, maar eerder een incrementele voortschrijding op verschillende aspecten van het eiwitvouwingsproces.

Ten eerste kan de efficiëntie van het uitvoeren van klassieke MD simulaties van biologische, waterstof- en waterrijke systemen verhoogd worden met ten minste een factor twee zonder significant verlies in nauwkeurigheid, en een factor drie met slechts een gering verlies in nauwkeurigheid.

Ten tweede overheersen sequentieel gecorreleerde bewegingen in de dynamica van eiwitten in de natieve toestand over ruimtelijke correlaties. Dit opent de mogelijkheid om een vereenvoudigde beschrijving van eiwitdynamica te bedenken in termen van lokaal, sequentieel samenhangende fragmenten. Voortbouwend op het idee dat eiwitdynamica en -vouwing verwant zijn, kunnen deze sequentiële fragmenten en de waargenomen correlaties daartussen, gebruikt worden als richtlijn bij het voorstellen van een opeenvolging van vouwingsgebeurtenissen. Men kan zich een methode voorstellen die gebruik maakt van deze informatie om een eiwit *ab initio* te vouwen door eerst de stabiele conformatie van een of enkele van de stabielste elementen te vinden en vervolgens die conformaties te gebruiken als een raamwerk met behulp waarvan de naburige elementen zich kunnen vouwen. De vraag die daarbij open blijft is of zo'n aanpak een feitelijke verbetering in termen van gesimuleerde tijdsschalen ten opzichte van de gebruikte rekentijd zal opleveren. Ongeacht deze mogelijke winst in efficiëntie van het simuleren van het vouwingsproces, zal het implementeren en toepassen van de methode een leerzame test inhouden van de hier beschreven aannames.

Tot slot, zoals we lieten zien voor peptide MD simulaties en NMR experimenten, blijft de kloof tussen theoretische en experimentele resultaten op veel vlakken van eiwitdynamica (en -vouwing) zo groot dat vooralsnog definitieve conclusies moeten uitblijven. Aan de andere kant, zoals getoond is voor FAD simulaties en fluorescentie experimenten, zijn er veel andere problemen waarvoor interessante vergelijkingen wel gemaakt kunnen worden. Gelukkig wordt er veel vooruitgang geboekt om de resterende obstakels te slechten, zowel van de kant van experimenten en simulaties, als in termen van methodes om de vergelijking mee uit te voeren.

Aldus hebben we laten zien dat het mogelijk is om een eiwit te beschrijven als een set van min of meer onafhankelijk bewegende eenheden. Of het mogelijk zal zijn om een accurate beschrijving te geven van complexe processen, zoals eiwitvouwing, met het vereenvoudigde model op grond van de onafhankelijke eenheden, zal nader bezien moeten worden. De complexiteit van de interacties tussen de eenheden geeft zeker aan dat voor zo'n reductionistische aanpak succes niet gegarandeerd is. Toch is een stap gemaakt in de richting van het ontsluiten van de intrinsieke complexiteit van eiwitdynamica.



# Appendices

## Appendix A

# Redistribution of Forces on Dummy Atoms

$$\mathbf{F}'_{ix} = - \frac{\partial \mathbf{r}_d}{\partial x_i} \cdot \frac{\partial V}{\partial \mathbf{r}_d}$$

Dummy atoms are virtual particles with position  $\mathbf{r}_d$ , which are constructed from the positions of the real particles  $\mathbf{r}_i$ . Therefore every  $\mathbf{r}_d$  is a known function of  $\mathbf{r}_i$ 's. Any force  $\mathbf{F}_d$  on a dummy atom is redistributed to the real atoms on which  $\mathbf{r}_d$  depends. When a linear combination of three atoms is used in constructing the dummy atom, the weights for redistributing the forces are equal to those used in the linear combination. The redistribution becomes nontrivial if normalization is used in constructing the dummy atom, as is the case for the dummy types used for aromatic, amide and amine hydrogens (see Fig. 3.1B and C), which use a fixed distance and/or angle with respect to the constructing atoms.

The force acting on atom  $i$  ( $\mathbf{F}'_i$ ) as a result of the force on the dummy atom must be calculated from the partial derivative of the position of the dummy atom with respect to the position of atom  $i$ :<sup>13</sup>

$$\mathbf{F}'_{ix} = - \frac{\partial \mathbf{r}_d}{\partial x_i} \cdot \frac{\partial V}{\partial \mathbf{r}_d} = \frac{\partial \mathbf{r}_d}{\partial x_i} \cdot \mathbf{F}_d \quad (\text{A.1})$$

Here  $V$  is the potential energy expressed in positions of real *and* dummy atom positions. Analogous expressions are valid for the  $y$  and  $z$  component.

For type B (see Fig. 3.1B) with a fixed distance to atom  $i$ , the position of the dummy atom  $d$  is calculated from the position of the constructing atoms  $i$ ,  $j$  and  $k$  as follows:

$$\mathbf{r}_d = \mathbf{r}_i + b \frac{\mathbf{r}_{ij} + a\mathbf{r}_{jk}}{|\mathbf{r}_{ij} + a\mathbf{r}_{jk}|} \quad (\text{A.2})$$

where  $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ . Using eqn. A.1 to calculate the redistributed force for atoms  $i$ ,  $j$  and  $k$  yields:

$$\begin{aligned} \mathbf{F}'_i &= \mathbf{F}_d - \gamma(\mathbf{F}_d - \mathbf{F}_1) & \gamma &= \frac{b}{|\mathbf{r}_{ij} + a\mathbf{r}_{jk}|} \\ \mathbf{F}'_j &= (1-a)\gamma(\mathbf{F}_d - \mathbf{F}_1) & \text{where} & \\ \mathbf{F}'_k &= a\gamma(\mathbf{F}_d - \mathbf{F}_1) & \mathbf{F}_1 &= \frac{\mathbf{r}_{id} \cdot \mathbf{F}_d}{\mathbf{r}_{id} \cdot \mathbf{r}_{id}} \mathbf{r}_{id} \end{aligned} \quad (\text{A.3})$$

For type C (see Fig. 3.1C) with a fixed distance to atom  $i$  and a fixed angle  $d$ - $i$ - $j$  the position is calculated using:

$$\mathbf{r}_d = \mathbf{r}_i + b \cos \alpha \frac{\mathbf{r}_{ij}}{|\mathbf{r}_{ij}|} + b \sin \alpha \frac{\mathbf{r}_\perp}{|\mathbf{r}_\perp|} \quad \text{where} \quad \mathbf{r}_\perp = \mathbf{r}_{jk} - \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}}{\mathbf{r}_{ij} \cdot \mathbf{r}_{ij}} \mathbf{r}_{ij} \quad (\text{A.4})$$

with corresponding forces, using eqn. A.1:

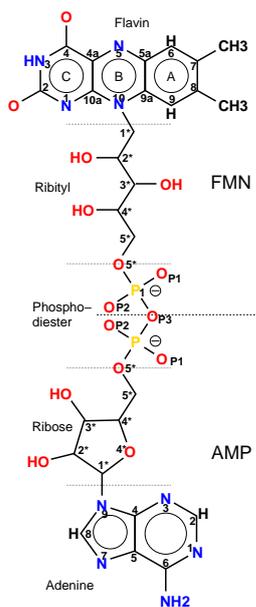
$$\begin{aligned} \mathbf{F}'_i &= \mathbf{F}_d - \frac{b \cos \alpha}{|\mathbf{r}_{ij}|} \mathbf{F}_1 + \frac{b \sin \alpha}{|\mathbf{r}_\perp|} \left( \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}}{\mathbf{r}_{ij} \cdot \mathbf{r}_{ij}} \mathbf{F}_2 + \mathbf{F}_3 \right) \\ \mathbf{F}'_j &= \frac{b \cos \alpha}{|\mathbf{r}_{ij}|} \mathbf{F}_1 - \frac{b \sin \alpha}{|\mathbf{r}_\perp|} \left( \mathbf{F}_2 + \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}}{\mathbf{r}_{ij} \cdot \mathbf{r}_{ij}} \mathbf{F}_2 + \mathbf{F}_3 \right) \\ \mathbf{F}'_k &= \frac{b \sin \alpha}{|\mathbf{r}_\perp|} \mathbf{F}_2 \end{aligned} \quad (\text{A.5})$$

$$\text{where } \mathbf{F}_1 = \mathbf{F}_d - \frac{\mathbf{r}_{ij} \cdot \mathbf{F}_d}{\mathbf{r}_{ij} \cdot \mathbf{r}_{ij}} \mathbf{r}_{ij}, \quad \mathbf{F}_2 = \mathbf{F}_1 - \frac{\mathbf{r}_\perp \cdot \mathbf{F}_d}{\mathbf{r}_\perp \cdot \mathbf{r}_\perp} \mathbf{r}_\perp, \quad \mathbf{F}_3 = \frac{\mathbf{r}_{ij} \cdot \mathbf{F}_d}{\mathbf{r}_{ij} \cdot \mathbf{r}_{ij}} \mathbf{r}_\perp$$

and  $\mathbf{r}_\perp$  as defined in eqn. A.4

## Appendix B

# FAD Force Field Parameters



## B.1 Atoms

moiety*	#	type†	name*	cgnr‡	charge (e)	mass (u)
Flavin	1	CB	C <sub>9A</sub>	1	0.14	12.011
	2	NR6*	N <sub>10</sub>	2	-0.05	14.0067
	3	CB	C <sub>10</sub>	3	0.27	12.011
	4	NR6	N <sub>1</sub>	4	-0.43	14.0067
	5	CB	C <sub>2</sub>	5	0.51	12.011
	6	O	O <sub>2</sub>	6	-0.45	15.9994
	7	NR6*	N <sub>3</sub>	7	-0.35	14.0067
	8	H	H <sub>3</sub>	8	0.35	1.008
	9	CB	C <sub>4</sub>	9	0.43	12.011
	10	O	O <sub>4</sub>	10	-0.43	15.9994
	11	CB	C <sub>4A</sub>	11	0.06	12.011
	12	NR6	N <sub>5</sub>	12	-0.19	14.0067
	13	CB	C <sub>5A</sub>	13	0.06	12.011
	14	CR6	C <sub>6</sub>	14	-0.12	12.011
	15	HCR	H <sub>6</sub>	15	0.19	1.008
	16	CB	C <sub>7</sub>	16	0	12.011
	17	CH3	C <sub>7M</sub>	17	0	15.035
	18	CB	C <sub>8</sub>	18	0	12.011
	19	CH3	C <sub>8M</sub>	19	0	15.035
	20	CR6	C <sub>9</sub>	20	-0.18	12.011
	21	HCR	H <sub>9</sub>	21	0.19	1.008
Ribityl	22	CH2	C <sub>1*</sub>	22	0	14.027
	23	CH1	C <sub>2*</sub>	23	0.15	13.019
	24	OA	O <sub>2*</sub>	23	-0.548	15.9994
	25	HO	H <sub>2*</sub>	23	0.398	1.008
	26	CH1	C <sub>3*</sub>	24	0.15	13.019
	27	OA	O <sub>3*</sub>	24	-0.548	15.9994
	28	HO	H <sub>3*</sub>	24	0.398	1.008
	29	CH1	C <sub>4*</sub>	25	0.15	13.019
	30	OA	O <sub>4*</sub>	25	-0.548	15.9994
	31	HO	H <sub>4*</sub>	25	0.398	1.008
	32	CH2	C <sub>5*</sub>	26	0.13	14.027
	33	OS	O <sub>5*</sub>	27	-0.52	15.9994
Phospho- di-ester	34	P	P	28	1.85	30.9738
	35	OS	O <sub>P3</sub>	29	-0.92	15.9994
	36	OM	O <sub>P1</sub>	30	-1	15.9994
	37	OM	O <sub>P2</sub>	31	-1	15.9994
	38	P	<sup>A</sup> P	32	1.85	30.9738
39	OM	<sup>A</sup> O <sub>P1</sub>	33	-1	15.9994	
40	OM	<sup>A</sup> O <sub>P2</sub>	34	-1	15.9994	
Ribose	41	OS	<sup>A</sup> O <sub>5*</sub>	35	-0.52	15.9994
	42	CS2	<sup>A</sup> C <sub>5*</sub>	36	0.13	14.027
	43	CS1	<sup>A</sup> C <sub>4*</sub>	37	0.16	13.019
	44	OS	<sup>A</sup> O <sub>4*</sub>	37	-0.36	15.9994
	45	CS1	<sup>A</sup> C <sub>1*</sub>	37	0.2	13.019

(continued ...)

\* names and labels correspond to Fig. 7.1

† GROMACS atom type label

‡ charge group number

## Atoms (... continued)

moiety*	#	type†	name*	cgmr‡	charge (e)	mass (u)
Adenine	46	NR5*	<sup>A</sup> N <sub>9</sub>	38	-0.2	14.0067
	47	CB	<sup>A</sup> C <sub>4</sub>	38	0.2	12.011
	48	NR6	<sup>A</sup> N <sub>3</sub>	39	-0.36	14.0067
	49	CR6	<sup>A</sup> C <sub>2</sub>	39	0.22	12.011
	50	HCR	H <sub>C2</sub>	39	0.14	1.008
	51	NR6	<sup>A</sup> N <sub>1</sub>	40	-0.36	14.0067
	52	CB	<sup>A</sup> C <sub>6</sub>	40	0.36	12.011
	53	NT	<sup>A</sup> N <sub>6</sub>	41	-0.83	14.0067
	54	H	H <sub>N61</sub>	41	0.415	1.008
	55	H	H <sub>N62</sub>	41	0.415	1.008
	56	CB	<sup>A</sup> C <sub>5</sub>	42	0	12.011
	57	NR5	<sup>A</sup> N <sub>7</sub>	43	-0.36	14.0067
	58	CR5	<sup>A</sup> C <sub>8</sub>	43	0.22	12.011
	59	HCR	H <sub>C8</sub>	43	0.14	1.008
Ribose	60	CS1	<sup>A</sup> C <sub>2*</sub>	44	0.15	13.019
	61	OA	<sup>A</sup> O <sub>2*</sub>	44	-0.548	15.9994
	62	HO	H <sub>O2*</sub>	44	0.398	1.008
	63	CS1	<sup>A</sup> C <sub>3*</sub>	45	0.15	13.019
	64	OA	<sup>A</sup> O <sub>3*</sub>	45	-0.548	15.9994
	65	HO	H <sub>O3*</sub>	45	0.398	1.008

## B.2 Bonds

$a_i$	$a_j$	$r_0$ (nm)	$f_c$ (kJ mol <sup>-1</sup> nm <sup>-1</sup> )	$a_i$	$a_j$	$r_0$ (nm)	$f_c$ (kJ mol <sup>-1</sup> nm <sup>-1</sup> )
C <sub>9A</sub>	N <sub>10</sub>	0.14	334720	C <sub>9A</sub>	C <sub>5A</sub>	0.139	418400
C <sub>9A</sub>	C <sub>9</sub>	0.139	418400	N <sub>10</sub>	C <sub>10</sub>	0.14	334720
N <sub>10</sub>	C <sub>1*</sub>	0.134	418400	C <sub>10</sub>	N <sub>1</sub>	0.134	418400
C <sub>10</sub>	C <sub>4A</sub>	0.139	418400	N <sub>1</sub>	C <sub>2</sub>	0.134	418400
C <sub>2</sub>	O <sub>2</sub>	0.123	502080	C <sub>2</sub>	N <sub>3</sub>	0.14	334720
N <sub>3</sub>	H <sub>3</sub>	0.1	374468	N <sub>3</sub>	C <sub>4</sub>	0.14	334720
C <sub>4</sub>	O <sub>4</sub>	0.123	502080	C <sub>4</sub>	C <sub>4A</sub>	0.139	418400
C <sub>4A</sub>	N <sub>5</sub>	0.134	418400	N <sub>5</sub>	C <sub>5A</sub>	0.134	418400
C <sub>5A</sub>	C <sub>6</sub>	0.139	418400	C <sub>6</sub>	H <sub>6</sub>	0.108	292880
C <sub>6</sub>	C <sub>7</sub>	0.139	418400	C <sub>7</sub>	C <sub>7M</sub>	0.153	334720
C <sub>7</sub>	C <sub>8</sub>	0.139	418400	C <sub>8</sub>	C <sub>8M</sub>	0.153	334720
C <sub>8</sub>	C <sub>9</sub>	0.139	418400	C <sub>9</sub>	H <sub>9</sub>	0.108	292880
C <sub>1*</sub>	C <sub>2*</sub>	0.153	334720	C <sub>2*</sub>	O <sub>2*</sub>	0.143	334720
C <sub>2*</sub>	C <sub>3*</sub>	0.153	334720	O <sub>2*</sub>	H <sub>2*</sub>	0.1	313800
C <sub>3*</sub>	O <sub>3*</sub>	0.143	334720	C <sub>3*</sub>	C <sub>4*</sub>	0.153	334720
O <sub>3*</sub>	H <sub>3*</sub>	0.1	313800	C <sub>4*</sub>	O <sub>4*</sub>	0.143	334720
C <sub>4*</sub>	C <sub>5*</sub>	0.153	334720	O <sub>4*</sub>	H <sub>4*</sub>	0.1	313800
C <sub>5*</sub>	O <sub>5*</sub>	0.143	251040	O <sub>5*</sub>	P	0.161	251040
P	O <sub>P3</sub>	0.161	251040	P	O <sub>P1</sub>	0.148	376560
P	O <sub>P2</sub>	0.148	376560	O <sub>P3</sub>	<sup>A</sup> P	0.161	251040
<sup>A</sup> P	<sup>A</sup> O <sub>P1</sub>	0.148	376560	<sup>A</sup> P	<sup>A</sup> O <sub>P2</sub>	0.148	376560

(continued ...)

