

UNIVERSITY OF CINCINNATI

Date: 18-Nov-2009

I, Roman Petrenko ,

hereby submit this original work as part of the requirements for the degree of:

Doctor of Philosophy

in Physics

It is entitled:

Computer Simulations of Resilin-like Peptides

Student Signature: Roman Petrenko

This work and its defense approved by:

Committee Chair: Thomas Beck, PhD
Thomas Beck, PhD

Jaroslav Meller, PhD
Jaroslav Meller, PhD

Rostislav Serota, PhD
Rostislav Serota, PhD

COMPUTER SIMULATIONS OF RESILIN-LIKE PEPTIDES

BY

ROMAN PETRENKO

B.S., Kharkiv National University, 2000

M.S., University of Cincinnati, 2003

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Physics Department of the
University of Cincinnati, 2009

March 4, 2010

Cincinnati, Ohio

COMPUTER SIMULATIONS OF RESILIN-LIKE PEPTIDES

Roman Petrenko, Ph.D.
Physics Department
University of Cincinnati, 2009
T.L. Beck, Advisor

Resilin is an elastomeric protein characterized by rubber-like elasticity, very high resilience and high fatigue lifetime. These outstanding material properties are conferred by multiple elastic repeats, similar to those found in other elastomeric proteins. In this thesis I use molecular dynamics to elucidate the effect of amino-acid sequence variation on the mechanical properties of resilin-like peptides. In particular, I address the role of disorder in the relaxed (unstretched) state and the amount of conformational entropy lost upon extension. I simulate model systems comprising multiple identical repeats from single elastic units observed in *fruit fly* and *mosquito* resilin gene products. The length of the simulated peptides ranges from 11 to 176 residues. In order to study the nature of the restoring force in resilin I use steered molecular dynamics (SMD) and fixed end simulations. I find a high level of disorder and lack of stable secondary structure for the well solvated relaxed state in all simulated peptides; these results are consistent with conclusions from circular dichroism spectra of resilin-like peptides. Structural parameters, computed from molecular dynamics trajectories, are compared with experimental NMR and SAXS results. While upon stretching the conformational entropy is significantly decreased, the enthalpy is estimated to remain essentially unchanged. I conclude that the restoring force is primarily of entropic origin and largely insensitive to the amino-acid composition of resilin-like elastic repeats. Finally, I build a coarse-grained model from all-atomic simulation of two repeats in mosquito resilin and apply it larger peptides in order to assess flexibility and the effect of cross-linking in multiple resilin-like polypeptides.

Acknowledgements

I want to express gratitude to my supervisors, Tom Beck and Jarek Meller, for introducing me to the field of computational modeling and to research on elastomeric proteins. Their encouragement kept me on the right track and their valuable suggestions were always helpful to better understand underlying concepts. This work was partially supported from MURI grant and by grant from US Air Force.

This work would be impossible without computer resources from Biomedical Informatics Division. Particularly, I am thankful to their system administrator, Prakash Velayutham, for his prompt solutions to any kind of computer-related problems I had.

Also, I am grateful to Alexey Porollo for fruitful discussions on chemistry and biology related questions and to Slava Serota for sharing his experience in graduate life and research.

Finally, I am deeply grateful to my wife, Valeriya Petrenko, for her patience, understanding, encouragement and support during the time spent on this dissertation.

Table of Contents

List of Tables.....	viii
List of Figures.....	x
Chapter 1 Introduction.....	1
1.1 Proteins	1
1.1.1 Structure	2
1.1.2 Dynamics	3
1.1.3 Disorder	4
1.1.4 Elastomeric proteins	5
1.2 Resilin	6
1.2.1 Crosslinking and resilin fibers	7
1.2.2 Resilience	7
1.2.3 Resilin gene and elastic repeats	9
1.3 Experimental approaches for studies of elasticity	10
1.4 Theoretical models of elasticity in proteins	11
1.4.1 Restoring force	11
1.4.2 Radius of gyration of unstructured proteins	15
1.5 Computational approach to study elasticity in resilin-like peptides . .	15
Chapter 2 Identification of putative resilin-like peptides in genomic databases	17
2.1 Homology-based search for resilin-like peptides	17
2.2 Resilin-like peptides with chitin-binding domain	18
2.3 Resilin-like peptides without chitin-binding domain	22
2.4 Chitin-binding domain	24
2.5 Conclusion	25
Chapter 3 Molecular dynamics and related simulation methods.....	26
3.1 Molecular dynamics	26
3.1.1 Newtonian dynamics	27
3.1.2 Numerical integration of the equations of motion	28
3.1.3 Force-field	29
3.1.4 Statistical ensembles	31
3.1.5 Langevin dynamics	31
3.2 Energy-entropy decomposition	32

3.2.1	Thermodynamic integration and perturbation methods	33
3.2.2	Jarzynski equality	33
3.3	Order parameters	35
3.4	Coarse-grained methods and models	36
3.4.1	Radial distribution function	36
3.4.2	Inverse MC	37
Chapter 4	Simulated systems and protocols.....	39
4.1	Idealized repeats and resilin-like peptides	39
4.2	Simulation protocols	39
4.2.1	Unconstrained molecular dynamics	41
4.2.2	Constrained molecular dynamics	41
4.2.3	Steered molecular dynamics	41
4.3	Measures and quantities derived from simulations	43
4.3.1	Hydrogen bonds	43
4.3.2	Secondary structures	43
4.3.3	Conformational entropy using quasiharmonic approximation .	44
4.3.4	Order parameters	46
Chapter 5	Simulations of idealized resilin-like peptides from anopheles gambiae.....	47
5.1	Fluctuations and disorder in AN2 peptides from unconstrained MD simulations	47
5.1.1	Radius of gyration	47
5.1.2	Secondary structures	49
5.1.3	Order parameters	50
5.1.4	Tyrosine self-recognition	50
5.2	Effect of cross-linking on local structure	55
5.3	Structural parameters and entropic force in AN2 from constrained MD simulations	56
5.3.1	Methods	56
5.3.2	Results	56
Chapter 6	Assessment of disorder in drosophila resilin peptide.....	60
6.1	Unconstrained molecular dynamics	60
6.2	Constrained molecular dynamics	64
Chapter 7	Comparison of entropic force in other resilin-like peptides	71
7.1	Relaxation of multiple chains	71
7.2	SMD simulations	73
Chapter 8	Coarse-grained model of RLP.....	75
8.1	PYP-model	75
8.1.1	Cross-linking	81

Chapter 9 Conclusions	84
9.1 Disorder	84
9.2 Entropic force	85
9.3 Cross-linked biopolymers	85
References	86

List of Tables

1.1	Summary of the experimental data on resilin-like peptides.	13
2.1	Predicted resilin-like peptides, that are highly homologous to CG15920 gene product and have the same domain structure (N-terminal elastic domain, CBD, and C-terminal elastic domain).	18
2.2	Predicted resilin-like peptides that are homologous to AGAP002367-PA gene product (mosquito).	22
4.1	Amino acid sequences of single repeats in resilin-like peptides. While the first two repeats (AN and DN) are well established resilin-like repeats with experimentally known physical properties, the other three repeats (DC, TN, AP) were derived by sequence homology and their physical properties have not been studied yet.	40
4.2	Secondary structure 1-letter codes.	44
6.1	Comparison of the thermodynamic variables from different trajectories, each 100 ns long.	63
6.2	Secondary structure fractions from the combined vacuum (500 ns) and water (500 ns) simulations.	64
6.3	Setup for the constrained MD simulation of DN2 peptide at different extensions.	65
6.4	Fractions of the secondary structures in DN2 peptide from the constrained MD simulations.	66
8.1	Spring constants and average bond lengths for PY virtual bond in the coarse-grained model, obtained by Boltzmann inversion of virtual bond distribution from all-atom simulations.	77
8.2	Spring constants and average bond lengths for YP virtual bond in the coarse-grained model, obtained by Boltzmann inversion of virtual bond distribution from all-atom simulations.	77
8.3	Spring constants and average bond lengths for linkers PP in the coarse-grained model, obtained by Boltzmann inversion of virtual bond distribution from all-atom simulations.	77
8.4	Spring constants and average PYP bond angles in the coarse-grained model, obtained by Boltzmann inversion of virtual bond distribution from all-atom simulations.	78

8.5	Radius of gyration for a system of four coarse-grained AN16 chains with and without cross-linking. Average and standard deviations are reported.	82
-----	--	----

List of Figures

1.1	Force-extension curves in loading/unloading experiments. The resilience is defined as a ratio of the areas under loading and unloading forces. .	8
1.2	Schematic representation of atomic force microscope. The base of cantilever moves is approaching the sample and once tip touches the sample, cantilever bends and photodiode detects deflection in the reflected laser beam. From the data of laser deflections from tip that moves in and out of the sample, resilience can be computed from the resulting force-extension curves obtained.	8
1.3	Schematic representation of the domain structure of the gene product CG15920 from <i>Drosophila</i> species, which consists of three domains. From left to right these domains are: N-terminal elastic repeats, chitin-binding domain, and C-terminal elastic repeat units.	9
1.4	Consensus sequences of elastic repeats from N-terminal (top) and C-terminal (bottom) domain in CG15920 gene product. More conserved residues are higher. Tyrosines, for example, are completely conserved in both repeats.	9
1.5	CD spectra of 15-residue DN1 and DC1 peptides, constructed as consensus repeats from N- (solid line) and C-terminal (dots) elastomeric domains in <i>D. melanogaster</i> resilin. Pronounced minimum at the wavelength around 195 nm indicates on the dominance of disorder and PPII helices.	12
1.6	Example of damping chain dynamics upon extension of a 30-residue peptide from 50 to 100 Å. For clarity, only backbone α -carbons are shown. Spheres on the left and right are drawn for the constrained atoms. Significant reduction of entropy is achieved by the fact that the peptide is highly disordered in the relaxed state (top).	14
2.1	Disorder predictions from 7 servers of <i>Drosophila</i> resilin CG15920 gene product. Residues with black color were predicted to be disordered by all servers. Highlighted is the chitin binding domain that is predicted to form stable secondary structures.	20
2.2	Consensus sequences for elastic repeats in RLP-1 peptides (from top to down: DN, PHa, NN, BN, TN, PHb, AP). High conservation of proline and tyrosine residues is observed.	21

2.3	Consensus sequences of the elastic repeats in RLP-2 peptides. From top to down: DNb, CN, AN, TNb, BNb. Note conservation of proline positions with respect to tyrosine: four and three residues before and after tyrosine.	23
2.4	Sequence alignment of CBD domain in RLP-1 peptides. The VLLPDGR-motif is completely conserved in species. As opposed to sequence variations in the elastic repeats, there are only a few mutations in the CBD domains of different species.	24
2.5	Secondary structure predictions of CBD domain in <i>D. melanogaster</i> using Sable[1], Porter[48], PSIPRED[31] servers. Fully buried residues as predicted by Sable server are in bold. Most likely, the highlighted residues form hydrophobic core in the folded CBD domain. The image was generated using POLYVIEW server [49].	25
4.1	Schematic representation of typical SMD simulation setup. One end of the protein is kept fixed, while to the other end a harmonic constraint is applied. The constraint moves with constant speed along end-to-end vector.	42
5.1	Radius of gyration in AN2 peptide from 200 ns trajectory in unconstrained simulations. Fluctuations in the radius of gyration in AN2 peptide (left panel) and its distribution (right panel).	48
5.2	End-to-end distance in AN2 peptide from 200 ns trajectory in unconstrained simulations. Fluctuations in the end-to-end distance in AN2 peptide (left panel) and its distribution (right panel).	49
5.3	Dynamics of the secondary structures in AN2 peptide from the unconstrained MD simulations. From top to bottom, each line represents peptide's conformation. No stable secondary structure is formed, indicating on large fluctuations in the peptide. α -helices are observed only around SSQY-motif and PPII helices are found around proline residues.	50
5.4	The $1-S^2$ parameters, calculated from the experimental C^α and H^α chemical shifts (points) and from MD simulations of AN2 peptide (line).	51
5.5	Interaction of YGAP-motifs from different chains. The YGAP-motif from the first repeat (left, blue) interacts with the YGAP-motif from the second repeat of AN2 image (right, red).	51
5.6	Distance between backbone α -carbons of tyrosine and proline that are part of YGAP motif within one peptide. Correlated reduction of fluctuations of YGAP-motif in each peptide is observed once the two motifs start interacting with each other.	52
5.7	Interaction of YGAP-motifs from simulations of AN16 peptide.	53
5.8	Anti-parallel arrangement of two helices (sequence PSSQY) from different chains, observed in the AN16 simulations.	54

5.9	Interaction of tyrosines (from left to right: residues 74 and 85) from the adjacent repeats in one chain, observed in the simulation of AN16 peptide. This kind of interaction, that could potentially result in dityrosine bond formation, is not desired, since it "shortcuts" the fluctuations of the segment between them.	54
5.10	The 1-S ² parameters the 15-residue AN peptides, cross-linked at tyrosine site. While, with overall reduction of fluctuations as compared to the 1-S ² parameters from uncross-linked simulations, there is also a decrease in fluctuations in YGAP-motif (residues 8-11). The same decrease of fluctuations around YGAP-sequence was observed in the uncrosslinked simulations.	55
5.11	Conformational entropies of AN2 at different extensions. The values of conformational entropies from L75 and L85 simulations are not shown, since they are close to entropies from L80, L70 simulations. The entropic decrease with extension is captured by quasiharmonic approximation used to calculate conformational entropies of AN2 peptide. . .	57
5.12	Tensile forces in stretched to different extensions AN2 peptides. The force is higher at higher extensions.	58
5.13	Tensile forces in stretched to different extensions AN2 peptides. The sum of the entropic and energetic components of the restoring force is very close to the direct measurement of the restoring force. For extensions less than 92 Å, the AN2 peptide acts like an entropic spring, while at higher extensions it behaves more like a regular spring. . . .	59
6.1	The end-to-end distance of DN2 peptide from unconstrained molecular dynamics simulation in vacuum. Regardless of the input conformation, in all simulations the peptide collapsed into a compact conformation and remained in it with only minimal fluctuations.	61
6.2	The end-to-end distance of DN2 peptides from unconstrained langevin dynamics in water with different values of damping constants (1, 5, 10 ps ⁻¹). In all three simulations a compact input structure swell in water within 10 ns and remained highly flexible for the following 90 ns of the simulation time.	61
6.3	End to end distance and radius of gyration observed in unconstrained MD simulation of DN2 peptide. Langevin damping constant 1 ps ⁻¹ .	62
6.4	End to end distance and radius of gyration observed in unconstrained MD simulation of DN2 peptide. Langevin damping constant 10 ps ⁻¹ .	63
6.5	Force-extension curve from vacuum SMD simulation of a DN2 peptide, that was originally at some compact state. Significant decrease in force fluctuations corresponds to distortion of peptide's bonds and bond angles is observed for end-to-end distances larger than 100 Å. . .	65
6.6	Secondary structure fractions observed in the constrained simulations of DN2 peptides stretched to different extensions. Secondary structure notations are given in Table 4.2. From top to down the level of stretching is increased. Simulation labels are: L50, L70, L90, and L100.	67

6.7	Average number of peptide-water hydrogen bonds for each residue. Results from L50 and L90 constrained MD simulations are shown in open and filled circles. The reduction of the number of the peptide/water hydrogen bonds is observed at higher extension.	68
6.8	Conformational entropies in DN2 peptide from the constrained MD simulations. Significant reduction in conformational entropy is found at 90 and 100 Å extensions. Since less conformational space is available at higher extensions, the conformational entropy convergence is better at higher extensions.	69
6.9	Comparison of the conformational entropies in DN2 peptide from constrained MD simulations in vacuum and in water. This figure shows significant role of water in providing high flexibility at the relaxed state in DN2 peptide. While peptide's entropies in vacuum are reduced a little bit for 90 and 100 Å extensions, the significantly low entropy in vacuum is observed when peptide is in the relaxed state (extensions around 50 Å).	70
6.10	Entropic force in DN2 peptide computed from splined version of the conformational entropies from the constrained MD simulations. . . .	70
7.1	I setup 16 AP2 peptides in a 4 by 4 grid in XY plane. The grid size in either direction was 10 Å and all peptides were highly stretched and oriented along Z-axis. The averaged over all chains end-to-end distance decreases to an equilibrium value of 55 Å within 10 ns of simulation. The formation of β-sheets would result in the end-to-end distances at least 80-90 Å.	72
7.2	Two extreme extensions of DN2 peptide used in SMD simulations. Black sphere shows fixed atom and red sphere – for SMD atom that was pulled.	73
7.3	Loading (red, dash) and unloading (black, solid) forces, averaged over several SMD simulations at temperature 355 K for three resilin-like peptides: DC2, DN2, AP2. With the exception that proline-lacking AP2 peptide has less restoring force than DC2 or DN2 peptides at high extensions, no significant differences between peptides was found. In all three peptides the resilience estimated to be 80±10%.	74
8.1	The coarse-grained and all-atom models of AN16 peptide. Three beads per repeat were assigned. The beads were placed at prolines and tyrosine backbone α-carbons.	76
8.2	Dependence of radius of gyration in the coarse-grained simulations of AN2 on the value of the van der Waals strength parameter. With the values up to 0.1 kcal/mol high fluctuations of the peptide (as observed in all-atom simulations) are reproduced.	79
8.3	Comparison of bond distributions from coarse-grained (solid) and all-atom simulations of AN16 peptide.	80

8.4	Comparison of bond distributions from coarse-grained (solid) and all-atom simulations of AN16 peptide.	80
8.5	Comparison of bond distributions from coarse-grained (solid) and all-atom simulations of AN16 peptide.	81
8.6	A system of four coarse-grained AN16 chains with four randomly placed cross-links. Each chain from the central cell is shown in different color and chains from periodic cell images are shown in gray.	82
8.7	Distribution of the radius of gyration of a single chain in simulations of four coarse-grained chains of AN16 peptide with and without cross-linking. Gaussian fit was used to the distributions from simulations. .	83

Notations

D. melanogaster	Drosophila melanogaster (fruit fly)
T. castaneum	Tribolium castaneum (red flour beetle)
Ap. mellifera	Apis mellifera (european honey bee)
N. vitripennis	Nasonia vitripennis (parasitoid jewel wasp)
Ac. pisum	Acyrtosiphon pisum (pea aphid)
MC	monter carlo
MD	molecular dynamics
REMD	replica exchange molecular dynamics
SMD	steered molecular dynamics
PDB	protein data bank
VdW	van der Waals
RMSD	root-mean-square deviation
NMR	nuclear magnetic resonance
SAXS	small angle X-ray scattering
AFM	atomic force microscope
CD	circular dichroism
RLP	resilin-like peptides
PPII	poly-proline II secondary structure
HB	hydrogen bond
DSSP	dictionary of secondary structures in proteins
VMD	visual molecular dynamics
NAMD	nanoscale molecular dynamics
RCI	random coil index
IDP	intrinsically disordered proteins
PME	particle-mesh Ewald
NVE	constant number of particles, volume and energy
NVT	constant number of particles, volume and temperature
NPT	constant number of particles, pressure and temperature
PBC	periodic boundary conditions
BLAST	basic local alignment search tool
CBD	chitin-binding domain
NOE	Nuclear Overhauser effect
NCBI	National Center for Biotechnology Information
SHO	simple harmonic oscillator

Chapter 1

Introduction

1.1 Proteins

Proteins are polypeptides that are built in living cells from the 20 standard amino acids. Amino acids are linked to each other by peptide bonds, forming linear chains of unique sequence. The amino acid sequence, called *the primary structure* of the protein, is encoded in a gene, and it determines the protein's 3D structure together with its function.

All amino acids have a common chemical structure: amino group, α -carbon, carboxyl group, and a side chain. The first three groups of each amino acid form the protein backbone, where the carboxyl group of one amino acid is linked to the amino group of the next amino acid by a peptide bond. By convention, the numbering of the protein residues goes from amino group to carboxyl group, in other words, from N- to C-terminal. The side chains are unique for each amino acid and are attached to the backbone central α -carbon. There are two exceptions - proline, where an additional bond between its side chain atom and the backbone amide is formed; and glycine, that does not have a side chain. These two amino acids present extreme cases of the backbone largest and smallest flexibility.

1.1.1 Structure

During or after synthesis, proteins usually adopt a well-defined three-dimensional (3D) structure, known as the native state. The process of the structural changes from initial conformation to the native state is called *protein folding*. It depends on many factors, including the protein environment, the concentration of salts, temperature, the presence of folding assisting proteins, etc. Usually, the native state is much more energetically favorable than other states and is stabilized by a hydrophobic core, covalent bonding, and intra-protein hydrogen bonds. Certain 3D forms of the local arrangement in the proteins, known as *the secondary structures* have been recognized. The most common secondary structures include: α -helix, β -sheet, turn, and polyproline II helix (PPII). The PPII helix is the only secondary structure that is not stabilized by the internal hydrogen bonding, it has been a subject of many debates regarding whether or not one should consider PPII as a secondary structure. Although proline is very common in PPII helices, in half of the cases the PPII helices are formed by non-proline residues [14]. For example, it was shown with NMR and CD spectroscopies that for seven residue polyalanines the dominant structure is the PPII helix at the temperatures close to zero degrees of Celcius [57].

While *the secondary structures* are local arrangement in proteins, the overall 3D structure is called *the tertiary structure*. Since the process of experimentally determining protein's primary structure is much faster than the tertiary structure determination, the prediction of the tertiary structure from sequence alone, known as *protein structure prediction*, is one of the most important goals pursued by bioinformatics and theoretical chemistry.

Proteins that adopt a well defined 3D shape can be crystallized, so that 3D structure can be deduced from the resulting X-ray pattern. For highly flexible proteins, X-ray crystallography is not applicable, so other methods like NMR or CD spectroscopies are used to get insights into the protein's structure.

1.1.2 Dynamics

During folding, the proteins adopt different conformations before the native state is reached. At the equilibrium, the system occupies a state that corresponds to the minimal free energy. Such a state for the proteins is called the native state. Some proteins are intrinsically disordered, which means that many conformations are available to them with similar free energies. An intrinsically disordered protein is therefore characterized by a *broad* ensemble of states, as opposed to structural proteins that are represented by a much smaller ensemble of similar states.

Conformational changes

As it was mentioned earlier, proteins adopt many conformations in solution forming a *conformational ensemble*. The thermal fluctuations and low free energy barriers allow many transitions between conformations in this ensemble. Computational approaches, such as Monte Carlo (MC) or molecular dynamics (MD), allow one to model such an ensemble in atomic detail. Since the protein's energy landscape is very rough, simple MC or MD algorithms suffer from so called *local minimum trapping* problem. This happens because there are many energy barriers between the conformations in the ensemble. Therefore, in order to have a statistically meaningful ensemble, a proper sampling of the conformational space is required.

Sampling

The most common practices for effective sampling of the conformational space include: replica exchange molecular dynamics (REMD) or Monte Carlo (REMC) simulations, steered molecular dynamics (SMD) simulations, and free energy methods.

