NOTE TO USERS

Models of the stability of proteins

Cristiano L. Dias
Doctor of Philosophy
July 2007

Department of Physics, McGill University
Montreal,Quebec

A Thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirement for the degree of Doctor of Philosophy.

Your file Votre référence
ISBN: 978-0-494-38581-4
Our file Notre référence
ISBN: 978-0-494-38581-4

To my parents, *Alaerte and Eunice.*

ACKNOWLEDGEMENTS

It is a great pleasure for me to acknowledge all those who have assisted me in one way or another during my Ph.D degree. First and foremost, I would like to express gratitude to my supervisor, Martin Grant, for his thoughtful suggestions, continual support and guidance throughout my degree.

During the course of this work, I have benefitted from discussions with great minds like Martin Dubé, Anirban Sain, Norberto Majilisn and more recently Daniel Vernon. I am thankful to Tapio Ala-Nissila for his warm hospitality and thoughtful guidance during my stay in Finland. Mikko Karttunen has proven to be a wise leader and a good friend.

I wish to thank the members of the coffee club: Saeid Asgharizadeh, Jens Kröger (thanks for the French translation), David Montiel, Vincent Tabard-Cossa, James Hedberg, Mikhail and Nikolai Sergey and most specially Simiso Mkhonta. Without the many coffee cups we had, I would have finished this thesis sooner but life would certainly be much poorer. I also had pleasant discussions during dinner with Yongxiang Gao and his wife Ying, Quanyong Wang and Dominique Pouliot.

A special thanks goes to my officemates Robert Gagnon and Khosrow Hassani. Not only have I borrowed a large number of espresso-capsules from them but I also had many insights from the discussions we had. I also wish to thank Michael Gerry for the many passionate conversations we had. My friends and colleagues Josivaldo and Lindomar – it is always refreshing, for the body and soul, to have a cold beer with you guys in the hot evenings of Brasilia.

I would like to thank everyone in the department for making it such a nice place. In particular thanks to Juan Gallego and Paul Mercure for their support with computer-related issues, and to Paula, Elizabeth, Louise, Diane and Sonia for their administrative help.

Montreal would not have been such a great place without the gang, and in particular Tijana, Alexandra and Denise. I will also miss my capoeira friends and the cold beer we used to have after an exhaustive workout. It was nice talking about life with Helena Martin during a cold and sad winter. Thanks to Claudine Couture for the good time spent together and to Stella Maris for her long-lasting and supportive friendship.

Finally, I would like to thank my familly: parents, brother and sisters (Tatiana and Cecilia) for their unconditional support, love and cheering.

ABSTRACT

Although the native conformation of a protein is thermodynamically its most stable form, this stability is only marginal. As a consequence, globular proteins have a certain amount of flexibility in their backbones which allows for conformational changes in the course of their biological function. In the course of this thesis, we study protein models at the edge of stability in different contexts:

- First, we use molecular dynamics to determine the force needed to rupture a chain molecule (an unfolded protein) being stretched at constant loading rate and temperature. When all energy bonds of the molecule are identical, we find that the force F depends on the pulling rate r and temperature T according to $F \sim const - T^{1/3}|\ln(r/T)|^{1/3}$. When a single weak bond is introduced, this result is modified to $F \sim const - T^{2/3}|\ln(r/T)|^{2/3}$. This scaling, which is model independent, can be used with force-spectroscopy experiment to quantitatively extract relevant microscopic parameters of biomolecules.
- Second, we study the structural stability of models of proteins for which the selected folds are unusually stable to mutation, that is, designable. A two-dimensional hydrophobic-polar lattice model is used to determine designable folds and these folds were investigated under shear through Langevin dynamics. We find that the phase diagram of these proteins depends on their designability. In particular, highly designable folds are found to be weaker, *i.e.* easier to unfold, than low designable ones. This is argued to be related to protein flexibility.

- Third, we study the mechanism of cold denaturation through constant-pressure simulations for a model of hydrophobic molecules in an explicit solvent. We find that the temperature dependence of the hydrophobic effect is the driving force for cold denaturation. The physical mechanism underlying this phenomenon is identified as the destabilization of hydrophobic contact in favor of solvent separated configurations, the same mechanism seen in pressure induced denaturation. A phenomenological explanation proposed for the mechanism is suggested as being responsible for cold denaturation in real proteins.

ABRÉGÉ

La structure native des protéines est leur état le plus stable thermodynamiquement. Néanmoins, cette stabilité n'est que marginale et ces molécules ont un certain degré de flexibilité qui leur permet des changements structuraux afin de réaliser leurs fonctions biologiques. Dans cette thèse, nous étudions, dans différents contextes, des modèles de protéines qui sont à la limite de leur stabilité:

- Premièrement, nous utilisons la méthode de la dynamique moléculaire pour déterminer la force nécessaire afin de rompre une molécule en forme de chaîne (protéine dépliée) qui est étirée à vitesse constante et à temperature constante. Quand tous les liens sont identiques, on observe que la force F dépend du taux d'étirement r et de la temperature T selon $F \sim const - T^{1/3}|\ln(r/T)|^{1/3}$. Si un lien faible est introduit dans la molécule, ce résultat devient $F \sim const - T^{2/3}$ $|\ln(r/T)|^{2/3}$. Cette loi qui est indépendante du modèle, peut être utilisée conjointement avec des expériences de spectroscopie-de-force pour obtenir quantitativement des paramètres microscopiques de biomolecules.
- Deuxièmement, nous étudions la stabilité structurale pour des modèles de protéine dont les configurations sont résistantes aux mutations, c'est-à-dire "designables". Un modèle bidimensionel du genre hydrophobique-polaire est utilisé pour déterminer le degré de "designabilité" des configurations et celles-ci sont étudiées sous cisaillement avec une méthode de dynamique de Langevin. Nous trouvons que le diagramme de phase de ces protéines dépend de leurs "designabilité". En particulier, les configurations qui sont hautement "designées"

sont plus faibles, c'est-à-dire plus faciles à déplier que les configurations qui ont un faible degré de designabilité. Ce phénomène est possiblement relié à la flexibilité des protéines.

- Troisièmement, nous étudions le méchanisme de dénaturalisation-à-froid avec des simulations à pression constante pour un modèle de molécule hydrophobique dans un solvant explicite. Nous trouvons que la dépendance en température de l'effet hydrophobique est la force motrice de la dénaturalisation-à-froid. Le méchanisme derrière ce phénomène est la déstabilisation des contacts hydrophobiques en faveur de configurations entremêlées avec des molécules de solvant – le même méchanisme est observé dans la dénaturalisation à haute pression. Le modéle phénoménologique de ce méchanisme est proposé comme explication de la dénaturalisation-à-froid dans de vraies protéines.

TABLE OF CONTENTS

LIST OF TABLES

Table page

LIST OF FIGURES

Contribution to the literature

- Cristiano L. Dias, Martin Dube, Fernando Oliveira, and Martin Grant. **Scaling in Force Spectroscopy.** Physical Review E, 72:011918, 2005.

- Anirban Sain, Cristiano L. Dias, and Martin Grant. **Rupture of an extended object: A many-body Kramers calculation.** Physical Review E, 74:046111, 2006.

- Cristiano L. Dias and Martin Grant. **Designable structures are easy to unfold.** Physical Review E, 74:042902, 2006.

- C. L. Dias and Martin Grant. **Unfolding designable structures.** The European Physical Journal B, 50:265, 2006.

- Cristiano L. Dias, Tapio Ala-Nissila, Mikko Karttunen, Ilpo Vattulainen, and Martin Grant. **Microscopic mechanism for cold denaturation.** Submitted for publication, 2007.

Models of the stability of proteins

I do not know what I may appear to the world; but to myself I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.

– Isaac Newton

Chapter 1

Introduction

Everything should be made as simple as possible, but not simpler.

– A. Einstein

Systems like gases and liquids comprise a large number of particles of the order of Avogadro's number (6.02×10^{23}). Although microscopic constituents account for macroscopic properties, it is impossible to compute all the microscopic states theoretically due to their large number. The common approach to study liquids and gases is therefore to use statistical arguments. Based on the partition function, a general statistical framework has been developed for equilibrium systems from which thermodynamical properties can be derived.

Although general concepts for systems far from thermal equilibrium are lacking, some structure has emerged with concepts like universality. This concept, which was first developed for equilibrium systems, also applies to non-equilibrium processes. Universality expresses the notion that certain properties of physical systems do not depend on microscopic details and furthermore are equivalent for seemingly unrelated processes in systems. Those unrelated systems are said to form a universality class.

The usefulness of this concept lies in the fact that, in general, to understand a non-equilibrium system it is enough to study the simplest model falling within a universality class.

Simple models, also known as minimal or toy models, have only those mechanisms that are considered essential for the understanding of what one observes in more complex systems. Through various mappings and different physical interpretations of the observables, those models are able to describe a wide variety of phenomena in physics and beyond. The large degree of simplification performed while going from a complex system to a simple model leads to inaccuracies that can be corrected by adding other mechanisms to the model. In this way, one can proceed from a qualitative to a quantitative description of the phenomena by progressively adding other contributions to the dominant behaviour. This is usually referred to as the onion-like strategy for modelling [91] – a core which provides a first basic understanding of the dominant behaviour, coated with subsequent layers of decreasing significance.

A different approach for modelling, which is usually referred to as the traditional picture [41], consists in including as many of the fine details of the real system as possible. This strategy seeks to reproduce quantitative results found in real systems and is extensively employed in quantum chemistry. The difficulty is extracting meaningful information on realistic length and time scales.

In this thesis we study non-equilibrium biomolecular systems through minimal models. We are interested in a qualitative description. The models we study include only the dominant behaviour of the system being studied. Those models are studied

either in one or two dimensions and a first order fine-graining would be to include the third dimension. An additional refinement would be to add more realistic interactions between residues of the protein, like hydrogen bonding.

In particular, we study conformational changes of proteins in three different contexts: i) bond rupture of a stretched macromolecule, ii) conformational changes under shear flow, and, iii) cold denaturation. Bond rupture is the common ingredient relating those projects. In the first project we study how a bond, which accounts for the sequential alignment of monomers, breaks. In the second and third projects, the conformational changes of the protein occurs through the rupture of weak bonds which hold different segments of the protein together.

After the seminal work of Kauzmann [51], the minimal ingredient for modelling protein structure is to include hydrophobicity. This effect is when a protein minimizes the amount of hydrophobic amino acids which are left to interact with the solvent at the protein's surface, by burying them within the dry protein core. This effect explains the hydrophobic core seen in crystallographic data of globular proteins, as well as their high density packing – see section 2.1.2. One minimal model, that accounts for those structural features, is the hydrophobic-polar model which will be used to describe protein structures in chapter 4.