Bonds (... continued)

$a_i$	$a_j$	$r_0$ (nm)	$f_c$ (kJ mol <sup>-1</sup> nm <sup>-1</sup> )	$a_i$	$a_j$	$r_0$ (nm)	$f_c$ (kJ mol <sup>-1</sup> nm <sup>-1</sup> )
A <sub>P</sub>	A <sub>O<sub>5*</sub></sub>	0.161	251040	A <sub>O<sub>5*</sub></sub>	A <sub>C<sub>5*</sub></sub>	0.1435	251040
A <sub>C<sub>5*</sub></sub>	A <sub>C<sub>4*</sub></sub>	0.152	251040	A <sub>C<sub>4*</sub></sub>	A <sub>O<sub>4*</sub></sub>	0.1435	251040
A <sub>C<sub>4*</sub></sub>	A <sub>C<sub>3*</sub></sub>	0.152	251040	A <sub>O<sub>4*</sub></sub>	A <sub>C<sub>1*</sub></sub>	0.1435	251040
A <sub>C<sub>1*</sub></sub>	A <sub>N<sub>9</sub></sub>	0.148	251040	A <sub>C<sub>1*</sub></sub>	A <sub>C<sub>2*</sub></sub>	0.152	251040
A <sub>N<sub>9</sub></sub>	A <sub>C<sub>4</sub></sub>	0.133	418400	A <sub>N<sub>9</sub></sub>	A <sub>C<sub>8</sub></sub>	0.133	418400
A <sub>C<sub>4</sub></sub>	A <sub>N<sub>3</sub></sub>	0.134	418400	A <sub>C<sub>4</sub></sub>	A <sub>C<sub>5</sub></sub>	0.139	418400
A <sub>N<sub>3</sub></sub>	A <sub>C<sub>2</sub></sub>	0.132	418400	A <sub>C<sub>2</sub></sub>	H <sub>C<sub>2</sub></sub>	0.108	292880
A <sub>C<sub>2</sub></sub>	A <sub>N<sub>1</sub></sub>	0.132	418400	A <sub>N<sub>1</sub></sub>	A <sub>C<sub>6</sub></sub>	0.134	418400
A <sub>C<sub>6</sub></sub>	A <sub>N<sub>6</sub></sub>	0.133	376560	A <sub>C<sub>6</sub></sub>	A <sub>C<sub>5</sub></sub>	0.139	418400
A <sub>N<sub>6</sub></sub>	H <sub>N<sub>61</sub></sub>	0.1	374468	A <sub>N<sub>6</sub></sub>	H <sub>N<sub>62</sub></sub>	0.1	374468
A <sub>C<sub>5</sub></sub>	A <sub>N<sub>7</sub></sub>	0.133	418400	A <sub>N<sub>7</sub></sub>	A <sub>C<sub>8</sub></sub>	0.133	418400
A <sub>C<sub>8</sub></sub>	H <sub>C<sub>8</sub></sub>	0.108	292880	A <sub>C<sub>2*</sub></sub>	A <sub>O<sub>2*</sub></sub>	0.143	251040
A <sub>C<sub>2*</sub></sub>	A <sub>C<sub>3*</sub></sub>	0.152	251040	A <sub>O<sub>2*</sub></sub>	H <sub>O<sub>2*</sub></sub>	0.1	313800
A <sub>C<sub>3*</sub></sub>	A <sub>O<sub>3*</sub></sub>	0.143	251040	A <sub>O<sub>3*</sub></sub>	H <sub>O<sub>3*</sub></sub>	0.1	313800

### B.3 1-4 Lennard-Jones

$a_i$	$a_j$	$C_6$ (kJ mol <sup>-1</sup> nm <sup>6</sup> )	$C_{12}$ (kJ mol <sup>-1</sup> nm <sup>12</sup> )	$a_i$	$a_j$	$C_6$ (kJ mol <sup>-1</sup> nm <sup>6</sup> )	$C_{12}$ (kJ mol <sup>-1</sup> nm <sup>12</sup> )
C <sub>9A</sub>	N <sub>1</sub>	0.0023877	2.3896 · 10 <sup>-6</sup>	C <sub>9A</sub>	C <sub>4A</sub>	0.0023402	3.374 · 10 <sup>-6</sup>
C <sub>9A</sub>	H <sub>6</sub>	0.00044528	.22581 · 10 <sup>-6</sup>	C <sub>9A</sub>	C <sub>7</sub>	0.0023402	3.374 · 10 <sup>-6</sup>
C <sub>9A</sub>	C <sub>8M</sub>	0.0040046	6.3798 · 10 <sup>-6</sup>	C <sub>9A</sub>	C <sub>2*</sub>	0.0026103	3.5506 · 10 <sup>-6</sup>
N <sub>10</sub>	C <sub>2</sub>	0.0023877	2.3896 · 10 <sup>-6</sup>	N <sub>10</sub>	C <sub>4</sub>	0.0023877	2.3896 · 10 <sup>-6</sup>
N <sub>10</sub>	N <sub>5</sub>	0.0024362	1.6924 · 10 <sup>-6</sup>	N <sub>10</sub>	C <sub>6</sub>	0.0036699	3.7547 · 10 <sup>-6</sup>
N <sub>10</sub>	C <sub>8</sub>	0.0023877	2.3896 · 10 <sup>-6</sup>	N <sub>10</sub>	H <sub>9</sub>	0.00045432	.15993 · 10 <sup>-6</sup>
N <sub>10</sub>	O <sub>2*</sub>	0.0023473	1.1203 · 10 <sup>-6</sup>	N <sub>10</sub>	C <sub>3*</sub>	0.0026633	2.5147 · 10 <sup>-6</sup>
C <sub>10</sub>	O <sub>2</sub>	0.0023006	1.5818 · 10 <sup>-6</sup>	C <sub>10</sub>	N <sub>3</sub>	0.0023877	2.3896 · 10 <sup>-6</sup>
C <sub>10</sub>	O <sub>4</sub>	0.0023006	1.5818 · 10 <sup>-6</sup>	C <sub>10</sub>	C <sub>5A</sub>	0.0023402	3.374 · 10 <sup>-6</sup>
C <sub>10</sub>	C <sub>9</sub>	0.0035969	5.3015 · 10 <sup>-6</sup>	C <sub>10</sub>	C <sub>2*</sub>	0.0026103	3.5506 · 10 <sup>-6</sup>
N <sub>1</sub>	H <sub>3</sub>	0	0	N <sub>1</sub>	C <sub>4</sub>	0.0023877	2.3896 · 10 <sup>-6</sup>
N <sub>1</sub>	N <sub>5</sub>	0.0024362	1.6924 · 10 <sup>-6</sup>	N <sub>1</sub>	C <sub>1*</sub>	0.0033923	3.47 · 10 <sup>-6</sup>
C <sub>2</sub>	O <sub>4</sub>	0.0023006	1.5818 · 10 <sup>-6</sup>	C <sub>2</sub>	C <sub>4A</sub>	0.0023402	3.374 · 10 <sup>-6</sup>
O <sub>2</sub>	H <sub>3</sub>	0	0	O <sub>2</sub>	C <sub>4</sub>	0.0023006	1.5818 · 10 <sup>-6</sup>
N <sub>3</sub>	N <sub>5</sub>	0.0024362	1.6924 · 10 <sup>-6</sup>	H <sub>3</sub>	O <sub>4</sub>	0	0
H <sub>3</sub>	C <sub>4A</sub>	0	0	C <sub>4</sub>	C <sub>5A</sub>	0.0023402	3.374 · 10 <sup>-6</sup>
O <sub>4</sub>	N <sub>5</sub>	0.0023473	1.1203 · 10 <sup>-6</sup>	C <sub>4A</sub>	C <sub>6</sub>	0.0035969	5.3015 · 10 <sup>-6</sup>
C <sub>4A</sub>	C <sub>1*</sub>	0.0033248	4.8994 · 10 <sup>-6</sup>	N <sub>5</sub>	H <sub>6</sub>	0.00045432	.15993 · 10 <sup>-6</sup>
N <sub>5</sub>	C <sub>7</sub>	0.0023877	2.3896 · 10 <sup>-6</sup>	N <sub>5</sub>	C <sub>9</sub>	0.0036699	3.7547 · 10 <sup>-6</sup>
C <sub>5A</sub>	C <sub>7M</sub>	0.0040046	6.3798 · 10 <sup>-6</sup>	C <sub>5A</sub>	C <sub>8</sub>	0.0023402	3.374 · 10 <sup>-6</sup>
C <sub>5A</sub>	H <sub>9</sub>	0.00044528	.22581 · 10 <sup>-6</sup>	C <sub>5A</sub>	C <sub>1*</sub>	0.0033248	4.8994 · 10 <sup>-6</sup>
C <sub>6</sub>	C <sub>8M</sub>	0.006155	10.024 · 10 <sup>-6</sup>	C <sub>6</sub>	C <sub>9</sub>	0.0055284	8.33 · 10 <sup>-6</sup>
H <sub>6</sub>	C <sub>7M</sub>	0.00076197	.42698 · 10 <sup>-6</sup>	H <sub>6</sub>	C <sub>8</sub>	0.00044528	.22581 · 10 <sup>-6</sup>
C <sub>7</sub>	H <sub>9</sub>	0.00044528	.22581 · 10 <sup>-6</sup>	C <sub>7M</sub>	C <sub>8M</sub>	0.0068526	12.063 · 10 <sup>-6</sup>
C <sub>7M</sub>	C <sub>9</sub>	0.006155	10.024 · 10 <sup>-6</sup>	C <sub>8M</sub>	H <sub>9</sub>	0.00076197	.42698 · 10 <sup>-6</sup>
C <sub>9</sub>	C <sub>1*</sub>	0.0051102	7.6983 · 10 <sup>-6</sup>	C <sub>1*</sub>	H <sub>2*</sub>	0	0
C <sub>1*</sub>	O <sub>3*</sub>	0.0032685	2.2969 · 10 <sup>-6</sup>	C <sub>1*</sub>	C <sub>4*</sub>	0.0037086	5.1559 · 10 <sup>-6</sup>
C <sub>2*</sub>	H <sub>3*</sub>	0	0	C <sub>2*</sub>	O <sub>4*</sub>	0.0025662	1.6646 · 10 <sup>-6</sup>

(continued ...)

## 1-4 Lennard-Jones (... continued)

$a_i$	$a_j$	$C_6$ (kJ mol <sup>-1</sup> nm <sup>6</sup> )	$C_{12}$ (kJ mol <sup>-1</sup> nm <sup>12</sup> )	$a_i$	$a_j$	$C_6$ (kJ mol <sup>-1</sup> nm <sup>6</sup> )	$C_{12}$ (kJ mol <sup>-1</sup> nm <sup>12</sup> )
C2*	C5*	0.0037086	5.1559 · 10 <sup>-6</sup>	O2*	O3*	0.0022617	.74158 · 10 <sup>-6</sup>
O2*	C4*	0.0025662	1.6646 · 10 <sup>-6</sup>	H2*	C3*	0	0
C3*	H4*	0	0	C3*	O5*	0.0025662	1.6646 · 10 <sup>-6</sup>
O3*	O4*	0.0022617	.74158 · 10 <sup>-6</sup>	O3*	C5*	0.0032685	2.2969 · 10 <sup>-6</sup>
H3*	C4*	0	0	C4*	P	0.0065507	9.1058 · 10 <sup>-6</sup>
O4*	O5*	0.0022617	.74158 · 10 <sup>-6</sup>	H4*	C5*	0	0
C5*	OP3	0.0032685	2.2969 · 10 <sup>-6</sup>	C5*	OP1	0.0032685	2.2969 · 10 <sup>-6</sup>
C5*	OP2	0.0032685	2.2969 · 10 <sup>-6</sup>	O5*	AP	0.0057734	4.0567 · 10 <sup>-6</sup>
P	AOP1	0.0057734	4.0567 · 10 <sup>-6</sup>	P	AOP2	0.0057734	4.0567 · 10 <sup>-6</sup>
P	AO5*	0.0057734	4.0567 · 10 <sup>-6</sup>	OP3	AC5*	0.0032685	2.2969 · 10 <sup>-6</sup>
OP1	AP	0.0057734	4.0567 · 10 <sup>-6</sup>	OP2	AP	0.0057734	4.0567 · 10 <sup>-6</sup>
AP	AC4*	0.0065507	9.1058 · 10 <sup>-6</sup>	AOP1	AC5*	0.0032685	2.2969 · 10 <sup>-6</sup>
AOP2	AC5*	0.0032685	2.2969 · 10 <sup>-6</sup>	AO5*	AO4*	0.0022617	.74158 · 10 <sup>-6</sup>
AO5*	AC3*	0.0025662	1.6646 · 10 <sup>-6</sup>	AC5*	AC1*	0.0037086	5.1559 · 10 <sup>-6</sup>
AC5*	AC2*	0.0037086	5.1559 · 10 <sup>-6</sup>	AC5*	AO3*	0.0032685	2.2969 · 10 <sup>-6</sup>
AC4*	AN9	0.0026633	2.5147 · 10 <sup>-6</sup>	AC4*	AO2*	0.0025662	1.6646 · 10 <sup>-6</sup>
AC4*	HO3*	0	0	AO4*	AC4	0.0023006	1.5818 · 10 <sup>-6</sup>
AO4*	AC8	0.0035361	2.4854 · 10 <sup>-6</sup>	AO4*	AO2*	0.0022617	.74158 · 10 <sup>-6</sup>
AO4*	AO3*	0.0022617	.74158 · 10 <sup>-6</sup>	AC1*	AN3	0.0026633	2.5147 · 10 <sup>-6</sup>
AC1*	AC5	0.0026103	3.5506 · 10 <sup>-6</sup>	AC1*	AN7	0.0026633	2.5147 · 10 <sup>-6</sup>
AC1*	HC8	0.00049668	.23763 · 10 <sup>-6</sup>	AC1*	HO2*	0	0
AC1*	AO3*	0.0025662	1.6646 · 10 <sup>-6</sup>	AN9	AC2	0.0036699	3.7547 · 10 <sup>-6</sup>
AN9	AC6	0.0023877	2.3896 · 10 <sup>-6</sup>	AN9	AO2*	0.0023473	1.1203 · 10 <sup>-6</sup>
AN9	AC3*	0.0026633	2.5147 · 10 <sup>-6</sup>	AC4	HC2	0.00044528	.22581 · 10 <sup>-6</sup>
AC4	AN1	0.0023877	2.3896 · 10 <sup>-6</sup>	AC4	AN6	0.0023877	2.3896 · 10 <sup>-6</sup>
AC4	HC8	0.00044528	.22581 · 10 <sup>-6</sup>	AC4	AC2*	0.0026103	3.5506 · 10 <sup>-6</sup>
AN3	AC6	0.0023877	2.3896 · 10 <sup>-6</sup>	AN3	AN7	0.0024362	1.6924 · 10 <sup>-6</sup>
AN3	AC8	0.0036699	3.7547 · 10 <sup>-6</sup>	AC2	AN6	0.0036699	3.7547 · 10 <sup>-6</sup>
AC2	AC5	0.0035969	5.3015 · 10 <sup>-6</sup>	HC2	AC6	0.00044528	.22581 · 10 <sup>-6</sup>
AN1	HN61	0	0	AN1	HN62	0	0
AN1	AN7	0.0024362	1.6924 · 10 <sup>-6</sup>	AC6	AC8	0.0035969	5.3015 · 10 <sup>-6</sup>
AN6	AN7	0.0024362	1.6924 · 10 <sup>-6</sup>	HN61	AC5	0	0
HN62	AC5	0	0	AC5	HC8	0.00044528	.22581 · 10 <sup>-6</sup>
AC8	AC2*	0.0040121	5.5789 · 10 <sup>-6</sup>	AC2*	HO3*	0	0
AO2*	AO3*	0.0022617	.74158 · 10 <sup>-6</sup>	HO2*	AC3*	0	0