The REMD method, introduced by Sugita and Okamoto [59], is an example of a method that considers trajectories (replicas) of the system at different temperatures. The non-interacting replicas of the same system, each simulated at different tem-

perature, can be exchanged between different trajectories at random points in time, enabling effective crossing of energy barriers. For the success of the algorithm, one must ensure that the rate of replica exchange is sufficiently high. This is achieved by choosing enough replicas to cover the simulation temperature range, and by choosing the temperatures that result in the high overlap of the energy distributions between two replicas. Recently, an automatic adjustment of the temperatures was proposed [60]. Starting with an arbitrary choice of the temperatures, the method readjusts them in such a way that replica diffusion in the temperature space is maximized.

1.1.3 Disorder

Not long time ago, it was believed that only the proteins with well-defined 3D structure are important. In the last decade, a new class of proteins that are unstructured and functional, has become an active area of research. *Intrinsically disordered proteins* (IDP) are mostly found in eukaryotes (see review on the field [18, 65]). Some of the functions include: molecular recognition and assembly and protein modification. Some IDPs become structured upon binding to certain ligands, while others have flexible linkers and participate in macromolecular assembly [16]. Advances have been made in the prediction of the disordered regions in proteins. There is a database (DisProt) of the disordered proteins [58], and more than dozen web-servers predict disordered regions in proteins from sequence alone.

Experimental methods to study disordered proteins

Since disordered proteins do not adopt particular 3D structure, they can not be determined by X-ray crystallography and therefore are missing from the database of protein structures (PDB). The primary methods to obtain information about disordered proteins are NMR spectroscopy, circular dichroism (CD), fluorescence spectroscopy, small angle X-ray scattering (SAXS), and Raman spectroscopy (for a review see [17]). In

the NMR experiments such properties as chemical shifts, J-coupling constants, and distance constraints are typically measured. The constraints are derived from the transfer of spin polarization from one spin population to another – a phenomenon known as the Nuclear Overhauser Effect (NOE).

In CD spectroscopy the difference in absorbency spectra of left and right circularly polarized light is measured. Protein's secondary structure fractions can be estimated from CD spectra profiles. For example, proteins with dominant disordered and PPII helical secondary structures have CD spectra with a deep minimum at the wavelengths around 195 nm (illustrated in Fig. 1.5). Unfortunately, CD spectra do not reveal any information on how the secondary structures are distributed in the protein sequence.

Finally, from the SAXS experiments one can estimate the size of macromolecules, such as the radius of gyration of the molecule. Although SAXS experiments do not reveal many details about structural properties of the molecules, it is primarily used when other crystallographic methods can not be applied.

1.1.4 Elastomeric proteins

Elastomeric proteins are found in many biosystems, including plants, insects, humans, etc. For example, elasticity in wheat grain is attributed to a protein called *gluten*. When flour is mixed with water, the resulting dough becomes soft and elastic, whereas if it is dried, it loses the elastic properties and becomes stiff. In clams, the elastomeric protein *abductin* is involved in the primitive swimming action of the mollusks. Being compressed by a muscle that closes the shell, it can rapidly open the shell once the muscle is relaxed.

Spiders use two kinds of the elastomeric silk in their webs: dragline silk, which is stiff and forms radial spikes in the prey catching web, and soft flagelliform silk, that forms web orbs. The latter absorbs the energy of the impacting insect, preventing the insect from bouncing off the web.

The most studied example of elastomeric proteins is *elastin*, which is responsible for the elasticity of the skin and aorta. When the heart ejects a portion of blood, the aorta is stretched, thus smoothing blood flow, and then it relaxes back. This mechanism is more efficient in blood circulation if compared to discrete propulsions of blood. Elastin is also remarkable for its unusual property to become less disordered with the increase of temperature. Elastin is a cross-linked protein in the extracellular matrix. There are two models of elastin's elasticity: the *random-chain model*, that is based on rubber-like (or entropic) elasticity and the *beta-spiral model*, which is based on regular organization of the repetitive structural elements, such as beta-turns [62]. In either model the entropic elastic force arises upon stretching and is the result of damping of internal chain dynamics. Elastin is found only in mammals, while in insects there is a protein with similar elastic properties, called resilin. Recently, the interest in studying resilin is due to synthetic production of this material in the lab, opening great possibilities for industrial and biomedical applications of synthetic resilin, or resilin-like materials [20].

1.2 Resilin

Resilin is an elastomeric protein with remarkable mechanical properties, including high resilience and long fatigue lifetime. Identified originally in the flight systems of desert locusts and dragonflies by Weis-Fogh [63, 64], resilin has since become the subject of intense experimental efforts to elucidate and potentially enhance its rubber-like properties. Resilin was also found in jumping organs of fleas and froghoppers and sound producing thymbals of cicadas.

1.2.1 Crosslinking and resilin fibers

In insects the individual resilin peptides undergo light-induced crosslinking whereby tyrosines form stable di- and trityrosine covalent bonds [3]. Random network of highly flexible chains provide resilin with entropic elasticity, similar to that observed in rubber. The level of disorder in resilin is so high that, even stretched to nearly the breaking point and then slowly dried, resilin did not show any trace of crystallization in X-ray diffraction and electron microscopy experiments [19]. Swollen and unstrained resilin completely returned to its amorphous state, suggesting that resilin might have a long fatigue life. Studying the resilin tendon from the dragonfly, Elvin et al showed that it is highly expressed only during the pupal stage of the insect [20], emphasizing that resilin must function for the rest of the insect's life, undergoing many deformation cycles.

1.2.2 Resilience

Resilience is the ability of a material to recover deformation energy after the stretching (compressing) force has been removed. Quantitatively, resilience is defined as the ratio of the areas under loading and unloading force-extension curves (Fig. 1.1). Thus, the higher the resilience of a material, the less energy is irreversibly transformed into heat upon deformation.

Experimentally, resilience of materials can be measured with the atomic force microscope (AFM) [29]. The experimental setup is illustrated in Figure 1.2, where the bending in cantilever results in the laser beam deflection. Thus, knowing the spring constant of the cantilever, one can compute the force exerted on it at the different levels of tip penetration into the sample.

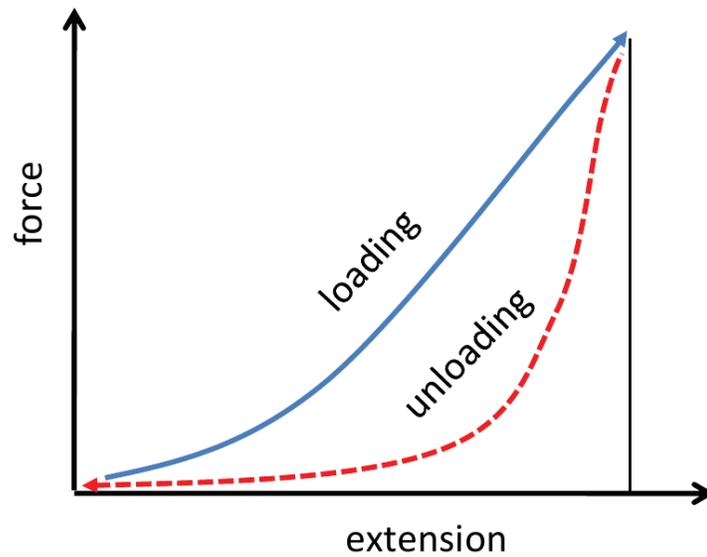


Figure 1.1: Force-extension curves in loading/unloading experiments. The resilience is defined as a ratio of the areas under loading and unloading forces.

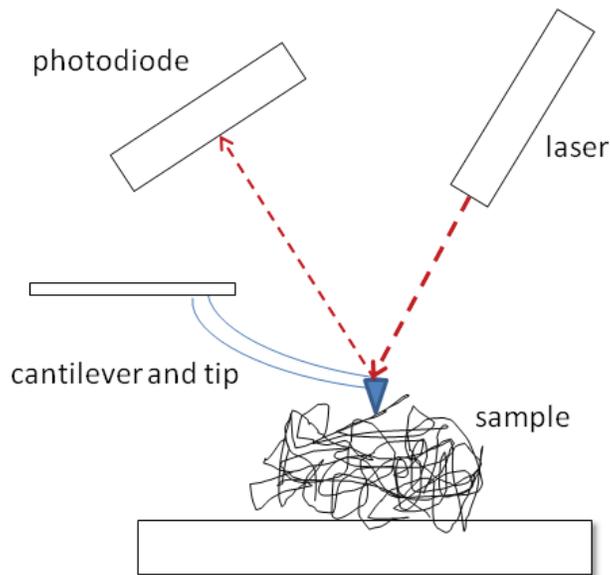


Figure 1.2: Schematic representation of atomic force microscope. The base of cantilever moves is approaching the sample and once tip touches the sample, cantilever bends and photodiode detects deflection in the reflected laser beam. From the data of laser deflections from tip that moves in and out of the sample, resilience can be computed from the resulting force-extension curves obtained.



Figure 1.3: Schematic representation of the domain structure of the gene product CG15920 from *Drosophila* species, which consists of three domains. From left to right these domains are: N-terminal elastic repeats, chitin-binding domain, and C-terminal elastic repeat units.

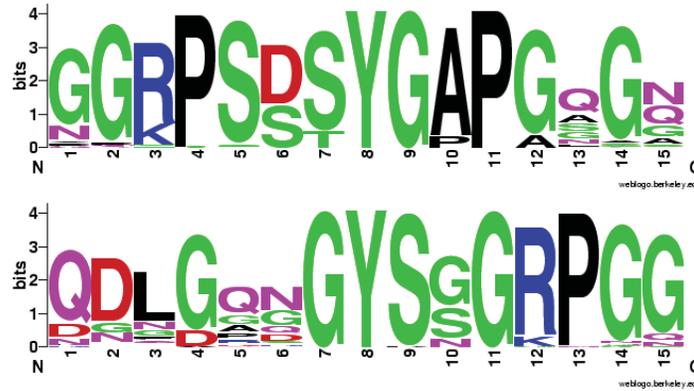


Figure 1.4: Consensus sequences of elastic repeats from N-terminal (top) and C-terminal (bottom) domain in CG15920 gene product. More conserved residues are higher. Tyrosines, for example, are completely conserved in both repeats.

1.2.3 Resilin gene and elastic repeats

Since the early 1960s resilin had remained unnoticed until Ardell et al identified the gene product CG15920 as a tentative *D. melanogaster* (fruit fly) resilin precursor [5]. Comparing the amino-acid sequence of the tentative pro-resilin with the known proteins, the domain structure was proposed. It consists of three regions: N-terminal domain, chitin-binding domain, and C-terminal domain (Fig. 1.3). Only N- and C-terminal domains have repetitive sequence structure and are believed to be responsible for the resilin's elasticity. Although the elastic repeats vary in sequence and length, two consensus motifs (GGRPSDSYGAPGGN and GYSGRPGGQDLG) (Fig. 1.4) were proposed to be the putative units of the elasticity in the protein [5]. The chitin-binding domain is a variant of the so-called RR-2 consensus sequence in proteins from solid cuticles [52] and it doesn't play a role in the elasticity of resilin [5].

1.3 Experimental approaches for studies of elasticity

In a seminal study, Elvin et al cloned and expressed the first exon of the CG15920 *D. melanogaster* resilin gene, which corresponds to the N-terminal domain. The soluble recombinant protein, named Rec1-resilin, was then converted into a resilin-like matrix by photochemically induced crosslinking of tyrosine residues. Elvin and coworkers showed that highly stretched synthetic material has high resilience (90-97%), similar to that of the native resilin [20]. In order to distinguish individual uncrosslinked resilin-like protein/peptides from the crosslinked matrix of resilin, Elvin et al referred to the individual resilin-like peptide as pro-resilin [20].

In another study, a resilin-like elastomer, called RLP12, was designed and synthesized [13]. The constructed peptides had a modular structure, comprising an elastic domain together with some biologically active domains, such as: cell adhesion, growth factor, and material degradation. Resilin-like domains comprised 12 elastic repeats, with the amino-acid sequence based on the N-terminal elastic domain in *D. melanogaster* resilin. Remarkably, similar CD and Fourier Transform Infrared (FTIR) spectra were obtained for RLP12 before and after crosslinking, suggesting that the secondary structures do not significantly change upon crosslinking. CD spectra of modular RLP12 peptides were similar to those reported by Bochicchio et al [11] for a few short resilin-like peptides and to those reported for a single 15-residue repeat from N-terminal domain of *D. melanogaster* [47].

Nairn et al proposed a synthetic construct based on a consensus repeat in malaria mosquito (*A. gambiae*) pro-resilin [42]. Recombinant peptides, labeled as AN16, comprised 16 ideal copies of the repeat sequence AQTPSSQYGAP. Raman spectra of the AN16 before and after crosslinking were similar, suggesting similar distribution of the secondary structures. Particularly, secondary structures from CD and NMR

spectra indicate a lack of stable α -helices and β -sheets. Chemical shifts from the NMR spectra indicate on random-coil configurations. In addition to local structural properties, the radius of gyration of AN16 was estimated from SAXS experiments. The value of $50 \pm 5 \text{ \AA}$ was obtained for a 185-residue AN16 peptide, which is in the range of R_g reported by Kohn et al for denatured proteins [36].

Comparative analysis of two recombinant resilin-like proteins, An16 and Dros16 were shown to have similar material properties to that of Rec1-resilin [38]. These peptides were constructed as 16 ideal copies of consensus repeats in malaria mosquito and drosophila. Using CD, AFM, and tensile testing, the authors did not find significant differences in modulus, elasticity, resilience, or dityrosine content of the synthetic ideal constructs compared with Rec1-resilin.

Based on the consensus sequence of N- and C-terminal elastic domains in *D. melanogaster* proresilin, the level of disorder in two kinds of 15-residue peptides (labeled as DN1 and DC1) was assessed [47]. CD spectra for DN1 and DC1 peptides both had pronounced minimum at the wavelength 195 nm (Fig. 1.5), indicating a high level of disorder and a high level of PPII secondary structures.

In Table 1.1, I summarize relevant experimental data about resilin and synthetic resilin-like peptides.

1.4 Theoretical models of elasticity in proteins

1.4.1 Restoring force

For a polypeptide stretched by an extension δL , the force f has two components, namely, the internal energy component f_U and the entropic component f_S . The change in the Helmholtz free energy due to a stretching force f is $\delta F = f\delta L$. Since

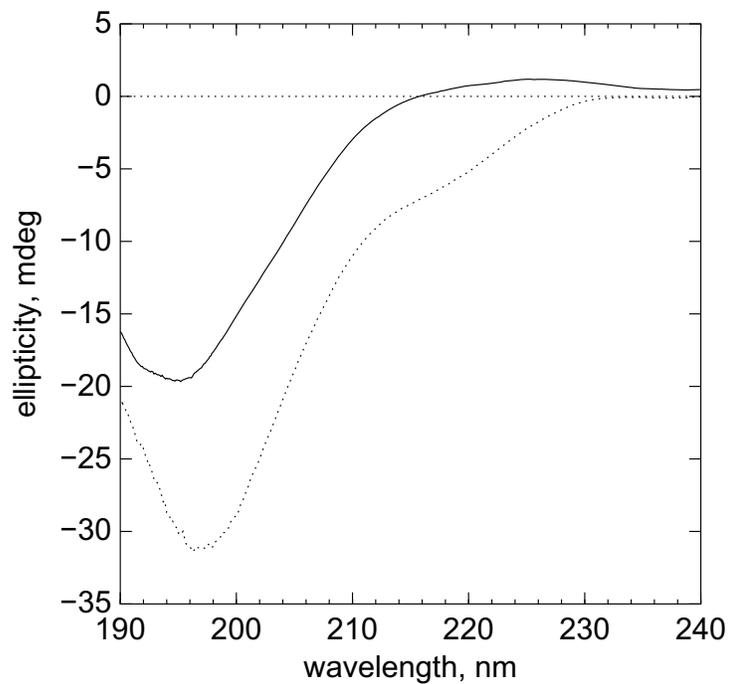


Figure 1.5: CD spectra of 15-residue DN1 and DC1 peptides, constructed as consensus repeats from N- (solid line) and C-terminal (dots) elastomeric domains in *D. melanogaster* resilin. Pronounced minimum at the wavelength around 195 nm indicates on the dominance of disorder and PPII helices.

Table 1.1: Summary of the experimental data on resilin-like peptides.

Name (authors)	Experimental data
rec1-resilin	degree of crosslinking
(Elvin et al, 2005)	high resilience
AN16 synthetic construct	$^3JH_NH_\alpha$ coupling constants
(Nairn et al, 2008)	chemical shifts
	NOE restraints
	radius of gyration (from SAXS)
	sec. structures (from CD spectra)
AN16, Dros16, rec1-resilin	resilience
(Lyons et al, 2009)	
RLP12 synthetic constructs	CD spectra
with other biological domains	
(Charati et al, 2009)	
DN1, DC1 synthetic constructs	sec. structures (from CD spectra)
(Petrenko et al, submitted)	

free energy change is $\delta F = \delta U - T\delta S$, we have

$$f_U = \left(\frac{\partial U}{\partial L} \right)_{N,V,T} \quad (1.1)$$

$$f_S = -T \left(\frac{\partial S}{\partial L} \right)_{N,V,T} \quad (1.2)$$

These two components present two different mechanisms of the restoring force in stretched polypeptides. The energetic component f_U is a result of an increase in po-

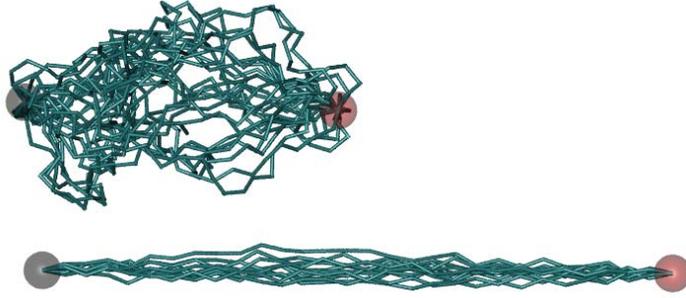


Figure 1.6: Example of damping chain dynamics upon extension of a 30-residue peptide from 50 to 100 Å. For clarity, only backbone α -carbons are shown. Spheres on the left and right are drawn for the constrained atoms. Significant reduction of entropy is achieved by the fact that the peptide is highly disordered in the relaxed state (top).

tential energy upon stretching, and it always becomes dominant for large enough extensions. The entropic component is significant if in the relaxed state the polypeptide chains are highly agitated in thermal fluctuations that are damped upon stretching. Materials in which the entropic mechanism of elasticity is dominant are said to have *rubber-like* (or entropic) elasticity. Such rubber-like elastomeric proteins are the focus of this dissertation. In Figure 1.6 the damping of chain dynamics with extension is shown.

Following Flory *et al* [24] the entropic component of the force can also be written as

$$f_S = T \left(\frac{\partial f}{\partial T} \right)_{N,V,L} \quad (1.3)$$

From the temperature dependence of the tensile force at a given extension one might find energetic and entropic components from a linear fit of $f = a + bT$. The slope and y-intercept from such a fit would give f_S/T and f_U . This approach is the basis of the energy-entropy decomposition of the restoring force in the thermoelastic experiments.

1.4.2 Radius of gyration of unstructured proteins

Recently Kohn et al [36] showed that for the denaturated proteins there is a similar to the random-coil theory behavior between the number of the residues and the ensemble average radius of gyration

$$R_g = R_0 N^\nu \quad (1.4)$$

with $R_0 = 1.93 \pm 0.25$ and $\nu = 0.598 \pm 0.028$ obtained by least-squares fitting (with 95% confidence level) of experimentally derived R_g from 17 proteins and peptides (ranging from 8 to 549 residues) under the denaturation conditions. By fitting SAXS intensity profiles to power-law model of random coil model, Nairn et al [42] estimated the radius of gyration of AN16 peptide to be $50 \pm 5 \text{ \AA}$ that falls in the range of values described by Eqn. 1.4.

1.5 Computational approach to study elasticity in resilin-like peptides

The primary goal of this thesis is to elucidate the nature of the restoring force in small resilin-like peptides and estimate its strength upon different extensions. Due to the limitations of computational power, all-atom simulations of the whole cross-linked resilin matrix is not feasible. We used the repetitive nature of the resilin-like peptides to reduce the entire protein sequence to smaller peptides with a couple of the elastic repeats. Particularly, the structural properties from the simulations of one and two repeats in mosquito resilin were compared with those from the simulation of 16 repeat units.

For stretched peptides, the entropic part of the restoring force was computed by subtracting energetic restoring force from the total force. Several simulations were performed to ensure quasi-equilibrium process.

In order to investigate further the entropic force in resilin-like peptides, I performed fixed-end simulations at different values of stretching, similar to my computational study with fruit fly peptide [46]. The entropic force is proportional to the entropy change upon extension, and therefore it can be calculated from entropy changes. I ignore solvent entropy and only conformational entropy of the peptide's backbone is considered. The conformational entropy is computed using the quasi-harmonic approximation, in which heavy backbone atoms are considered as coupled quasi-classical harmonic oscillators.

Long simulation times are required for either method of finding the entropic force. While in the fixed-end simulations long trajectories are required for conformational entropy convergence (especially, at small extensions); in the SMD simulations slow pulling velocities are needed in order to keep the system in equilibrium.

Then, I compare the entropic forces between resilin-like peptides from mosquito, fruit fly and pea aphid. While resilin-like peptides from other insects were also identified, the interest in the putative pea aphid resilin is motivated by observation that its sequence lacks proline residues. Proline residues are believed to play a crucial role in providing rubber-like elasticity [51].

The coarse-grained simulations were used to simulate larger time and length scales than those that are typical for the all-atom simulations. Particularly, I simulated a system of four mosquito peptides, each comprising 16 repeat units. The coarse-grained model was built from the all-atom simulations of mosquito resilin-like peptide. The force-field had the same structure as in the all-atom simulations, only with different parameters. The parameters were derived from all-atom simulations of one and two repeats. From the coarse-grained molecular dynamics simulations I find better agreement of the radius of gyration with experiment. The effect of cross-linking is discussed.

Chapter 2

Identification of putative resilin-like peptides in genomic databases

2.1 Homology-based search for resilin-like peptides

A homology search of CG15920 *D. melanogaster* (gi:75026432) was performed using the BLAST[2] (Basic Local Alignment Search Tool) program, which finds regions of local similarity between a query sequence and all sequences from a non-redundant protein database. Using a threshold of E-value 1.0 and BLOSUM62 amino acid substitution matrix [27], the BLAST search resulted in 27 sequences. The best matches with E-value up to 10^{-36} were among all 12 known species of *Drosophila* (FlyBase [61]). Therefore, only one sequence (from *D. melanogaster*) was used to represent all *Drosophila* homologs. After removing redundancy to the 90% maximum similarity level, the remaining sequences were separated into two groups, depending whether or not they possessed a chitin-binding domain [8]. I will refer to the group with the chitin-binding domain as RLP-1, and to the another group as RLP-2, respectively.