One should point out that minimal models are not designed to describe all features of the system, only a particular aspect of the system. Minimal models are designed to address specific issues [91]. Therefore, for example, the hydrophobic-polar model, which was designed to describe the relation between amino acid sequence and structure, fails to describe the temperature-dependent stability of protein structures.

While real proteins are maximally stable at an intermediate temperature and unfold at both high and low temperatures, any particular structure in the hydrophobic-polar model becomes continuously more stable as temperature decreases.

To study the temperature dependent stability of proteins, the solvent which accounts for the hydrophobic effect has to be described explicitly. In chapter 5, we use the Mercedes-Benz model to account for the solvent. This minimal two-dimensional model reproduces many of the anomalies of water and also accounts for the hydrophobic effect. Those anomalies are the result of a competition between hydrogen bonding and Van-der-Waals interactions. Coupled to a bead-and-spring model for the protein, we reproduce and explain the temperature dependent behaviour of proteins.

1.1 Thesis Overview

This thesis is organised as follows. In Chapter 2 we introduce the theoretical background which is relevant to the bulk of this thesis. In particular, we review proteins, their microscopic constituents and characteristic features of their structures. The thermodynamics of proteins is then presented, showing their temperature and pressure dependent stability. Subsequently, we discuss the hydrophobic effect – which is the main driving force for protein folding. Finally, we discuss minimal models which inspired the work in chapters 4 and 5.

Chapters 3, 4 and 5 correspond to original work. In chapter 3 we study scaling laws of bond rupture under an external mechanical force. This phenomena is thermally activated, and two scalings for the force-dependent energy barrier are obtained. Those scalings result from different shapes of the energy landscape at the

critical force where the energy barrier vanishes. Langevin simulations on model systems confirm those scalings. Scaling laws for the experimentally relevant situation where force increases linearly with time is also studied. This is again confirmed with Langevin simulations.

In chapter 4, we study how protein conformations are affected by external perturbations. First we use the hydrophobic-polar model to classify protein conformations in a lattice according to their stability to mutation, that is, designability. Geometrically we show that structures with greater designability have more covalent bonds connecting surface and core monomers. Second, designable structures are subjected to shear flow at different temperatures. We find that conformations which are highly designable unfold at lower shear rates. A phenomenological explanation for this behaviour is given based on the geometrical properties of highly designable structures and a relationship to protein flexibility is suggested.

In Chapter 5 we study cold denaturation for a model of proteins in an explicit solvent. First we introduce the model and describe how simulations are carried out. Results are then presented showing that proteins are more compact at an intermediate temperature, and that they open up at high and low temperatures – in agreement with heat and cold denaturation, respectively. A change in the energetics of water molecules neighbouring the protein is demonstrated to correlate with cold denaturation. Characteristic configurations of proteins at different temperatures are, then, discussed and a phenomenological explanation for cold denaturation is provided. We conclude this thesis in chapter 6.

Chapter 2

Theoretical Background

2.1 Protein

The importance of proteins for living systems is highlighted by their name, which derives from the Greek word *proteios* and means of first importance. Proteins are linear polymers made of 20 types of monomers, called amino acids. The combination of those amino acids account for the variety of proteins found in nature: more than 100,000 different proteins are estimated to exist in the human body alone [45].

The native structure of a protein corresponds to the configurations taken by its amino acids in the three-dimensional space under physiological conditions. This structure minimizes the Gibbs free energy of the system which depends both on the interaction between amino-acids, and between amino-acids and the solvent. Therefore amino acid sequences, which are very different from each other, are likely to assume different native structures. Nowadays much effort is being spent to determine the relationship between sequence and structure. The solution to this problem will greatly improve the design of new proteins and, hence, new drugs.

Proteins have evolved through natural selection to perform functions in living systems. Those functions are the result of a protein's interaction with other macromolecules in the cell. Roughly speaking, proteins interact with macromolecules which have complementary structures such that protein function is intimately related to protein structure. Therefore, determining the structure of a protein is an important step towards understanding its role in the cell. Examples of function are:

- Structural - many proteins serve as a support to give biological structures strength or protection. One example is keratin which is found in hair and fingernails.
- Defense - some proteins are responsible for defending an organism against invasion of other species. Example: antibodies.
- Transport - many proteins are responsible for the transport of specific molecules from one organ to another. Hemoglobin is one example.

Proteins can be of two types: fibrous and globular. Fibrous proteins are long-chains which play mostly a structural role in cells. Examples are collagen fibers and *alpha*-helical fibers of wool. The former are strong, resistant to stretching and relatively rigid. They are one the most abundant fibrous proteins, being $3,000$ Å long and having 15 Å of diameter [19]. The latter are flexible and can be elastically extended to twice their original length. In contrast, globular proteins are arranged in a compact globule-like shape. In this thesis we deal with globular proteins, in particular single subunit globular proteins.

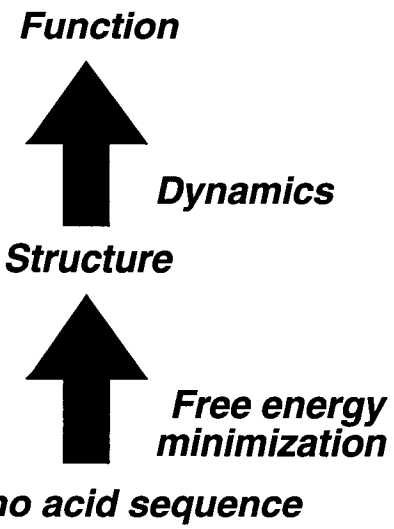

Figure 2–1: Schematic representation of the relation between amino acid sequence, structure and function. The mechanism relating those levels of organization is also shown.

2.1.1 Building Blocks of Proteins

Amino acids are the building blocks of proteins. Although there are 20 amino acids, they all share a common form. It consists of a central carbon atom, called the α carbon (C_α) bound to an amino-group (NH_2), a carboxylic acid group (COOH), a hydrogen atom (H), and a variable side chain (R) - see Fig. 2–2. It is this side chain, which can be one of twenty different structures, that distinguishes the various amino acids. Amino acids differ in size, shape, hydrogen bonding capacity, chemical reactivity, and, particularly important in the context of this study, their hydrophobic character, i.e., their affinity to the aqueous environment (see table 2–1).

To form proteins, amino acids bind through peptide bonds. Those are covalent bonds between the carboxyl group of one amino acid and the amino group of another amino acid (see Fig. 2–2). Amino acids in the chain are called residues and proteins

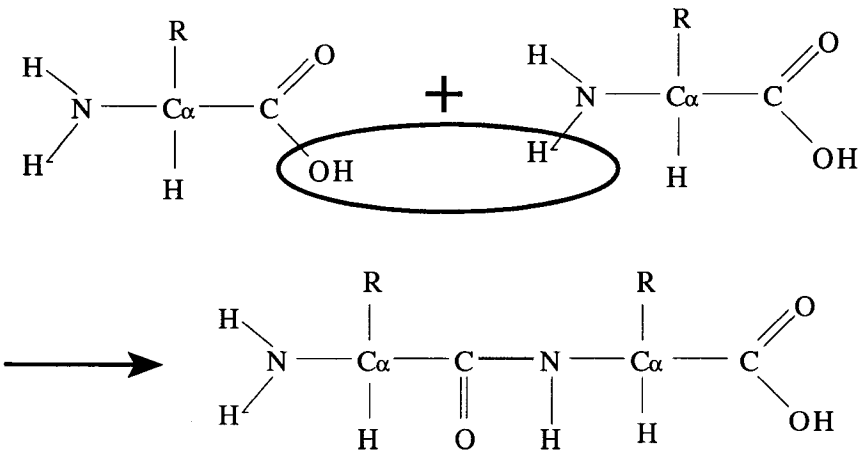

Figure 2–2: Chemical reaction involved in the formation of a peptide bond. This formation results in the loss of one oxygen, which is removed in the form of water (blue balloon). The peptide bond between the carbonyl carbon of one amino acid and the amino nitrogen of the other is represented in red. The left end of the chain is the N-terminus and the right is the C-terminus.

with fewer than 30 residues are referred to as peptides. One of the shortest amino acid sequences that has a biological function is the pentapeptide YGGFM [89]. Proteins with more than 4,000 residues are also found in nature while their average size is around 367 residues [100]. The two ends of a protein are different from each other: one end contains the amino-group while the other end contains a carboxyl group. By convention, the amino terminal (N-terminal) residue is taken to be the beginning of a polypeptide chain while the carboxyl-terminal (C-terminal) residue is its end. Thus, for example, CKQTW and WTQKC are two different pentapeptides with different chemical properties. Inside the cell, protein synthesis takes place sequentially from the N-terminus to the C-terminus.

Nature	Amino Acids	3-letter code	Single letter code
Hydrophobic	Alanine	Ala	A
	Valine	Val	V
	Phenylalanine	Phe	F
	Proline	Pro	P
	Methionine	Met	M
	Leucine	Leu	L
	Isoleucine	Ile	I
Charged	Aspartic acid	Asp	D
	Glutamic acid	Glu	E
	Lysine	Lys	K
	Arginine	Arg	R
Polar	Serine	Ser	S
	Threonine	Thr	T
	Tyrosine	Tyr	Y
	Histidine	His	H
	Cysteine	Cys	C
	Asparagine	Asn	N
	Glutamine	Gln	Q
	Tryptophan	Trp	W
	Glycine	Gly	G

Table 2–1: List of the 20 amino acids. The full name, 3-letter and single letter codes are given as well as the chemical nature of the amino acids (hydrophobic, polar and charged). The side chain of glycine contains only one hydrogen atom. It is the smallest amino acid and has special chemical properties being usually considered either to belong to the hydrophobic group or to form a separate group.

2.1.2 Structure of Globular Proteins

The structure of proteins is determined either by X-ray crystallography or by Nuclear-Magnetic-Resonance (NMR) methods [101]. Nowadays, more than 30,000 protein structures are known. Those native structures possess a number of common properties:

Packing: Globular proteins are highly packed objects. Their average packing density is about 0.75 [85, 61]. For comparison, hard spheres of equal sizes are believed to have a maximum packing density of approximately 0.74. It may seem paradoxical that residues, which are constrained to maintain their sequential position in the backbone, pack better than hard spheres with no constraint. However, proteins are not composed of equal-size objects, and one can usually pack a set of unequal-size spheres more densely than a set of equal-size ones.

Hydrophobic core: The first X-ray crystallographic structure of a globular protein was determined for myoglobin in 1958 [52, 53]. It was noticed that amino acids in the core of the protein had almost exclusively hydrophobic side chains. This burying of hydrophobic residues is believed to be the primary driving force for protein folding [51, 58, 24]. The interaction accounting for the formation of an hydrophobic core, i.e., the hydrophobic interaction, is discussed in more detail latter in this chapter.