## B.4 Angles

$a_i$	$a_j$	$a_k$	$\theta_0$ (degr)	$f_c$ (kJ mol <sup>-1</sup> rad <sup>-2</sup> )	$a_i$	$a_j$	$a_k$	$\theta_0$ (degr)	$f_c$ (kJ mol <sup>-1</sup> rad <sup>-2</sup> )
N10	C9A	C5A	120	418.4	N10	C9A	C9	120	418.4
C5A	C9A	C9	120	418.4	C9A	N10	C10	120	418.4
C9A	N10	C1*	120	418.4	C10	N10	C1*	120	418.4
N10	C10	N1	120	418.4	N10	C10	C4A	120	418.4
N1	C10	C4A	120	418.4	C10	N1	C2	120	418.4
N1	C2	O2	120	418.4	N1	C2	N3	120	418.4

(continued ...)

## Angles (... continued)

$a_i$	$a_j$	$a_k$	$\theta_0$ (degr)	$f_c$ (kJ mol <sup>-1</sup> rad <sup>-2</sup> )		$a_i$	$a_j$	$a_k$	$\theta_0$ (degr)	$f_c$ (kJ mol <sup>-1</sup> rad <sup>-2</sup> )
O <sub>2</sub>	C <sub>2</sub>	N <sub>3</sub>	120	418.4		C <sub>2</sub>	N <sub>3</sub>	H <sub>3</sub>	120	376.56
C <sub>2</sub>	N <sub>3</sub>	C <sub>4</sub>	120	418.4		H <sub>3</sub>	N <sub>3</sub>	C <sub>4</sub>	120	376.56
N <sub>3</sub>	C <sub>4</sub>	O <sub>4</sub>	120	418.4		N <sub>3</sub>	C <sub>4</sub>	C <sub>4A</sub>	120	418.4
O <sub>4</sub>	C <sub>4</sub>	C <sub>4A</sub>	120	418.4		C <sub>10</sub>	C <sub>4A</sub>	C <sub>4</sub>	120	418.4
C <sub>10</sub>	C <sub>4A</sub>	N <sub>5</sub>	120	418.4		C <sub>4</sub>	C <sub>4A</sub>	N <sub>5</sub>	120	418.4
C <sub>4A</sub>	N <sub>5</sub>	C <sub>5A</sub>	120	418.4		C <sub>9A</sub>	C <sub>5A</sub>	N <sub>5</sub>	120	418.4
C <sub>9A</sub>	C <sub>5A</sub>	C <sub>6</sub>	120	418.4		N <sub>5</sub>	C <sub>5A</sub>	C <sub>6</sub>	120	418.4
C <sub>5A</sub>	C <sub>6</sub>	H <sub>6</sub>	120	376.56		C <sub>5A</sub>	C <sub>6</sub>	C <sub>7</sub>	120	418.4
H <sub>6</sub>	C <sub>6</sub>	C <sub>7</sub>	120	376.56		C <sub>6</sub>	C <sub>7</sub>	C <sub>7M</sub>	120	418.4
C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	120	418.4		C <sub>7M</sub>	C <sub>7</sub>	C <sub>8</sub>	120	418.4
C <sub>7</sub>	C <sub>8</sub>	C <sub>8M</sub>	120	418.4		C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	120	418.4
C <sub>8M</sub>	C <sub>8</sub>	C <sub>9</sub>	120	418.4		C <sub>9A</sub>	C <sub>9</sub>	C <sub>8</sub>	120	418.4
C <sub>9A</sub>	C <sub>9</sub>	H <sub>9</sub>	120	376.56		C <sub>8</sub>	C <sub>9</sub>	H <sub>9</sub>	120	376.56
N <sub>10</sub>	C <sub>1*</sub>	C <sub>2*</sub>	111	460.24		C <sub>1*</sub>	C <sub>2*</sub>	O <sub>2*</sub>	109.5	460.24
C <sub>1*</sub>	C <sub>2*</sub>	C <sub>3*</sub>	111	460.24		O <sub>2*</sub>	C <sub>2*</sub>	C <sub>3*</sub>	109.5	460.24
C <sub>2*</sub>	O <sub>2*</sub>	H <sub>2*</sub>	109.5	397.48		C <sub>2*</sub>	C <sub>3*</sub>	O <sub>3*</sub>	109.5	460.24
C <sub>2*</sub>	C <sub>3*</sub>	C <sub>4*</sub>	111	460.24		O <sub>3*</sub>	C <sub>3*</sub>	C <sub>4*</sub>	109.5	460.24
C <sub>3*</sub>	O <sub>3*</sub>	H <sub>3*</sub>	109.5	397.48		C <sub>3*</sub>	C <sub>4*</sub>	O <sub>4*</sub>	109.5	460.24
C <sub>3*</sub>	C <sub>4*</sub>	C <sub>5*</sub>	111	460.24		O <sub>4*</sub>	C <sub>4*</sub>	C <sub>5*</sub>	109.5	460.24
C <sub>4*</sub>	O <sub>4*</sub>	H <sub>4*</sub>	109.5	397.48		C <sub>4*</sub>	C <sub>5*</sub>	O <sub>5*</sub>	111	460.24
C <sub>5*</sub>	O <sub>5*</sub>	P	120	397.48		O <sub>5*</sub>	P	OP <sub>3</sub>	103	397.48
O <sub>5*</sub>	P	OP <sub>1</sub>	109.6	397.48		O <sub>5*</sub>	P	OP <sub>2</sub>	109.6	397.48
OP <sub>3</sub>	P	OP <sub>1</sub>	109.6	397.48		OP <sub>3</sub>	P	OP <sub>2</sub>	109.6	397.48
OP <sub>1</sub>	P	OP <sub>2</sub>	120	585.76		P	OP <sub>3</sub>	<sup>A</sup> P	120	397.48
OP <sub>3</sub>	<sup>A</sup> P	<sup>A</sup> OP <sub>1</sub>	109.6	397.48		OP <sub>3</sub>	<sup>A</sup> P	<sup>A</sup> OP <sub>2</sub>	109.6	397.48
OP <sub>3</sub>	<sup>A</sup> P	<sup>A</sup> O <sub>5*</sub>	103	397.48		<sup>A</sup> OP <sub>1</sub>	<sup>A</sup> P	<sup>A</sup> OP <sub>2</sub>	120	585.76
<sup>A</sup> OP <sub>1</sub>	<sup>A</sup> P	<sup>A</sup> O <sub>5*</sub>	109.6	397.48		<sup>A</sup> OP <sub>2</sub>	<sup>A</sup> P	<sup>A</sup> O <sub>5*</sub>	109.6	397.48
<sup>A</sup> P	<sup>A</sup> O <sub>5*</sub>	<sup>A</sup> C <sub>5*</sub>	120	397.48		<sup>A</sup> O <sub>5*</sub>	<sup>A</sup> C <sub>5*</sub>	<sup>A</sup> C <sub>4*</sub>	109.5	284.512
<sup>A</sup> C <sub>5*</sub>	<sup>A</sup> C <sub>4*</sub>	<sup>A</sup> O <sub>4*</sub>	109.5	284.512		<sup>A</sup> C <sub>5*</sub>	<sup>A</sup> C <sub>4*</sub>	<sup>A</sup> C <sub>3*</sub>	109.5	251.04
<sup>A</sup> O <sub>4*</sub>	<sup>A</sup> C <sub>4*</sub>	<sup>A</sup> C <sub>3*</sub>	109.5	284.512		<sup>A</sup> C <sub>4*</sub>	<sup>A</sup> O <sub>4*</sub>	<sup>A</sup> C <sub>1*</sub>	109.5	334.72
<sup>A</sup> O <sub>4*</sub>	<sup>A</sup> C <sub>1*</sub>	<sup>A</sup> N <sub>9</sub>	109.5	284.512		<sup>A</sup> O <sub>4*</sub>	<sup>A</sup> C <sub>1*</sub>	<sup>A</sup> C <sub>2*</sub>	109.5	284.512
<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>1*</sub>	<sup>A</sup> C <sub>2*</sub>	109.5	284.512		<sup>A</sup> C <sub>1*</sub>	<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>4</sub>	126	418.4
<sup>A</sup> C <sub>1*</sub>	<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>8</sub>	126	418.4		<sup>A</sup> C <sub>4</sub>	<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>8</sub>	108	418.4
<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> N <sub>3</sub>	132	418.4		<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>5</sub>	108	418.4
<sup>A</sup> N <sub>3</sub>	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>5</sub>	120	418.4		<sup>A</sup> C <sub>4</sub>	<sup>A</sup> N <sub>3</sub>	<sup>A</sup> C <sub>2</sub>	120	418.4
<sup>A</sup> N <sub>3</sub>	<sup>A</sup> C <sub>2</sub>	H <sub>C2</sub>	120	376.56		<sup>A</sup> N <sub>3</sub>	<sup>A</sup> C <sub>2</sub>	<sup>A</sup> N <sub>1</sub>	120	418.4
H <sub>C2</sub>	<sup>A</sup> C <sub>2</sub>	<sup>A</sup> N <sub>1</sub>	120	376.56		<sup>A</sup> C <sub>2</sub>	<sup>A</sup> N <sub>1</sub>	<sup>A</sup> C <sub>6</sub>	120	418.4
<sup>A</sup> N <sub>1</sub>	<sup>A</sup> C <sub>6</sub>	<sup>A</sup> N <sub>6</sub>	120	418.4		<sup>A</sup> N <sub>1</sub>	<sup>A</sup> C <sub>6</sub>	<sup>A</sup> C <sub>5</sub>	120	418.4
<sup>A</sup> N <sub>6</sub>	<sup>A</sup> C <sub>6</sub>	<sup>A</sup> C <sub>5</sub>	120	418.4		<sup>A</sup> C <sub>6</sub>	<sup>A</sup> N <sub>6</sub>	H <sub>N61</sub>	120	292.88
<sup>A</sup> C <sub>6</sub>	<sup>A</sup> N <sub>6</sub>	H <sub>N62</sub>	120	292.88		H <sub>N61</sub>	<sup>A</sup> N <sub>6</sub>	H <sub>N62</sub>	120	334.72
<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> C <sub>6</sub>	120	418.4		<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> N <sub>7</sub>	108	418.4
<sup>A</sup> C <sub>6</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> N <sub>7</sub>	108	418.4		<sup>A</sup> C <sub>5</sub>	<sup>A</sup> N <sub>7</sub>	<sup>A</sup> C <sub>8</sub>	108	418.4
<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>8</sub>	<sup>A</sup> N <sub>7</sub>	108	418.4		<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>8</sub>	H <sub>C8</sub>	120	376.56
<sup>A</sup> N <sub>7</sub>	<sup>A</sup> C <sub>8</sub>	H <sub>C8</sub>	120	376.56		<sup>A</sup> C <sub>1*</sub>	<sup>A</sup> C <sub>2*</sub>	<sup>A</sup> O <sub>2*</sub>	109.5	284.512
<sup>A</sup> C <sub>1*</sub>	<sup>A</sup> C <sub>2*</sub>	<sup>A</sup> C <sub>3*</sub>	109.5	251.04		<sup>A</sup> O <sub>2*</sub>	<sup>A</sup> C <sub>2*</sub>	<sup>A</sup> C <sub>3*</sub>	109.5	284.512
<sup>A</sup> C <sub>2*</sub>	<sup>A</sup> O <sub>2*</sub>	H <sub>O2*</sub>	109.5	397.48		<sup>A</sup> C <sub>4*</sub>	<sup>A</sup> C <sub>3*</sub>	<sup>A</sup> C <sub>2*</sub>	109.5	251.04
<sup>A</sup> C <sub>4*</sub>	<sup>A</sup> C <sub>3*</sub>	<sup>A</sup> O <sub>3*</sub>	109.5	284.512		<sup>A</sup> C <sub>2*</sub>	<sup>A</sup> C <sub>3*</sub>	<sup>A</sup> O <sub>3*</sub>	109.5	284.512
<sup>A</sup> C <sub>3*</sub>	<sup>A</sup> O <sub>3*</sub>	H <sub>O3*</sub>	109.5	397.48						