Table 2.1: Predicted resilin-like peptides, that are highly homologous to CG15920 gene product and have the same domain structure (N-terminal elastic domain, CBD, and C-terminal elastic domain).

Label	Insect	GI code	NCBI annotation
DN	<i>D. melanogaster</i>	75026432	CG15920-PA
TN	<i>T. castaneum</i>	189235130	predicted
BN	<i>Ap. mellifera</i>	66557459	predicted
NN	<i>N. vitripennis</i>	156554483	predicted
PHa	<i>P. humanus corporis</i>	212513300	putative
PHb	<i>P. humanus corporis</i>	212516240	putative
AP	<i>Ac. pisum</i>	193652801	partial

2.2 Resilin-like peptides with chitin-binding domain

The well characterized CG15920 gene product from *D. melanogaster* serves as a representative protein from the RLP-1 group. The other four tentative resilin sequences were from *T. castaneum*, *Ap. mellifera*, *N. vitripennis*, *Ac. pisum*, as shown in Table 2.1. Surprisingly, two predictions from *P. humanus corporis* (human louse) were also homologous to the CG15920 gene. As opposed to other species, lice are the only insects in the group that do not have wings, making the function of resilin in this insect questionable.

The domain structure of the identified peptides is similar to that of the CG15920 gene product (Fig. 1.3), starting and ending with the N- and C-terminal elastic domains that contain multiple repeats. There is a chitin-binding domain (CBD) in the middle of the protein sequence that is believed to play no role in the protein's

elasticity. While the sequence of each elastic repeat (as well as the number of repeats) varies in each elastic domain, the CBD domain is very well conserved among species (described below).¹

The elastic N- and C-domains in all species are highly disordered. This fact is illustrated in Figure 2.1 on the example of the disorder prediction in *D. melanogaster* resilin.

All prediction servers consistently identify disorder in the N-terminal elastomeric domain, and only some servers predict the C-terminal domain to be disordered. Signal peptides (at the beginning of each sequence) and chitin-binding domains were predicted to be most ordered.

The number of repeats in the N-terminal domain in all species varied from 10 to 20 and they were equally distributed in the sequence. By comparing consensus sequences of the elastic repeats, I found that the YGAP-motif, highly conserved in fruit flies, is not well conserved in other species. In fact, the YGPP-motif is more common in PHa, NN, BN, TN peptides. Another evolutionary conservation is the location of prolines with respect to tyrosine (four residues before and three residues after). Surprisingly, in pea aphids (AP), there are no prolines, raising questions about the importance of prolines in elastic repeats. It could also mean that the identified AP sequence is not a resilin-like peptide and has a different function (or properties) in pea aphids. The presence of valines is observed only in PHa and PHb at the fourth and first positions within repeat (Fig. 2.2). The importance of this amino-acid mutation in lice is unknown.

¹The exception is isoform B of resilin CG15920-PB in *D. melanogaster*, which has incomplete sequence of CBD domain with a large gap in the middle.

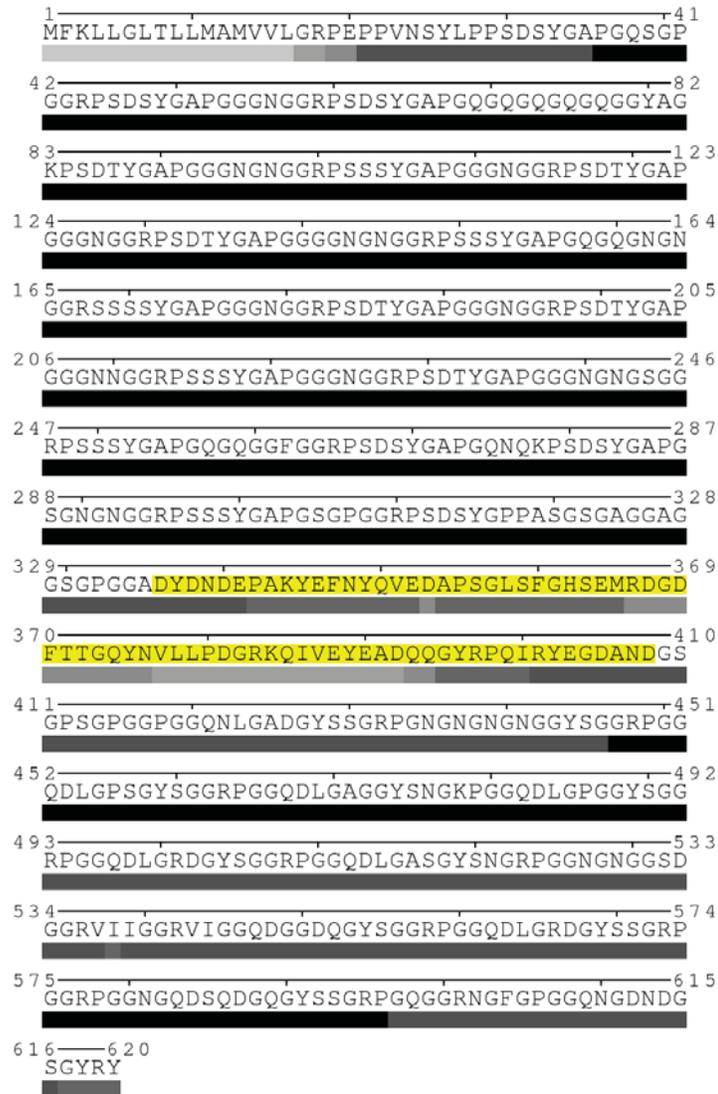


Figure 2.1: Disorder predictions from 7 servers of *Drosophila* resilin CG15920 gene product. Residues with black color were predicted to be disordered by all servers. Highlighted is the chitin binding domain that is predicted to form stable secondary structures.

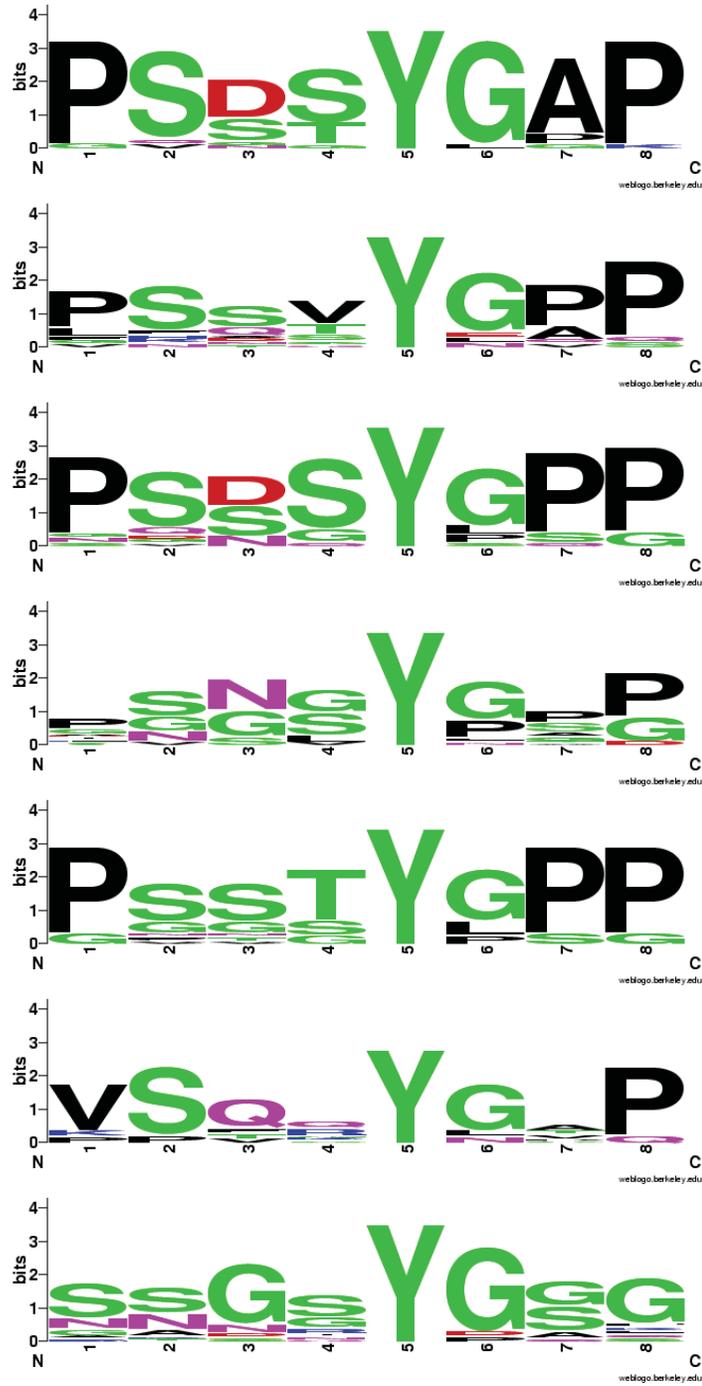


Figure 2.2: Consensus sequences for elastic repeats in RLP-1 peptides (from top to down: DN, PHa, NN, BN, TN, PHb, AP). High conservation of proline and tyrosine residues is observed.

Table 2.2: Predicted resilin-like peptides that are homologous to AGAP002367-PA gene product (mosquito).

Name	Insect	GI code	NCBI annotation
DNb	<i>D. melanogaster</i>	24648035	CG7709-PA
CN	<i>C. quinquefasciatus</i>	170030538	pro-resilin
AN	<i>An. gambiae</i>	158291080	AGAP002367-PA
TNb	<i>T. castaneum</i>	91076994	predicted
BNb	<i>Ap. mellifera</i>	110758137	predicted

2.3 Resilin-like peptides without chitin-binding domain

Despite similarity in consensus repeats of the sequences from this group with the RLP-1 peptides, the chitin-binding domain was missing in all proteins in this group. It was shown that 16 ideal repeats from the consensus sequence of *An. gambiae* peptide from this group have very similar properties [42] to those in *Drosophila* pro-resilin [20].

From eight proteins in this group four proteins belong to the *Drosophila* family and only the representative *D. melanogaster* is retained. Thus, this group consists of five sequences (Table 2.2), with only mosquito resilin-like peptide that is known to have high resilience [38].

In general, the peptides from this group have more elastic repeats than those from the RLP-1 group. Consensus sequences of all repeats in this group (Fig. 2.3) are very similar to those in the RLP-1 group.

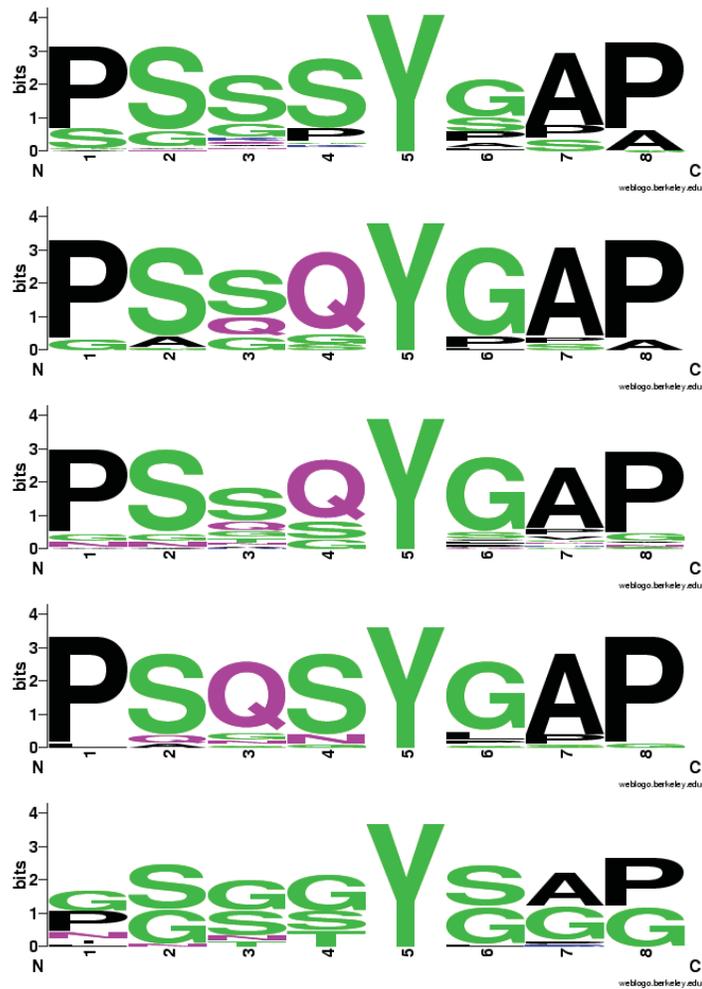


Figure 2.3: Consensus sequences of the elastic repeats in RLP-2 peptides. From top to down: DNb, CN, AN, TNb, BNb. Note conservation of proline positions with respect to tyrosine: four and three residues before and after tyrosine.

```

          330          340          350
Dros1mel1  DYDND E P A K Y E F N Y Q V E D A . P S G L S F G H S E M R D G D
Wasplvit   DEGND E P A K Y E F S Y E V K D D . Q S G S N F G H T E M R D G D
Beelmel    DE . N N E P A K Y E F S Y E V K D E . Q S G A D Y G H T E S R D G D
Beetle1cas EEEST E P A K Y E F E Y Q V D D D . E H N T H F G H Q E S R D G D
Louse1cor1 E D P F A E P A K Y E Y D Y K V Q A S D E T G T E F G H K E S R E N E
Louse1cor2 . D S . V E P A N Y E Y T Y E V Q A A Q Q S G V Q F G H Q E A R D N E
Aphid1pis  . D P L S E P A N Y E F S Y E V N A P . E S G A I F G H K E S R Q G E

```

```

          360          370          380
Dros1mel1  F T T G Q Y N V L L P D G R K Q I V E Y E A D Q Q G Y R P Q I R Y E .
Wasplvit   R A Q G E F N V L L P D G R K Q I V E Y E A D Q D G F K P Q I R Y E .
Beelmel    R A Q G E F N V L L P D G R K Q I V E Y E A D Q D G F K P Q I R Y E .
Beetle1cas K A T G E Y N V L L P D G R K Q V V O Y E A D S E G Y K P K I S Y E .
Louse1cor1 S A R G A Y H V L L P D G R M Q I V O Y E A D E T G Y R P Q I R Y E .
Louse1cor2 S A K G A Y H V L L P D G R T Q V V E Y H A D E Q G F O P Q I R Y E .
Aphid1pis  E A T G V Y H V L L P D G R T Q I V E Y E A D E D G Y K P K I T Y T D

```

Figure 2.4: Sequence alignment of CBD domain in RLP-1 peptides. The VLLPDGR-motif is completely conserved in species. As opposed to sequence variations in the elastic repeats, there are only a few mutations in the CBD domains of different species.

2.4 Chitin-binding domain

Finally, I highlight the sequence and structural properties of the chitin-binding domain, found in all RLP-1 peptides. While this domain is not elastic and therefore does not play a direct role in elasticity, it bears a close similarity to the *D. melanogaster* pro-resilin and makes stronger the assumption that the selected tentative peptides have similar function to resilin in fruit flies. This domain is very well conserved among species (Table 2.4). Particularly, the sequence VLLPDGR (residues 363-369 in *D. melanogaster*) can be used as a new query to look for other RLP-1 peptides. I searched non-redundant protein database for at least one occurrence of the sequence VLLPDGR, and after filtering results, so that each sequence had at least eight segments with the pattern A-XXX-Y-XX-B (where A is either P, G, V, or S; X is any amino acid letter, and B is either P or G), I reproduced a BLAST search of tentative proteins with the CBD domain.

To my knowledge, there is no crystal structure of this domain, so the only structural information I could obtain is from the amino-acid sequence by using several

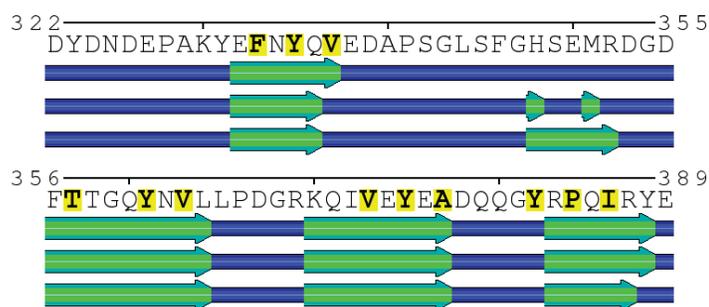


Figure 2.5: Secondary structure predictions of CBD domain in *D. melanogaster* using Sable[1], Porter[48], PSIPRED[31] servers. Fully buried residues as predicted by Sable server are in bold. Most likely, the highlighted residues form hydrophobic core in the folded CBD domain. The image was generated using POLYVIEW server [49].

secondary structure prediction servers. All of them consistently agree on the presence of several β -strands (Fig. 2.5).

2.5 Conclusion

I have presented a systematic classification of resilin-like peptides based on the annotated genes CG15920 in *D. melanogaster* and the AGAP002367-PA gene in *An. gambiae*. Also, I have identified novel putative resilin-like peptides based on sequence similarity and sequence pattern search. The results of this analysis helped me identify mutations in consensus sequences to address the effect of sequence variation on structural and mechanical properties of resilin-like peptides using molecular dynamics simulations.

Chapter 3

Molecular dynamics and related simulation methods

3.1 Molecular dynamics

Since the 1970s, molecular dynamics (MD) has been widely used to study the structure and dynamics of macromolecules, including proteins, nucleic acids, membranes and their complexes. There are two main families of MD methods, defined by the model (and the resulting mathematical formalism) which is used to represent a physical system. In classical MD simulations, molecules are treated as classical objects, resembling very much the “ball and stick” model. Atoms correspond to soft balls and elastic sticks correspond to bonds. The laws of classical mechanics define the dynamics of the system. The quantum or first-principles MD simulations, on the other hand, which started in the 1980’s with the seminal work of Car and Parinello, take explicitly into account the quantum nature of the chemical bond. The electron density functional for the valence electrons that determine bonding in the system is computed using equations of quantum mechanics, whereas the dynamics of ions (nuclei with their inner electrons) is followed classically. Quantum MD simulations represent an important improvement over the classical approach, and they provide insights into a

number of biological problems, including those that involve chemical changes, such as enzymatic reactions. However, they require more computational resources. At present, only classical MD is practical for simulations of large biomolecular systems comprising thousands of atoms over time scales of nanoseconds. In this dissertation only classical MD simulations are considered.

3.1.1 Newtonian dynamics

Molecular dynamics is a simulation method to study properties of many-body systems. It uses computational algorithms to numerically solve the system of Newtonian equations of motion. Normally, a simulation is run for a certain number of time steps that represent real time. Depending on the algorithm used, a time step is usually either 1 or 2 fs, which corresponds to a typical frequency of bond fluctuations. At every k-th time step the positions of atoms are updated according to the forces acting on them (Eq. 3.1).

$$m_i \frac{dv_i^k}{dt^k} = \sum_{j \neq i} F_{i,j}^k \quad (3.1)$$

$$F_{i,j} = -\frac{\partial U_i}{\partial x_j} \quad (3.2)$$

During a simulation a snapshot of the simulated system is saved every n-th time step for later analysis. At the end of simulation one obtains a number of snapshots which is called the trajectory. Since molecular systems are stochastic in nature, the point of molecular dynamics simulation is mostly to observe average values of different quantities, instead of a finding a particular value. Due to the small value of the time step, and relatively large size of the simulated systems, the values computed from the same trajectory are not statistically independent. This results in poor sampling for some observables. One way to evaluate statistical independence is by computing the autocorrelation function, $c(\tau) = \langle (x_{t+\tau} - \mu)(x_t - \mu) \rangle / \sigma^2$, from which the autocorrelation time, τ , can be found using the relation $c(\tau) \approx \exp(-t/\tau)$. Statistical

independence is achieved either by saving frames separated in time by more than τ or by obtaining several trajectories of the same system and then averaging over all trajectories.

3.1.2 Numerical integration of the equations of motion

The aim of the numerical integration of Newton's equations of motion is to find an expression that defines positions vs. time in terms of the already known positions at time t . Because of its simplicity and stability, the Verlet algorithm is commonly used in MD simulations. The basic formula of this algorithm can be derived from the Taylor expansions for the positions $r_i(t + \Delta t)$ and $r_i(t - \Delta t)$ and it reads:

$$r_i(t + \Delta t) \cong 2r_i(t) - r_i(t - \Delta t) + \frac{F_i(t)}{m_i} \Delta t^2 \quad (3.3)$$

The corresponding velocities can be calculated from the positions or propagated explicitly as in the alternative leap-frog or velocity Verlet schemes. The exact trajectories correspond to the limit of an infinitesimally small integration step. It is, however, desirable to use larger time steps to sample longer trajectories. In practice, the time step is determined by the fastest motion in the system. Bonds involving light atoms vibrate with periods of several femtoseconds, implying that the time step should be on a sub-femtosecond scale to ensure stability of the integration. Although the fastest and not crucial vibrations can be eliminated by imposing constraints on the bond length in the integration algorithm, a time step of more than 5 fs rarely can be achieved in all-atom simulations of biomolecules. On the other hand, coarse-grained simulations deal with groups of atoms which (due to larger mass) move slower, allowing for a significant increase of the integration time step and the overall length of the trajectories.

3.1.3 Force-field

Atoms are represented by points with certain masses and partial charges. The force-field consists of bonded and non-bonded interactions. Bonded interactions include: bond, angle, dihedral angle and improper angle terms. Non-bonded interactions include the van der Waals potential and Coulomb interaction due to partial charges on atoms. The van der Waals potential consists of two terms: repulsion, arising from the Pauli exclusion principle, and an attraction term from induced dipole interactions. While the potential energy, describing bond vibrations and bond angle vibrations, have a unique set of spring constants, for the dihedral angles multiple interactions can be defined.