Secondary structure: The formation of a hydrophobic core requires the backbone to be buried inside the protein. However the backbone is highly polar with one hydrogen donor, NH, and one hydrogen receptor, $C{=}O$, for each peptide unit. In a hydrophobic environment, polar molecules are hydrophilic and they

tend to stay in contact with water molecules - as opposed to the core of the protein. Therefore, for the backbone to be buried inside the protein, it must be neutralized by the formation of hydrogen bonds. This neutralization takes place through the formation of regular *secondary structures.* The most important secondary structures are alpha helix and beta sheets.

The *alpha helix* was first described by Linus Pauling in 1951 [77]. It is the most important secondary structure in globular proteins. In this structure, amino acids are displayed around a main axis in an helical form - Fig. 2–3. Each turn of the helice has 3.6 residues with hydrogen bonds between $C{=}O$ of residue n and NH of residue $n{+}4$. In globular proteins, the length of α helices varies from five to over forty residues. The side chains of residues in an α helix are placed in the outer part of the helix.

Beta sheets are the second major structural element found in globular proteins [76]. They are formed by binding two parts of the backbone, which are not necessarily adjacent, through hydrogen bonds - Fig. 2–4. β sheets are aligned parallel to each other such that a hydrogen bond can form between a $C{=}O$ group of one segment and a NH group of the parallel segment and vice-versa. The length of each sheet ranges from 5 to 10 residues. Beta sheets can be parallel or antiparallel. Parallel beta sheets occurs when the N-terminal residue of the two segments of the sheet are facing each other such that the two ends run from a N to C terminus. When the two ends run in opposite direction the β sheet is said to be antiparallel - see Fig. 2–4.

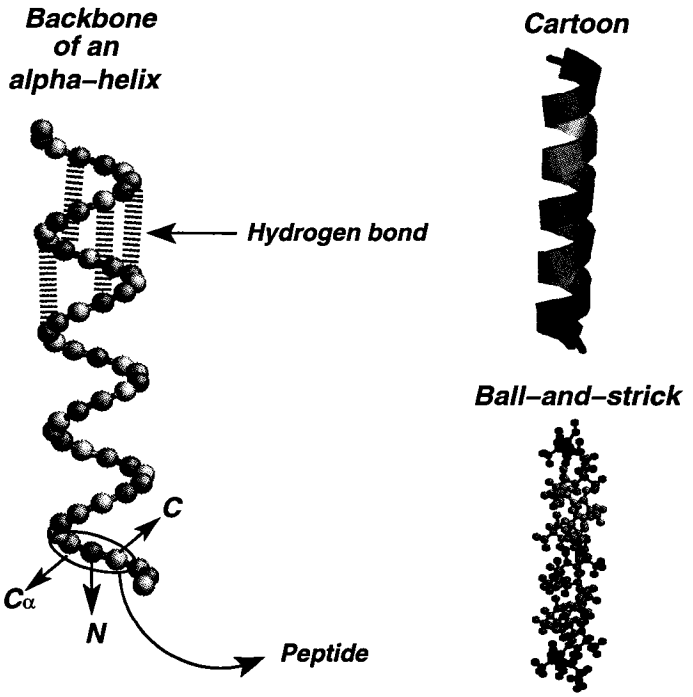

Figure 2–3: (LEFT) Representation of Hydrogen bonds in an α-helix . (RIGHT) Cartoon and ball-and-stick representation of an hypothetical α-helix made of 22 ALA amino acids.

The structure of proteins can be summarized, in a naive manner, as a combination of secondary structures which are efficiently packed together in order to bury the hydrophobic amino acids inside the core of the protein. A variation in the combination and length of secondary structures, as well as the type of connections (loops) between those structures, account for the different structures of proteins.

It is convenient to organize the structure of proteins into four hierarchical levels – see Fig. 2–5:

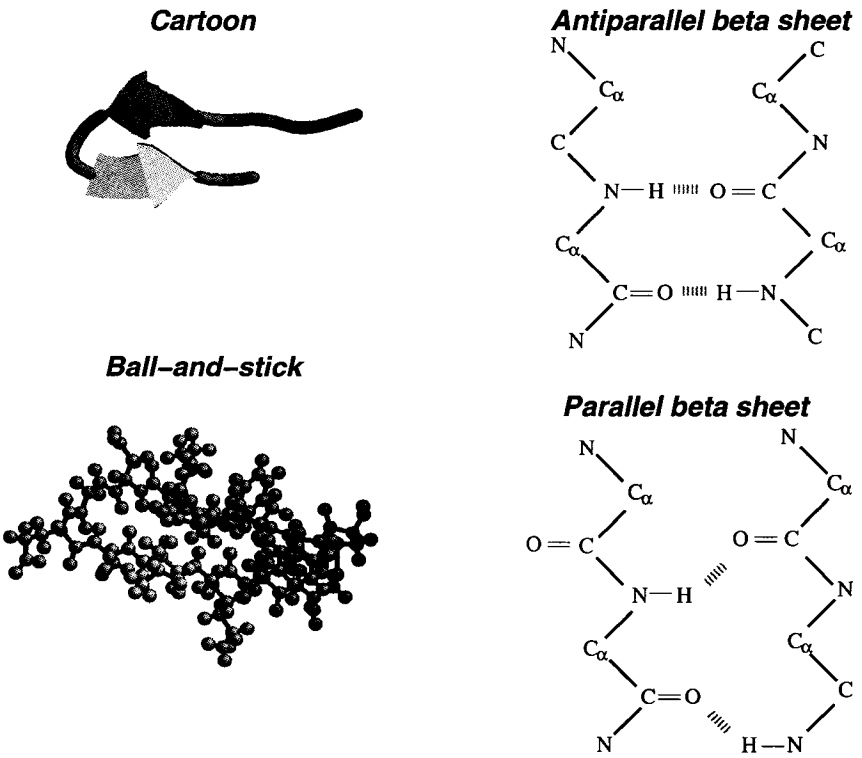

Figure 2–4: β-hairpin. (LEFT) Cartoon and Ball-and-stick representations of beta sheets. (RIGHT) Antiparallel and parallel β-sheets.

Primary structure refers to the sequence of amino-acids. This level of structure is determine by protein synthesis. Covalent bonding (see Fig. 2–2) is the interaction responsible for holding amino acids together.

Secondary structure is the recurring arrangement of adjacent residues in space. The most important secondary structures are alpha-helix and beta-sheet. This secondary level of structure is stabilized by hydrogen bonds between amino acids.

Tertiary structure refers to the arrangement of secondary structures in space. Hydrophobicity is believed to be the main driving force at this level of packing.

Quaternary structure deals with proteins that are made of several polypeptides or subunits.

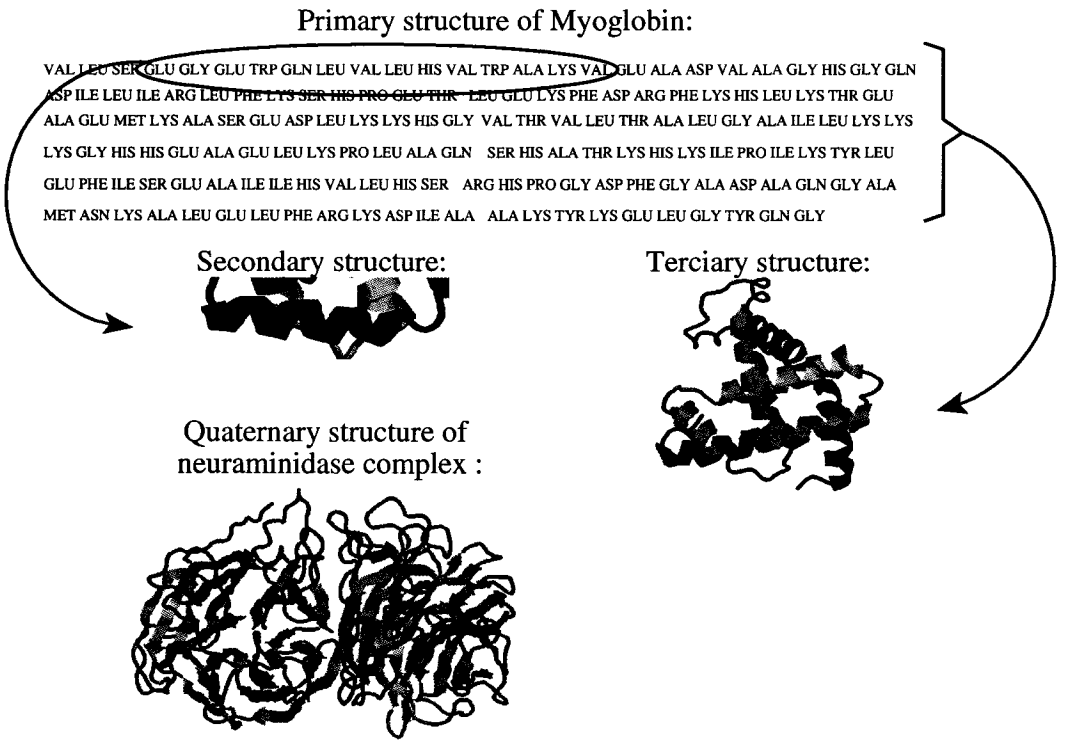

Figure 2–5: Hierarchical levels of structural organization.

2.1.3 Classification of Protein Structure and Designability

The importance of structure to protein function has led to the development of classification schemes for protein structure. Those schemes are useful in determining common evolutionary origins and, therefore, common function of proteins. The Structural Classification Of Proteins (SCOP) database organizes proteins within 11 main classes [71]. Example of these classes are: all α (which contains proteins made

exclusively of alpha helix secondary structures) and α/β (which includes proteins with alpha and beta secondary structures which are largely interspersed). Those classes are subdivided into folds. Proteins are said to have a common fold if they have the "same major secondary structures in same arrangement with the same topological connections" [71] – see Fig. 2–6 for an example of fold. In a naive manner, folds can be viewed as structural templates in which amino acid sequences fold.

In 1997, the SCOP database comprised 3,179 protein domains grouped into 279 folds. In 2005, the number of domains in the database was 75,930 and the number of folds 971. Althought the number of domains grew by a factor of 24, the number of protein folds increased only 3.5 times in the same period of time. Therefore folds are recurrent among proteins such that it is increasingly less likely that a newly solved protein structure would involve a new fold. Recurrence seems to suggest that folds are not merely an arbitrary outcome of evolution but there is a reason behind their selection.

A *designability principle* has been proposed [57] to explain nature's selection mechanism for protein structure. This mechanism states that folds are selected in nature because they are readily designed and stable against mutation, and that such a mechanism simultaneously leads to thermodynamical stability. This mechanism has been tested using the hydrophobic-polar lattice model and some of its variant models (see section 2.4.1). We will devote Chapter 4 of this thesis to investigate some implications of designability for protein dynamics.