## B.5 Proper Dihedrals

$a_i$	$a_j$	$a_k$	$a_l$	$\theta_{max}^{\S}$ (degr)	$f_c$ (kJ mol <sup>-1</sup> )	mult <sup>¶</sup> (#)
C <sub>9A</sub>	N <sub>10</sub>	C <sub>1*</sub>	C <sub>1*</sub>	0	0.418	6
N <sub>10</sub>	C <sub>1*</sub>	C <sub>2*</sub>	C <sub>2*</sub>	0	5.858	3
C <sub>1*</sub>	C <sub>2*</sub>	O <sub>2*</sub>	O <sub>2*</sub>	0	1.255	3
C <sub>1*</sub>	C <sub>2*</sub>	C <sub>3*</sub>	C <sub>3*</sub>	0	5.858	3
C <sub>2*</sub>	C <sub>3*</sub>	O <sub>3*</sub>	O <sub>3*</sub>	0	1.255	3
C <sub>2*</sub>	C <sub>3*</sub>	C <sub>4*</sub>	C <sub>4*</sub>	0	5.858	3
C <sub>3*</sub>	C <sub>4*</sub>	O <sub>4*</sub>	O <sub>4*</sub>	0	1.255	3
C <sub>3*</sub>	C <sub>4*</sub>	C <sub>5*</sub>	C <sub>5*</sub>	0	5.858	3
C <sub>4*</sub>	C <sub>5*</sub>	O <sub>5*</sub>	O <sub>5*</sub>	0	3.766	3
C <sub>5*</sub>	O <sub>5*</sub>	P	P	0	3.138	2
O <sub>5*</sub>	P	O <sub>P3</sub>	O <sub>P3</sub>	0	3.138	2
P	O <sub>P3</sub>	AP	AP	0	3.138	2
O <sub>P3</sub>	AP	AO <sub>5*</sub>	AO <sub>5*</sub>	0	3.138	2
AP	AO <sub>5*</sub>	AC <sub>5*</sub>	AC <sub>5*</sub>	0	3.766	3
AO <sub>5*</sub>	AC <sub>5*</sub>	AC <sub>4*</sub>	AC <sub>4*</sub>	0	2.092	2
AC <sub>5*</sub>	AC <sub>4*</sub>	AO <sub>4*</sub>	AO <sub>4*</sub>	0	3.766	3
AC <sub>5*</sub>	AC <sub>4*</sub>	AC <sub>3*</sub>	AC <sub>3*</sub>	0	2.092	2
AC <sub>4*</sub>	AO <sub>4*</sub>	AC <sub>1*</sub>	AC <sub>1*</sub>	0	3.766	3
AO <sub>4*</sub>	AC <sub>1*</sub>	AN <sub>9</sub>	AN <sub>9</sub>	0	0	2
AO <sub>4*</sub>	AC <sub>1*</sub>	AC <sub>2*</sub>	AC <sub>2*</sub>	0	2.092	2
AN <sub>1</sub>	AC <sub>6</sub>	AN <sub>6</sub>	AN <sub>6</sub>	180	33.472	2
AC <sub>1*</sub>	AC <sub>2*</sub>	AO <sub>2*</sub>	AO <sub>2*</sub>	0	1.255	3
AC <sub>1*</sub>	AC <sub>2*</sub>	AC <sub>3*</sub>	AC <sub>3*</sub>	0	2.092	2
AC <sub>4*</sub>	AC <sub>3*</sub>	AO <sub>3*</sub>	AO <sub>3*</sub>	0	1.255	3

## B.6 Improper Dihedrals

$a_i$	$a_j$	$a_k$	$a_l$	$\theta_0^{\parallel}$ (degr)	$f_c$ (kJ mol <sup>-1</sup> rad <sup>-2</sup> )	$a_i$	$a_j$	$a_k$	$a_l$	$\theta_0^{\parallel}$ (degr)	$f_c$ (kJ mol <sup>-1</sup> rad <sup>-2</sup> )
C <sub>9A</sub>	N <sub>10</sub>	C <sub>10</sub>	C <sub>10</sub>	0	167.36	C <sub>9A</sub>	N <sub>10</sub>	C <sub>9</sub>	C <sub>9</sub>	0	167.36
C <sub>9A</sub>	C <sub>5A</sub>	C <sub>6</sub>	C <sub>6</sub>	0	167.36	N <sub>10</sub>	C <sub>9A</sub>	C <sub>5A</sub>	C <sub>5A</sub>	0	167.36
N <sub>10</sub>	C <sub>10</sub>	C <sub>4A</sub>	C <sub>4A</sub>	0	167.36	N <sub>10</sub>	C <sub>10</sub>	C <sub>9A</sub>	C <sub>9A</sub>	0	167.36
C <sub>10</sub>	N <sub>1</sub>	C <sub>2</sub>	C <sub>2</sub>	0	167.36	C <sub>10</sub>	N <sub>1</sub>	N <sub>10</sub>	N <sub>10</sub>	0	167.36
C <sub>10</sub>	C <sub>4A</sub>	N <sub>5</sub>	N <sub>5</sub>	0	167.36	N <sub>1</sub>	C <sub>10</sub>	C <sub>4A</sub>	C <sub>4A</sub>	0	167.36
N <sub>1</sub>	C <sub>2</sub>	N <sub>3</sub>	N <sub>3</sub>	0	167.36	C <sub>2</sub>	N <sub>1</sub>	N <sub>3</sub>	N <sub>3</sub>	0	167.36
C <sub>2</sub>	N <sub>3</sub>	C <sub>4</sub>	C <sub>4</sub>	0	167.36	N <sub>3</sub>	C <sub>4</sub>	C <sub>4A</sub>	C <sub>4A</sub>	0	167.36
N <sub>3</sub>	C <sub>2</sub>	C <sub>4</sub>	C <sub>4</sub>	0	167.36	C <sub>4</sub>	C <sub>4A</sub>	C <sub>10</sub>	C <sub>10</sub>	0	167.36
C <sub>4</sub>	N <sub>3</sub>	C <sub>4A</sub>	C <sub>4A</sub>	0	167.36	C <sub>4A</sub>	N <sub>5</sub>	C <sub>5A</sub>	C <sub>5A</sub>	0	167.36
C <sub>4A</sub>	C <sub>4</sub>	N <sub>5</sub>	N <sub>5</sub>	0	167.36	C <sub>4A</sub>	C <sub>10</sub>	N <sub>1</sub>	N <sub>1</sub>	0	167.36
N <sub>5</sub>	C <sub>5A</sub>	C <sub>9A</sub>	C <sub>9A</sub>	0	167.36	C <sub>5A</sub>	N <sub>5</sub>	C <sub>6</sub>	C <sub>6</sub>	0	167.36
C <sub>5A</sub>	C <sub>9A</sub>	N <sub>10</sub>	N <sub>10</sub>	0	167.36	C <sub>5A</sub>	C <sub>9A</sub>	C <sub>9</sub>	C <sub>9</sub>	0	167.36

(continued ...)

<sup>§</sup>  $\theta$  is the angle between the planes  $i-j-k$  and  $j-k-l$ ;  $\theta = 0$  corresponds to the *cis* conformation with  $a_i$  and  $a_l$  on the same side. <sup>¶</sup> potential multiplicity <sup>||</sup>  $\theta$  is the angle between the planes  $i-j-k$  and  $j-k-l$ .

## Improper Dihedrals (... continued)

$a_i$	$a_j$	$a_k$	$a_l$	$\theta_0^{\parallel}$ (degr)	$f_c$ (kJ mol <sup>-1</sup> rad <sup>-2</sup> )	$a_i$	$a_j$	$a_k$	$a_l$	$\theta_0^{\parallel}$ (degr)	$f_c$ (kJ mol <sup>-1</sup> rad <sup>-2</sup> )
C <sub>5A</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>7</sub>	0	167.36	C <sub>6</sub>	C <sub>5A</sub>	C <sub>7</sub>	C <sub>7</sub>	0	167.36
C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>8</sub>	0	167.36	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>9</sub>	0	167.36
C <sub>7</sub>	C <sub>6</sub>	C <sub>8</sub>	C <sub>8</sub>	0	167.36	C <sub>8</sub>	C <sub>9</sub>	C <sub>9A</sub>	C <sub>9A</sub>	0	167.36
C <sub>8</sub>	C <sub>7</sub>	C <sub>9</sub>	C <sub>9</sub>	0	167.36	C <sub>9</sub>	C <sub>9A</sub>	C <sub>5A</sub>	C <sub>5A</sub>	0	167.36
C <sub>9</sub>	C <sub>9A</sub>	H <sub>9</sub>	H <sub>9</sub>	0	167.36	C <sub>2*</sub>	O <sub>2*</sub>	C <sub>3*</sub>	C <sub>3*</sub>	35.264	334.72
C <sub>3*</sub>	O <sub>3*</sub>	C <sub>4*</sub>	C <sub>4*</sub>	35.264	334.72	C <sub>4*</sub>	O <sub>4*</sub>	C <sub>5*</sub>	C <sub>5*</sub>	35.264	334.72
<sup>A</sup> C <sub>4*</sub>	<sup>A</sup> O <sub>4*</sub>	<sup>A</sup> C <sub>5*</sub>	<sup>A</sup> C <sub>5*</sub>	35.264	334.72	<sup>A</sup> C <sub>1*</sub>	<sup>A</sup> N <sub>9</sub>	<sup>A</sup> O <sub>4*</sub>	<sup>A</sup> O <sub>4*</sub>	35.264	334.72
<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>8</sub>	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>4</sub>	0	167.36	<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>8</sub>	<sup>A</sup> N <sub>7</sub>	<sup>A</sup> N <sub>7</sub>	0	167.36
<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> C <sub>5</sub>	0	167.36	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> N <sub>3</sub>	<sup>A</sup> C <sub>2</sub>	<sup>A</sup> C <sub>2</sub>	0	167.36
<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> C <sub>6</sub>	<sup>A</sup> C <sub>6</sub>	0	167.36	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> N <sub>9</sub>	<sup>A</sup> N <sub>3</sub>	<sup>A</sup> N <sub>3</sub>	0	167.36
<sup>A</sup> C <sub>4</sub>	<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>8</sub>	<sup>A</sup> C <sub>8</sub>	0	167.36	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> N <sub>7</sub>	<sup>A</sup> N <sub>7</sub>	0	167.36
<sup>A</sup> N <sub>3</sub>	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> C <sub>5</sub>	0	167.36	<sup>A</sup> N <sub>3</sub>	<sup>A</sup> C <sub>2</sub>	<sup>A</sup> N <sub>1</sub>	<sup>A</sup> N <sub>1</sub>	0	167.36
<sup>A</sup> C <sub>2</sub>	<sup>A</sup> N <sub>1</sub>	H <sub>C2</sub>	H <sub>C2</sub>	0	167.36	<sup>A</sup> C <sub>2</sub>	<sup>A</sup> N <sub>1</sub>	<sup>A</sup> C <sub>6</sub>	<sup>A</sup> C <sub>6</sub>	0	167.36
<sup>A</sup> N <sub>1</sub>	<sup>A</sup> C <sub>6</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> C <sub>5</sub>	0	167.36	<sup>A</sup> C <sub>6</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> N <sub>1</sub>	<sup>A</sup> N <sub>1</sub>	0	167.36
<sup>A</sup> N <sub>6</sub>	H <sub>N61</sub>	H <sub>N62</sub>	H <sub>N62</sub>	0	167.36	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> N <sub>7</sub>	<sup>A</sup> C <sub>8</sub>	<sup>A</sup> C <sub>8</sub>	0	167.36
<sup>A</sup> C <sub>5</sub>	<sup>A</sup> C <sub>6</sub>	<sup>A</sup> N <sub>7</sub>	<sup>A</sup> N <sub>7</sub>	0	167.36	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> N <sub>3</sub>	<sup>A</sup> N <sub>3</sub>	0	167.36
<sup>A</sup> C <sub>5</sub>	<sup>A</sup> C <sub>6</sub>	<sup>A</sup> N <sub>1</sub>	<sup>A</sup> N <sub>1</sub>	0	167.36	<sup>A</sup> C <sub>8</sub>	<sup>A</sup> N <sub>7</sub>	<sup>A</sup> C <sub>5</sub>	<sup>A</sup> C <sub>5</sub>	0	167.36
<sup>A</sup> C <sub>8</sub>	<sup>A</sup> N <sub>9</sub>	<sup>A</sup> C <sub>4</sub>	<sup>A</sup> C <sub>4</sub>	0	167.36	<sup>A</sup> C <sub>8</sub>	<sup>A</sup> N <sub>9</sub>	H <sub>C8</sub>	H <sub>C8</sub>	0	167.36
<sup>A</sup> C <sub>2*</sub>	<sup>A</sup> O <sub>2*</sub>	<sup>A</sup> C <sub>3*</sub>	<sup>A</sup> C <sub>3*</sub>	35.264	334.72	<sup>A</sup> C <sub>3*</sub>	<sup>A</sup> C <sub>2*</sub>	<sup>A</sup> O <sub>3*</sub>	<sup>A</sup> O <sub>3*</sub>	35.264	334.72



# Bibliography

- [1] Feenstra, K. A., Hess, B., Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comp. Chem.* 20:786–798, 1999.
- [2] Feenstra, K. A., Berendsen, H. J. C., Mark, A. E. Analysis of the hierarchy of motion of small proteins with implications for protein folding. *Proteins: Struct. Funct. Gen.* (submitted Okt. 2001).
- [3] Kraulis, P. J. MOLSCRIPT a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 24:946–950, 1991.
- [4] Esnouf, R. M. An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* 15:132–134, 1997.
- [5] Bacon, D. J., Anderson, W. F. A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics* 6:219–220, 1988.
- [6] Merritt, E. A., Murphy, M. E. P. Raster3D version 2.0: A program for photorealistic molecular graphics. *Act. Cryst. D.* 50:869–873, 1994.
- [7] Feenstra, K. A., Berendsen, H. J. C. The domain decomposition of a single-domain protein. In *Monte Carlo approach to biopolymers and protein folding* (London, 1998). Grassberger, P., Barkema, G. T., Nadler, W., eds. Höchstleistungsrechenzentrum Jülich, Germany. World Scientific. 255–267. ISBN 981-02-3658-1.
- [8] Mazur, A. K. Common molecular dynamics algorithms revisited: Accuracy and optimal time steps of störmmer-leapfrog integrators. *J. Comp. Phys.* 136:354–365, 1997.
- [9] Mazur, A. K. Hierarchy of fast motions in protein dynamics. *J. Phys. Chem.* 102:473–479, 1998.
- [10] Hess, B., Bekker, H., Berendsen, H. J. C., Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* 18:1463–1472, 1997.
- [11] Berendsen, H. J. C. Molecular dynamics simulations: The limits and beyond. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*. Deuffhard, P., Hermans, J., Leimkuhler, B., Mark, A., Reich, S., Skeel, R. D., eds. Springer-Verlag. ISBN 3-540-63242-5.
- [12] Bennet, C. M. Mass tensor molecular dynamics. *J. Comp. Phys.* 19:267–279, 1975.
- [13] Berendsen, H. J. C., van Gunsteren, W. F. Molecular dynamics simulations: Techniques and approaches. In *Molecular Liquids-Dynamics and Interactions* (Dordrecht, The Netherlands, 1984). Barnes, A. J. *et al.* eds. NATO ASI C 135. Reidel. 475–500.
- [14] Jacucci, G., Rahman, A. The possibility of using a larger timestep in molecular dynamics studies of water. In *Report on Workshop Methods in Molecular Dynamics - Long Timescale Events* (Orsay, August 1974). C.E.C.A.M. 32–40.
- [15] Wood, D. W. Computer simulation of water and aqueous solutions. In: *Water: A Comprehensive Treatise*.