The force is calculated as a partial derivative of a set of classical potentials, called *the force field* (Eqn. 3.2). There are several force field parameterizations, based on experimental data and quantum mechanical computations, and I use the CHARMM27 force-field [40]. While bond lengths and bond angles are governed by quadratic potentials (Eqns. 3.4, 3.5), the dihedral angle interaction term has the form (Eqn. 3.6).

$$U_l = k_l (l - l_0)^2 \quad (3.4)$$

$$U_\theta = k_\theta (\theta - \theta_0)^2 \quad (3.5)$$

$$U_\phi = k_\phi [1 + \cos(n\phi - \phi_0)] \quad (3.6)$$

The Lennard-Jones potential is used for the van der Waals interactions (Eqn. 3.7)

and electrostatic interactions are described by the Coulomb potential.

$$U_{i,j}^{LJ} = \epsilon_{i,j} \left[\left(\frac{R_{i,j}^{min}}{R_{i,j}} \right)^{12} - 2 \left(\frac{R_{i,j}^{min}}{R_{i,j}} \right)^6 \right] \quad (3.7)$$

$$\epsilon_{i,j} = \sqrt{\epsilon_i \epsilon_j} \quad (3.8)$$

$$R_{i,j}^{min} = \frac{1}{2} (R_i^{min} + R_j^{min}) \quad (3.9)$$

$$U_{i,j}^{Coulomb} = \frac{q_i q_j}{4\pi\epsilon_0 r_{i,j}} \quad (3.10)$$

Due to the limitations imposed by currently available computational speed, the simulated systems of interest are usually much smaller than the real system. For example, in the simulations of a protein in solution, it is not feasible to create a water box with dimensions much larger than the size of the protein, since the majority of the computations will be spent on water interactions. On the other hand, taking a water box of small size introduces artificial boundary effects. One way to overcome such effects is by using periodic boundary conditions (PBC). A protein of interest is placed in a water box (unit cell), and the whole system is infinitely replicated in space along unit cell vectors. Since van der Waals interactions decrease faster (r^{-6}) than the electrostatic interactions (r^{-1}), one can use a cutoff distance to truncate the interactions at certain distance. But such a truncation introduces a discontinuity in the van der Waals force at the cutoff distance since force is a derivative of the corresponding potential (Eq. 3.2). One of the solutions is to use a switching function that smoothly brings the potential to zero at the cutoff distance. With the electrostatic potential, the usage of a cutoff distance introduces large errors in force calculations due to the long-range nature of the potential. For simulations in PBC one can use the particle-mesh Ewald (PME) algorithm [15, 21] that is based on Ewald summation [22]. The idea behind Ewald summation is to separate the electrostatic potential into a short-ranged term that converges quickly in real space and a long-ranged term that converges quickly in reciprocal (Fourier) space. For a system of N particles the method has better scaling ($N \log N$) when compared to direct pairwise summation

(N^2) .

3.1.4 Statistical ensembles

A statistical ensemble model considers many microscopic states that correspond to the same macroscopic conditions of the system. This formalism is related to physical experiments, where macroscopic conditions, like temperature or pressure, can be imposed on a system, but microscopic conditions are not under control.

Three important ensembles considered in thermodynamics are: the microcanonical (NVE), the canonical (NVT), and the grand canonical ensembles. The NVE ensemble is an ensemble of systems with the same total energy, that is, when the system is not interacting with the surrounding. In the NVT ensemble heat transfer between the system and coupled to it heat reservoir is possible. In the grand canonical ensemble exchange of particles is allowed, thus, volume, temperature and chemical potential are fixed. In the thermodynamic limit, that is for large systems when fluctuations become negligible, the thermodynamic observables from all ensembles converge to the same values.

3.1.5 Langevin dynamics

Langevin dynamics (LD) is governed by the following stochastic differential equation

$$m_i \frac{dv_i}{dt} = \sum_{j \neq i} F_{i,j} - \gamma v_i + \eta_i(t) \quad (3.11)$$

The advantage of using stochastic equations over the deterministic Newtonian dynamics is in the random noise term, which can help crossing the potential energy barriers. A Langevin equation is an extension of the Newtonian equation (Eq. 3.1) in which two additional terms are added to the total force on a particle. One term, $-\gamma v$, represents a viscosity of a solvent and the other term is a stochastic force. For

the i -th particle, the stochastic force is

$$\eta_i(t) = \sqrt{2m_i\gamma kT} R(t) \quad (3.12)$$

where $R(t)$ is a Gaussian random number with zero mean and variance of 1. The factor before $R(t)$ comes from the fluctuation-dissipation theorem

$$\langle \eta_i(t) \rangle = 0 \quad (3.13)$$

$$\int_0^\infty \langle \eta_i(0)\eta_i(t) \rangle dt = 2m_i\gamma kT \quad (3.14)$$

$$(3.15)$$

If there were only a viscosity term, the system would eventually be damped down to zero Kelvin motion. Conversely, the stochastic force, without the damping force, would continuously add extra energy to the simulated system, thus raising the temperature to infinity. The balance of two Langevin terms allows one to control the simulated system by coupling it to a constant temperature reservoir. Therefore, a constant temperature ensemble (NVT) can be generated.

3.2 Energy-entropy decomposition

For the NVT ensemble the maximum work that can be extracted from system is limited by the difference of the Helmholtz free energy (A), defined as

$$A = U - TS \quad (3.16)$$

In general, if one is interested in a system transition from state 0 to state 1, a switching parameter λ can be introduced that varies smoothly between 0 and 1 during such transition. An intermediate state is described by a potential

$$V_\lambda = (1 - \lambda)V_0 + \lambda V_1 \quad (3.17)$$

The thermodynamic relationships are

$$\Delta U = \partial(\beta\Delta A)/\partial\beta \quad (3.18)$$

$$T\Delta S = \beta(\partial\Delta A/\partial\beta) \quad (3.19)$$

which satisfy $\Delta A = \Delta U - T\Delta S$. Using these equations, there are three methods which can be used to decompose the free energy change into the energetic and the entropic components - thermodynamic perturbation (TP), thermodynamic integration (TI), and a method derived from the Jarzynski equality (JE).

3.2.1 Thermodynamic integration and perturbation methods

In TI an integral expression for the ΔA is used

$$\Delta A = \int_0^1 \langle \partial V_{\lambda'} / \partial \lambda \rangle_{\lambda} d\lambda \quad (3.20)$$

$$(3.21)$$

In TP the free energy difference between two values of λ is computed as an ensemble average at one λ

$$\Delta A = -\frac{1}{\beta} \ln \langle \exp(-\beta[V_{\lambda'} - V_{\lambda}]) \rangle_{\lambda} \quad (3.22)$$

$$(3.23)$$

3.2.2 Jarzynski equality

For a quasi-static process from the second law of thermodynamics the average work done on a system $\langle W \rangle$ equals the free energy difference ΔF , and the work is always greater than ΔA for a non-equilibrium process

$$\langle W \rangle \geq \Delta A \quad (3.24)$$

The Jarzynski equality (JE) allows one to calculate free energy differences from non-equilibrium work

$$\Delta A = -kT \ln \langle \exp(-\beta W) \rangle \quad (3.25)$$

$$W = \int_0^1 \partial V / \partial \lambda d\lambda \quad (3.26)$$

Due to the presence of the exponent in this equation, only low values of work contribute significantly to the free energy differences. Since one usually deals with a limited number of trajectories, the estimate of the average is accurate only if the fluctuations of work are comparable to the temperature fluctuations kT . Indeed, from the cumulant expansion of the logarithmic expression in the right hand side of Eqn.3.25, one gets up to the terms of the second order

$$\Delta A \approx \langle W \rangle - \frac{\text{var}(W)}{2kT} + \dots \quad (3.27)$$

I follow recently published method, where Jarzynski equality is used to calculate free energy differences from pulling (SMD) computational experiments [43]. In JE formalism one generates N finite time trajectories of 0-1 transition and records the non-equilibrium work W . The free energy difference is computed from arithmetic average of W_i from each trajectory as

$$\Delta A = -\frac{1}{\beta} \ln \langle \exp(-\beta W_i) \rangle \quad (3.28)$$

$$= -\frac{1}{\beta} \ln \left[\frac{1}{N_s} \sum_i \exp(-\beta W_i) \right] \quad (3.29)$$

$$(3.30)$$

The main difference of this formalism from the above two is the computation of the non-equilibrium work from each trajectory.

For pulling experiment one can introduce a constraining potential

$$V_\lambda = V_0 + \frac{k}{2} (L - \lambda_t)^2 \quad (3.31)$$

where L is actual end-to-end distance and λ_t is an external parameter that represents a virtual point moving with constant speed along certain direction $\lambda_t = \lambda_0 + vt$. Since λ_t is a function of time, we can change the integration variable in Eqn. 3.26 from λ to t and obtain the same expression for the work as in [44]

$$W = \int \frac{\partial V_\lambda}{\partial \lambda} \frac{\partial \lambda}{\partial t} dt \quad (3.32)$$

$$= -kv \int (L - \lambda_t) dt \quad (3.33)$$

$$(3.34)$$

3.3 Order parameters

Generalized Lipari-Szabo S^2 order parameters [37] from trajectories for backbone N-H vectors [12] are computed,

$$1 - S^2 = \frac{4\pi}{5} \sum_{m=-2}^2 cov(Y_2^m, Y_2^{-m}) \quad (3.35)$$

$$cov(a, b) = \langle ab \rangle - \langle a \rangle \langle b \rangle \quad (3.36)$$

where $cov(Y_2^m, Y_2^{-m})$ covariance of the normalized spherical harmonics $Y_l^m(\theta, \phi)$ of the second order ($l = 2$). In the absence of motion S^2 is maximal and equal to one. This approximation is valid when local fluctuations are separable from global motion.

Recently, a more general method, termed isotropic reorientational eigenmode dynamics (iRED), for order parameter calculation from MD trajectories was introduced [50]. For a system of N normalized N-H bond vectors $\{e_i\}$ covariance matrix M_{ij} is computed as

$$M_{ij} = \frac{1}{2} \langle 3(e_i e_j)^2 - 1 \rangle \quad (3.37)$$

where vectors $\{e_i\}$ are measured with respect to laboratory frame. Since only the angle between bonds is present in the above equation, this method has an advantage

in that no assumption about motion separability is needed. By solving eigenvalue problem $M|m \rangle = \lambda_m|m \rangle$, order parameters can be computed

$$1 - S_j^2 = \sum_{m=1}^{N-5} \lambda_m ||m \rangle_j|^2 \quad (3.38)$$

where the sum is over all eigenvalues except the five largest ones [50].

3.4 Coarse-grained methods and models

For any coarse-grained model the main problem is to derive the governing potential that has many unknown parameters for each kind of atomic interaction. The repetitive nature of the studied peptides simplifies the construction of such potential by reducing the number of the unique interaction pairs. From the coarse-grained model one can also get structural parameters, such as the radius of gyration or the end-to-end distance. Indeed, the goal of any coarse-grained model is to integrate high-frequency modes into a coarse-grained model, so that the global peptide's properties can be studied.

3.4.1 Radial distribution function

For a system with N particles, a radial distribution function, $g(r) = g^{(2)}(r_{1,2})$, is a special case of many-particle correlation function, such that if $\rho = N/V$ is the particle density, the quantity $\rho g(r)dr$ is the probability of finding two molecules separated in space by distances between r and $r + dr$. Integrating over all space one gets the normalization condition of $g(r)$

$$N - 1 = \int_0^\infty \rho g(r) 4\pi r^2 dr \quad (3.39)$$

since the probability to find the remaining $N - 1$ particles at any distance away from the particle at the origin is equal to one.

3.4.2 Inverse MC

A method of automatic adjustment of the coarse-grained interaction potential from known radial distribution functions (RDF) was proposed [39]. The RDFs can be either known from experiments or computed from the all-atomic simulations (MD or MC). Without loss of generality, the idea can be illustrated on a system of N identical particles in a cube with size A . Suppose the dynamics of the particles is governed by a pairwise potential $V_{ij} = V(r_{i,j})$ which depends only on the relative distance between particles $r_{i,j}$. The total potential is a sum of the number of pairwise interactions $N_p = N(N - 1)/2$ excluding self-interactions

$$H = \sum_{i=1}^N \sum_{j=i+1}^N V(r_{i,j}) \quad (3.40)$$

To avoid fitting of the potentials V_{ij} to any analytical form it is discretized into M subintervals

$$H = \sum_{\alpha=1}^M S_{\alpha} V_{\alpha} \quad (3.41)$$

where S_{α} ($0 \leq S_{\alpha} \leq N_p$) is the number of particle interactions that belong to the α -th subinterval, and the corresponding interaction potential $V_{\alpha} = V(r_{\alpha})$.

The application of this mechanism to the molecular systems consists of the following steps:

1. Define N interaction centers (coarse-grained points);
2. Define a working distance interval $[0, R_{max}]$ on which the coarse grained potential is applied; then divide it into M subintervals; a suitable choice of R^{max} is to set it to half of the simulation box;
3. Run simulations on fine scale and collect distances between coarse points r_{ij}
4. For each α -th subinterval of $[0, R_{max}]$, compute the number of the corresponding interactions, averaged over trajectory $S_{\alpha}^{ref} = \langle S_{\alpha} \rangle$;

5. Set an initial guess of the interaction potentials V_α^0 ; a reasonable approximation is to use a potential of mean force calculated from RDFs, $g(r_\alpha)$

$$V_\alpha^0 = -kT \ln g(r_\alpha) \quad (3.42)$$

6. for each k -th iteration

- (a) run simulations of coarse scale to obtain a set of $\langle S_\alpha^k \rangle$ and covariance matrix

$$C_{\alpha,\beta}^k = \text{cov}(S_\alpha^k, S_\beta^k) \quad (3.43)$$

$$= \langle S_\alpha^k S_\beta^k \rangle - \langle S_\alpha^k \rangle \langle S_\beta^k \rangle \quad (3.44)$$

- (b) from the differences of S_α between reference value and the one from the coarse-grain simulations

$$\delta \langle S_\alpha^k \rangle = \langle S_\alpha^k \rangle - S_\alpha^{ref} \quad (3.45)$$

find the correction δV_α^k to the initial guess of the interaction potential by solving a system of linear algebraic equations

$$\delta S_\alpha^k = -\frac{1}{kT} \sum_{\beta} C_{\alpha,\beta}^k \delta V_\beta^k \quad (3.46)$$

- (c) update potentials $V_\alpha^{k+1} = V_\alpha^k + \delta V_\alpha^k$ to be used in the next iteration
 (d) quit iterations if the convergence is reached.

Chapter 4

Simulated systems and protocols

4.1 Idealized repeats and resilin-like peptides

In addition to the DN2 and AN2 peptides, that are known to be resilin-like, I simulated other elastomeric proteins (4.1). There is experimental evidence that DC is a little more ordered than DN. The consensus repeat in TN is different from DN by a single mutation of alanine to proline in the YGAP-motif. Two repeats from *Ac. pisum* (AP) were chosen to address the role of prolines in the elasticity mechanism of resilin-like peptides, since there are no prolines in the AP consensus repeat (4.1).

4.2 Simulation protocols

The parallel molecular dynamics (MD) package NAMD 2.6 [34] was used in all simulations. The system coordinates were saved every 0.01 ns for later analysis. The all-atom CHARMM force field [40] was used for peptides and the TIP3P model for water [32]. Constraining covalently linked hydrogen atoms to fixed lengths with SHAKE algorithm [54] allowed us to use the integration time step of 2 fs. For van der Waals interactions the cutoff distance was 12 Å and the switching function started at 10 Å in order to bring the interactions to zero at the cutoff distance. To compute

Table 4.1: Amino acid sequences of single repeats in resilin-like peptides. While the first two repeats (AN and DN) are well established resilin-like repeats with experimentally known physical properties, the other three repeats (DC, TN, AP) were derived by sequence homology and their physical properties have not been studied yet.

Name	Sequence
AN	AQTPSSQYGAP
DN	NGGRPSDSYGAPGGG
DC	DLGQNGYSGGRPGGQ
TN	NGGRPSDSYGPPGGG
AP	GGSGSSGSYGGGSSG

long-range electrostatic forces the particle-mesh Ewald method (PME) was used. Depending on the periodic cell size, the grid point spacing for PME was adjusted to be approximately at 1 Å.

For the simulations in water I performed simulations in two ensembles: NPT ensemble was used for the unconstrained peptides and NVT ensemble – for the peptide with fixed end-to-end distance (L). Periodic boundary conditions (PBC) were applied. For constant pressure simulations, the pressure (1 atm) was controlled by a Langevin piston Nose-Hoover method [41, 23]. For temperature control I used Langevin dynamics (from which all hydrogen atoms were excluded) with friction coefficient of 5 ps^{-1} .

The following protocols were used: unconstrained MD, fixed end-points MD, and constrained MD. The first protocol was used to assess flexibility of the simulated peptides, while the second protocol was used to assess the energetic and entropic components of the elastomeric force. The last protocol was used to directly measure the tensile force corresponding to different extensions and to double check the energy-entropy decomposition of the elastomeric force from the thermoelastic behavior of the

simulated peptides.

4.2.1 Unconstrained molecular dynamics

In the unconstrained simulations the peptides were placed in a water box simulated in the NPT ensemble at a pressure of 1 atm and temperature of 310 K. A minimum number of sodium or chloride ions was added in order to neutralize the system's charge.

4.2.2 Constrained molecular dynamics

In the constrained MD simulations I applied harmonic potentials (with spring constant 10 kcal/mol) to the backbone carbons at the peptide's terminal residues. NVT ensembles were simulated at default temperatures of 310 K. In some cases additional simulations at 355 K were performed to test the temperature dependence of the reported results.

4.2.3 Steered molecular dynamics

Steered molecular dynamics (SMD) consists of fixing one end of the peptide and applying a pulling force at the other end. I used a constant velocity protocol in which a fictitious point is attached by a spring with constant $k=694.79 \text{ pN}/\text{\AA}$ and is pulled with constant velocity $v=1.5 \text{ \AA}/\text{ns}$ along z-axis. While this value of the pulling speed is much higher than that used in AFM experiments, it is small enough to ensure a quasi-equilibrium process during pulling as was tested by the Jarzynski formula (Eqn. 3.27).

The resilience depends on the rate and extent of deformation [26]. While the stretching rate in the experiments is about 5 mm/min (or $0.001 \text{ \AA}/\text{ns}$) [20], the natural stretching rates in insects are much higher. It was shown that energy loss is less than 5% even at 200 Hz (normal frequency of wing beats in insects) [30]. Assuming that the

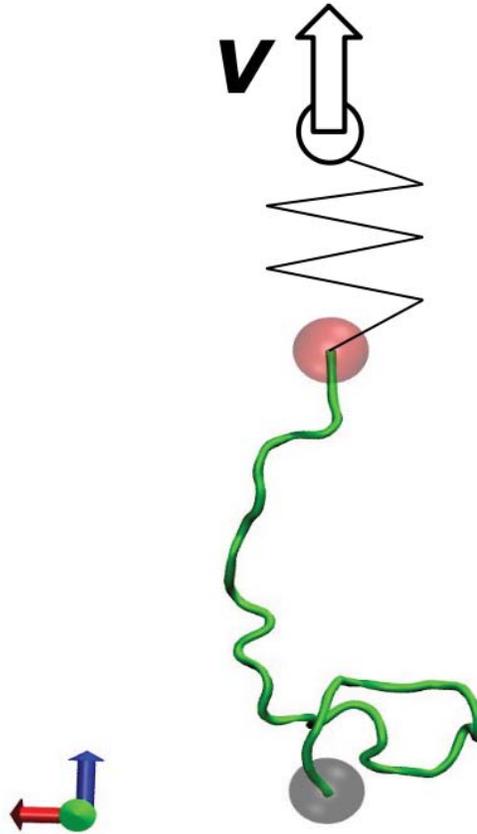


Figure 4.1: Schematic representation of typical SMD simulation setup. One end of the protein is kept fixed, while to the other end a harmonic constraint is applied. The constraint moves with constant speed along end-to-end vector.

tendon is stretched 100-200% of the length at rest, the speed of extension/contraction is around 400-800 mm/s (or 4-8 /ns). The experimentally used pulling speeds are not practical for all-atom MD simulations due to the very small integration step, one femtosecond, in the Newtonian equations, that requires very long simulation times. I used a pulling speed of 1.5 Å/ns that is several times less than the estimated above rate of resilin extension in insects.

4.3 Measures and quantities derived from simulations

The end-to-end distance (L) is defined as the distance between backbone α -carbons of the terminal residues. For a system with N atoms, the radius of gyration was computed as:

$$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^N (\vec{r}_i - \vec{r}_c)^2} \quad (4.1)$$

where \vec{r}_c is the center of mass.

4.3.1 Hydrogen bonds

A hydrogen bond was defined as having the acceptor-hydrogen distance smaller than 2.7 and donor-hydrogen-acceptor angle less than 50 degrees.¹ Computation of the hydrogen bonds was done using VMD package [28]. The end-to-end distance was measured as a distance between backbone carbon atoms of the terminal residues.

4.3.2 Secondary structures

Secondary structure assignments were performed using the Xtlsstr program [35]. The advantage of the Xtlsstr program over widely used DSSP [33] program lies in its ability to identify polyproline II (PPII) type helices in addition to other secondary structures. The comparison of the assignment of helices, β -strands and turns of Xtlsstr and DSSP for the unconstrained MD simulations yield similar results. Small letters in the Xtlsstr assignment denote the end residues of secondary structure stretches. From the comparison of secondary structures from MD simulations with the ones obtained by deconvolution of circular dichroism (CD) spectra, I concluded that the terminal PPII

¹This choice is similar to the molecular dynamics simulations of elastin [7].

Table 4.2: Secondary structure 1-letter codes.

Letter code	Xtlsstr code	Color code	Secondary structure
C	-,p,N	gray	disordered
E	E,e	blue	β -strand
H	H,h,G,g	red	α -helix
P	P	black	polyproline II helix
T	T	green	hydrogen-bonded turn

residue (p) and non-hydrogen bonded turn (N) should be considered as a disordered state (C). For α -helices and β -strands, the terminal secondary structures (e, h, g) were merged with corresponding secondary structure assignment (E, H, G). Finally, the 3_{10} -helix (G) was not distinguished from α -helix (H). Consequently, I have reduced the 10-state secondary structure assignments to a 5-state model (Table 4.2).

4.3.3 Conformational entropy using quasiharmonic approximation

For a one-dimensional simple quantum harmonic oscillator (SHO) with frequency ω and at temperature T (inverse temperature $\beta = 1/kT$), and with the energy levels $E_n = \hbar\omega(n + 1/2)$, the canonical partition function Q can be computed analytically. Summation over all energies yields

$$Q(\beta) = \sum_{n=0}^{\infty} e^{-\beta E_n} \quad (4.2)$$

$$Q = \frac{e^{-\frac{\alpha}{2}}}{1 - e^{-\alpha}} \quad (4.3)$$

where $\alpha = \hbar\omega/kT$ and k is Boltzmann constant. Using standard formulas for the Helmholtz free energy A , entropy S and average energy U

$$A = -kT \ln Q \quad (4.4)$$

$$U = -\frac{\partial}{\partial \beta} \ln Q \quad (4.5)$$

$$S = \frac{U - A}{T} \quad (4.6)$$

one gets

$$A_{sho} = kT \left(\frac{\alpha}{2} + \ln(1 - e^{-\alpha}) \right) \quad (4.7)$$

$$U_{sho} = kT \left(\frac{\alpha}{2} + \frac{\alpha}{e^{\alpha} - 1} \right) \quad (4.8)$$

$$S_{sho} = k \left(\frac{\alpha}{e^{\alpha} - 1} - \ln(1 - e^{-\alpha}) \right) \quad (4.9)$$

The classical limit of the coordinate variance,

$$m \langle x^2 \rangle = \frac{kT}{\omega^2} \quad (4.10)$$

which holds in the limit of small frequencies $\alpha \ll 1$, can be applied since it is the low-frequency motion that contributes significantly to the entropy. Then, the dimensionless variable α can be expressed as

$$\alpha = \sqrt{\frac{\hbar^2}{mkT \langle x^2 \rangle}} \quad (4.11)$$

Andricioaei and Karplus generalized this problem to many degrees of freedom using the quasiharmonic approximation [4]. Replacing equation (4.10) with an eigenvalue problem for mass-weighted covariance matrix of atomic coordinates, one gets

$$0 = \det(\sigma_{i,j} - \lambda I) \quad (4.12)$$

$$\sigma_{i,j} = \text{cov}(x_i \sqrt{m_i}, x_j \sqrt{m_j}) \quad (4.13)$$

where $\lambda = kT/\omega^2$ and I is the identity matrix of rank $3N$. Then, from the $3N - 6$ nonzero eigenvalues λ_i one can compute the entropy and the energy from Eqs. (4.8,4.9) by summing over oscillators with $\alpha_i = \sqrt{\hbar^2/(\lambda_i kT)}$.