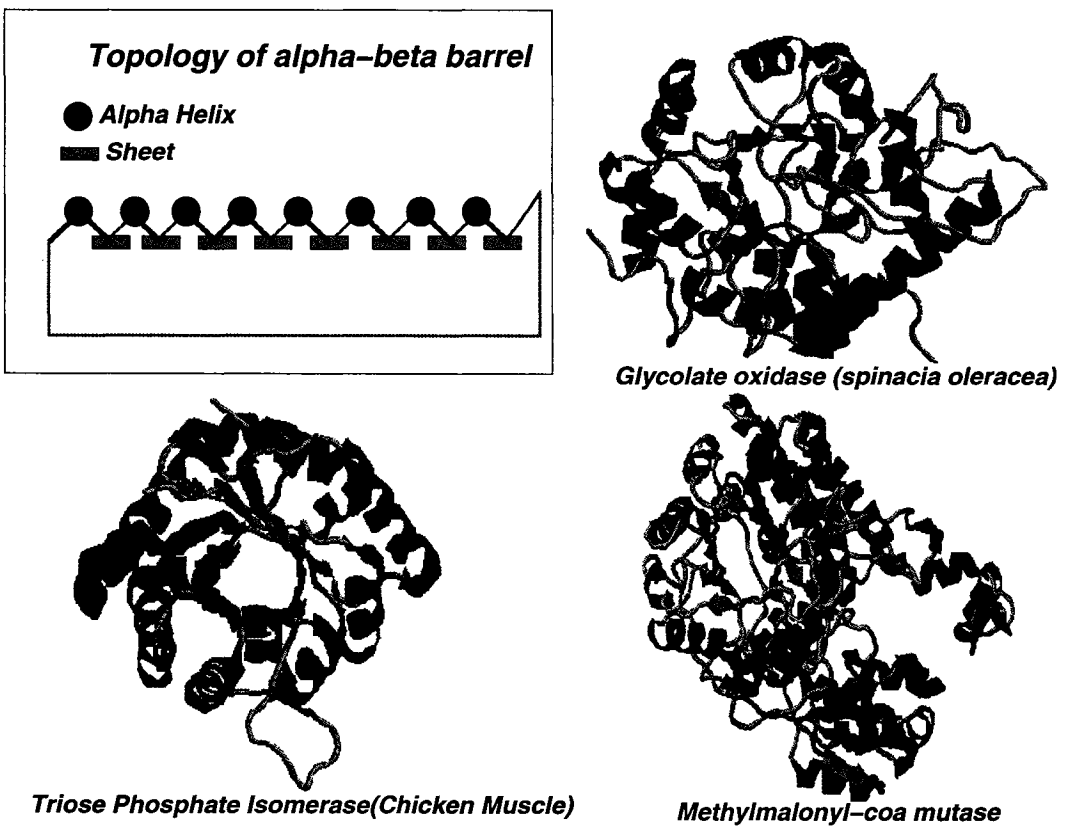

Figure 2–6: (Up-Left)Topological representation of the alpha-beta barrel fold. Example of three proteins containing the alpha-beta barrel fold: Glucolate oxidase (Up-Right), Thiose Phosphate Isomerase (Bottom-Left) and Methylmalonyl-coa mutase (Bottom-Right).

2.2 Thermodynamics of proteins

To study the stability of proteins pressure and temperature are useful quantities to be tuned. By adjusting temperature, we control the kinetic energy of each degree of freedom of the system, and a phase transition can be observed as temperature is reduced: the protein goes from an unfolded state into a folded one. Pressure can be used to both destabilize the structure of proteins, and as an antifreeze[1] . By increasing pressure, proteins can denature. The transition pressure varies from a low value ($<$ 200 MPa) to a very high value ($>$ 700 MPa), depending on the structure and nature of each protein.

2.2.1 Temperature dependent denaturation

Denaturation of proteins occurs at both high ($> 54\ ^oC$) and low (subzero) temperatures. Thermodynamically, this can be understood by studying the temperature dependence of the Gibbs energy of transition [55, 106, 78]: $\Delta G = G^D - G^N$, D and N upperscripts are used to describe the Denatured and Native states of the protein. At the transition temperature T_c, both the native and denatured states have the same energy:

$$\Delta G(T_c) = \Delta H(T_c) - T_c \Delta S(T_c) = 0, \tag{2.1a}$$

such that

$$\Delta S(T_c) = \Delta H(T_c)/T_c, \tag{2.1b}$$

[1] The freezing point of water decreases as pressure increases up to 200 MPa, where water freezes at about $-22\ ^oC$.

where ΔS is the change in entropy and ΔH is the change in enthalpy. The enthalpy of transition can be written in terms of the heat capacity at constant pressure:

$$\Delta H(T) = \Delta H(T_c) + \int_{T_c}^{T} \Delta C_P(T) dT, \tag{2.2}$$

since $\Delta C_p \equiv (\partial \Delta H(T)/\partial T)_P$. Whenever $\Delta C_P(T)$ can be assumed to be a constant:

$$\Delta H(T) = \Delta H(T_c) + (T - T_c)\Delta C_P. \tag{2.3}$$

For the entropy of transtion, we have:

$$\begin{aligned} \Delta S(T) &= \Delta S(T_c) + \int_{T_c}^{T} (\partial \Delta S(T)/\partial T)\, dT \\ &= \frac{\Delta H(T_c)}{T_c} + \int_{T_c}^{T} \Delta C_P(T) d(\ln T) \\ &= \frac{\Delta H(T_c)}{T_c} + \Delta C_P \ln\left(\frac{T}{T_c}\right). \end{aligned} \tag{2.4}$$

Now, using Equations 2.3 and 2.4, we obtain the Gibbs energy of transition:

$$\begin{aligned} \Delta G(T) &= \Delta H(T) - T\Delta S(T) \\ &= \frac{(T_c - T)}{T_c}\Delta H(T_c) + (T - T_c)\Delta C_P - T\Delta C_p \ln\left(\frac{T}{T_c}\right). \end{aligned} \tag{2.5}$$

Therefore, the Gibbs energy of transtion can be computed from a knowledge of T_c, $\Delta H(T_c)$, and ΔC_P. These quantities can be determined experimentally. Typical numbers for real proteins are: $T_c = 60$ °C, $\Delta H(T_c) = 500$ kJ mol^{-1}, and $\Delta C_P = 10$ kJ mol^{-1} K^{-1}. Fig. 2–7 shows the Gibbs energy for these values. ΔG has a convex shape which leads to two transition temperatures – one at T_c and the other at

-30 °C, corresponding to cold denaturation. Notice that the convex shape results from a subtle balance of the ΔH and ΔS terms.

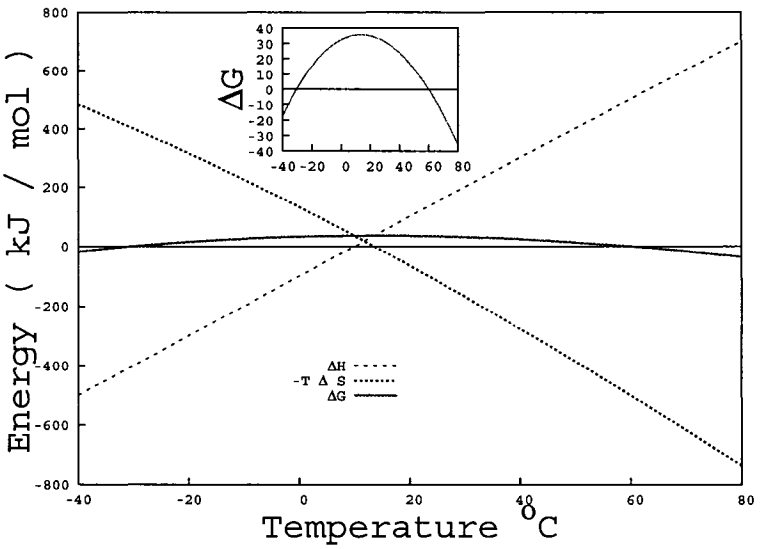

Figure 2–7: Gibbs Free energy (red/solid line) of a typical protein. The entropic (blue/dotted line) and enthalpic (grey/dashed line) contributions to the free energy are also shown.

From Equation 2.5, the convexity of Gibbs energy can be computed:

$$\partial^2 \Delta G(T)/\partial T^2 = -\Delta C_P/T. \tag{2.6}$$

Therefore cold denaturation can only occur if $\Delta C_P > 0$. This heat capacity of transition has been measured for several globular proteins [79]. Peaks of C_P were observed [78] at each end (high and low temperatures) of the native state and ΔC_P was positive: C_P showed a larger value in the denatured states.

2.2.2 Pressure dependent cold denaturation

As seen in the example above, cold denaturation occurs at subzero (−30 °C) temperatures. Hence it is difficult to observe cold denaturation experimentally: the solvent freezes before the protein denatures. This difficulty can however be overcome by studying the system at high pressure. By increasing the pressure, the protein becomes less stable: the red curve in Fig. 2–7 shifts downwards. By doing so, the subzero temperature, at which $\Delta G = 0$ shifts to a higher value, becomes accessible experimentally. Also, the freezing point of water decreases (up to −22 °C) as pressure increases (up to 200 MPa) – such that experiments can be performed at subzero temperature (see Fig. 2–8). Therefore increasing pressure works as both a destabilizer of the protein structure and as an antifreeze [55].

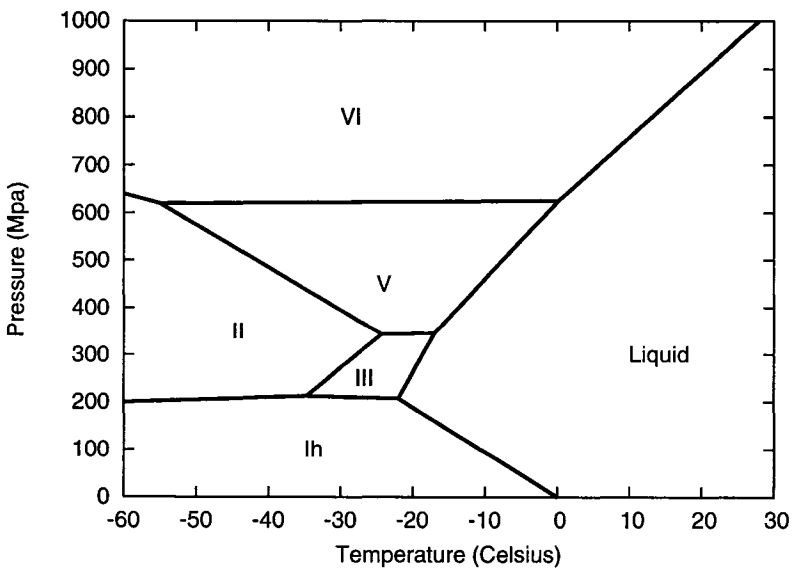

Figure 2–8: Phase diagram of water. This figure was obtained by connecting triple points of water [1] by straight lines. Roman numbers characterize the different states of ice.