- Vol. 6. Franks, F. ed. Vol. 6. Plenum Press New York 1979 279.
- [16] Pomès, R., McCammon, J. A. Mass and step length optimization for the calculation of equilibrium properties by molecular dynamics simulations. *Chem. Phys. Lett.* 166:425–428, 1990.
- [17] Egberts, E., Marrink, S. J., Berendsen, H. J. C. Molecular dynamics simulation of a phospholipid membrane. *Eur. Biophys. J.* 22:423–436, 1994.
- [18] Mao, B., Maggiora, G. M., Chou, K. C. Mass-weighted molecular dynamics simulation of cyclic polypeptides. *Biopolymers* 31:1077–1086, 1991.
- [19] Verlet, L. *Phys. Rev.* 34:1311–1327, 1967.
- [20] van Gunsteren, W. F., Berendsen, H. J. C. GROMOS-87 manual. Biomos BV Nijenborgh 4, 9747 AG Groningen, The Netherlands 1987.
- [21] van Buuren, A. R., Marrink, S. J., Berendsen, H. J. C. A molecular dynamics study of the decane/water interface. *J. Phys. Chem.* 97:9206–9212, 1993.
- [22] Daura, X., Oliva, B., Querol, E., Avilés, F. X., Tapia, O. On the sensitivity of MD trajectories to changes in water-protein interaction parameters: The potato carboxypeptidase inhibitor in water as a test case for the GROMOS force field. *Proteins: Struct. Funct. Gen.* 25:89–103, 1996.
- [23] Miyamoto, S., Kollman, P. A. SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comp. Chem.* 13:952–962, 1992.
- [24] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690, 1984.
- [25] Lide, D. R., ed. *CRC Handbook of chemistry and physics : a ready-reference book of chemical and physical data.* 72 Ed. Boca Raton [etc.]: CRC Press. 1991.
- [26] Montrose, C. J., Bucaro, J. A., Marshall-Coakley, J., Litovitz, T. A. Depolarized Rayleigh scattering and hydrogen bonding in liquid water. *J. Chem. Phys.* 60:5025–5029, 1974.
- [27] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces* (Dordrecht, 1981). Pullman, B., ed. D. Reidel Publishing Company. 331–342.
- [28] Lau, K. F., Alpher, H. E., Thacher, T. S., Stouch, T. R. Effects of switching functions on the behaviour of liquid water in molecular dynamics simulations. *J. Phys. Chem.* 98:8785–8792, 1994.
- [29] Smith, P. E., Pettitt, B. M. Peptides in ionic solutions: A comparison of the Ewald and switching function techniques. *J. Chem. Phys.* 95:8430–8441, 1991.
- [30] Gabdoulline, R. R., Zheng, C. Effective charges for macromolecules in solvent. *J. Chem. Phys.* 16:1428, 1996.
- [31] van der Spoel, D., van Maaren, P. J., Berendsen, H. J. C. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *J. Chem. Phys.* 108:10220–10230, 1998.
- [32] van Nuland, N., Hangyi, I. W., van Schaik, R. C., Berendsen, H. J. C., van Gunsteren, W. F., Scheek, R. M., Robillard, G. T. The high-resolution structure of the histidine-containing phosphocarrier protein HPr from *Escherichia coli* determined by restrained molecular dynamics from Nmr Nuclear Overhauser Effect data. *J. Mol. Biol.* 237:544–559, 1994.

- [33] Hockney, R. W., Eastwood, J. W. *Computer simulation using particles*. New York: McGraw-Hill. 1981.
- [34] Luty, B. A., Davies, M. E., Tironi, I. G., van Gunsteren, W. F. A comparison of particle-particle, particle-mesh and Ewald methods for calculating electrostatic interactions in periodic molecular systems. *Mol. Sim.* 14:11–20, 1994.
- [35] Berendsen, H. J. C., van der Spoel, D., van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* 91:43–56, 1995.
- [36] van der Spoel, D., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., van Drunen, R., Berendsen, H. J. C. *GROMACS User Manual version 1.5*. Nijenborgh 4, 9747 AG Groningen, The Netherlands. Internet: <http://md.chem.rug.nl/~gmx> 1997.
- [37] Allen, M. P., Tildesley, D. J. *Computer Simulations of Liquids*. Oxford: Oxford Science Publications. 1987.
- [38] Berendsen, H. J. C. Transport properties computed by linear response through weak coupling to a bath. In *Computer Simulations in Material Science*. Meyer, M., Pontikis, V., eds. Kluwer. 139–155.
- [39] Kabsch, W., Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biophys. J.* 22:2577–2637, 1983.
- [40] van Gunsteren, W. F., Berendsen, H. J. C. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 34:1311–1327, 1977.
- [41] Eaton, W. A., Noz, V. M., Thompson, P. A., Henry, E. R., Hofrichter, J. Kinetics and dynamics of loops,  $\alpha$ -helices,  $\beta$ -hairpins and fast-folding proteins. *Acc. Chem. Res.* 31:745–753, 1998.
- [42] Jamin, M., Yeh, S.-R., Rousseau, D. L., Baldwin, R. L. Submillisecond unfolding kinetics of apomyoglobin and its pH 4 intermediate. *J. Mol. Biol.* 292:731–740, 1999.
- [43] Daura, X., Jaun, B., Seebach, D., van Gunsteren, W. F., Mark, A. E. Reversible peptide folding in solution by molecular dynamics simulation. *J. Mol. Biol.* 280:925–932, 1998.
- [44] Duan, Y., Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282:740–744, 1998.
- [45] Fersht, A. R. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* 7:3–9, 1997.
- [46] Ptitsyn, O. B. Protein folding and protein evolution: Common folding nucleus in different subfamilies of c-type cytochromes? *J. Mol. Biol.* 278:655–666, 1998.
- [47] Poupon, A., Morion, J.-P. Predicting the protein folding nucleus from a sequence. *FEBS Lett.* 452:283–289, 1999.
- [48] Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E., Shakhnovich, E. I. Identifying the protein folding nucleus using molecular dynamics. *J. Mol. Biol.* 296:1183–1188, 2000.
- [49] Panchenko, A. R., Luthey-Schulten, Z., Wolynes, P. G. Foldons as independently folding units of proteins. *Physica* 107(2-4):312–315, 1997.
- [50] Furois-Corbin, S., Smith, J. C., Kneller, G. R. Picosecond timescale rigid-helix and side-chain motions in deoxymyoglobin. *Proteins: Struct. Funct. Gen.* 16:141–154, 1993.
- [51] Bruscolini, P. A coarse-grained, realistic model for protein folding. *J. Chem. Phys.* 107:7512–7529, 1997.
- [52] Bruscolini, P. Testing the helix model for protein folding on four simple

- proteins. *Modern Phys. Lett. B* 11(16-17):691–702, 1997.
- [53] Fischer, K. F., Marqusee, S. A rapid test for identification of autonomous folding units in proteins. *J. Mol. Biol.* 302:701–712, 2000.
- [54] Yue, K., Dill, K. A. Constraint-based assembly of tertiary protein structures from secondary structure elements. *Prot. Sci.* 9:1935–1946, 2000.
- [55] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. The protein data bank. *Nucl. Acids Res.* 28:235–242, 2000.
- [56] Noble, M. E., Musacchio, A., Saraste, M., Courtneidge, S. A., Wierenga, R. K. Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *EMBO J.* 12:2617, 1993.
- [57] Morton, C. J., Pugh, D. J., Brown, E. L., Kahmann, J. D., Renzoni, D. A., Campbell, I. D. Solution structure and peptide binding of the SH3 domain from human Fyn. *Structure* 4:705, 1996.
- [58] Weaver, L. H., Matthews, B. W. Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* 193:189, 1987.
- [59] Jia, Z., Quail, J. W., Waygood, E. B., Delbaere, L. T. J. The 2.0 Å-resolution structure of *Escherichia coli* histidine-containing phosphocarrier protein HPr. A redermination. *J. Biol. Chem.* 268:22490–22501, 1993.
- [60] Kovacs, H., Comfort, D., Lord, M., Campbell, I. D., Yudkin, M. D. Solution structure of spoIIAA, a phosphorylatable component of the system that regulates transcription factor  $\sigma^F$  of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 95:5067–5071, 1998.
- [61] Fucini, P., Renner, C., Herberhold, C., Noegel, A. A., Holak, T. A. The repeating segments of the f-actin cross-linking gelation factor (abp-120) have an immunoglobulin-like fold. *Nature Struct. Biol.* 4:223, 1997.
- [62] Hess, B. Convergence of sampling in protein simulations. *Phys. Rev. E* 65:031910, 2002.
- [63] van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P., Tironi, I. G. Biomolecular simulation: GROMOS96 manual and user guide. BIOMOS b.v. Zürich, Groningen 1996.
- [64] van der Spoel, D., Hess, B., Feenstra, K. A., Lindahl, E., Berendsen, H. J. C. GROMACS User Manual version 2.0. Nijenborgh 4, 9747 AG Groningen, The Netherlands. Internet: <http://md.chem.rug.nl/~gmx> 1999.
- [65] Hayward, S., Kitao, A., Berendsen, H. J. C. Model-free methods of analyzing domain motions in proteins from simulation: A comparison of a normal mode analysis and a molecular dynamics simulation of lysozyme. *Proteins: Struct. Funct. Gen.* 27:425–437, 1997.
- [66] Hayward, S., Berendsen, H. J. C. Systematic analysis of domain motions in proteins conformational change; new results on citrate synthase and T4 lysozyme. *Proteins: Struct. Funct. Gen.* 30:144–154, 1998.
- [67] Amadei, A., Linssen, A. B. M., Berendsen, H. J. C. Essential dynamics of proteins. *Proteins: Struct. Funct. Gen.* 17:412–425, 1993.
- [68] Hayward, S. Structural principles governing domain motions in proteins. *Proteins: Struct. Funct. Gen.* 36:425–435, 1999.
- [69] de Groot, B. L., Hayward, S., van Aalten, D. M. F., Amadei, A., Berendsen, H. J. C. Domain motions in bacteriophage t4 lysozyme; a

- comparison between molecular dynamics and crystallographic data. *Proteins: Struct. Funct. Gen.* 31:116–127, 1998.
- [70] Taylor, W. R. Defining linear segments in protein structure. *J. Mol. Biol.* 310:1135–1150, 2001.
- [71] Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., Baker, D. Experiment and theory highlight role of native state topology in SH3 folding. *Nature Struct. Biol.* 6:1016–1024, 1999.
- [72] Hilser, V. J., Townsend, B. D., Freire, E. Structure-based statistical thermodynamic analysis of T4 lysozyme mutants: Structural mapping of cooperative interactions. *Bioph. Chem.* 64:69–79, 1997.
- [73] Lu, J., Dahlquist, F. W. Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry* 31:4749–4756, 1992.
- [74] Najbar, L. V., Craik, D. J., Wade, J. D., McLeish, M. J. Identification of initiation sites for T4 lysozyme folding using CD and NMR spectroscopy of peptide fragments. *Biochemistry* 39:5911–5920, 2000.
- [75] Chen, L., Hodgson, K. O., Doniach, S. A lysozyme folding intermediate revealed by solution x-ray scattering. *J. Mol. Biol.* 261:658–671, 1996.
- [76] Vriend, G. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8:52–56, 1990.
- [77] Maiorov, V. N., Crippen, G. M. Size-independent comparison of protein three-dimensional structures. *Proteins: Struct. Funct. Gen.* 22:273–283, 1995.
- [78] Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., Mark, A. E. Peptide folding: When simulation meets experiment. *Angew. Chem. Intl. Ed.* 38(1/2):236–240, 1999.
- [79] Jardetzki, O. Nature of molecular-conformations inferred from high-resolution NMR. *Biochim. Biophys. Acta* 621:227–232, 1980.
- [80] Tropp, J. Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances. *J. Chem. Phys.* 72:6035–6043, 1980.
- [81] Lipari, G., Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559, 1982.
- [82] Daura, X., Antes, I., van Gunsteren, W. F., Thiel, W., Mark, A. E. The effect of motional averaging on the calculation of NMR-derived structural properties. *Proteins* 36:542–555, 1999.
- [83] Bürgi, R., Pitera, J., van Gunsteren, W. F. Assessing the effect of conformational averaging on the measured values of observables. *J. Biomol. NMR* 19:305–320, 2001.
- [84] Peter, C., Daura, X., van Gunsteren, W. F. Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations. *J. Biomol. NMR* 20:297–310, 2001.
- [85] Daura, X., Mark, A. E., van Gunsteren, W. F. Parametrisation of aliphatic CH<sub>n</sub> united atoms of GROMOS96 force field. *J. Comp. Chem.* 19:535–547, 1998.
- [86] Bax, A., Davis, D. G. An improved method for two-dimensional heteronuclear relayed-coherence-transfer NMR spectroscopy. *J. Magn. Reson.* 63:203–206, 1985.
- [87] Braunschweiler, L., Ernst, R. R. Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* 53:521–528, 1983.
- [88] Macura, S., Ernst, R. R. Elucidation of cross relaxation in liquids by

- two-dimensional NMR spectroscopy. *Mol. Phys.* 41:95–117, 1980.
- [89] Bax, A., Davis, D. G. Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn. Reson.* 63:207–213, 1985.
- [90] Bothner-By, A. A., Stephons, R. L., Lee, J., Warren, C. D., Jeanloz, R. W. Structure determination of a tetrasaccharide: Transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc.* 106:811–813, 1984.
- [91] Piotto, M., Sandek, V., Sklenár, V. Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2:661–665, 1992.
- [92] States, D. J., Haberkorn, R. A., Ruben, D. J. A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson.* 48:286–292, 1982.
- [93] van Hoesel, F. H. J. SNARF v. 0.8.9. University of Groningen Groningen 2000.
- [94] AB, E., Schuurman-Wolters, G. K., Nijlant, D., Dijkstra, K., Saier, M. H., Robillard, G. T., Scheek, R. M. NMR structure of cysteinyl-phosphorylated enzyme IIB of the N,N<sup>2</sup>-diacetylchitobiose-specific phosphoenolpyruvate-dependent phosphotransferase system of *Escherichia coli*. *J. Mol. Biol.* 308:933–1009, 2001.
- [95] van den Berg, P. A. W., Visser, A. J. W. G. Tracking molecular dynamics of flavoproteins with time-resolved fluorescence spectroscopy. In *New Trends in Fluorescence Spectroscopy. Applications to Chemical and Life Sciences* (Berlin, 2001). Valeur, B., Brochon, J. C., eds. Springer. 457–485.
- [96] van den Berg, P. A. W., Mulrooney, S. B., Gobets, B., van Stokkum, I. H. M., van Hoek, A., Williams, C. H., Visser, A. J. W. G. Exploring the conformational equilibrium of *E. coli* thioredoxin reductase: Characterization of two catalytically important states by ultra-fast flavin fluorescence spectroscopy. *Prot. Sci.* 10:2037–2049, 2001.
- [97] Song, P. S. Molecular orbital treatment of the intramolecular complexing in flavin adenine dinucleotide (FAD). In *Quantum aspects of heterocyclic compounds*. Bergmann, E. D., Pullman, B., eds. Israel Academy of Sciences. 358–384.
- [98] Leijonmarck, M. On the structures of flavin derivatives. crystallographic studies and a review. *Chem. Commun.* 8:1–70., 1977.
- [99] Weber, G. Fluorescence of riboflavin and flavin-adenine dinucleotide. *Biochem. J.* 47:114–121, 1950.
- [100] Spencer, R. D., Weber, G. Thermodynamics and kinetics of the intramolecular complex in flavin-adenine dinucleotide. In *Structure and Function of Oxidation Reduction Enzymes*. Åkeson, Å., Ehrenberg, A., eds. Pergamon. 393–399.
- [101] Penzer, G. R., Radda, G. K. The chemistry and biological functions of isoalloxazines (flavines). *Quart. Rev.* 21:43–65, 1967. London.
- [102] Voet, D., Rich, A. A hydrogen-bonded complex between riboflavin and an adenosine derivative, and its possible relation to FAD function. In *Flavins and Flavoproteins* (Baltimore, 1971). Kamin, H., ed. University Park Press. 23–35.
- [103] Voet, D., Rich, A. The crystal and molecular structure of an intermolecular complex between riboflavin and an adenosine derivative. *Proc. Natl. Acad. Sci. USA* 68:1151–1156, 1971.