It should be noted that, while the entropy S_{sho} defines an upper limit to true entropy ($S < S_{sho}$) [55], the entropy estimation is also dependent on the simulation length of a system due to logarithmic convergence. In my simulations the differences between entropies calculated with this method and with commonly used Schlitter’s heuristic formula [56] were much less when compared to the differences from trajectories of different simulation lengths.

Only heavy backbone atoms were used for entropy calculations as in [53]. The peptide’s rotations and translations were removed before entropy calculations. I used a modified version of CARMA software [25], which implements only entropy calculations, to include internal and free energies into the trajectory analysis.

4.3.4 Order parameters

To calculate order parameters from molecular dynamic trajectories, I used an empirical formula that relies on the number of the close contacts that the backbone amide hydrogen and the carbonyl oxygen have with the protein’s heavy atoms from the other residues [66].

$$1 - S_i^2 = 1.1 - \tanh \left(2.656 \sum_k e^{-r_k^O} + 0.8e^{-r_k^H} \right) \quad (4.14)$$

To compute order parameters from NMR chemical shifts, the random coil index (RCI) was used [9, 10].

Chapter 5

Simulations of idealized resilin-like peptides from anopheles gambiae

5.1 Fluctuations and disorder in AN2 peptides from unconstrained MD simulations

Two repeats of idealized anopheles gambiae peptides were simulated in a water box at the temperature of 310 K. The simulations were carried out in the NPT ensemble for 200 ns out of which the first 5 ns were considered as equilibration and therefore excluded from the analysis.

In order to get insights into the structural properties of AN-based peptides, I compute global structural properties, such as, radius of gyration and end-to-end distance. Then, I compute local properties, such as secondary structures and order parameters S^2 . The results are compared with experimental data.

5.1.1 Radius of gyration

The radius of gyration fluctuates significantly around the mean value of 12 Å (Fig. 5.1). The high level of fluctuations indicates that the free energy landscape is rela-

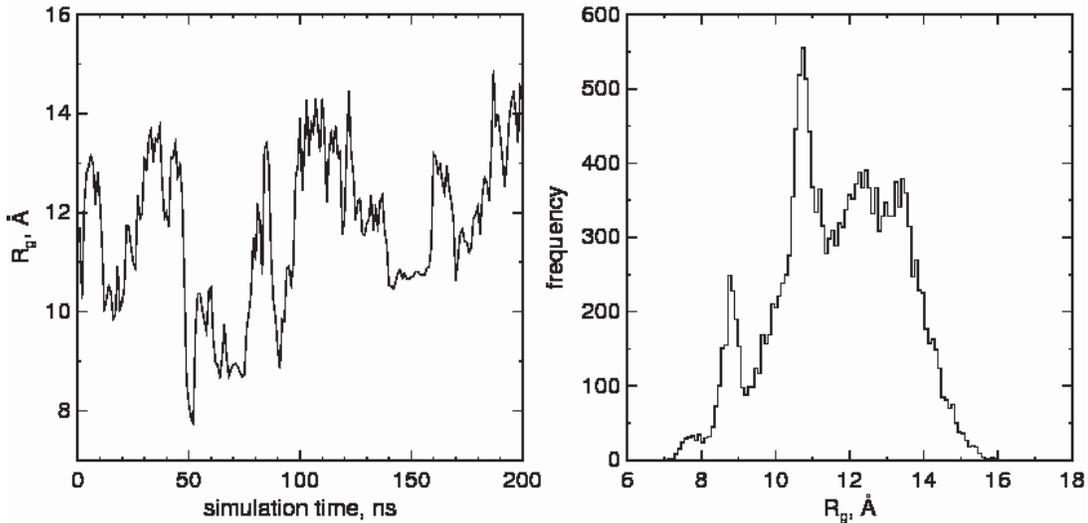


Figure 5.1: Radius of gyration in AN2 peptide from 200 ns trajectory in unconstrained simulations. Fluctuations in the radius of gyration in AN2 peptide (left panel) and its distribution (right panel).

tively flat and has low energy barriers. There are two pronounced peaks corresponding to 9 and 11 Å radii of gyration.

Similarly to the large fluctuations of R_g , I observe large fluctuations of the peptide’s end-to-end distance in the range from 5 to 50 Å with mean value 28 Å (Fig. 5.2).

I also simulated single peptides comprising 1 and 16 repeat units. The input structures were in random conformations. These peptides also exhibited large fluctuations of the radius of gyration. While the average values of the R_g all fit power-law (Eqn. 1.4), the relative fluctuations of small peptides (AN1 and AN2) are much higher than those for large AN16 peptide. Namely, the standard deviations of R_g for AN1, AN2, and AN16 peptides were 10, 16, and 3% of the corresponding averages. Relatively small fluctuations of R_g in AN16 peptide clearly indicate on inadequate sampling of the available conformational space or short sampling time.

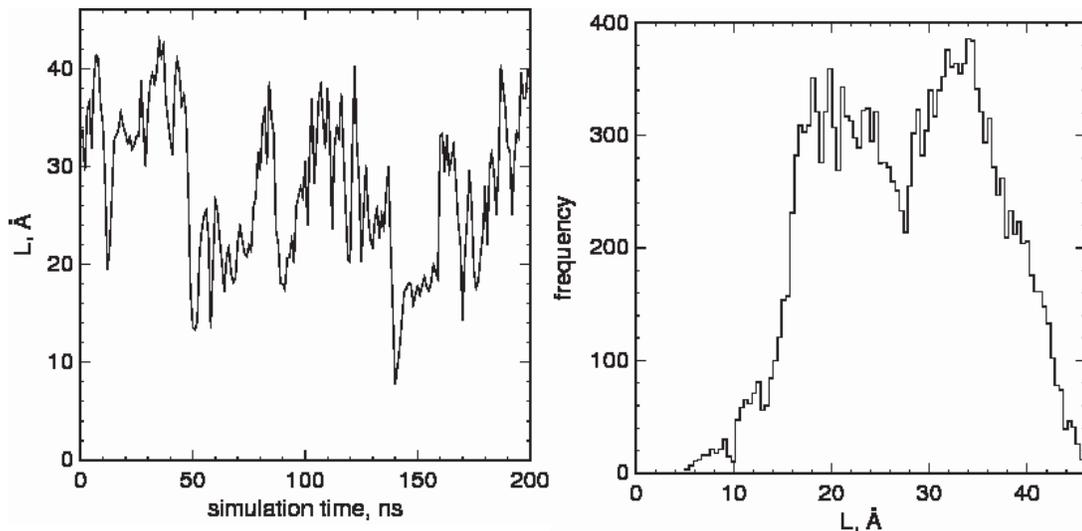


Figure 5.2: End-to-end distance in AN2 peptide from 200 ns trajectory in unconstrained simulations. Fluctuations in the end-to-end distance in AN2 peptide (left panel) and its distribution (right panel).

5.1.2 Secondary structures

The fractions of the secondary structures, observed in the trajectories, are in good agreement with each other and with those from the CD spectra [42, 38]. The highest fraction of the secondary structures comes from the disordered/coiled structure (70%), then the polyproline II helices, turns, and beta-sheets account for 10% each. The smallest fraction of the secondary structures corresponds to the α -helical conformations. Such secondary structures as turns, helices and beta-sheets were not stable. A slightly lower fraction of beta-sheet formation observed in simulations, when compared to that from CD spectra, could be due to a single chain approximation. The proline residues and neighboring residues had most stable PPII helices than other residues.

The observed high level of fluctuations is supported by a very small number of intra-peptide hydrogen bonds, observed in all simulations. Most peptide's hydrogen bonds were formed with water molecules. The fraction of peptide/water hydrogen bonds was computed as the average number of peptide/water hydrogen bonds, nor-

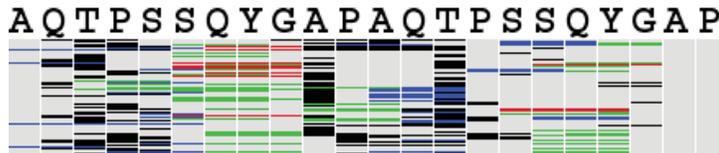


Figure 5.3: Dynamics of the secondary structures in AN2 peptide from the unconstrained MD simulations. From top to bottom, each line represents peptide’s conformation. No stable secondary structure is formed, indicating on large fluctuations in the peptide. α -helices are observed only around SSQY-motif and PPII helices are found around proline residues.

malized by the maximal possible number of HB for a given peptide.

In order to get further insights into structural properties of AN-based peptides I compute order parameters S^2 and compare them with experimental data for AN16 peptides [42]. The simulation temperature of 310 K differs from that used in experiments 278 K.

5.1.3 Order parameters

Computation of C^α and H^α chemical shifts from MD trajectories is not trivial. Therefore, direct comparison of computed chemical shifts with experimentally available data is not possible. Instead, Lipari-Szabo order parameters S^2 are predicted from chemical shifts using the RCI server and compared with S^2 parameters computed from MD trajectories using empirical formula proposed by Zhang et al [66].

5.1.4 Tyrosine self-recognition

After 100 ns of simulation the YGAP motif from the first repeat of the central simulation cell (residues 8-11) interacted with the YGAP motif from the second repeat of the image (residues 19-22). This could indicate the importance of the conservation of YGAP motif among species (as described in Chapter 2) for self association of AN chains and potential further cross-linking. The relative orientation of YGAP-motifs is antiparallel, where tyrosine from one chain interacts with proline from another as

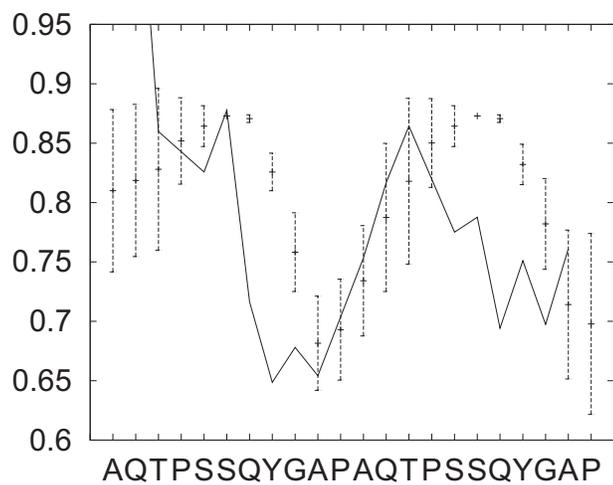


Figure 5.4: The $1-S^2$ parameters, calculated from the experimental C^α and H^α chemical shifts (points) and from MD simulations of AN2 peptide (line).

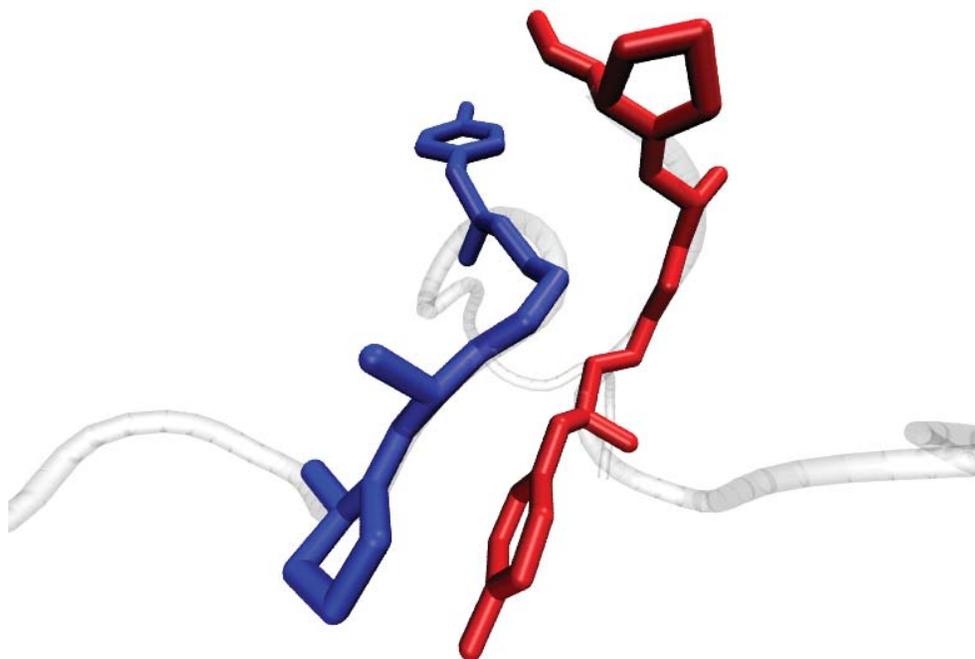


Figure 5.5: Interaction of YGAP-motifs from different chains. The YGAP-motif from the first repeat (left, blue) interacts with the YGAP-motif from the second repeat of AN2 image (right, red).

shown in Figure 5.5. The interaction of YGAP motifs can be observed by significant decrease in the fluctuations of the virtual bonds between tyrosine and proline in YGAP motifs (Fig. 5.6).

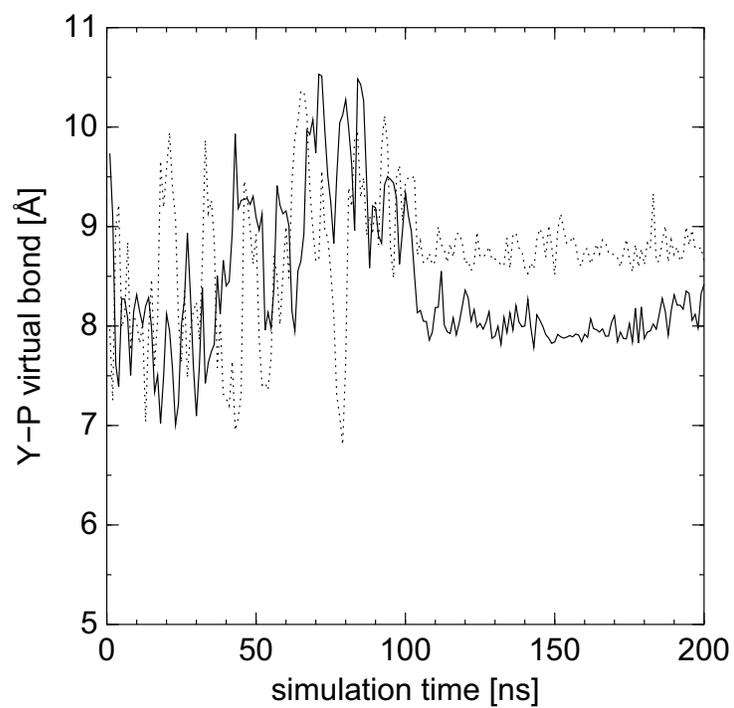


Figure 5.6: Distance between backbone α -carbons of tyrosine and proline that are part of YGAP motif within one peptide. Correlated reduction of fluctuations of YGAP-motif in each peptide is observed once the two motifs start interacting with each other.

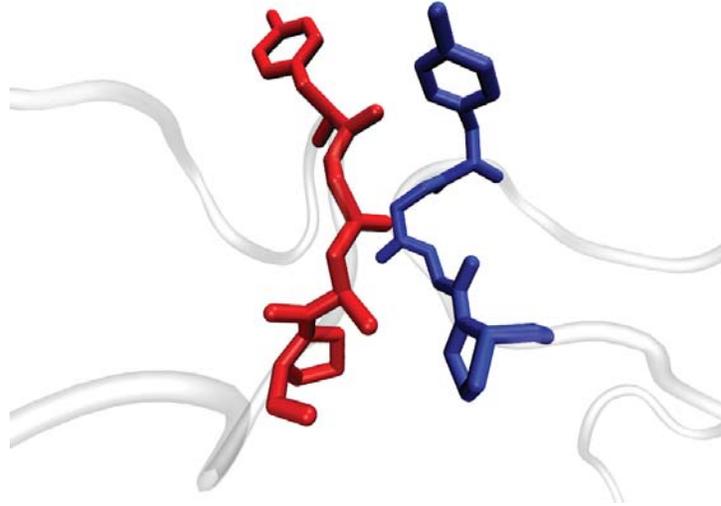


Figure 5.7: Interaction of YGAP-motifs from simulations of AN16 peptide.

In the simulation of 16 AN-repeats there were two interactions of the peptide with its image: above described the YGAP-motifs and anti-parallel arrangement of PSSQY helices. In detail, the YGAP-motif from the fifth repeat (residues 52-55) interacted with the YGAP-motif of the second repeat in the peptide's image (residues 19-22) in a similar orientation as the one observed in AN2 simulation (Fig. 5.7).

The interaction of PSSQY motifs was not observed in AN2 simulations. In AN16 simulation PSSQY motif (residues 92-98) interacted with the same motif in the image (residues 169-173) as shown.

The importance of tyrosine self-recognition is that it gives structural insight into two facts observed in sequence similarity searches: conservation of the YGAP motif and conservation of proline location in the sequence, 4 residues before tyrosine (Chapter 2). My data suggests, that these two evolutionary conservations are designed to facilitate tyrosine self-recognition with the goal to enhance subsequent di- and trityrosine cross-link formation.

I also observe from simulation of the AN16 peptide an interaction of tyrosines (residues 74 and 85) from the adjacent repeats of the same chain (Fig. 5.9). This



Figure 5.8: Anti-parallel arrangement of two helices (sequence PSSQY) from different chains, observed in the AN16 simulations.

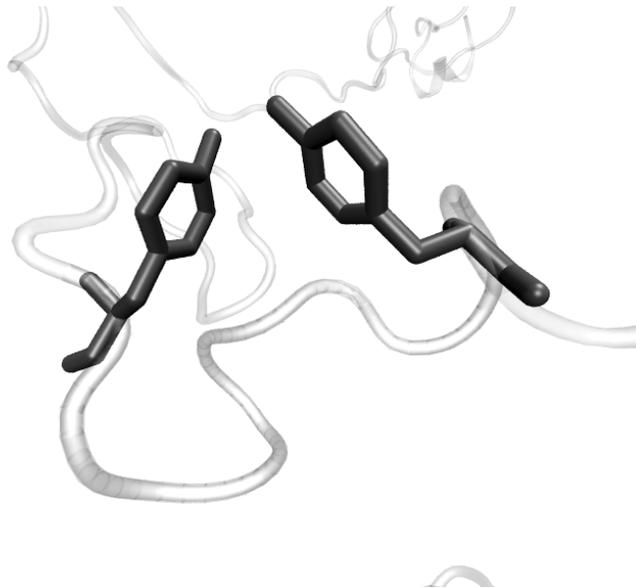


Figure 5.9: Interaction of tyrosines (from left to right: residues 74 and 85) from the adjacent repeats in one chain, observed in the simulation of AN16 peptide. This kind of interaction, that could potentially result in dityrosine bond formation, is not desired, since it "shortcuts" the fluctuations of the segment between them.

kind of interaction is not desirable, since, when cross-linked, it effectively "shortcuts" fluctuations of the segment between two tyrosines and does not contribute to the entropic force upon stretching.

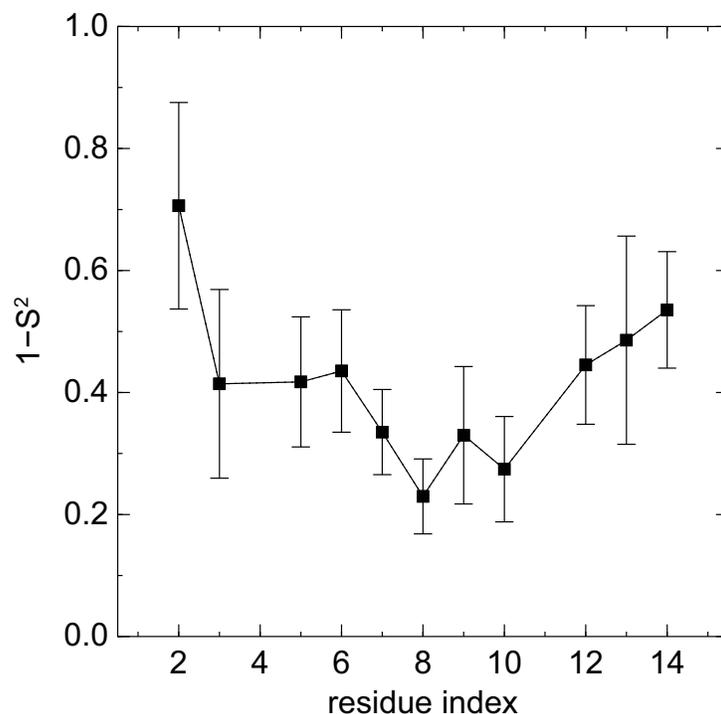


Figure 5.10: The $1-S^2$ parameters the 15-residue AN peptides, cross-linked at tyrosine site. While, with overall reduction of fluctuations as compared to the $1-S^2$ parameters from uncross-linked simulations, there is also a decrease in fluctuations in YGAP-motif (residues 8-11). The same decrease of fluctuations around YGAP-sequence was observed in the uncrosslinked simulations.

5.2 Effect of cross-linking on local structure

In order to investigate the effect of cross-linking on structural properties, I cross-linked two 15-residue AN-based peptides at the tyrosine sites. To make the cross-linking site in the middle of the simulated sequence, I added to the 11-residue consensus repeat an extra 4 residues. Similar to free end simulations of AN peptides (Fig. 5.4), some local order is observed around YGAP motifs (residues 8-11) in both chains, although the overall level of fluctuations was reduced in the cross-linked system (Fig. 5.10).

As expected, the $1 - S^2$ parameters from uncross-linked simulations are in better agreement with experimental parameters, since the latter were derived from the chemical shifts of AN16 in solution, before cross-linking. Secondary structures were similar to the uncross-linked peptides in my previous simulations.