The pressure dependent phase diagram of proteins can be described [44] using the Gibbs energy of transition $\Delta G = G^D - G^N$. By definition:

$$d(\Delta G) = -\Delta S dT + \Delta V dP, \tag{2.7}$$

where to first-order $\Delta S(T,P) = \int (\partial \Delta S/\partial T)_P \, dT + \int (\partial \Delta S/\partial P)_T \, dP$ and $\Delta V(T,P) = \int (\partial \Delta V/\partial T)_P \, dT + \int (\partial \Delta V/\partial P)_T \, dP$. The temperature and pressure dependence of ΔG can be obtained by integrating Equation 2.7 from a chosen reference point (T_o, P_o) to (T, P):

$$\begin{aligned} \Delta G &= \frac{\Delta\beta}{2}(P-P_o)^2 + \\ &\quad \Delta\alpha(P-P_o)(T-T_o) - \Delta C_p\left[T\left(\ln(T/T_o)-1\right)+T_o\right] + \\ &\quad \Delta V_o(P-P_o) - \Delta S_o(T-T_o) + \Delta G_o, \end{aligned} \tag{2.8}$$

where $\Delta\beta = (\partial\Delta V/\partial P)_T$ is the compressibility factor, $\Delta\alpha = (\partial\Delta V/\partial T)_P = -(\partial\Delta S/\partial P)_T$ is the thermal expansion factor and $\Delta C_P = T(\partial\Delta S/\partial T)_P$ is the heat capacity at constant pressure[2] . Close to the reference point $T\left(\ln(T/T_o)-1\right)+T_o \approx (T-T_o)^2/2T_o$, such that the equation above reads:

$$\begin{aligned} \Delta G &= \frac{\Delta\beta}{2}(P-P_o)^2 + \Delta\alpha(P-P_o)(T-T_o) - \frac{\Delta C_p}{2T_o}(T-T_o)^2 + \\ &\quad \Delta V_o(P-P_o) - \Delta S_o(T-T_o) + \Delta G_o. \end{aligned} \tag{2.9}$$

[2] Notice that Eq. 2.8 reduces to the case of constant pressure (Eq. 2.5) if one sets $P = P_o$, $T_o = T_c$ and on making use of Eqs. 2.1

This energy is shown in Figure 2–9 for Chymotrypsinogen at pH 2.07 [44]. The transition line between native and denatured states can now be obtained by solving $\Delta G(P,T) = 0$. When equated to zero, Equation 2.9 becomes the equation of a conic section which can be either a hyperbola or an ellipse. For proteins [44], one finds that $\Delta\alpha^2 > \Delta C_P \Delta\beta / T_o$ such that Equation 2.9 has an elliptic shape (see Fig. 2–9). Notice that in the previous section the free energy was shown to have the correct curvature only if $\Delta C_P > 0$.

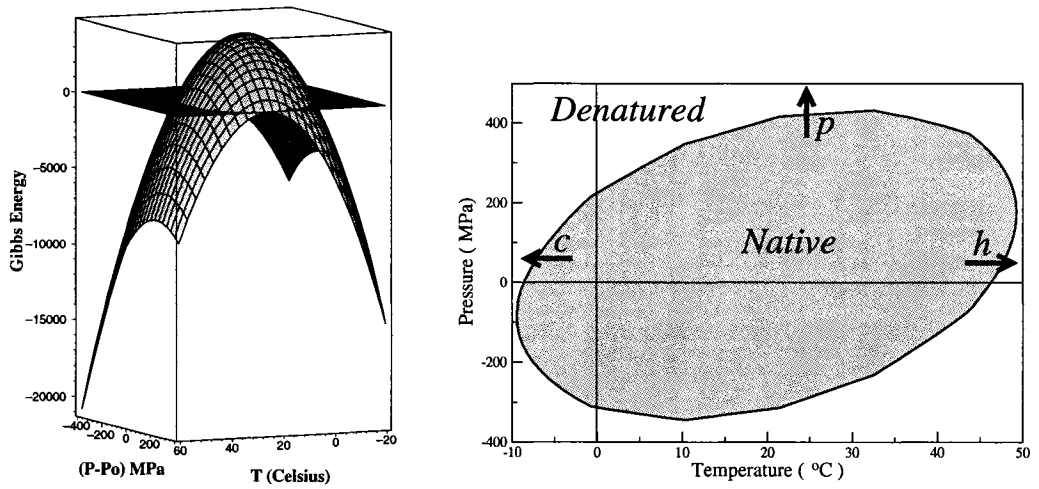

Figure 2–9: (LEFT) Pressure and temperature dependence of Gibbs energy (Eq. 2.9) for Chymotrypsinogen [44]. (RIGHT) Transition line separating the native and denatured state of the protein ($\Delta G = 0$). The arrows represent the three types of transition: pressure (p), cold (c) and heat (h) denaturation. Data taken according to reference [44]: $\Delta\beta = -1.24$ cm^6/cal mol^{-1}, $\Delta V_o = -14.3$ cm^3/mol, $\Delta\alpha = 1.32$ cm^3/mol K^{-1}, $\Delta S_o = -227$ cal/mol K^{-1}, $\Delta C_P = 3800$ cal/mol K^{-1}, $\Delta G_o = 2530$ cal/mol and the reference point is ($T_o = 273$ K, $P_o = 1$ atm). These data were obtained at pH 2.07.

The right panel of Figure 2–9 shows the values of pressure and temperature for which the protein is in the native (grey) and in the denatured states. If one follows a line of constant temperature towards higher pressure, the native state of the

protein becomes unstable and the protein unfolds. This is called pressure denaturation. At constant pressure, the protein unfolds at both a high temperature (heat denaturation) and a low temperature (cold denaturation) – see Fig. 2–9.

2.3 Hydrophobic effect

The hydrophobic effect accounts for the strong attraction among non-polar molecules immersed in water. It is considered the main driving force for folding and a satisfactory model for proteins has to take this effect into account. We will start this section describing some peculiarities of this effect. A thermodynamical model of this interaction is then described followed by a microscopic picture.

2.3.1 Signature of the hydrophobic effect

The hydrophobic effect can be inferred from the Gibbs energy of transfer of non-polar molecules from their bulk liquid state into dilute water solution: $\Delta G = G^w - G^b$. If ΔG is positive, the solute prefers to be surrounded by other solutes as opposed to water. The larger ΔG becomes, the higher the tendency of solutes to cluster together. In contrast, a negative ΔG implies a molecule that is soluble in water.

Non-polar molecules have a large positive energy of transfer. For example [48] the free energy of transfer of n-butane molecules at 25 °C is about 24.5 kJ mol^{-1}. Some insights can be gained by computing the enthalpy ΔH and entropy ΔS of transfer separately. For n-butane, the enthalpic contribution to ΔG is negative (-4.3 kJ mol^{-1}) while the entropic contribution $-T\Delta S$ is positive (28.7 kJ mol^{-1}) and corresponds to 85% of the interaction. Thus, non-polar solutes attract each other mostly to increase the entropy of the system.

The temperature dependence of ΔG, ΔH and $-T\Delta S$ for methane is shown in Fig. 2–10. As temperature decreases, bulk methane becomes less stable since ΔG decreases. Enthalpy can be held responsible for this destabilization as ΔH decreases

with decreasing temperature. Entropy has the opposite behavior: it stabilizes bulk methane since $-T\Delta S$ increases with decreasing temperature.

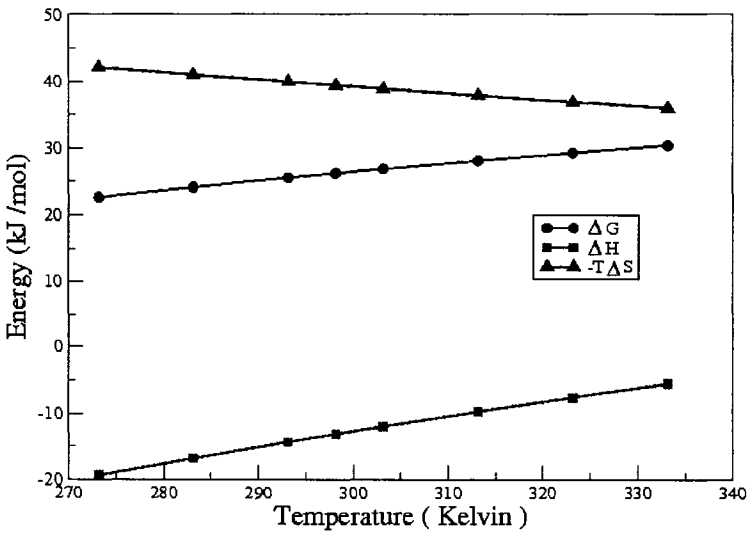

Figure 2–10: Experimental data (taken from reference [84]) of methane's free energy (black circle), entropy (triangles) and enthalpy (squares).

The effects of hydrophobicity can also be measured by the heat capacity of transfer from liquid to water $\Delta C_p = C_p^w - C_p^b$. For simple solutions, $\Delta C_p = C_p^w - C_p^b$ is small while for non-polar solutes it is large and positive [29]. This large heat capacity of transfer was shown to be proportional to the surface area around non-polar solutes accessible to water [50], and thus proportional to the number of solvent molecules around the solute [40].

2.3.2 Thermodynamical model: Muller's model

To describe these features of the hydrophobic effect, Muller's model [70, 56] focuses on hydrogen bonds among water molecules. These can exist in two states in mutual equilibrium:

$$\text{H bond (intact)} \rightleftharpoons \text{H bond (broken)}, \tag{2.10}$$

and the equilibrium constant K is given by:

$$K \equiv \frac{f}{1-f} = \exp(-\Delta G^o/RT), \tag{2.11}$$

where f is the fraction of broken hydrogen bonds, R is the gas constant and ΔG^o is the Gibbs free energy difference between the two states ($\Delta G^o = G^{intact} - G^{broken}$). The specific heat is given by:

$$C_p = \left(\frac{\partial \Delta H}{\partial T}\right)_P = \left(\frac{\partial f}{\partial T}\right)_P \Delta H^o, \tag{2.12}$$

where $\Delta H = f H^{broken} - (1-f) H^{intact}$ and ΔH^o is the enthalpic difference between broken and intact states. ΔH^o is considered to be independent of temperature. Equation 2.11 can be solved for f and its derivative with respect to temperature becomes:

$$\left(\frac{\partial f}{\partial T}\right)_P = \Delta G^o \frac{f(1-f)}{RT^2}, \tag{2.13}$$

such that

$$C_p = (\Delta H^o)^2 f(1-f)/RT^2. \tag{2.14}$$

To describe the hydrophobic effect one has to distinguish between bulk water and water in the hydration shell. The difference of these two states will contribute

ΔH_b^o	ΔS_b^o	ΔH_s^o	ΔS_s^o	ΔH_u	ΔS_u
9.80	21.6	10.696	27.36	0	0

Table 2–2: Parameters for Muller's model. ΔH and ΔS are given in units of J/mol and J/K/mol. For the number of hydrogen bonds in the shell, Muller uses $n = 3N/2$, where N is the number of solvent molecules in the shell. For propane, butane and isobutane, N is [40] 25,28 and 28.

to the experimentally accessible enthalpy, entropy and heat capacity changes upon hydration. Quantities referring to bulk and hydration shell will receive lower-scripts b and s respectively: ΔH_s^o, ΔS_s^o, f_s (for the first shell) and ΔH_b^o, ΔS_b^o, f_b (for the bulk).