- [104] Copeland, R. A., Spiro, T. G. Ultraviolet resonance Raman spectroscopy of flavin mononucleotide and flavin adenine dinucleotide. *J. Phys. Chem.* 90:6648–6654, 1986.
- [105] Sarma, R. H., Dannies, P., Kaplan, N. O. Investigations of inter- and intramolecular interactions in flavin-adenine dinucleotide by proton magnetic resonance. *Biochemistry* 7:4359–4367, 1968.
- [106] Kotowycz, G., Teng, N., Klein, M. P., Calvin, M. The 220 MHz nuclear magnetic resonance study of a solvent-induced conformational change in flavin adenine dinucleotide. *J. Biol. Chem.* 244:5656–5662, 1969.
- [107] Raszka, M., Kaplan, N. O. Intramolecular hydrogen bonding in flavin adenine dinucleotide. *Proc. Natl. Acad. Sci. USA* 71:4546–4550, 1974.
- [108] Visser, A. J. W. G. Kinetics of stacking interactions in flavin adenine dinucleotide from time-resolved fluorescence. *Photochem. Photobiol.* 40:703–706, 1984.
- [109] Ungar, L. W., Scherer, N. F., Voth, G. A. Classical molecular dynamics simulation of the photo-induced electron transfer dynamics of plastocyanin. *Biophys. J.* 72:5–17, 1997.
- [110] Whitby, L. G. A new method for preparing flavin-adenine dinucleotide. *Biochem. J.* 54:437–442, 1953.
- [111] van den Berg, P. A. W., van Hoek, A., Walentas, C. D., Perham, R. N., Visser, A. J. W. G. Flavin fluorescence dynamics and photoinduced electron transfer in *Escherichia coli* glutathione reductase. *Biophys. J.* 74:2046–2058, 1998.
- [112] Visser, A. J. W. G., van den Berg, P. A. W., Visser, N. V., van Hoek, A., van den Burg, H. A., Parsonage, D., Claiborne, A. Time-resolved fluorescence of flavin adenine dinucleotide in wild-type and mutant NADH peroxidase. Elucidation of quenching sites and discovery of a new fluorescence depolarization mechanism. *J. Phys. Chem.* 102:10431–10439, 1998.
- [113] Bastiaens, P. I. H., van Hoek, A., Wolkers, W. F., Brochon, J. C., Visser, A. J. W. G. Comparison of the dynamical structures of lipoamide dehydrogenase and glutathione reductase by time-resolved polarized flavin fluorescence. *Biochemistry* 31:7050–7060, 1992.
- [114] Digris, A. V., Skakoun, V. V., Novikov, E. G., van Hoek, A., Claiborne, A., Visser, A. J. W. G. Thermal stability of a flavoprotein assessed from associative analysis of polarized time-resolved fluorescence spectroscopy. *Eur. Biophys. J.* 28:526–531, 1999.
- [115] van der Spoel, D., de Groot, B. L., Hayward, S., Berendsen, H. J. C., Vogel, H. J. Bending of the calmodulin central helix: A theoretical study. *Prot. Sci.* 5:2044–2053, 1996.
- [116] Zheng, Y. J., Ornstein, R. L. A theoretical study of structures of flavin in different oxidation and protonation states. *J. Am. Chem. Soc.* 118:9402–9408, 1996.
- [117] Meyer, M., Hartwig, H., Schomburg, D. Semiempirical and *ab initio* study of closed and open shell derivatives of 10-methylisoalloxazine: a model for flavin redox states. *J. Mol. Struct.* 364:139–149, 1996.
- [118] Platenkamp, R. J., Palmer, M. H., Visser, A. J. W. G. *Ab initio* molecular orbital studies of flavin radicals and the lowest triplet state of isoalloxazine. *J. Mol. Struct.* 67:45–64, 1980.
- [119] Teitell, M. F., Suck, S. H., Fox, J. L. A MINDO/3 MO study of oxidized flavins. *Theor. Chim. Acta* 60:127–141, 1981.

- [120] Hall, L. H., Orchard, B. J., Tripathy, S. K. The structure and properties of flavins: molecular orbital study based on totally optimized geometries. II. Molecular orbital structure and electron distribution. *Int. J. Quantum Chem.* 31:217–242, 1987.
- [121] van der Spoel, D., Feenstra, K. A., Hemminga, M. A., Berendsen, H. J. C. Molecular modelling of the RNA binding N-terminal part of CCMV coat protein in solution with phosphate ions. *Biophys. J.* 71:2920–2932, 1996.
- [122] Mittl, P. R. E., Schulz, G. E. Structure of glutathione reductase from *Escherichia coli* at 1.86 Å resolution: Comparison with the enzyme from human erythrocytes. *Protein Sci.* 3:799–809, 1994.
- [123] Waksman, G., Krishna, T. S. R., C. H. Williams, J., Kuriyan, J. Crystal structure of *Escherichia coli* thioredoxin reductase refined at 2 Å resolution. Implications for a large conformational change during catalysis. *J. Mol. Biol.* 236:800–816, 1994.
- [124] Serre, L., Vellieux, F. M. D., Medina, M., Gomez-Moreno, C., Fontecilla-Camps, J. C., Frey, M. X-ray structure of the ferredoxin:NADP<sup>+</sup> reductase from the *Cyanobacterium anabaena* PCC 7119 at 1.8 Å resolution, and crystallographic studies of NADP<sup>+</sup> binding at 2.25 Å resolution. *J. Mol. Biol.* 263:20–39, 1996.
- [125] Berendsen, H. J. C., Grigera, J. R., Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* 91:6269–6271, 1987.
- [126] Hockney, R. W., Eastwood, J. W. Computer simulation using particles. Bristol, England: IOP Publishing Ltd.. 1988.
- [127] Luty, B. A., van Gunsteren, W. F. Calculating electrostatic interactions using the particle-particle, particle-mesh method with non-periodic long-range interactions. *J. Phys. Chem.* 100:2581–2587, 1996.
- [128] Bastiaens, P. I. H., van Hoek, A., Benen, J. A. E., Brochon, J. C., Visser, A. J. W. G. Conformational dynamics and intersubunit energy transfer in wild-type and mutant lipoamide dehydrogenase from *Azotobacter vinelandii*. *Biophys. J.* 63:839–853, 1992.
- [129] Mark, A. E. Free energy perturbation calculations. In: *Encyclopaedia of Computational Chemistry*. Vol. 2. von Rague Schleyer, P. ed. Vol. 2. John Wiley & Sons 1998 1070–1083.
- [130] Beechem, J. M. Global analysis of biochemical and biophysical data. *Methods Enzymol.* 210:37–54, 1992.
- [131] Heelis, P. F. The photochemistry of flavins. In *Chemistry and Biochemistry of Flavoenzymes* (Boca Raton, 1991). Müller, F., ed. Vol. I. CRC Press. 171–193.
- [132] Tsibris, J. C. M., McCormick, D. B., Wright, L. D. Studies on donor-acceptor complexes relating to the intramolecular association of the riboflavin and adenosine moieties of flavin-adenine dinucleotide. *Biochemistry* 4:504–509, 1965.
- [133] Karen, A., Ikeda, N., Tanaka, F., Mataga, N. Picosecond laser photolysis studies of fluorescence quenching mechanisms of flavin: a direct observation of indole-flavin singlet charge transfer state formation in solution and flavoenzymes. *Photochem. Photobiol* 37:495–502, 1983.
- [134] Karen, A., Sawada, M. T., Tanaka, F., Mataga, N. Dynamics of excited flavoproteins-picosecond laser photolysis studies. *Photochem. Photobiol.* 45:49–53, 1987.
- [135] Zhong, D., Zewail, A. H. Femtosecond dynamics of flavoproteins: charge separation and recombination in riboflavin (vitamin b2)-binding

- protein and in glucose oxidase enzyme. Proc. Natl. Acad. Sci. USA 98:11867–11872, 2001.
- [136] Mataga, N., Chrosrowjan, H., Shibata, Y., Tanaka, F. Ultrafast fluorescence quenching dynamics of flavin chromophores in protein nanospace. J. Phys. Chem. B 102:7081–7084, 1998.
- [137] Mataga, N., Chrosrowjan, H., Shibata, Y., Tanaka, F., Nishina, Y., Shiga, K. Dynamics and mechanisms of ultrafast fluorescence quenching reactions of flavin chromophores in protein nanospace. J. Phys. Chem. B 104:10667–10677, 2000.
- [138] Rehm, D., Weller, A. Kinetics of fluorescence quenching by electron and H-atom transfer. Israel J. Chem. 21st Farkas Memorial Symp. 8,:259–271, 1970.
- [139] Marcus, R. A., Sutin, N. Electron transfers in chemistry and biology. Biochim. Biophys. Acta 811:265–322, 1985.
- [140] Seidel, C. M., Schulz, A., Sauer, M. H. M. Nucleobase-specific quenching of fluorescent dyes. 1. Nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. J. Phys. Chem. 100:5541–5553, 1996.
- [141] Moser, C. C., Keske, J. M., Warncke, K., Farid, R. S., Dutton, P. L. Nature of biological electron transfer. Nature 355:796–802, 1992.
- [142] Stanley, R. J., MacFarlane IV, A. W. Ultrafast excited state dynamics of oxidized flavins: direct observation of quenching by purines. J. Phys. Chem. A 104:6899–6906, 2000.
- [143] Park, H. W., Kim, S. T., Sancar, A., Deisenhofer, J. Crystal structure of DNA photolyase from *Escherichia coli*. Science 268:1866–1872, 1995.

# Index

- $\pi$ -system 139  
 $\rho_{sc}$  90, 92, 93  
 lauz 66  
 lhdn 48, 66, 90  
 lksr 66, 90  
 lnyf 66  
 lpoh 66  
 lshf 66  
 2lzm 66  
 7,8,10-trimethylisoalloxazine  
     *see* trimethylisoalloxazine, 7,8,10-
- A**
- ab initio  
     folding method  
         *see* folding method, ab initio  
     molecular orbital calculation  
         *see* molecular orbital calculation, ab initio  
 abp-120 66  
 absorption, light *see* light absorption  
 acquisition, data *see* data acquisition  
 actin 33  
 active site *see* site, active  
 adenine 128, 139  
 adenosine 123, 125  
 afm 35  
 algorithm  
     cluster *see* cluster algorithm  
     compression *see* compression algorithm  
     constraint *see* constraint algorithm  
     steepest descent  
         *see* steepest descent algorithm  
 all-atom md  
     *see* dynamics, all-atom molecular  
 amino acid sequence 32  
 amplitude, fluorescence  
     *see* fluorescence amplitude  
 amyloid plaque 33  
 anabaena, cyanobacterium  
     *see* cyanobacterium anabaena  
 analysis  
     cluster *see* cluster analysis  
     principal components  
         *see* principal components analysis  
     rigid body *see* rigid body analysis  
 analyzer, multichannel 124  
 anchoring point 34  
 angle  
     bond *see* bond angle  
     constraint *see* constraint, angle  
 anisotropy, fluorescence  
     *see* fluorescence anisotropy  
 apparent  
     distance, reciprocal 109  
     poor convergence 93  
 approach, efficient *see* efficient approach  
 assignment, unambiguous 107  
 atomic  
     force microscope *see* afm  
     level 34, 123  
     model *see* model, atomic  
 atp 125  
 average  
     distance *see* distance average  
     ensemble *see* ensemble average  
 azide 102
- B**
- b. subtilis 66  
 backbone hydrogen *see* hydrogen, backbone  
 barrier, complexity *see* complexity barrier  
 baseline correction 103  
 basis, molecular *see* molecular basis  
 behavior  
     dynamic *see* dynamic behavior  
     rotational *see* rotational behavior  
 biochemistry 31–33  
 biological  
     macromolecule 33  
     system 45, 47  
 biology 31, 32  
     molecular *see* molecular biology  
     structural *see* structural biology  
 biomacromolecules 121

- biophysics, computational  
*see* computational biophysics
- block  
 building 125  
 off-diagonal 92  
 structure 92
- bobscript 19, 73
- bond  
 angle 47  
 covalent 47, 67, 89, 100  
 hydrogen *see* hydrogen bond  
 length 47  
 stretching vibration 45
- bound  
 cofactor *see* cofactor, bound  
 fad *see* fad, bound
- boundary, periodic *see* periodic boundary
- box  
 cubic 48, 67, 90, 101  
 edge 127  
 length 54, 68, 100  
 rectangular 47  
 size 126
- breakthroughs 32
- bridge  
 diphosphate *see* diphosphate bridge  
 salt- 34  
 sulfur 34
- building block *see* block, building
- buried hydrogen bond  
*see* hydrogen bond, buried
- C**
- calculated intensity *see* intensity, calculated
- calculation  
 free energy *see* free energy calculation  
 quantum-mechanical  
*see* quantum-mechanical calculation
- catalytic mechanism 121
- cavity-dumped dye laser  
*see* laser, cavity-dumped dye
- cell 31, 32  
 living *see* living cell
- central structure 91
- charge  
 distribution 123  
 ground state *see* ground state charge  
 partial *see* partial charge  
 shielding 126  
 transfer  
 classical *see* classical charge transfer  
 photo-induced  
*see* photo-induced charge transfer
- chemical  
 process 33  
 shift 107, 108
- chemistry, quantum *see* quantum chemistry
- classical  
 charge transfer 140  
 statistical mechanics 46  
 treatment 45
- closed conformation 121
- cluster  
 algorithm 91, 102, 129  
 analysis 89, 105, 128, 137  
 conformational 89, 91, 92  
 consensus 92  
 dominant 93, 105  
 largest 106  
 minor 106  
 non-overlapping 91
- code, genetic *see* genetic code
- cofactor  
 bound 121  
 flavin *see* flavin cofactor  
 free 121
- coherent element, motionally  
*see* motionally coherent element
- collective mode 65, 68
- combining simulation and experiment  
*see* experiment, combining simulation and
- compact conformation 139
- complex  
 data sets 32  
 intramolecular 121  
 meta- 32  
 model 32  
 protein 33  
 system 32, 33
- complexity 31, 146  
 barrier 31  
 high 31  
 minimum 146
- component, lifetime *see* lifetime component
- components, principal  
*see* principal components
- compression algorithm 31
- computational  
 biophysics 33  
 cost 65  
 work 146
- computationally inaccessible 129
- computer 31

- computing power 32  
 concerted motion *see* motion, concerted  
 confirmation, experimental  
     *see* experimental confirmation  
 conformation  
     closed *see* closed conformation  
     compact *see* compact conformation  
     extended *see* extended conformation,  
         126, 127  
     native *see* native conformation  
     open *see* open conformation  
     protein *see* protein conformation  
     stacked *see* stacked conformation, 127,  
         128  
 conformational  
     cluster *see* cluster, conformational  
     dynamics  
         *see* dynamics, conformational  
     lifetime 140  
     space 93, 104, 106, 139  
     state 92, 140  
     substates 140  
 connected constraint  
     *see* constraint, connected  
 consensus  
     cluster *see* cluster, consensus  
     dynamics *see* dynamics, consensus  
 constant  
     diffusion *see* diffusion constant  
     dissociation *see* dissociation constant  
 constraint 45  
     algorithm 89, 100  
     angle 46  
     connected 46  
     harmonic 127  
 context, evolutionary 33  
 conventional analysis of nmr spectra  
     *see* nmr spectra, conventional analysis of  
 convergence 92, 93, 129  
 conversion, internal *see* internal conversion  
 correction, baseline *see* baseline correction  
 correlated motion *see* motion, correlated  
 correlation 32  
     dynamical *see* dynamical correlation  
     motional *see* motional correlation  
     rotational *see* rotational correlation  
     sequential *see* sequential correlation  
     spatial *see* spatial correlation  
     time  
         effective 103, 109  
         rotational 138  
 cosmology 31  
 cost, computational *see* computational cost  
 coulomb interaction  
     *see* interaction, coulomb  
 coupling  
     temperature *see* temperature coupling  
     weak 47, 67, 89, 100, 127  
 covalent bond *see* bond, covalent  
 covariance matrix 101, 104  
 cross  
     over 32  
     peak intensity *see* intensity, cross-peak  
     relaxation  
         dipolar 99  
         rate 99, 103  
 crystal  
     packing 123  
     structure 126  
 crystallographic 123  
 cubic box *see* box, cubic  
 cut- off 49, 127  
     effect 48, 53  
     rmsd *see* rmsd cut-off  
     twin-range 48, 67, 89, 100  
 cyanobacterium anabaena 126  
 cycle, thermodynamic  
     *see* thermodynamic cycle  
  