5.3 Structural parameters and entropic force in AN2 from constrained MD simulations

5.3.1 Methods

I simulated a 30-residue peptide, based on the AN consensus sequence in a water box in the NVT ensemble at a temperature of 310 K. I will refer to this peptide as AN2e to distinguish it from the two-repeat AN2 peptide. The extension of AN2 peptide to 30 residues was done to enable later consistent comparison between AN-based peptide with 30-residue DN-peptide. Seven simulations were performed for extensions that correspond to end-to-end distances of 70-98 Å. To the first and last backbone carbons harmonic constraints were applied with spring constant of 10 kcal/mol. The chosen value of spring constant resulted in 0.24 Å fluctuations of end-to-end distance at any extension. Each simulated system was minimized for 200 steps and production run for 15 ns was carried out. Coordinates and energies were saved every 10 ps, forces were saved every ps. The tensile force was calculated as an arithmetic average of the forces in the virtual harmonic springs.

The entropic force was calculated from the conformational entropies. For each extension, the quasiharmonic conformational entropies and peptide bonded energies were computed. Then, a spline function with 100 intermediate points was used to get values for other extensions. Energetic and entropic components of the restoring force were computed using Eqs. 1.2 and 1.1, where the derivative was approximated by the finite difference method from spline points.

5.3.2 Results

Since more stretched peptides sample less conformational space, the conformational entropies are better converged at higher extensions (Fig. 5.11). Conformational en-

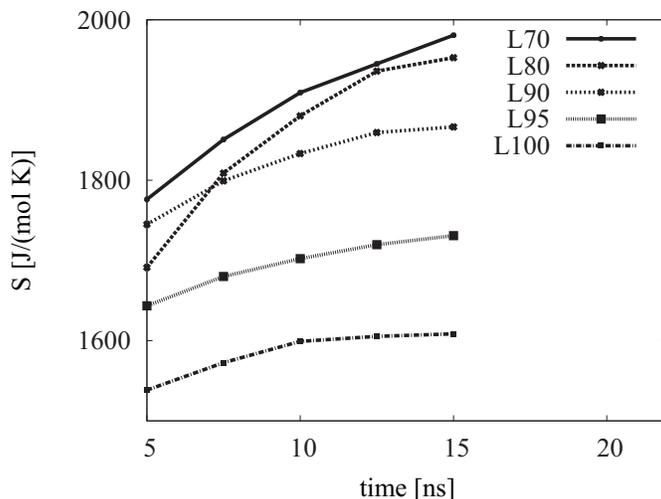


Figure 5.11: Conformational entropies of AN2 at different extensions. The values of conformational entropies from L75 and L85 simulations are not shown, since they are close to entropies from L80, L70 simulations. The entropic decrease with extension is captured by quasiharmonic approximation used to calculate conformational entropies of AN2 peptide.

tropy loss is captured by the quasiharmonic approximation, particularly, the dramatic decrease in the conformational entropy is found when AN2e peptide is stretched to end-to-end distances of more than 85 \AA . Conversely, the change in conformational entropies is only marginal in the range from 70 to 85 \AA .

As opposed to entropies, the tensile forces converge much faster, and within 15 ns of simulations they seem to fluctuate around the average values (Fig. 5.12). Consistent with the changes in the conformational entropies, tensile force is greatly increased in peptides stretched to more than 85 \AA . Under small extensions (70 - 85 \AA) the tensile force fluctuates between 30 and 60 pN .

To further explore the nature of the restoring force I computed entropic and energetic contributions to the total force. Since the number of protein-water hydrogen bonds were the same (around 10 ± 3) for all extensions and practically no protein-protein hydrogen bonds were formed, I assumed that protein-solvent interaction energies do not depend upon extension. Conversely, the peptide bonded interactions, especially backbone bond angle and dihedral interaction terms, strongly depend on

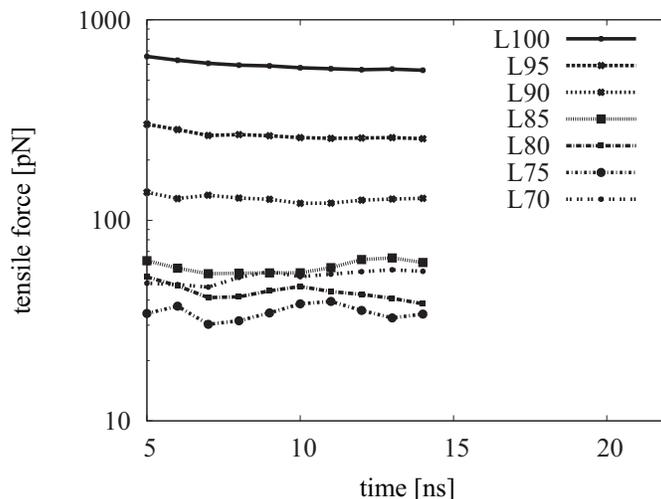


Figure 5.12: Tensile forces in stretched to different extensions AN2 peptides. The force is higher at higher extensions.

extension. Thus, I assumed that only bonded interactions contribute to f_U . From the dependence of f_U on extension I observe that the peptide's bonds are not significantly distorted up to extensions of 93 \AA (Fig. 5.13). The entropic force f_S is always larger than f_U and it is the dominant contribution to the restoring force up to 93 \AA . As one might expect, further extensions, such as $L > 100 \text{ \AA}$, would result in totally energetic nature of the restoring force. The results for smaller extensions ($L < 80 \text{ \AA}$,) are not shown, since both entropic and energetic forces highly fluctuated in that region, resulting in negative forces. This would correspond to expansion force upon compression and it is not studied here.

The combined entropic and energetic forces are in very good agreement with the total tensile force, calculated directly from harmonic constraints (Fig. 5.13).

Finally, I compared these results with the results from the steered molecular dynamics (SMD) simulations. Two sets of SMD simulations were set up with constant pulling speed 1.5 \AA/ns . Five simulations were used in each SMD set. The initial structure from the extended conformation was compressed to the relaxed state. The restoring force was computed using Eqn. 3.27 with up to the second cumulant ex-

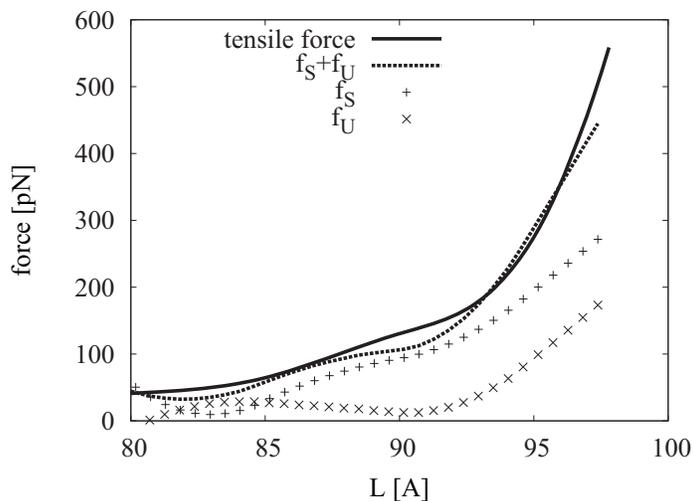


Figure 5.13: Tensile forces in stretched to different extensions AN2 peptides. The sum of the entropic and energetic components of the restoring force is very close to the direct measurement of the restoring force. For extensions less than 92 \AA , the AN2 peptide acts like an entropic spring, while at higher extensions it behaves more like a regular spring.

pression of the Jarzynski equality. Work was computed as an average force times ΔL , which was chosen to be 3.6 \AA . The correction to the free energy difference (second term in Eqn. 3.27) was usually less than 10%. The restoring force from SMD simulations was consistently smaller by around 30 pN than the entropic component of the restoring force from the constrained MD simulations.

Chapter 6

Assessment of disorder in drosophila resilin peptide

6.1 Unconstrained molecular dynamics

The main reason for the flexibility of resilin-like peptides is believed to be due to high solubility of the peptides in water and lack of such flexibility in vacuum. To assess the extent of the solvation effect I performed several vacuum simulations. An input structure in PPII conformation collapses after about 5-10 ns of simulations to a random coil and stays in that conformation with only minimal fluctuations of the end-to-end distance (Fig. 6.1). A qualitatively similar picture was observed when the original peptide was solvated in a 5 Å water shell resulting in only little increase in the fluctuations, but the overall shape of the peptide did not change.

After the solvation of the peptide in bulk water, within the first 5-10 ns of simulations virtually all intra-protein hydrogen bonds were replaced by protein-water hydrogen bonds. As a result, the peptide is highly flexible and remains mostly disordered for the rest of the simulation. The level of fluctuations can be seen, for example, from the time evolution of the end-to-end distances (Fig. 6.2). This observation is consistent with one of the functions of water in this kind of proteins as a plasticizer

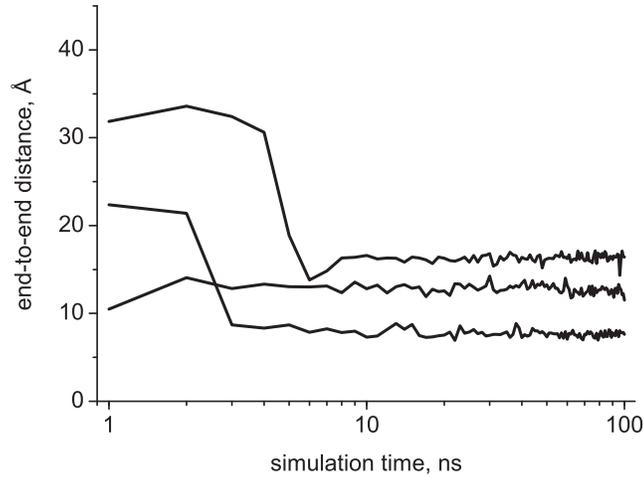


Figure 6.1: The end-to-end distance of DN2 peptide from unconstrained molecular dynamics simulation in vacuum. Regardless of the input conformation, in all simulations the peptide collapsed into a compact conformation and remained in it with only minimal fluctuations.

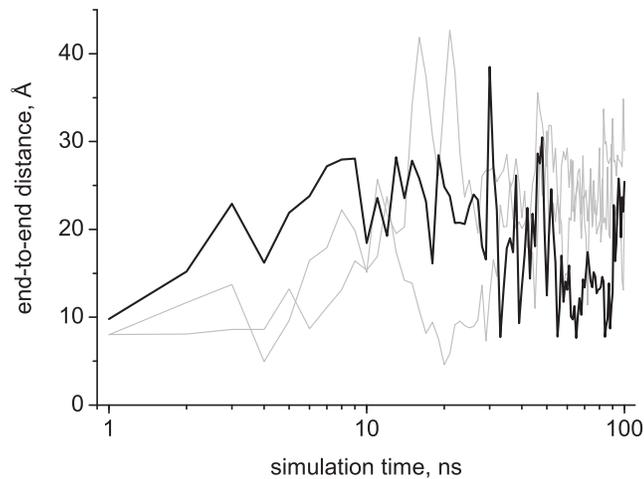


Figure 6.2: The end-to-end distance of DN2 peptides from unconstrained langevin dynamics in water with different values of damping constants (1, 5, 10 ps⁻¹). In all three simulations a compact input structure swell in water within 10 ns and remained highly flexible for the following 90 ns of the simulation time.

that facilitates peptide transitions between different states.

The observed secondary structures correspond mostly to the disordered and PPII

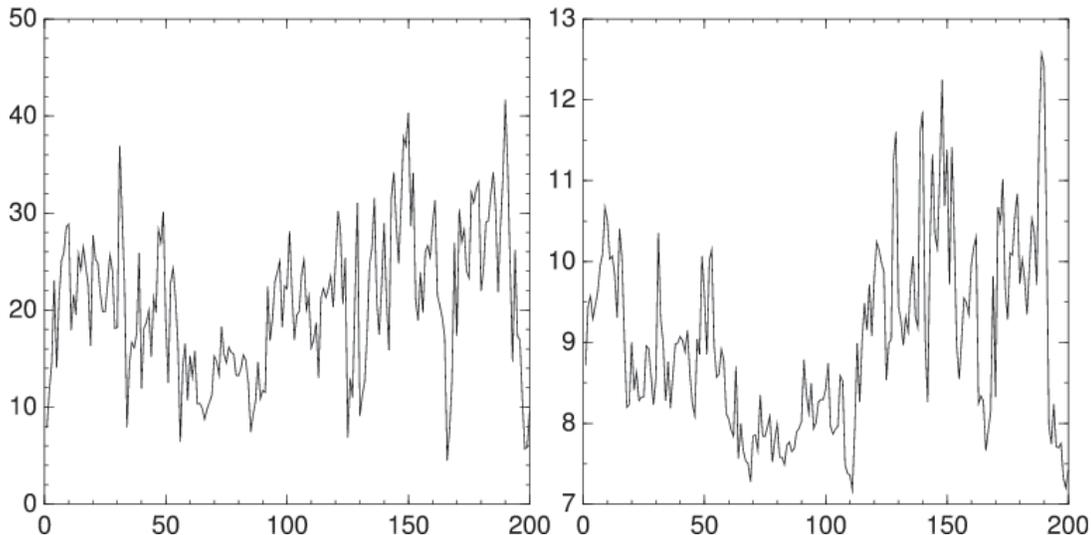


Figure 6.3: End to end distance and radius of gyration observed in unconstrained MD simulation of DN2 peptide. Langevin damping constant 1 ps^{-1} .

formations, that is qualitatively consistent with the secondary structures from CD spectra (to be published). The high flexibility of the peptide is maintained at the local level as well, since I haven't observed any formation of the stable secondary structures within a particular segment. I observed only slight preferences of SSS segment towards β -strand formation, XPX-motif (where X is either R,S,A, or G) towards PPII formation, and GNGG segment towards turn formation.

From the SHO analysis the estimated S , U , A are reported in Table 6.1. I observe that free energy is considerably lowered upon solvation, and the major changes come from the entropy term in the free energy expression. The lowest free energy correspond to $\lambda = 5 \text{ ps}^{-1}$ (Fig. 6.1), which is consistent with the original langevin dynamics study [45] where the optimal value of damping was reported to be around 5 ps^{-1} . Larger values turned the simulated system into non-canonical ensemble while smaller values provided too weak temperature coupling so that the protein and solvent had different temperatures. The dependence of the conformational entropy on the langevin damping constant shows that larger values of damping constant reduce the entropy.

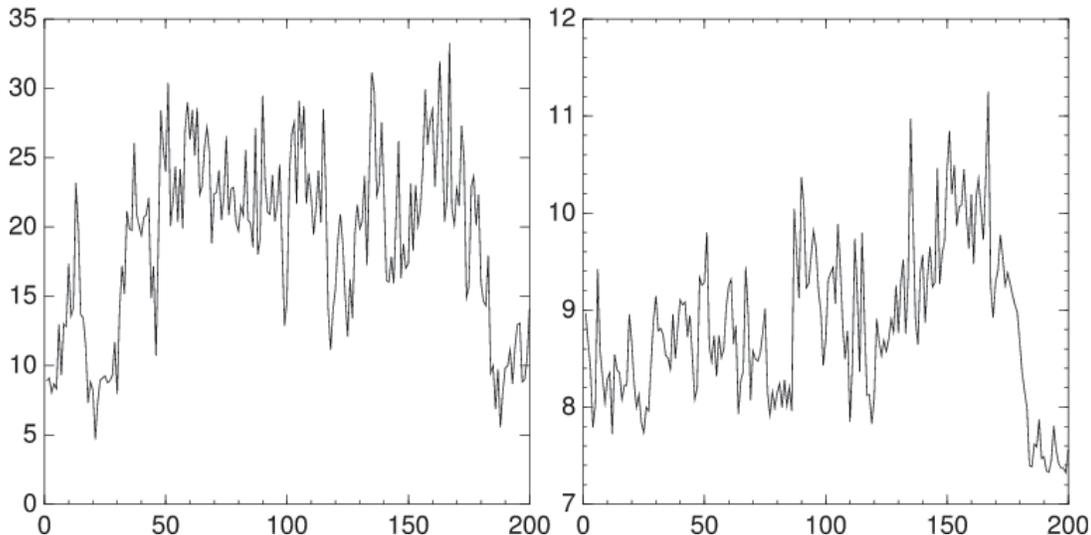


Figure 6.4: End to end distance and radius of gyration observed in unconstrained MD simulation of DN2 peptide. Langevin damping constant 10 ps⁻¹.

Table 6.1: Comparison of the thermodynamic variables from different trajectories, each 100 ns long.

Trajectory	U,(kJ/mol)	TS,(kJ/mol)	A,(kJ/mol)
V1-5	238±6	328±26	-90±32
D1	221	529	-308
D5	218	624	-406
D10	219	585	-366
D100	224	472	-249

From vacuum simulations V1-V5 and water simulations D1,D5,D10 combined trajectories were created. Average end-to-end distance and radius of gyration in water (20.8 ± 7.4 and 16.0 ± 0.2) are both about twice larger than those in vacuum (11.6 ± 3.0 and 7.8 ± 0.3). As seen from the Table 6.2, the secondary structure fractions do not capture peptide’s flexibility in water. This can be explained by different nature of disorder from vacuum and water simulations. While in each of the vacuum simulations the peptide collapses into certain conformation with certain fraction of residues in coil

Table 6.2: Secondary structure fractions from the combined vacuum (500 ns) and water (500 ns) simulations.

Trajectory	C,%	E,%	H,%	P,%	T,%
in vacuum	67	5	4	5	16
in water	54	1	23	5	15

state, in contrast, in water simulations the peptide is highly flexible and residues interchange secondary structure states while still maintaining a constant overall fraction of disorder.

6.2 Constrained molecular dynamics

It was experimentally shown that the elastic force is connected with entropy changes when an external force is applied to an elastic tendon resilin in dragonflies [64]. Stretching a tendon swollen in water to more than twice of its resting length resulted in a restoring force of 20 kg/cm², which was dominantly entropic. In this section I used MD simulation to estimate the entropic part of the elastomeric force in DN2 peptide fixed at the peptide's ends with different extensions. Fixing the peptide's ends greatly reduces the available conformational space, allowing us to estimate entropy loss upon extension.

By stretching the DN2 peptide in vacuum (using SMD protocol) I defined the maximum stretching to correspond to 100-105 Å of end-to-end distance (Fig. 6.5). Since I was looking only at the distortion of the bonded interactions with extension, adding water would not change the results. The average distance between adjacent backbone carbon atoms is around 3.5 Å making the upper limit of L around 105 Å for a 30-residue peptide. Further stretching results in the distortions of backbone angles and bonds as can be seen from almost linear rapid increase of the stretching force with extension (Fig. 6.5).

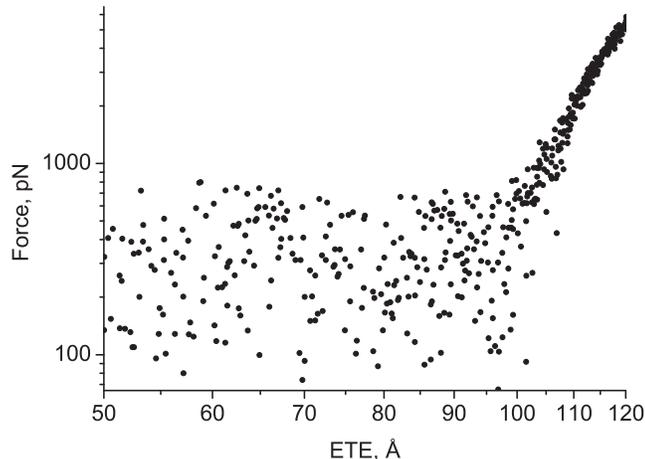


Figure 6.5: Force-extension curve from vacuum SMD simulation of a DN2 peptide, that was originally at some compact state. Significant decrease in force fluctuations corresponds to distortion of peptide’s bonds and bond angles is observed for end-to-end distances larger than 100 Å.

Table 6.3: Setup for the constrained MD simulation of DN2 peptide at different extensions.

Label	Time, ns	L, Å	System size, Å ³
L50	200	50.4	27x27x63
L70	200	68.3	25x26x81
L90	200	88.8	21x21x101
L100	200	100.6	21x21x114

A peptide conformation with the end-to-end distance 100 Å was chosen to represent stretched state and to probe intermediate and relaxed states additional three simulations were performed: L90, L70, L50 (set Table 6.3 for details). All further simulations were performed in water.

First, the level of fluctuations was assessed by comparing secondary structure fractions of DN2 peptide at different extensions. As expected, the fraction of the extended (or β -strands) conformations increased with the increase in stretching; the

Table 6.4: Fractions of the secondary structures in DN2 peptide from the constrained MD simulations.

Name	C	E	H	P	T
L50	71	2	5	15	8
L70	77	2	2	16	3
L90	62	10	0	28	0
L100	62	31	0	8	0

maximum value of 30% observed in L100 simulation. The highest fraction of PPII helices is observed in L90 simulation (Table 6.4), which is consistent with a crude estimate of 3.1 \AA for the distance between two adjacent residues in an ideal PPII helix. Surprisingly, the secondary structure fractions in L50 and L70 simulations turned out to be almost the same with only minor difference the formation of turns - higher fraction observed in L50 simulation as the peptide is more compact than in L70 simulation.

The secondary structures were more stable at higher extensions reflecting more order and entropy loss. Once again, the PSSSY-motif had the highest fraction of β -strand secondary structures, especially when the peptide is maximally stretched at 100 \AA extension (Fig. 6.6).

The number of intra-protein hydrogen bonds was extremely small. The number of protein/water hydrogen bonds was higher at larger extensions (Fig. 6.7). Interestingly, a central asparagine residue formed twice as many hydrogen bonds as serine, glycine or proline residues. Another observation is that the number of protein/water hydrogen bonds decreases for residues close to the terminals. This could suggest that lack of fluctuations suppresses the formation of protein/water hydrogen bonds.

Finally, I report entropic forces at different extensions of the peptide. Due to lack of the intra-protein HBs the internal energy is not expected to change much, therefore,

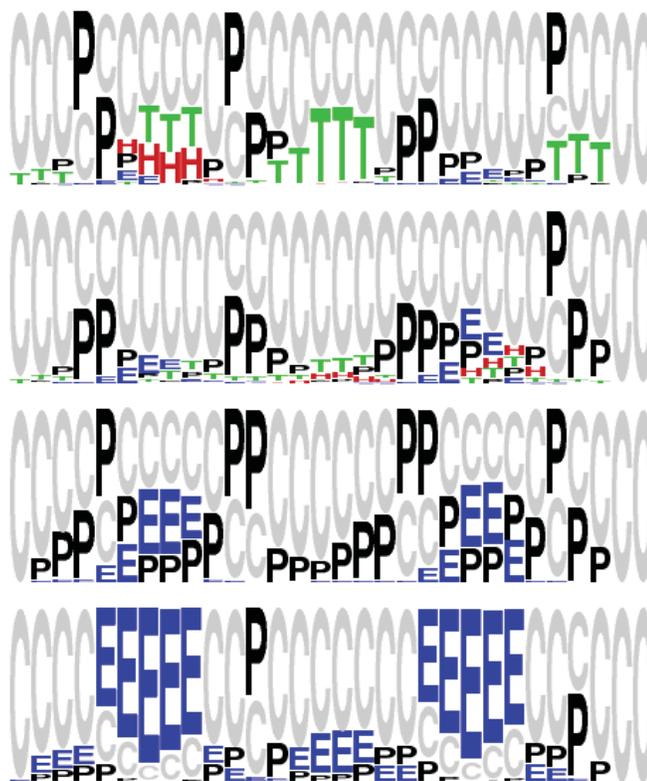


Figure 6.6: Secondary structure fractions observed in the constrained simulations of DN2 peptides stretched to different extensions. Secondary structure notations are given in Table 4.2. From top to down the level of stretching is increased. Simulation labels are: L50, L70, L90, and L100.