Three additional parameters are needed: two to measure the relative enthalpy ΔH_u and entropy ΔS_u levels of the hydration shell with respect to those in the bulk (see Fig. 2–11) and a third to measure the amount n of hydrogen atoms in the hydration shell. In terms of these parameters, changes in thermodynamic quantities upon hydration are given by:

$$\Delta H = n\left[\Delta H_u - (1-f_s)\Delta H_s^o + (1-f_b)\Delta H_b^o\right], \tag{2.15}$$

$$\Delta S = n\left[\Delta S_u - (1-f_s)\Delta S_s^o + (1-f_b)\Delta S_b^o - R\Delta F\right], \tag{2.16}$$

and

$$\Delta C_p = n\left[C_s - C_b\right], \tag{2.17}$$

where $\Delta F \equiv F_s - F_b$ with $F_b \equiv f_b \ln f_b + (1-f_b)\ln(1-f_b)$ and similarly for F_s. These are the "mixing" entropies characteristic of mixture models.

This model has seven parameters: ΔH_b^o, ΔS_b^o, ΔH_s^o,ΔS_s^o, n, ΔH_u, ΔS_u. Estimated values for these parameters are shown in Table 2–2. The entropy computed

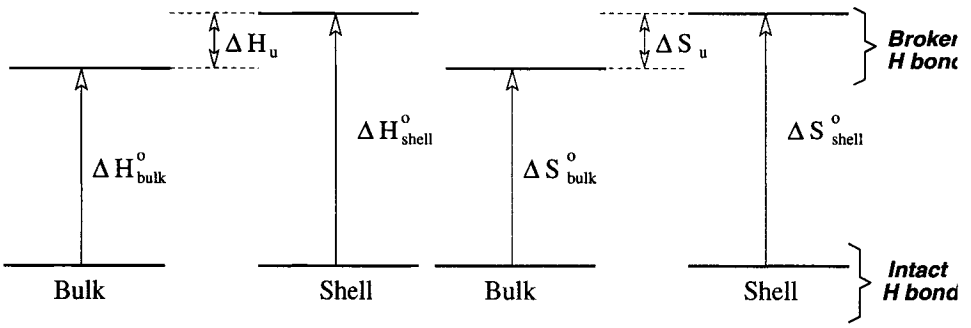

Figure 2–11: Schematic representation of the energetic (LEFT) and entropic (RIGHT) levels of broken and intact hydrogen bonds in the bulk and hydration shell.

from Muller's model (Equation 2.16) was compared to experiments for propane, butane, and isobutane at 25 °C giving -88.4 (-75.32), -101.4 (-93.20) and -98.9 (-89.14). Experimental values are inside parentheses and units are J/K mol^{-1}. Muller's model has been refined [56] to account for temperature dependent behavior of the entropy. A simplified version of this model has also been used to study cold denaturation of proteins - see section 2.4.2.

2.3.3 Microscopic model

The hydrophobic effect will now be discussed from a microscopic point of view. We first describe the hydrogen bond – which is the interaction responsible for the peculiar properties of water including the hydrophobic effect. A microscopic model of the hydrophobic effect is then described.

Hydrogen bond

The electronic distribution of the six valence electrons of oxygen in a H_2O molecule can be described by four sp^3 hybrids [48], two such hybrids being occupied by a single electron each and the other two hybrids being occupied by two

electrons each. Oxygen binds covalently with hydrogen by overlapping each half-occupied hybrid of oxygen with the hybrid of hydrogen atom - as represented in Fig. 2–12. The intramolecular distance between H and O atoms is about 0.10 nm.

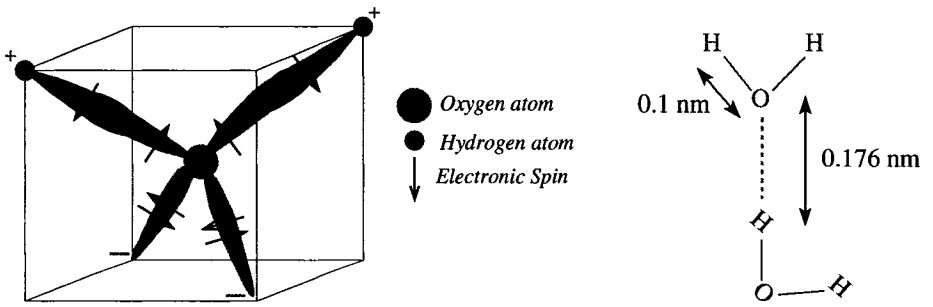

Figure 2–12: (LEFT) Schematic representation of the electronic distribution in a H_20 molecule. The polarization of the molecule is also represented by "+" and "-" signs. (RIGHT) Intermolecular bond (doted line) between two H_20 molecules.

Although water is a neutral molecule, it is polar due to the strong electronegativity of oxygen. The O atom attracts the electronic cloud belonging to H atoms to its neighborhood such that the net charge on the H atoms becomes slightly positive. On the other hand, the two fully-occupied hybrids of O are negative. As a result of this polarity, a strong electrostatic attraction between two water molecules occurs when a completely occupied hybrid of one molecule (negative hybrid) points towards a H atom (positive) of the other molecule. Since this type of bond requires the presence of a hydrogen atom, it is called a *hydrogen bond.* The intermolecular distance between O and H atoms is about 0.176 nm.

The binding energy of a H bond lies between 10 and 40 kJ mol^{-1}. This is much stronger than typical van de Waals bonds ($\sim$ 1 kJ mol^{-1}) but still weaker than covalent or ionic bonds ($\sim$ 500 kJ mol^{-1}). Hydrogen bonds are strongly directional

such that, upon solidification at ambient pressure, water does not assume a close-packed configuration but a hexagonal crystal structure.

Hydrogen bonds occur not only between water molecules, but between any highly electronegative atom (e.g., O, N, F and Cl) and a H atom which is covalently bound to another electronegative atom. For example, in a protein where two peptides come close together, the H atom which is covalently bound to the N atom forming one of the peptides can interact with the oxygen atom of the other peptide forming a hydrogen bond – see Fig. 2–4. These NH-O bonds are responsible for the secondary structures in proteins.

Iceberg Model

In the liquid phase water molecules form on average $3 - 3.5$ H bonds. These interactions are, however, disturbed when a non-polar molecule is introduced in water. The pertubation is mostly important close to the solute where, in most configurations, water has at least one broken H bond – i.e. at least one of the H bonds in the molecules points towards the solute. This is energetically unfavorable, and shell water, i.e. water molecules surrounding the solute, will favor the very few configurations, called clathrate cages [35, 10], where there is no broken bond – see Fig. 2–13. Since the number of clathrate cages is small, shell water can only occupy a small number of states and have a low entropy.

The free energy of shell water is thus minimized when those molecules assume clathrate-like configurations. This free energy per molecule is however larger than the one of bulk water, such that non-polar solutes tend to cluster together to reduce the amount of shell water in the system – see 2–14. Since shell water has a low entropy,

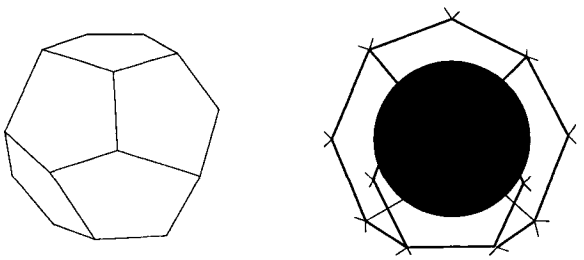

Figure 2–13: Representation of clathrate cages. The non-polar solute is shown in blue while water molecules rest at the edges of the polyhedra.

the transfer of shell water into the bulk during clustering of solutes is characterized by a large increase in the entropy of the system, as observed experimentally.

This picture of the hydrophobic effect was first suggested by Frank and Evand in 1945 [36]. They postulated that the tranfer of solutes into water perturb shell water in the direction of greater crystallinity. For obvious reasons, this picture is called iceberg model.

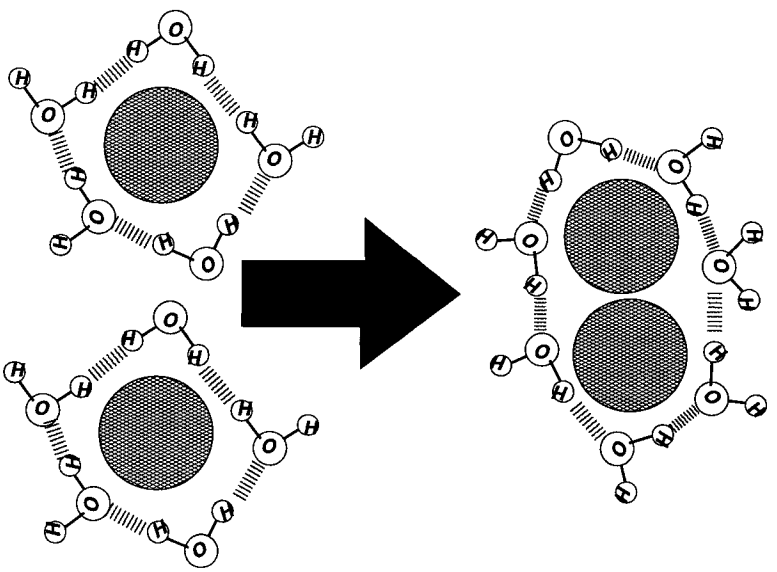

Figure 2–14: Schematic representation of the hydrophobic effect. Non-polar molecules attract each other in other to minimize their contact area with the solvent.