**D**  
 d-amino acid oxidase 140  
 d. discoideum 66  
 data  
     acquisition 103  
     experimental *see* experimental data  
     processing 103  
     sets, complex *see* complex data sets  
 de-excitation route 140  
 decay  
     fluorescence *see* fluorescence decay  
         anisotropy  
             *see* fluorescence anisotropy decay  
         intensity  
             *see* fluorescence intensity decay  
     free-induction *see* free-induction decay  
     nanosecond *see* nanosecond decay  
 degrees of freedom 45  
     fast hydrogen 67, 89, 100  
     out-of-plane 89  
     planar ring 67  
 dehydrogenase, lipoamide 121  
 demarcation, fragment  
     *see* fragment demarcation  
 density, spectral *see* spectral density

- depolarization, fluorescence  
*see* fluorescence depolarization
- description  
 quantum-dynamical  
*see* quantum-dynamical description  
 simplified *see* simplified description
- detection frequency 124
- dielectric  
 constant, relative 67, 90, 100, 126  
 relaxation time 47
- difference  
 factor, size-independent  
*see* size-independent difference factor  
 free energy *see* free energy difference  
 root-mean-square *see* rmsd
- different size 90
- diffusion 47, 48, 54, 56  
 constant 102, 104  
 random 146  
 spin *see* spin diffusion
- digestion, enzymatic 121
- dinucleotide, flavin adenine *see* fad
- diphosphate bridge 121
- dipolar cross relaxation  
*see* cross relaxation, dipolar
- dipole moment, emission 128, 138
- discrete exponential 124, 129
- dispersion force *see* force, dispersion
- displacement, mean square 54
- dissociation constant 121
- distance  
 apparent reciprocal  
*see* apparent distance, reciprocal  
 average 103  
 experimental 103  
 interatomic 99, 109  
 proton pair *see* proton pair distance,  
 108
- distribution  
 charge *see* charge distribution  
 equilibrium  
*see* equilibrium distribution  
 mass *see* mass distribution
- dna 33
- domain  
 dynamical *see* dynamical domain  
 identification 67  
 sh3 *see* sh3 domain
- dominant cluster *see* cluster, dominant
- drift, energy *see* energy drift
- dssp 57, 69, 80
- dye laser, cavity-dumped  
*see* laser, cavity-dumped dye
- dynamic  
 behavior 121  
 landscape 139  
 nature 35
- dynamical  
 correlation 99  
 domain 68
- dynamically distinct fragment 93
- dynamics  
 all-atom molecular 146  
 conformational 121, 123  
 consensus 66, 71, 72, 74, 76, 78–81, 93  
 excited state 129  
 internal 103, 104, 108  
 molecular 45–49, 59, 61, 65, 89
- dyndom 67–71
- E**
- e. coli 66, 99, 100, 126
- eco-system 31, 32
- ecology 31, 32
- economics 31, 32
- edge, box *see* box edge
- effect  
 cut-off *see* cut-off effect  
 electronic *see* electronic effect  
 nuclear overhauser *see* noe
- effective correlation time  
*see* correlation time, effective
- efficiency, simulation  
*see* simulation efficiency
- efficient approach 146
- eigenvector 104, 106
- electron microscope 35
- electronic  
 effect 123, 128, 132  
 state 129, 134
- electrostatic interaction  
*see* interaction, electrostatic, 127
- element  
 folding *see* folding element  
 motionally coherent  
*see* motionally coherent element  
 rigid *see* rigid element  
 secondary structure  
*see* secondary structure element  
 semi-rigid *see* semi-rigid element  
 sequential *see* sequential element
- elementary particle *see* particle, elementary

- emission dipole moment  
*see* dipole moment, emission
- energy  
 drift 47–49, 53, 55, 57, 58, 60  
 fluctuation 53  
 minimize 48, 67, 90, 101
- engineering 32
- ensemble 35  
 average 129  
 nmr *see* nmr ensemble
- environment 32  
 vacuum 49
- enzymatic digestion  
*see* digestion, enzymatic
- enzyme 33
- equilibrium  
 distribution 46  
 open/closed 129  
 simulation 65, 89  
 stacking *see* stacking equilibrium
- erythrosine b 124
- events, folding *see* folding events
- evolution 32  
 time, maximum 103
- evolutionary  
 context *see* context, evolutionary  
 history 32
- exchange, hydrogen/deuterium  
*see* hydrogen/deuterium exchange
- excitation intensity 124
- excited state 123, 132  
 dynamics *see* dynamics, excited state
- experiment 34  
 combining simulation and 34, 35  
 fluorescence  
*see* fluorescence experiment
- experimental  
 confirmation 95  
 data 114  
 raw 34  
 distance *see* distance, experimental  
 intensity *see* intensity, experimental  
 set-up 34
- explicit  
 hydrogen 47  
 solvent 65
- exploded view 91
- exponential, discrete  
*see* discrete exponential
- extended  
 conformation 90  
 structure *see* structure, extended
- extent of sampling *see* sampling, extent of
- ## F
- fad 35, 121, 125, 132, 139  
 bound 126
- fast hydrogen degrees of freedom  
*see* degrees of freedom, fast hydrogen
- fastest motion *see* motion, fastest
- ferredoxin reductase  
*see* reductase, ferredoxin
- fiber  
 keratin *see* keratin fiber  
 muscle *see* muscle fiber
- first excited  
 singlet state  
*see* singlet state, first excited  
 triplet state  
*see* triplet state, first excited
- flavin 139  
 adenine  
 dinucleotide *see* fad  
 stacking *see* stacking, flavin-adenine  
 cofactor 121, 123, 126  
 free 121  
 mononucleotide *see* fmn
- flavoenzyme 140
- flavoprotein 121, 126
- fluctuation 67, 68, 79  
 energy *see* energy fluctuation  
 mean square 68, 70  
 uncorrelated local 65
- fluorescence  
 amplitude 132  
 anisotropy 121  
 decay 124, 131  
 decay 129  
 heterogeneous 129  
 depolarization 132  
 experiment 35  
 intensity decay 124  
 lifetime 123, 124, 131, 132  
 component 140  
 intrinsic 132  
 polarized 123, 124  
 quantum yield  
*see* quantum yield, fluorescence  
 quenching 121, 123, 129, 132, 139, 140  
 picosecond  
*see* picosecond fluorescence quenching  
 spectroscopy 139  
 steady-state 132  
 time-resolved 121, 123

- fluorescent group 121  
 fmn 121, 125  
 folding  
   element 65, 84  
   events 82  
   hierarchy 84  
   of 84  
   method, ab initio 81  
   nucleation site 65, 89, 93, 99  
   pathway 84, 85, 94  
   process 65, 81  
   protein 32, 34, 35, 65, 80  
   rate 81, 83  
   time 65  
   state 81, 82  
   of *see* folding transition state  
   unit 65, 84, 91, 93  
 foldon 65, 89  
 force, dispersion 121  
 forcefield 51  
   gromos  
     87 47, 51  
     96 67, 89, 100, 125  
   parameter 125  
 four helix bundle *see* helix bundle, four  
 fragment  
   demarcation 95  
   dynamically distinct  
     *see* dynamically distinct fragment  
   isolated 89, 90, 92  
   peptide 90  
   protein *see* protein fragment  
   separate *see* separate fragment  
   stability 90, 92, 95  
   stable 93  
 free  
   cofactor *see* cofactor, free  
   energy  
     calculation 134  
     difference 129  
     relative 128, 129  
     perturbation 129  
   flavin *see* flavin, free  
   induction decay 103  
 freedom, degrees of *see* degrees of freedom  
 frequency, detection *see* detection frequency  
 function  
   protein *see* protein function  
   shift *see* shift function  
   time-correlation 47, 48, 55–58, 80, 102, 106, 128,  
     132, 139  
     *see* time-correlation function  
 fyn-sh3 66
- G**
- gene 32  
 genetic  
   code 32  
   information 33  
 globule, molten 34  
 glutathione reductase  
   *see* reductase, glutathione  
 green light *see* light, green  
 gromacs 19, 49, 50, 67, 89, 100, 125, 155  
 gromos 47, 51, 52, 67, 89, 100, 125  
   87 forcefield *see* forcefield, gromos-87  
   96 forcefield *see* forcefield, gromos-96  
 ground state *see* state, ground, 132  
   charge 128  
   quantum-mechanical  
     *see* quantum-mechanical ground state  
 group  
   fluorescent *see* fluorescent group  
   heme *see* heme group  
   redox-active *see* redox-active group
- H**
- h. sapiens 32, 39  
 harmonic constraint  
   *see* constraint, harmonic  
 height, peak *see* peak height  
 helix bundle, four 34  
 heme group 34  
 heterogeneous fluorescence decay  
   *see* fluorescence decay, heterogeneous  
 hierarchical  
   level 91  
   organization 65, 84  
 hierarchy  
   folding *see* folding hierarchy  
   motional 71, 73, 75, 77, 78, 84, 94  
   of folding *see* folding, hierarchy of  
 high  
   complexity *see* complexity, high  
   resolution structural information 99  
 history, evolutionary  
   *see* evolutionary history  
 hpr 48, 66, 90, 100  
 human 66  
 hydrogen  
   backbone 108  
   bond  
     47, 48, 55–58, 80, 102, 106, 128,  
       132, 139  
     *see* hydrogen bond

- buried 34  
 lifetime 102, 105  
 network 134, 137  
 bonded water bridge 135  
 bonding, intramolecular 123  
 degrees of freedom, fast  
   *see* degrees of freedom, fast hydrogen  
 exchange, pulsed 81, 83, 84  
 explicit *see* explicit hydrogen  
 mass 46, 48, 55  
 hydrogen/deuterium exchange 81  
 hydroxylase, para-hydroxybenzoate 121  
 hypochromism, raman  
   *see* raman hypochromism
- I**
- identification, domain  
   *see* domain identification  
 image, periodic *see* periodic image  
 in vacuo  
   simulation *see* simulation, in vacuo  
   stacking *see* stacking, in vacuo  
 inaccessible, computationally  
   *see* computationally inaccessible  
 independent motion  
   *see* motion, independent  
 information, genetic *see* genetic information  
 instrumental response  
   *see* response, instrumental  
 insufficient sampling  
   *see* sampling, insufficient  
 intensity  
   calculated 114  
   cross-peak 103, 107  
   excitation *see* excitation intensity  
   experimental 108, 114  
   noe *see* noe intensity  
   peak *see* peak intensity  
   proton pair 109, 114  
   roe *see* roe intensity  
   theoretical 108, 113, 114  
 interaction  
   coulomb 67, 89  
   electrostatic 90, 127  
   lennard-jones 127  
   non-bonded 48, 67, 89, 100  
   stacking *see* stacking interaction  
   van der waals 67, 89  
 interatomic distance  
   *see* distance, interatomic  
 interdigitation 34  
 internal
- conversion 140  
 dynamics *see* dynamics, internal  
 mobility 90  
 interproton vector 103  
 intramolecular  
   complex *see* complex, intramolecular  
   hydrogen bonding  
     *see* hydrogen bonding, intramolecular  
   stacking *see* stacking, intramolecular  
 intrinsic fluorescence lifetime  
   *see* fluorescence lifetime, intrinsic  
 isoalloxazine 125, 128  
   ring 121  
 isolated  
   fragment *see* fragment, isolated  
   peptide *see* peptide, isolated  
 isotropically tumbling  
   *see* tumbling, isotropically
- K**
- keratin fiber 33
- L**
- landscape, dynamic *see* dynamic landscape  
 large-scale simulation  
   *see* simulation, large-scale  
 larger molecule *see* molecule, larger  
 largest cluster *see* cluster, largest  
 laser, cavity-dumped dye 124  
 lattice model *see* model, lattice  
 laws of motion, newtons'  
   *see* newtons' laws of motion  
 length  
   bond *see* bond length  
   box *see* box length  
 lennard-jones interaction  
   *see* interaction, lennard-jones  
 level  
   atomic *see* atomic level  
   hierarchical *see* hierarchical level  
 levinthal paradox 33  
 life sciences 32  
 lifetime  
   component 123, 129  
   fluorescence  
     *see* fluorescence lifetime component  
   conformational  
     *see* conformational lifetime  
   fluorescence *see* fluorescence lifetime  
   hydrogen bond  
     *see* hydrogen bond lifetime  
   spectrum 140

- light  
 absorption 123, 128, 134  
 green 121
- lincs 46, 47, 59, 67, 89, 100
- lipid 33
- lipoamide dehydrogenase  
*see* dehydrogenase, lipoamide
- liquid nitrogen *see* nitrogen, liquid
- living  
 cell 33  
 organism *see* organism, living
- long  
 md simulation *see* md simulation, long  
 timescale *see* timescale, long
- lumiflavin 125
- lysozyme, t4 *see* t4 lysozyme
- M**
- macromolecule, biological  
*see* biological macromolecule
- mass  
 distribution 100  
 hydrogen *see* hydrogen mass
- mathematical method 31, 32
- matrix  
 covariance *see* covariance matrix  
 relaxation *see* relaxation matrix  
 rmsd *see* rmsd matrix
- maximum  
 entropy method *see* mem  
 evolution time  
*see* evolution time, maximum  
 time step *see* time step, maximum
- md *see* dynamics, molecular  
 simulation 127  
 long 99
- mean square  
 displacement  
*see* displacement, mean square  
 fluctuation  
*see* fluctuation, mean square
- mechanism, catalytic  
*see* catalytic mechanism
- mem 125, 131
- meta-complex *see* complex, meta-
- meteorology 32
- method  
 mathematical  
*see* mathematical method  
 maximum entropy *see* mem
- methyl proton *see* proton, methyl
- metric tensor 46
- microscope  
 atomic force *see* afm  
 electron *see* electron microscope
- mind-3, semi-empirical  
*see* semi-empirical mind-3
- minimize, energy *see* energy minimize
- minimum complexity  
*see* complexity, minimum
- minor cluster *see* cluster, minor
- mixing time 107, 109
- mobility, internal *see* internal mobility
- mode, collective *see* collective mode
- model  
 atomic 65, 146  
 complex *see* complex model  
 lattice 146  
 nucleation 65  
 simplified 65, 146
- modes of motion *see* motion, modes of
- molecular  
 basis 32  
 biology 32  
 dynamics *see* dynamics, molecular  
 package 67, 89, 100  
 orbital calculation, ab initio 125
- molecule  
 larger 99  
 single *see* single molecule  
 small flexible 99
- molsript 19, 73, 135
- molten globule *see* globule, molten
- mononucleotide, flavin *see* fmn
- motion  
 concerted 80, 89  
 correlated 65  
 fastest 45  
 independent 89  
 modes of 34  
 overall 99
- motional  
 correlation 34  
 hierarchy *see* hierarchy, motional
- motionally coherent element 35, 65, 67, 69,  
 71–73, 78–82, 84, 85, 94, 95
- msd *see* displacement, mean square
- msf *see* fluctuation, mean square
- multi-domain protein  
*see* protein, multi-domain
- multichannel analyzer  
*see* analyzer, multichannel
- multiple starting structures  
*see* starting structures, multiple