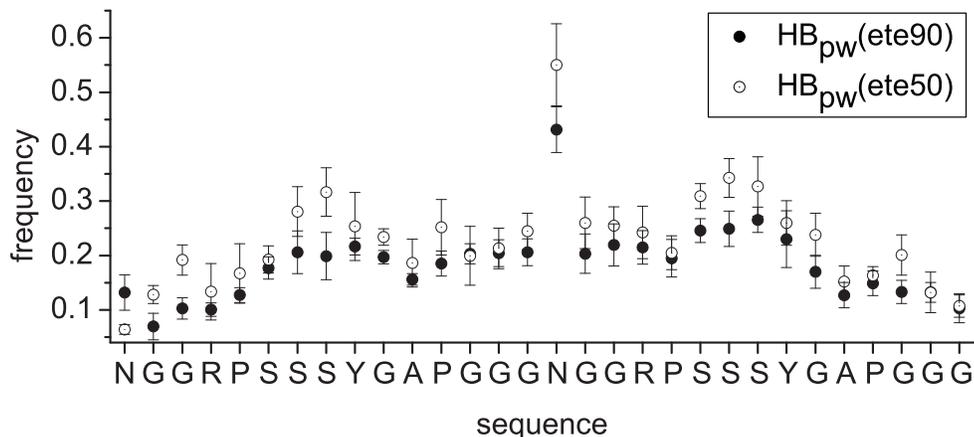


Figure 6.7: Average number of peptide-water hydrogen bonds for each residue. Results from L50 and L90 constrained MD simulations are shown in open and filled circles. The reduction of the number of the peptide/water hydrogen bonds is observed at higher extension.

the entropic component of the elastomeric force becomes dominant. Computing conformational entropies of the peptide from each simulation, I approximate derivative in entropic restoring force (Eq.1.2) by the finite difference method ($f_s \approx -T\Delta S/\Delta L$). Ideally, infinitely long simulation is required to sample the available conformational space. In addition, the conformational entropy of the peptide depends on how well the conformational space is sampled. Therefore, the question of entropy convergence should be addressed. In the unconstrained MD simulations the conformational entropy did not converge in all three simulations after 180 ns (data not shown). In contrast, in the constrained MD simulations, even though I did not get full convergence of the conformational entropies in the L50 (L70) simulations (Fig. 6.8), the entropy loss upon extension is clearly seen in Figure 6.9. As one might expect, the shortest time (around 10 ns) to reach convergence in the conformational entropy calculations was in L100 simulation; twice larger simulation time is required in L90 simulation. The conclusion drawn from the secondary structure distributions that L50 and L70 that these two simulations represent the same (relaxed) state of DN2 is

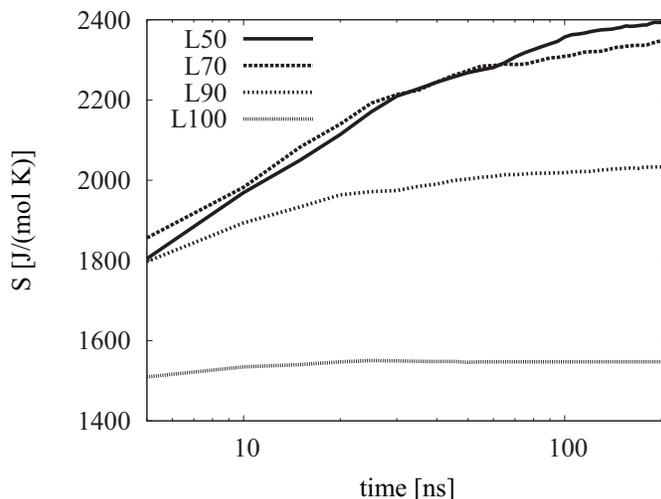


Figure 6.8: Conformational entropies in DN2 peptide from the constrained MD simulations. Significant reduction in conformational entropy is found at 90 and 100 Å extensions. Since less conformational space is available at higher extensions, the conformational entropy convergence is better at higher extensions.

further supported by insignificant difference in the conformational entropies in these two simulations.

To estimate entropic force in DN2 peptide I used spline function with a stride of 1 Å to get additional values of the conformational entropies between simulated extensions. Then, entropic force was computed by finite difference method. The results are shown in Figure 6.10. Since only one simulation was used per extension, I can not report error bars on the calculated entropic forces.

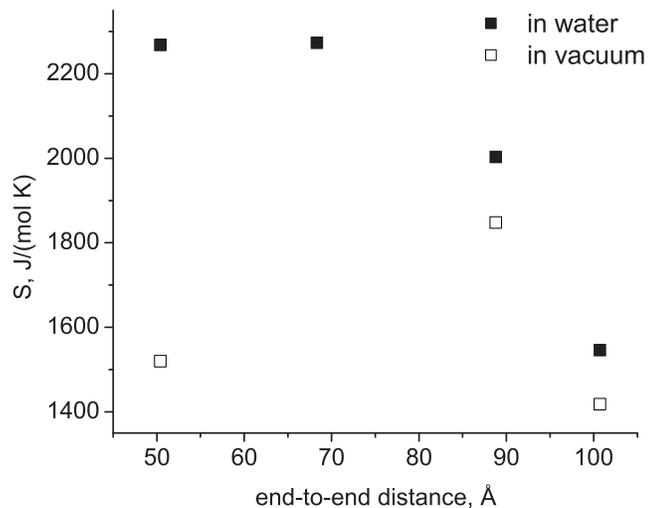


Figure 6.9: Comparison of the conformational entropies in DN2 peptide from constrained MD simulations in vacuum and in water. This figure shows significant role of water in providing high flexibility at the relaxed state in DN2 peptide. While peptide's entropies in vacuum are reduced a little bit for 90 and 100 Å extensions, the significantly low entropy in vacuum is observed when peptide is in the relaxed state (extensions around 50 Å).

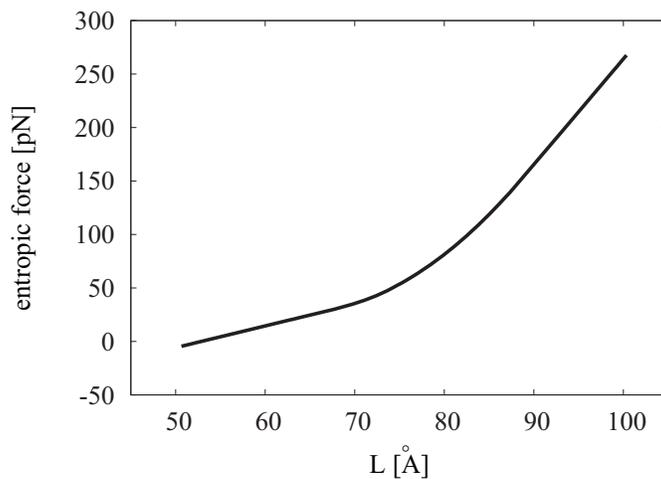


Figure 6.10: Entropic force in DN2 peptide computed from splined version of the conformational entropies from the constrained MD simulations.

Chapter 7

Comparison of entropic force in other resilin-like peptides

7.1 Relaxation of multiple chains

Based on the consensus repeats of DN, DC, and AP, I constructed the 30-residue peptides that were simulated in the NVT ensemble at the temperature of 355 K. The temperature was chosen to be high in order to decrease the equilibration times. Moreover, in my previous simulations of the DN2 peptides I have not observed the structural changes between the simulations at the temperatures of 310 and 355 K. The choice of the peptides is justified as follows: the DC peptide is chosen because it is not known about the elastic properties of this repeat; and the AP peptide is chosen since its consensus repeat, as opposed to all other repeats, does not contain any proline residues and it would be interesting to see the effect of prolines on the elastic properties of the peptides.

In order to look at the structural properties of other resilin-like peptides, first I verified that assembling multiple chains does not result in the peptide aggregation. By setting 16 chains of the AP2 and DC2 peptides along z-axis with 10 Å separation between chains, I observed that in all simulations the average end-to-end distance

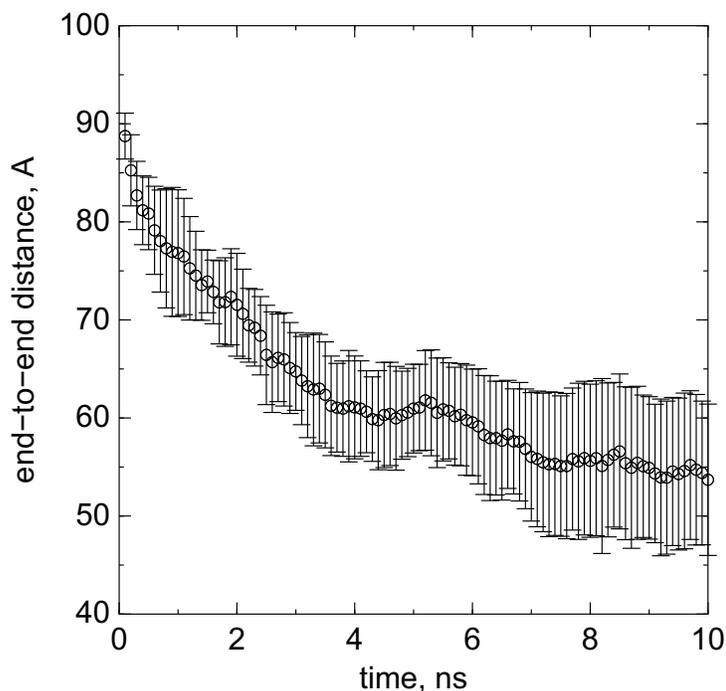


Figure 7.1: I setup 16 AP2 peptides in a 4 by 4 grid in XY plane. The grid size in either direction was 10 \AA and all peptides were highly stretched and oriented along Z-axis. The averaged over all chains end-to-end distance decreases to an equilibrium value of 55 \AA within 10 ns of simulation. The formation of β -sheets would result in the end-to-end distances at least $80\text{-}90 \text{ \AA}$.

decreased from original value (100 \AA) to $50\text{-}70 \text{ \AA}$. An example of such simulation for AP2 peptide is shown in Figure 7.1. Together with the secondary structure formations (data not shown) this indicates that neither β -sheets nor PPII nor α -helices were formed.

The simulation time (10 ns) was very small, so that one might argue that this time is not long enough to observe, say, β -sheet aggregations. Two additional simulations were performed with the collagen-like sequence $(GPP)_{10}$, known to form stable PPII structures, and 30-residue polyalanine peptides, that is known to form stable β -sheets in the extended state. Similar setup of 16 chains of each peptide type was used and within the same simulation time I observed the formation of the stable secondary structures (PPII helices in the collagen-like peptides, and β -sheets in the polyalanine peptides).

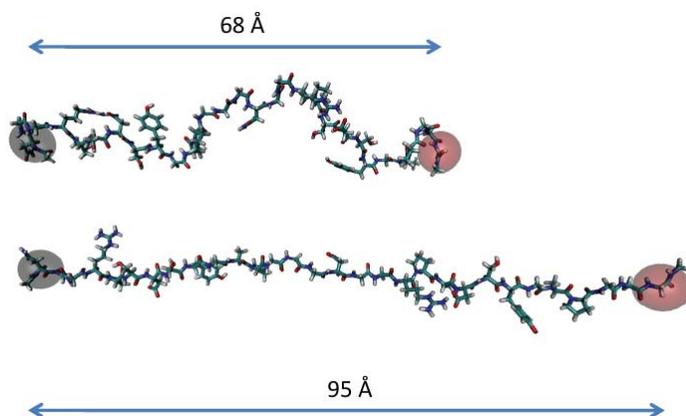


Figure 7.2: Two extreme extensions of DN2 peptide used in SMD simulations. Black sphere shows fixed atom and red sphere – for SMD atom that was pulled.

7.2 SMD simulations

Once the disorder of the DN2, DC2, and AP2 peptides was established in the multiple chain simulations, I used a single chain approximation and for each peptide and I ran 10 compression/extension SMD cycles at temperature of 355 K. A single cycle consisted of taking an extended structure of a peptide (around 100 Å for end-to-end distance) and compressing it by 30 Å. After that, the pulling direction was inverted and the peptide was extended by another 30 Å. Two typical conformations of the DN2 peptide at two extreme extensions is shown in Figure 7.2.

The force/extension curves in all peptides show small hysteresis (Fig. 7.3), suggesting high resilience. From Figure 7.3 one can see that the presence of prolines in DC2 and DN2 peptides require larger force for extreme extensions than that in AP2. At the same time, in the relaxed state all peptides have the same restoring force below 50 pN.

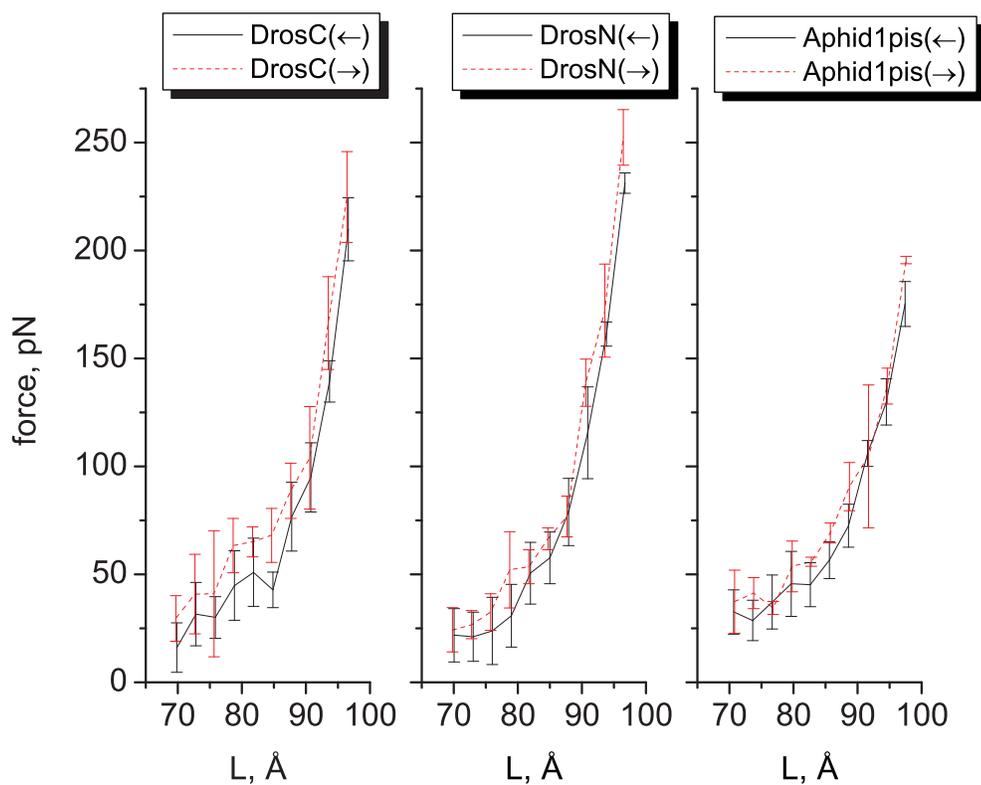


Figure 7.3: Loading (red, dash) and unloading (black, solid) forces, averaged over several SMD simulations at temperature 355 K for three resilin-like peptides: DC2, DN2, AP2. With the exception that proline-lacking AP2 peptide has less restoring force than DC2 or DN2 peptides at high extensions, no significant differences between peptides was found. In all three peptides the resilience estimated to be $80 \pm 10\%$.

Chapter 8

Coarse-grained model of RLP

8.1 PYP-model

I have defined a PYP-model as three beads per repeat. The elastic repeat is defined as PxxxYxxP sequence, where x is any residue. Since AN- and DN-based peptides differ mostly in the length of linker (AQT for AN-repeats and GGGNGGR for DN-repeats), the comparison of coarse-grained parameters is possible.

The beads were placed at the backbone C^α carbons of proline (P) and tyrosine (Y) residues, hence, the name of the model PYP. We chose tyrosine residues, since they are the sites of cross-linking, and the choice of prolines was arbitrary. It should be emphasized, that the coarse-grained model is built only using properties of a single repeat and a linker between two adjacent repeats and no parametrization as to global fluctuations of multiple repeats were used in the model. Atomic masses and charges were equally distributed among the coarse-grained beads.

The all-atom and coarse-grained representation of AN16 peptide is shown in Figure 8.1.

In the harmonic approximation, the integration time step for coarse-grained simulations τ_{cg} can be expressed in terms of the time step for all-atom simulations τ_f as

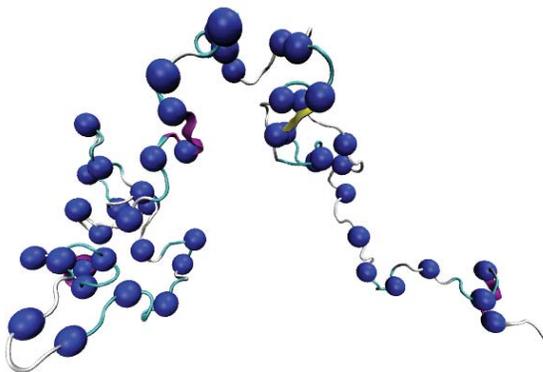


Figure 8.1: The coarse-grained and all-atom models of AN16 peptide. Three beads per repeat were assigned. The beads were placed at prolines and tyrosine backbone α -carbons.

follows

$$T_{cg} = T_f \sqrt{\frac{m_{cg} k_H}{m_H k_{cg}}} \quad (8.1)$$

For $m_{cg} \approx 300m_H$, as defined by choice of coarse beads, and the largest spring constant $k_{cg} \approx 0.001k_H$, as derived from all-atom simulations, we have $\tau_{cg} \approx 600\tau_H$, or about 600 fs.

The bonded interaction parameters are obtained by Boltzmann inversion (Eqn 8.2) from the distributions of virtual bonds in all repeats from all-atom simulations. These values for virtual bonds P-Y, Y-P, and P-P are shown in Tables 8.1, 8.2, 8.3. In a similar way, bond angle parameters were obtained, for example, P-Y-P bond angle parameters are shown in Table 8.4.

$$K = \frac{k_B T}{2(\langle R^2 \rangle - \langle R \rangle^2)} \quad (8.2)$$

Since there were no charged residues I did not assign charges to the coarse beads, so that the electrostatic interactions were not present in the coarse-grained force field. For the van der Waals interactions I assigned R_i^{min} to beads in such way, that sum

Table 8.1: Spring constants and average bond lengths for PY virtual bond in the coarse-grained model, obtained by Boltzmann inversion of virtual bond distribution from all-atom simulations.

Peptide	K, kcal/(mol Å ²)	$\langle R_0 \rangle$, Å
DN2	0.06	8.03
AN1	0.18	9.41
AN2	0.18	10.72
AN16	0.06	10.05

Table 8.2: Spring constants and average bond lengths for YP virtual bond in the coarse-grained model, obtained by Boltzmann inversion of virtual bond distribution from all-atom simulations.

Peptide	K, kcal/(mol Å ²)	$\langle R_0 \rangle$, Å
DN2	0.43	8.16
AN1	0.32	8.32
AN2	0.45	8.66
AN16	0.27	8.43

Table 8.3: Spring constants and average bond lengths for linkers PP in the coarse-grained model, obtained by Boltzmann inversion of virtual bond distribution from all-atom simulations.

Peptide	K, kcal/(mol Å ²)	$\langle R_0 \rangle$, Å
DN2	0.05	10.72
AN2	0.17	11.9
AN16	0.08	10.5

of these radii from two beads give average of the corresponding virtual bond. While this choice is not justified, varying van der Waals radii did not have significant effect

Table 8.4: Spring constants and average PYP bond angles in the coarse-grained model, obtained by Boltzmann inversion of virtual bond distribution from all-atom simulations.

Peptide	K, kcal/(mol rad ²)	$\langle\Theta\rangle$, degrees
DN2	0.9	89.2
AN1	1.4	55.5
AN2	2.0	85.5
AN16	0.8	100.0

on the results. The choice of van der Waals ϵ parameters were chosen uniformly and adjusted in such a way that fluctuations of radius of gyration in AN2 peptide are similar to those from all-atom simulations. The R_g from all-atom simulations of AN2 is $11.8 \pm 2.0 \text{ \AA}$. In the coarse-grained simulations ϵ in the range between 0.01 and 0.3 kcal/mol led to R_g comparable with the all-atomistic radius of gyration $R_g = 11.5 \pm 2 \text{ \AA}$ (Fig. 8.2). When I applied this model to coarse-grained simulations of AN16 I observe that, starting with $\epsilon > 0.1$ kcal/mol, the average radius of gyration started decreasing below 34 \AA . On the other hand, for range between 0.01 and 0.1 the same radius of gyration was obtained $42 \pm 8 \text{ \AA}$. Thus, final choice of 0.01 kcal/mol was made. The average kinetic and potential energies were 45 and 10 kcal/mol.

The parameters derived from AN2-segments were used in simulations of coarse-grained AN16 model. I ran NVT ensemble (with temperature set to 310 K) langevin dynamics with PBC (size $105 \times 105 \times 105 \text{ \AA}^3$) for $15e6$ steps that corresponds to $9 \mu\text{s}$. The choice of langevin damping constant was the same as in [6] and equal to 2 ps^{-1} . For van der Waals interactions cutoff distance was 20 \AA .

As seen on Figures 8.3, 8.4, 8.5 coarse-grained virtual bonds qualitatively match the results from all-atom simulations.