The iceberg model also explains the large heat capacity of transfer of non-polar solutes. At room temperatures, shell water occupies mostly clathrate-like configurations: low-energy and low-enthalpy state. As temperature increases, shell water increasingly populates a higher energy, higher entropy state: shell water becomes less ordered and has weakened attractions. The existence of 2 energetic states for shell water provides an energy storage mechanism which explains the large heat capacity of transfer. If ΔH is the difference in enthalpy between the 2 states and N is the number of molecules of shell water, then the temperature dependence of ΔC_p can be computed [40]:

$$\Delta C_p = N\frac{\Delta H^2}{RT^2}\left(\frac{\exp -(\Delta H/RT)}{[1+\exp(-\Delta H/RT)]^2}\right) \tag{2.18}$$

This model predicts that ΔC_p is proportional to the number of shell water and thus, to the the surface area of non-polar solutes. This result has been confirmed experimentally [50, 40].

2.4 Simplified Model for Proteins

Details of the atomic scale structure of proteins are provided by all-atom models. Simulations of those models with molecular dynamics or Monte Carlo methods permits the calculation of free energies, and also the characterization of structural changes in short periods of time. Many force fields have been developed for those simulations. AMBER, CHARMM and GROMOS are examples of widely used force fields. The best force field for a particular application depends upon the system being studied.

Coupled to molecular dynamics, this all-atom approach is limited by the integration time-step which must be smaller than the smallest atomic vibration period of the system. For proteins, the time-step is around one femtosecond (10^{-15}s) such that, to encompass a timescale of tens of microseconds, 10^9 integrations of the equations of motion are necessary. This is beyond reach of contemporary computational power for systems containing $\sim 10,000$ atoms. One of the longest simulation on biomolecular systems was conducted to study the early-stages of protein folding [26]. The protein in question was a small 36-residue peptide, the villin headpiece subdomain. The simulation was carried out for 1-μs at 300 K and contained 3000 water molecules. It used 256 processors for 4 months in a CRAY super-computer. To account for the complete folding, the simulation would have to be extended to a few tens of microseconds.

Another approach for studying proteins is to use minimal models. In such models, the degrees of freedom of the protein which are considered irrelevant to the phenomena being studied are averaged out either explicitly, or by phenomenological

arguments. Therefore minimal models lack the precision of all-atom systems but they are much faster, allowing simulations to be carried out for longer time. Those models are also easier to analyze. Hence they have proven of great importance in determining and highlighting the causes of phenomena under study. In this section I outline a few minimal models which are relevant for the work carried out in this thesis.

2.4.1 Hydrophobic model and designability

The hydrophobic force determines the tendency of hydrophobic amino acids to be buried inside the dry protein core, leaving polar amino acids to interact with the solvent at the surface of the protein. This force is considered the main driving force for protein folding. Models for studying the relation between sequence and structure have to take this force into account. One example of such a model is the Hydrophobic model, which provides a geometrical and therefore intuitive view of the designability principle described in section 2.1.3. This principle states that structures which survived evolution were more stable against mutation, i.e., more designable, and this feature leads to thermodynamical stability. Following Hao Li et al. [59] we provide geometrical arguments to explain the positive correlation between designability and thermodynamical stability.

In the hydrophobic model, the hydrophobicity of an amino acid sequence of length N is described by a N-dimensional vector $\vec{h} = (h_1, h_2, ..., h_N)$ where each element h_i is a number which accounts for the degree of hydrophobicity of the i-th amino acid. We take h_i to be normalized, $0 \leq h_i \leq 1$, such that the most hydrophobic amino acid has $h_i = 1$ while the most polar is defined by $h_i = 0$. The structure taken

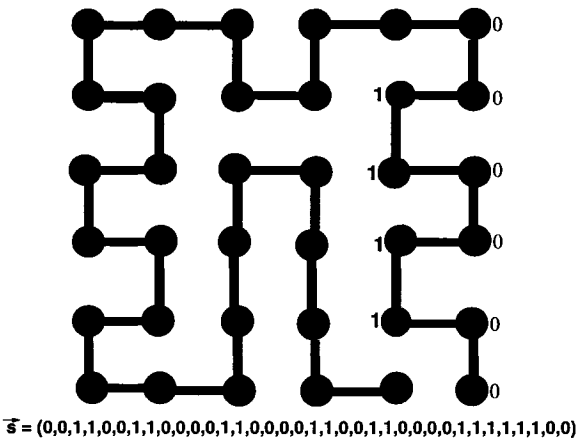

Figure 2–15: Schematic representation of a compact 6×6 lattice structure. Structures are represented by vectors where each element represents a lattice site and can be either 0 or 1 – according to whether the site is on the surface or in the core, respectively.

by a given sequence is also described by an N-dimensional vector $\vec{s} = (s_1, s_2, ..., s_N)$. Each element s_i of this vector is a number describing the degree of burial inside the protein of the i-th amino acid. For simplicity we consider that s_i is 1 if the i-th amino acid is in the protein core and 0 if it is in the surface of the protein. We also assume that sequences fold only in structures which are maximally compact. See Figure 2–15 for an example of structure.

The energy of a sequence which is folded in a particular structure is defined by:

$$\mathcal{H} = -\sum_{i=1}^{N} s_i h_i. \tag{2.19}$$

This energy is minimized when hydrophobic amino acids are buried inside the dry protein core. For the task of determining the minimal energy structure of a given

sequence, it is useful to rewrite Equation 2.19 in the form:

$$\mathcal{H} = \sum_{i=1}^{N} (h_i - s_i)^2. \tag{2.20}$$

Eqs. 2.19 and 2.20 are equivalent since for a fixed sequence, $\sum_{i=1}^{N} h_i^2$ is a constant and $\sum_{i=1}^{N} s_i^2$ has the same value for all compact structures.

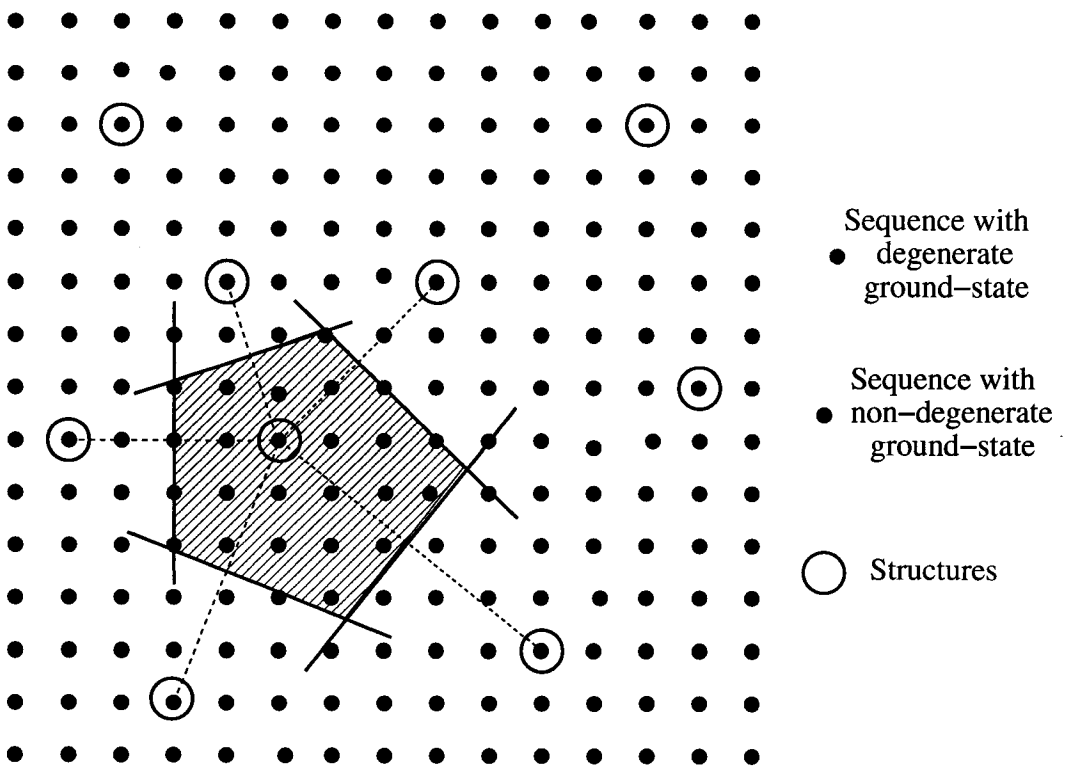

Figure 2–16: Schematic representation of structure and sequence spaces and the Voronoi construction. The shaded area corresponds to the Voronoi polytope.

Now since $\vec{h}$ and $\vec{s}$ are vectors spanning the same space they can both be represented in the same N-dimensional Euclidean space. The number of sequences of length N that can be formed with the 20 amino acids that exist in nature is 20^N. In principle all those sequences can exist and can be represented in the Euclidean

that, on average, the difference between the squared distance of $\vec{h}$ and their two closest structures is large. Thus, this geometrical representation explains the positive correlation between designability and thermodynamical stability: both properties require a large Voronoi polytope.

One can therefore conclude that designability of structures is directly related to the distribution of structures in the N-dimensional space such that designability can, in principle, be computed without having to sample through the space of sequences. Although the thermodynamical stability of a particular protein depends on its amino acid sequence, the average stability of structures does not depend on amino acid sequences and can also be computed from the distribution of structures only. Assuming that nature has chosen high designable structures as proteins folds, as suggested by the designability principle, it might be expected that other properties of proteins will also be favored by this choice. One example of such a property is the stability to forces – which we discuss in detail in Chapter 3.

2.4.2 Putting proteins back into water

Paolo De Los Rios and Guido Caldarelli [87] proposed a model to describe cold denaturation. In this model, the protein corresponds to a self-avoiding random walk where each amino acid occupies a site in a lattice. All other nodes i are occupied by water molecules which can be in the bulk or form the first shell around the polymer. The bulk state is considered to be q times degenerate and for simplicity the energy of this state is set to zero. The first shell can be in an ordered or a disordered state [70]. The disordered state is considered to be $q-1$ times degenerate while the ordered state is not degenerate. This model is therefore described by three parameters: J,

K and q – where J and K are the energies of the ordered and disordered shell states respectively. If s describes the state of first shell water molecules such that $s = 0$ represents the ordered state and $s = 1, ..., q-1$ corresponds to the disordered states, then the Hamiltonian of the system reads:

$$\mathcal{H} = \sum_j \left(-J\delta_{s_j,0} + K(1 - \delta_{s_j,0})\right), \tag{2.21}$$

where the sum is over all water molecules that are nearest neighbors of some hydrophobic monomer. Therefore, for each conformation C of the polymer its energy can be computed. The partition function of the system can be cast in the form: $Z_N = \sum_C Z_N(C)$ – where the partition function of a given conformation C reads:

$$Z_N(C) = q^{n_b(C)} \left[exp(\beta J) + (q-1)exp(-\beta K)\right]^{n_s(C)}, \tag{2.22}$$

n_s and n_b are the number of water molecules in the first shell and bulk respectively, and β is reciprocal of the thermal energy.