- muscle fiber 33  
 myosin 33
- N**
- nano-technology 31, 32  
 nanosecond decay 131  
 native  
   conformation 93  
   state 34, 65, 79, 84  
   structure 34  
 natural protein *see* protein, natural  
 nature, dynamic *see* dynamic nature  
 neighbor-list update 67, 89, 100, 128  
 network, hydrogen bond  
   *see* hydrogen bond network  
 newtons' laws of motion 34  
 nitrogen, liquid 124  
 nmr 80  
   ensemble 66, 90  
   solution structure  
     *see* solution structure, nmr  
   spectra, conventional analysis of 103  
   spectrometer 103  
   spectroscopy 99  
 noe 99  
   intensity 99  
 noesy 80, 103, 107  
 non  
   bonded interaction  
     *see* interaction, non-bonded  
   overlapping cluster  
     *see* cluster, non-overlapping  
   polar solvent *see* solvent, non-polar  
   roe *see* roe, non-  
 nuclear overhauser effect *see* noe  
 nucleation  
   model *see* model, nucleation  
   site, folding *see* folding nucleation site
- O**
- off-diagonal block *see* block, off-diagonal  
 open conformation 121  
 open/closed equilibrium  
   *see* equilibrium, open/closed  
 organ 32  
 organism 32, 33  
   living 32  
 organization, hierarchical  
   *see* hierarchical organization  
 orthogonal subspace 101  
 oscillation period 45  
 out-of-plane degrees of freedom  
   *see* degrees of freedom, out-of-plane
- overall motion *see* motion, overall  
 overlap 101, 104  
 oxidation potential 140
- P**
- package, molecular dynamics  
   *see* molecular dynamics package  
 packed structure *see* structure, packed  
 packing, crystal *see* crystal packing  
 para-hydroxybenzoate hydroxylase  
   *see* hydroxylase, para-hydroxybenzoate  
 paradox, levinthal *see* levinthal paradox  
 parameter, forcefield  
   *see* forcefield parameter  
 partial charge 132  
 particle  
   elementary 31  
   particle particle-mesh *see* ppm  
   virtual 31  
 pathway  
   folding *see* folding pathway  
   transition *see* transition pathway  
 pca *see* analysis, principal components, 101  
 pdb 48, 66–68, 72, 76, 90, 127  
 peak  
   height 103  
   intensity 114  
 peptide  
   fragment *see* fragment, peptide  
   isolated 100  
   stability 89, 90  
   synthesis 102  
 period, oscillation *see* oscillation period  
 periodic  
   boundary 47, 49  
   image 127  
 perturbation, free energy  
   *see* free energy perturbation  
 phosphate, potassium  
   *see* potassium phosphate  
 phosphotransferase system 66  
 photo-induced charge transfer 140  
 picosecond fluorescence quenching 140  
 planar ring degrees of freedom  
   *see* degrees of freedom, planar ring,  
   100  
 plaque, amyloid *see* amyloid plaque  
 point, anchoring *see* anchoring point  
 polar solvent *see* solvent, polar  
 polarized fluorescence  
   *see* fluorescence, polarized

- poor convergence, apparent  
     *see* apparent poor convergence
- population 105
- potassium phosphate 102
- potential, oxidation *see* oxidation potential
- power, computing *see* computing power
- ppm 49, 128
- preparation, sample *see* sample preparation
- pressure 47, 67, 89, 100, 127
- principal components 79  
     analysis 32, 65, 67, 68
- problem, protein-folding  
     *see* protein-folding problem
- process  
     chemical *see* chemical process  
     folding *see* folding process  
     stacking *see* stacking process
- processing, data *see* data processing
- projection 104, 106
- protein 32  
     complex *see* complex, protein  
     conformation 90, 92  
     folding *see* folding, protein  
     problem 33  
     fragment 89  
     function 32  
     multi-domain 69, 84  
     natural 34  
     sequence 95  
     system 146
- proton  
     methyl 108  
     pair  
         distance 99  
         intensity *see* intensity, proton pair
- pulsed hydrogen exchange  
     *see* hydrogen exchange, pulsed
- Q**
- quantum  
     chemical 33  
     chemistry 32  
     dynamical description 146  
     mechanical  
         calculation 45  
         ground state 45  
     physics 31, 32  
     yield, fluorescence 121
- quenching, fluorescence  
     *see* fluorescence quenching
- R**
- raman  
     hypochromism 123  
     resonance 123
- random diffusion *see* diffusion, random
- raster3d 19, 135
- rate  
     cross relaxation  
         *see* cross relaxation rate  
     folding *see* folding rate  
     unfolding *see* unfolding rate
- raw experimental data  
     *see* experimental data, raw
- reciprocal apparent distance  
     *see* apparent distance, reciprocal
- rectangular box *see* box, rectangular
- redox  
     active group 121  
     enzyme 121
- reductase  
     ferredoxin 121, 126  
     glutathione 121, 126  
     thioredoxin 121, 126
- relative  
     dielectric constant  
         *see* dielectric constant, relative  
     free energy difference  
         *see* free energy difference, relative
- relaxation  
     matrix 99, 103  
     time, dielectric  
         *see* dielectric relaxation time
- relevance, structural  
     *see* structural relevance
- resolution, time 129
- resonance raman *see* raman, resonance
- response, instrumental 124
- ribityl 125
- riboflavin 123  
     binding protein 140
- rigid  
     body analysis 89  
     element 65  
     structure *see* structure, rigid
- ring, isoalloxazine *see* isoalloxazine ring
- rmsd 69, 90, 91, 104, 105, 116, 128, 137  
     cut-off 91, 102, 105  
     matrix 92, 116
- rna 33
- roe  
     intensity 108  
     non- 108, 114
- roesy 80, 103  
     spectrum 107

- root-mean-square difference *see* rmsd  
 rosetta 81  
 rotational  
   behavior 123  
   correlation 102, 105, 128  
   time *see* correlation time, rotational  
 route, de-excitation *see* de-excitation route
- S**
- salt-bridge *see* bridge, salt-  
 sample  
   preparation 102, 121  
   temperature *see* temperature, sample,  
   124  
 sampled subspace *see* subspace, sampled  
 sampling  
   extent of 99  
   insufficient 93  
 scheme, time step *see* time step scheme  
 sciences, life *see* life sciences  
 secondary structure 57, 69, 80  
   element 65, 72, 76, 78, 80  
 semi  
   empirical mindo-3 125  
   rigid element 68–72, 78, 79  
 separate fragment 91  
 sequence  
   amino acid *see* amino acid sequence  
   protein *see* protein sequence  
 sequential  
   correlation 34  
   element 80  
 set-up, experimental *see* experimental set-up  
 sh3 domain 66  
 shake 59  
 shift  
   chemical *see* chemical shift  
   function 48  
 short timescale *see* timescale, short  
 simplification 32  
 simplified  
   description 35, 95  
   model *see* model, simplified  
 simulation  
   and experiment, combining 34, 35  
   efficiency 35  
   equilibrium *see* equilibrium simulation  
   in vacuo 126  
   large-scale 32  
   md *see* md simulation  
 single molecule 35
- singlet state, first excited 128  
 site, active 121  
 size  
   box *see* box size  
   different *see* different size  
   independent difference factor 90, 92  
 small flexible molecule  
   *see* molecule, small flexible  
 snarf 103  
 society 33  
 sociology 31, 32  
 solution  
   stacking *see* stacking, solution  
   structure 99  
   nmr 101  
 solvent 123  
   explicit *see* explicit solvent  
   non-polar 121  
   polar 121  
 space, conformational  
   *see* conformational space  
 spatial correlation 34  
 spc water *see* water, spc  
 spc/e water *see* water, spc/e  
 spectral  
   density 99, 103, 109  
   width 103  
 spectrometer, nmr *see* nmr spectrometer  
 spectroscopy  
   fluorescence  
     *see* fluorescence spectroscopy  
   nmr *see* nmr spectroscopy  
 spectrum 103  
   lifetime *see* lifetime spectrum  
   roesy *see* roesy spectrum  
   theoretical *see* theoretical spectrum  
 spin diffusion 103  
 spoiaa 66  
 spontaneous stacking  
   *see* stacking, spontaneous
- stability  
   fragment *see* fragment stability  
   peptide *see* peptide stability  
 stable fragment *see* fragment, stable  
 stacked conformation 121  
 stacking  
   equilibrium 121  
   flavin-adenine 121, 123  
   in vacuo 126, 135, 138  
   interaction 34, 132  
   intramolecular 123  
   process 123

- solution 135, 138  
 spontaneous 128  
 starting structures, multiple 79  
 state  
   conformational  
     *see* conformational state  
   electronic *see* electronic state  
   excited *see* excited state  
   folding transition  
     *see* folding transition state  
   ground 125  
   native *see* native state  
 statistical mechanics, classical  
   *see* classical statistical mechanics  
 steady-state fluorescence  
   *see* fluorescence, steady-state  
 steepest descent algorithm 67, 90, 101, 127  
 step, time *see* time step  
 stretching, bond-~ vibration  
   *see* bond-stretching vibration  
 structural  
   biology 33  
   information, high resolution  
     *see* high resolution structural information  
   relevance 109  
 structure  
   block *see* block structure  
   central *see* central structure  
   crystal *see* crystal structure  
   extended 101  
   native state *see* native state structure  
   packed 34  
   rigid 99  
   secondary *see* secondary structure  
   solution *see* solution structure  
   sub-nanosecond resolved time-correlated  
     single photon counting  
     *see* tcspc, sub-nanosecond resolved  
 subspace  
   orthogonal *see* orthogonal subspace  
   sampled 101  
 substates, conformational  
   *see* conformational substates  
 sulfur bridge *see* bridge, sulfur  
 synthesis, peptide *see* peptide synthesis  
 system  
   biological *see* biological system  
   complex *see* complex system  
   eco *see* eco-system  
   phosphotransferase  
     *see* phosphotransferase system  
   protein *see* protein system
- T**
- t4  
   lys 66  
   lysozyme 66  
 tcspc 124, 129, 140  
   sub-nanosecond resolved 123  
 temperature 47, 67, 89, 100, 123, 127  
   coupling 49  
   sample 103  
 tensor, metric *see* metric tensor  
 termini, uncharged *see* uncharged termini  
 theoretical  
   intensity *see* intensity, theoretical  
   spectrum 103  
 thermodynamic cycle 129, 134  
 thioredoxin reductase  
   *see* reductase, thioredoxin  
 through space 99  
 time  
   correlated single photon counting,  
   sub-nanosecond resolved  
     *see* tcspc, sub-nanosecond resolved  
   correlation function 99  
   folding *see* folding time  
   mixing *see* mixing time  
   resolution *see* resolution, time  
   resolved fluorescence  
     *see* fluorescence, time-resolved  
   step 48, 49, 55, 58, 59, 89  
     maximum 45, 56, 60  
     scheme 99, 104  
 timescale  
   long 34  
   short 34  
 tocsy 103, 107  
 transition  
   pathway 139  
   state of folding  
     *see* folding transition state  
 treatment, classical *see* classical treatment  
 trimethylisalloxazine, 7,8,10- 125  
 triplet state, first excited 125  
 tumbling, isotropically 99  
 twin-range cut-off *see* cut-off, twin-range
- U**
- unambiguous assignment  
   *see* assignment, unambiguous  
 uncharged termini 90  
 uncorrelated local fluctuation  
   *see* fluctuation, uncorrelated local  
 unfolding rate 81

unit, folding *see* folding unit  
unstacking 134, 139  
update, neighbor-list  
*see* neighbor-list update

## V

vacuum environment  
*see* environment, vacuum  
van der waals interaction  
*see* interaction, van der waals  
vector, interproton *see* interproton vector  
vibration, bond-stretching  
*see* bond-stretching vibration  
view, exploded *see* exploded view  
virtual particle *see* particle, virtual  
viscosity 47, 48, 54, 56

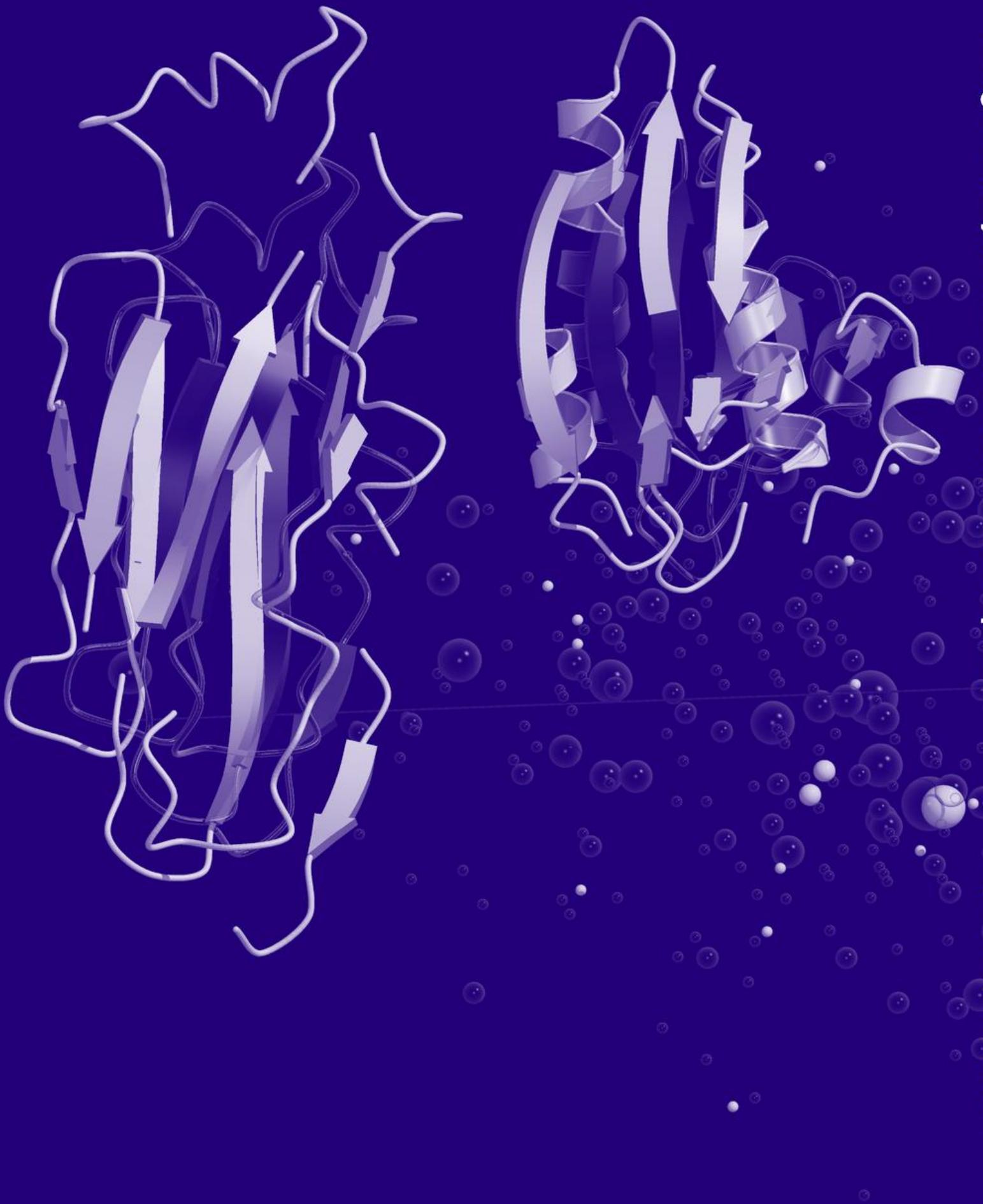
## W

water  
bridge, hydrogen-bonded  
*see* hydrogen-bonded water bridge  
spc 48, 55, 67, 90, 101  
spc/e 126  
watergate 103  
weak coupling *see* coupling, weak  
whatif 90  
width, spectral *see* spectral width  
work, computational  
*see* computational work

Please also visit my homepage:  
[md.chem.rug.nl/~anton](http://md.chem.rug.nl/~anton)

An online version of this thesis can be found at:  
[md.chem.rug.nl/~anton/thesis.pdf](http://md.chem.rug.nl/~anton/thesis.pdf)

Much of my time was spent building the old GROMACS homepage at:  
[md.chem.rug.nl/~gmx](http://md.chem.rug.nl/~gmx)  
which is now superseded by the new GROMACS homepage at:  
[www.gromacs.org](http://www.gromacs.org)



ISBN 90-6464-992-8