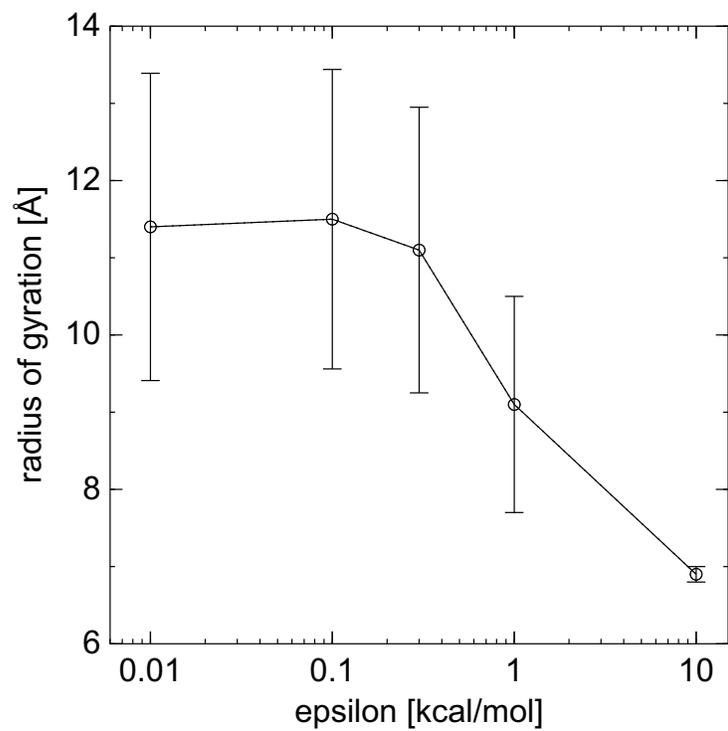


Figure 8.2: Dependence of radius of gyration in the coarse-grained simulations of AN2 on the value of the van der Waals strength parameter. With the values up to 0.1 kcal/mol high fluctuations of the peptide (as observed in all-atom simulations) are reproduced.

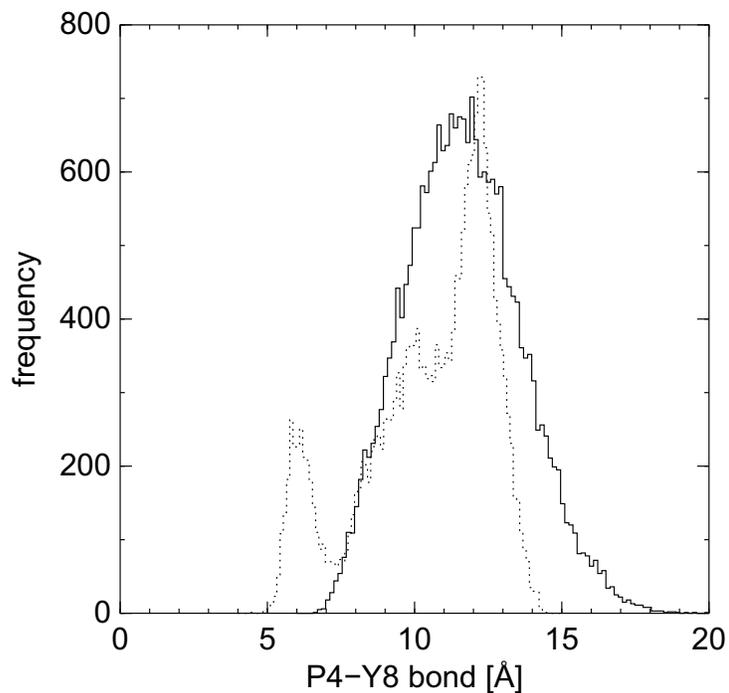


Figure 8.3: Comparison of bond distributions from coarse-grained (solid) and all-atom simulations of AN16 peptide.

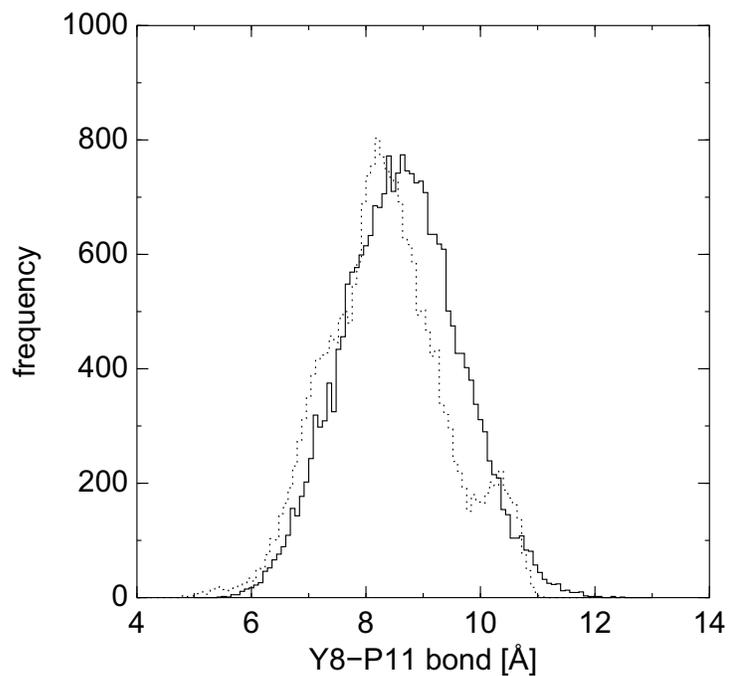


Figure 8.4: Comparison of bond distributions from coarse-grained (solid) and all-atom simulations of AN16 peptide.

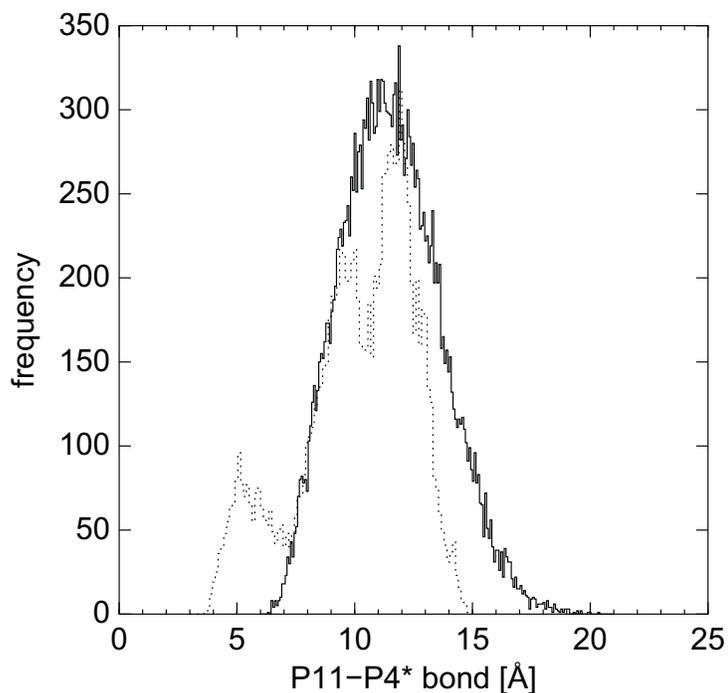


Figure 8.5: Comparison of bond distributions from coarse-grained (solid) and all-atom simulations of AN16 peptide.

8.1.1 Cross-linking

I assembled four AN16 coarse-grained peptides and created four cross-links at random places to connect all chains into one network. The parameters for cross-linking were derived from 200 ns simulations of two cross-linked peptides (two 15-residue peptides with AN-based repeat sequence). In principle, cross-linking bond length and related angles can be derived, I used only tyrosine-tyrosine bond length parameters in our coarse-grained cross-link model. The spring constant and equilibrium length for Y-Y cross-link were $1.2 \text{ kcal}/(\text{mol \AA})$ and 8.9 \AA . Two simulations were performed for $50 \mu\text{s}$, with and without crosslinks, in NVT ensemble at temperature 310 K. Periodic boundary conditions were applied to the cubic simulation cell with cube size of 100 \AA . A snapshot from trajectory of the cross-linked system is shown in Figure 8.6.

As seen from Table 8.5, moderate level of crosslinking (12.5% of tyrosines converted to cross-links) does not reduce peptides fluctuations. This fact is consistent

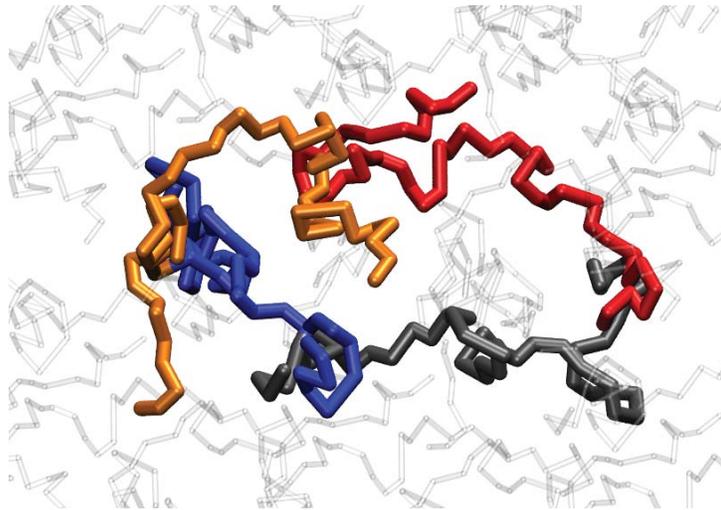


Figure 8.6: A system of four coarse-grained AN16 chains with four randomly placed cross-links. Each chain from the central cell is shown in different color and chains from periodic cell images are shown in gray.

Table 8.5: Radius of gyration for a system of four coarse-grained AN16 chains with and without cross-linking. Average and standard deviations are reported.

Chain	R_g (free), Å	R_g (cross), Å
1	41±8	42±8
2	41±8	41±7
3	41±8	40±7
4	42±8	42±8

with the experimental results (Raman scattering and CD spectra) that with the level of crosslinking at 16% in AN16 peptides the dynamics of peptides does not change ([38]). Also, it was shown for N-terminal domain of *D. melanogaster* resilin that the cross-linking reaches a plateau when 20% of tyrosines are transformed into di- and trityrosine cross-links, and even at that level resilience remains relatively high [20].

The distribution of the radius of gyration of a single chain in the cross-linked and uncross-linked systems show good fit with normal distribution (Fig. 8.7). A small deviation of the normal distribution of the maximum probable radius of gyration,

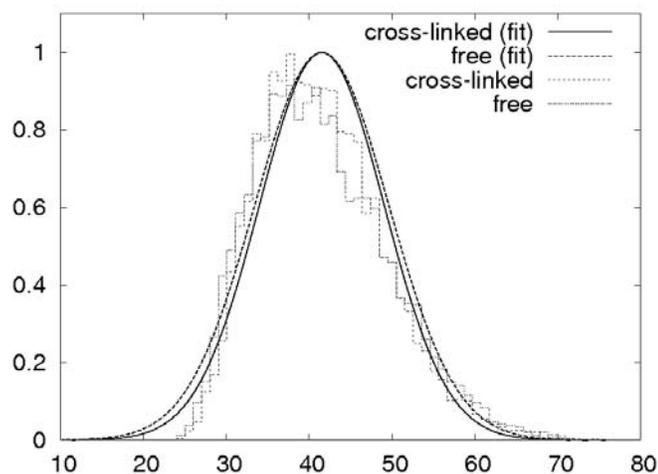


Figure 8.7: Distribution of the radius of gyration of a single chain in simulations of four coarse-grained chains of AN16 peptide with and without cross-linking. Gaussian fit was used to the distributions from simulations.

observed in simulations, is most likely due to weak van der Waals intra- and inter-peptide interactions.

Experimentally reported level of cross-link formation includes both, intra- and inter-chain cross-links and for high resilience the formation of the cross-links *between* chains is desired. Since the resilience was reported to be similar for different cross-linking levels [20], it means that there is some mechanism that favors inter-chain cross-links over intra-chain ones. Perhaps, the fact that resilin-like peptides do not collapse into a compact structure is sufficient to provide inter-chain cross-linking.

All the scripts to build coarse-grained model and run NAMD simulations are available at <http://bmiwiki.cchmc.org/index.php/Coarse-graining> web site.

Chapter 9

Conclusions

I have presented a comprehensive analysis of the resilin-like peptides using bioinformatics and molecular dynamics approaches. To the best of my knowledge this is the first attempt to relate elastomeric properties of the idealized resilin-like peptides to their structure using computational methods.

9.1 Disorder

Based on a number of molecular dynamics simulations with different systems and different conditions I conclude that all studied here peptides are highly agitated in thermal motion and possess rubber-like elasticity. High level of disorder is supported by the following observations: i) lack of stable secondary structure, particularly low content of α -helices and β -strands that are known to stabilize protein structures; ii) lack of stable intra-protein hydrogen bonds; iii) the $1 - S^2$ structural parameters are close to those for random coil proteins; iv) large fluctuations in radius of gyration. As was observed in several simulations one of the resilin-like peptides collapsed to some compact state, but that state was not stable and, after some time, a highly fluctuating behavior of the peptide resumed.

9.2 Entropic force

From the tensile and stretching computer experiments I have found the restoring force to be dominantly of the entropic nature over a range of extensions (up to 100%). For each peptide the resilience and the entropic part of the restoring force were computed.

9.3 Cross-linked biopolymers

Using the results from the all-atom simulations, I have built a coarse-grained model of the AN and DN-based peptides. This allows one to simulate larger peptides and explore larger timescales. Randomly cross-linked model of four AN16 peptides was built on the coarse level. Up to the level of cross-linking when 12% of tyrosine residues form dityrosine chemical bonds, no significant decrease in peptide fluctuations was observed.

I hope that this study would serve as a starting point in revealing the mechanism of high resilience in resilin-like peptides and can help design biomaterial with given mechanical properties.

References

- [1] R. Adamczak, A. Porollo, and J. Meller. Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins*, 59(3):467–475, May 2005.
- [2] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–410, October 1990.
- [3] S. O. Andersen. The cross-links in resilin identified as dityrosine and trityrosine. *Biochimica et biophysica acta*, 93:213–215, October 1964.
- [4] I. Andricioaei and M. Karplus. On the calculation of entropy from covariance matrices of the atomic fluctuations. *The Journal of Chemical Physics*, 115(14):6289–6292, 2001.
- [5] D. Ardell and S. O. Andersen. Tentative identification of a resilin gene in drosophila melanogaster. *Insect Biochemistry and Molecular Biology*, 31(10):965–970, September 2001.
- [6] A. Arkhipov, W. H. Roos, G. J. L. Wuite, and K. Schulten. Elucidating the mechanism behind irreversible deformation of viral capsids. *Biophys J*, 97(7):2061–2069, Oct 2009.
- [7] M. Baer, E. Schreiner, A. Kohlmeyer, R. Rousseau, and D. Marx. Inverse temperature transition of a biomimetic elastin model: Reactive flux analysis of folding/unfolding and its coupling to solvent dielectric relaxation†. *The Journal of Physical Chemistry B*, 110(8):3576–3587, March 2006.
- [8] A. Bateman et al. The pfam protein families database. *Nucleic Acids Res.*, 32:D138–D141, 2004.
- [9] M. V. Berjanskii and D. S. Wishart. A simple method to predict protein flexibility using secondary chemical shifts. *Journal of the American Chemical Society*, 127(43):14970–14971, November 2005.
- [10] M. V. Berjanskii and D. S. Wishart. The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. *Nucl. Acids Res.*, 35(suppl_2):W531–537, July 2007.

- [11] B. Bochicchio, A. Pepe, and A. M. Tamburro. Investigating by cd the molecular mechanism of elasticity of elastomeric proteins. *Chirality*, 20(9):985–994, Sep 2008.
- [12] R. Brueschweiler and P. E. Wright. Nmr order parameters of biomolecules: A new analytical representation and application to the gaussian axial fluctuation model. *Journal of the American Chemical Society*, 116(18):8426–8427, September 1994.
- [13] M. B. Charati, J. L. Ifkovits, J. A. Burdick, J. G. Linhardt, and K. L. Kiick. Hydrophilic elastomeric biomaterials based on resilin-like polypeptides. *Soft Matter*, 5(18):3412–3416, 2009.
- [14] M. V. Cubellis, F. Caillez, T. L. Blundell, and S. C. Lovell. Properties of polyproline ii, a secondary structure element implicated in protein-protein interactions. *Proteins*, 58(4):880–892, Mar 2005.
- [15] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An $n \log(n)$ method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.
- [16] H. J. Dyson and P. E. Wright. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol*, 12(1):54–60, Feb 2002.
- [17] H. J. Dyson and P. E. Wright. Unfolded proteins and protein folding studied by nmr. *Chem Rev*, 104(8):3607–3622, Aug 2004.
- [18] H. J. Dyson and P. E. Wright. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, 6(3):197–208, Mar 2005.
- [19] G. Elliott et al. On the structure of resilin. *J Mol Biol*, 13:791–795, 1965.
- [20] C. M. Elvin, A. G. Carr, M. G. Huson, J. M. Maxwell, R. D. Pearson, T. Vuocolo, N. E. Liyou, D. C. C. Wong, D. J. Merritt, and N. E. Dixon. Synthesis and properties of crosslinked recombinant pro-resilin. *Nature*, 437(7051):999–1002, October 2005.
- [21] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.
- [22] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.
- [23] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks. Constant pressure molecular dynamics simulation: The langevin piston method. *The Journal of Chemical Physics*, 103(11):4613–4621, 1995.
- [24] P. J. Flory, A. Ciferri, and C. A. J. Hoeve. The thermodynamic analysis of thermoelastic measurements on high elastic materials. *Journal of Polymer Science*, 45(145):235–236, 1960.

- [25] N. M. Glykos. Carma: a molecular dynamics analysis program. *J Comp Chem*, 27:1765–1768, 2006.
- [26] J. Gosline et al. Elastic proteins: biological roles and mechanical properties. *Phil Trans R Soc Lond B*, 357:121–132, 2002.
- [27] S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22):10915–10919, November 1992.
- [28] W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, February 1996.
- [29] M. Huson and J. Maxwell. The measurement of resilience with a scanning probe microscope. *Polymer Testing*, 25(1):2–11, February 2006.
- [30] M. Jensen and T. Weis-Fogh. Biology and physics of locust flight. V. Strength and elasticity of locust cuticle. *Phil Trans Roy Soc Lond B*, 245:137–169, 1962.
- [31] D. T. Jones. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology*, 292(2):195–202, September 1999.
- [32] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.
- [33] W. Kabsch and C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, December 1983.
- [34] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, and K. Varadarajan. NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, 151(1):283–312, May 1999.
- [35] S. M. King and W. C. Johnson. Assigning secondary structure from protein coordinate data. *Proteins: Structure, Function, and Genetics*, 35(3):313–320, 1999.
- [36] J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach, and K. W. Plaxco. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34):12491–12496, August 2004.
- [37] G. Lipari and A. Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *Journal of the American Chemical Society*, 104(17):4546–4559, August 1982.

- [38] R. E. Lyons, K. M. Nairn, M. G. Huson, M. Kim, G. Dumsday, and C. M. Elvin. Comparisons of recombinant resilin-like proteins: Repetitive domains are sufficient to confer resilin-like properties. *Biomacromolecules*, Oct 2009.
- [39] A. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Phys Rev E*, 52:3730–3737, 1995.
- [40] A. D. Mackerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-Mccarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, April 1998.
- [41] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, 1994.
- [42] K. M. Nairn, R. E. Lyons, R. J. Mulder, S. T. Mudie, D. J. Cookson, E. Lesieur, M. Kim, D. Lau, F. H. Scholes, and C. M. Elvin. A synthetic resilin is largely unstructured. *Biophys J*, 95(7):3358–3365, Oct 2008.
- [43] J. Nummela, F. Yassin, and I. Andricioaei. Entropy-energy decomposition from nonequilibrium work trajectories. *The Journal of chemical physics*, 128(2), January 2008.
- [44] S. Park et al. Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality. *J Chem Phys*, 119:3559–3566, 2003.
- [45] G. M. Paterlini and D. M. Ferguson. Constant temperature simulations using the langevin equation with velocity verlet integration. *Chemical Physics*, 236(1-3):243–252, September 1998.
- [46] R. Petrenko, M. Dickerson, R. Naik, S. Patnaik, T. Beck, and J. Meller. Entropic force in drosophila resilin: molecular dynamics study. *Proc Int’l Conf Bioinformatics and Comp Biology*, 2:598–603, 2009.
- [47] R. Petrenko, M. Dickerson, R. Naik, S. Patnaik, T. Beck, and J. Meller. Molecular dynamics study of entropic force in resilin-like peptides. *Biophys J*, submitted, 2010.
- [48] G. Pollastri and A. McLysaght. Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics*, 21:1719–1720, 2005.
- [49] A. A. Porollo, R. Adamczak, and J. Meller. Polyview: a flexible visualization tool for structural and functional annotations of proteins. *Bioinformatics*, 20(15):2460–2462, October 2004.

- [50] J. J. Prompers and R. Bruschweiler. General framework for studying the dynamics of folded and nonfolded proteins by nmr relaxation spectroscopy and md simulation. *Journal of the American Chemical Society*, 124(16):4522–4534, April 2002.
- [51] S. Rauscher et al. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure*, 14:1667–1676, 2006.
- [52] J. Rebers. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochemistry and Molecular Biology*, 31(11):1083–1093, October 2001.
- [53] R. Rousseau, E. Schreiner, A. Kohlmeyer, and D. Marx. Temperature-dependent conformational transitions and hydrogen-bond dynamics of the elastin-like octapeptide GVG(VPGVG): A molecular-dynamics study. *Biophys J*, 86(3):1393–1407, March 2004.
- [54] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, March 1977.
- [55] H. Schäfer, A. E. Mark, and W. F. van Gunsteren. Absolute entropies from molecular dynamics simulation trajectories. *The Journal of Chemical Physics*, 113(18):7809–7817, 2000.
- [56] J. Schlitter. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chemical Physics Letters*, 215(6):617–621, December 1993.
- [57] Z. Shi, C. A. Olson, G. D. Rose, R. L. Baldwin, and N. R. Kallenbach. Polyproline ii structure in a sequence of seven alanine residues. *Proc Natl Acad Sci U S A*, 99(14):9190–9195, Jul 2002.
- [58] M. Sickmeier, J. A. Hamilton, T. Legall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic, and A. K. Dunker. Disprot: the database of disordered proteins. *Nucleic Acids Res*, 35(Database issue), January 2007.
- [59] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314:141–151, 1999.
- [60] S. Trebst, M. Troyer, and U. H. E. Hansmann. Optimized parallel tempering simulations of proteins. *J Chem Phys*, 124(17):174903, May 2006.
- [61] S. Tweedie, M. Ashburner, K. Falls, P. Leyland, P. Mcquilton, S. Marygold, G. Millburn, D. Osumi-Sutherland, A. Schroeder, R. Seal, H. Zhang, and T. F. Consortium. Flybase: enhancing drosophila gene ontology annotations. *Nucl. Acids Res.*, 37(suppl_1):D555–559, January 2009.

- [62] D. Urry et al. Elastin: a representative ideal protein elastomer. *Phil. Trans. R. Soc. Lond. B*, 357:169–184, 2002.
- [63] T. Weis-Fogh. A rubber-like protein in insect cuticle. *J Exp Biol*, 37(4):889–907, December 1960.
- [64] T. Weis-Fogh. Thermodynamic properties of resilin, a rubber-like protein. *J Mol Biol*, 3:520–531, 1961.
- [65] P. E. Wright and H. J. Dyson. Linking folding and binding. *Curr Opin Struct Biol*, 19(1):31–38, Feb 2009.
- [66] F. Zhang and R. Bruschweiler. Contact model for the prediction of NMR NH order parameters in globular proteins. *Journal of the American Chemical Society*, 124(43):12654–12655, October 2002.