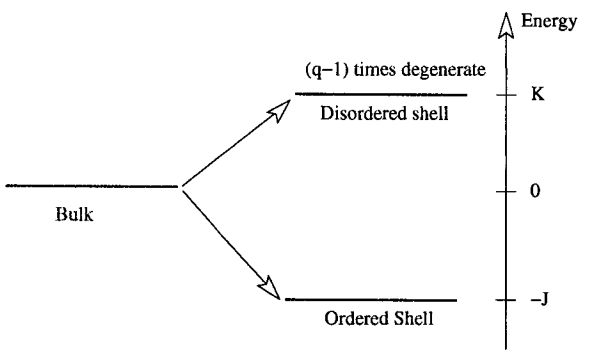

Figure 2–17: Representation of the energy of bulk (red) and first shell (blue) water molecules.

It is possible to classify the polymer according to n_s. For self-avoiding random walks of size N, the fraction of configurations of perimeter n_s is well approximated by a Poisson distribution of the form:

$$P_N(n_s) \sim e^{\delta(N-1)} \frac{[\delta(N-1)]^{2N+2-n_s}}{(2N+2-n_s)!}, \tag{2.23}$$

with $\delta \sim 0.75$. Using this distribution and equation 2.22, the partition function of the system reads:

$$Z_N(\beta) = \sum_{n_{min}}^{2N+2} P_N(n) q^{2N+2-n} \left[\exp(\beta J) + (q-1)\exp(-\beta K)\right]^n, \tag{2.24}$$

where the smallest perimeter n_{min}, assuming that in this configuration the system has a circular compact shape, is $2\sqrt{(\pi)}\sqrt{(N)}$. The maximum number of water sites in contact with the polymer is $2N+2$. All these sites where taken into account when computing $Z_N(C)$.

The specific heat, computed as $C_v = \beta^2 \left(\partial^2 \ln Z/\partial\beta^2\right)$, is shown in Fig. 2–18. Three peaks appear in the specific heat. From zero temperature to the first peak, the polymer is swollen – in agreement with cold denaturation. As the temperature is raised above the first peak, the polymer folds and the number of first shell water is approximately $2\sqrt{(\pi)}\sqrt{(N)}$. As temperature increases above the second peak, the polymer occupies globule-molten states. Beyond the third peak, the polymer reopens.

The free energy difference ΔF between the denatured and native states is convex – see section 2.2.1 – such that it becomes zero at both a high and a low temperature. These temperatures correspond to the transition point of hot and cold denaturation.

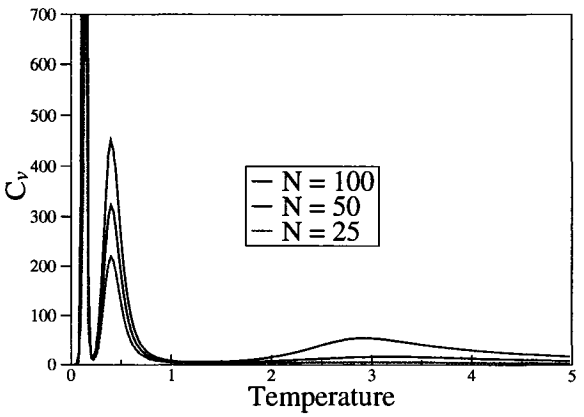

Figure 2–18: Heat capacity for different polymer sizes. Here, $K/J = 2$ and $q = 10^3$.

A rough estimation of ΔF was also computed for this model showing that it has the expected curvature – see Fig. 4 of reference [87].

2.4.3 Pressure dependent cold denaturation

In this section I describe the model proposed by Marques et al. [64] to describe cold denaturation. In this model, the protein corresponds to a self-avoid random walk where each amino acid sits in a site of a $N \times N$ lattice. All other sites i are occupied by one water molecule which can be in q different states σ_i. When two neighboring water molecules $< i, j >$ are in the same state ($\sigma_i = \sigma_j$), they are said to form a hydrogen bond that increases the volume of the system by ΔV. If the two molecules do not have the same orientation, the interaction of the particles does not imply any increment in volume. Therefore,

$$N_{HB} = \sum_{<i,j>} \delta_{\sigma_i,\sigma_j}, \tag{2.25}$$

corresponds to the number of hydrogen bond in the system. The Hamiltonian of water-water interaction is simply $\mathcal{H} = -JN_{HB}$, where $J > 0$ is the strength of the hydrogen bond interaction between water molecules. The enthalpy of the system is simply given by:

$$\mathcal{H} + PV = -(J - P\Delta V)N_{HB}. \tag{2.26}$$

Note that this solvent model has a critical pressure $P_c = J/\Delta V$. At this critical pressure, the system behaves as a paramagnetic system at zero field – see Fig. 2–19. Above that pressure, N_{HB} decreases as the temperature is lowered and water molecules undergo a transition to a state where all hydrogen bonds are broken. Below P_c, the number of hydrogen bonds increases as temperature decreases and water molecules undergo a sharp transition at $T = T_c = (J - P\Delta V)/[ln(1 + q)]$ to a state where all hydrogen bonds are formed. Therefore this model reproduces the freezing of water to low ($P > P_c$) and high ($P < P_c$) density ice.

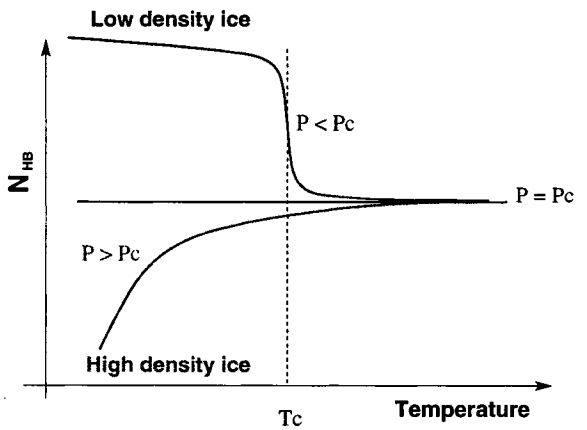

Figure 2–19: Schematic representation of the temperature dependence of N_{HB} at three different pressure.

The energetics of the protein is given by:

$$\mathcal{H}_p = J_r n_{HB}\big(N_{max} - \sum_{<i,j>} n_i n_j\big), \tag{2.27}$$

where $J_r > 0$ is the strength of the interaction and $n_{HB} \equiv N_{HB}/N_{water}$, N_{water} being the number of water molecules in the system. The term inside the brackets decreases as the protein becomes more compact: N_{max} being the maximal number of residue-residue contact and $n_i = 1$ if the lattice position i is occupied with a residue, and zero otherwise. The sum is over all first-neighbors residues. Note that $\mathcal{H}_p$ is not a local Hamiltonian: N_{HB} is the number of hydrogen bonds of the whole system.

According to Equation 2.27, the energetic cost of embedding a protein in water is less important when the protein assumes compact configurations. This cost is proportional to the amount of hydrogen bonds in the system such that, for $P < P_c$, it is continuously less likely to have an open protein as temperature decreases since N_{HB} increases. For $P > P_c$, N_{HB} decreases with decreasing temperature such that the cost of having an open protein in water also decreases. Close to the solidification point of water the protein unfolds – in agreement with cold denaturation. Thus, this model reproduces qualitatively the pressure and temperature dependent behavior of proteins.

In the model, the presence of water-water hydrogen bonds is the driving force for protein folding. This mechanism is in contrast to the results of experiments: protein folding is known to be entropically driven [79]. In the model, the occurrence of cold denaturation results from a lack of water-water hydrogen bonds at high pressure. Again, this is in contradiction with molecular dynamics simulation which find a

correlation [75] between water-water hydrogen bonds of shell water and cold denaturation. Finally, it is important to point out that cold denaturation can also occur at low pressure ($P < P_c$) if other mechanisms are introduced to destabilize the protein – such as increasing the solution's pH. This cannot be reproduced by the present model.

In conclusion, the present model is attractive due to its simplicity. It accounts qualitatively for the behavior of proteins. However it has some inconsistencies. In chapter 5 of this thesis we study cold denaturation and suggest a mechanism for cold denaturation which is in line with molecular dynamics simulations [75] and the model described in section 2.4.2.

2.5 Langevin Dynamics

Simulations in this thesis rest upon the study of Langevin's equation for Brownian motion [83]. This equation accounts for the motion of a particle in thermal equilibrium with the medium surrounding it. It assumes that the medium can be described by two types of forces: a rapidly fluctuating force F^r and a viscous-damping force F^d. The latter mimics the effect of the average collision of the atoms of the medium with the Brownian particle and the former, deviations from this average behavior. In the most simple case, the damping force is linearly dependent on the particle's velocity such that the equation of motion reads:

$$m\frac{dv}{dt} = -\gamma v + F^r(t), \tag{2.28}$$

where γ is the viscosity coefficient, m and v are the mass and the velocity of the Brownian particle. The formal solution of equation 2.28 is:

$$v(t) = v_o e^{-\gamma t} + e^{-\gamma t}\int_0^t dt' e^{\gamma t'} F^r(t'). \tag{2.29}$$

The first moment of F^r is chosen to be zero such as to avoid biased motion and this stochastic variable is uncorrelated in time:

$$< F^r(t)F^r(t+\tau) >= g\delta(\tau). \tag{2.30}$$

The intensity of this random force g is chosen such as to guarantee equilibrium. This is achieved whenever the work performed per unit of time by the random force, $E_r = lim_{\tau \longrightarrow 0} < F^r(t+\tau)v(t) >= g/2$, is equal to the work dissipated per unit of

time by the friction force, $E_f =< \gamma v^2 >= \gamma k_b T/m$. As a result:

$$g = 2\gamma k_b T/m. \tag{2.31}$$

This is known as the fluctuation-dissipation theorem since it accounts for how energy fluctuations are dissipated in systems close to equilibrium.

2.5.1 Numerical Integrators

Several algorithms have been used to solve Langevin's equation of motion, Eq. 2.28, numerically. The most straightforward, *Euler's method* [103], computes the position and velocity of each particle i at time t by iterating through the equations:

$$\vec{r_i}(t+\Delta t) = \vec{r_i}(t) + \vec{v_i}(t)\Delta t, \tag{2.32}$$

$$\vec{v_i}(t+\Delta t) = \vec{v_i}(t) + M^{-1}\vec{F_i}(t)\Delta t, \tag{2.33}$$

starting from the initial condition $\vec{r_i}(0)$ and $\vec{v_i}(0)$. Here, $\vec{F_i}$ is the sum of conservative, dissipative and random forces. Hence the equation of motions are non-linear. Conservative forces are computed from the interaction potential $\mathcal{V}$ between particles of the system. The random force is gaussian distributed with zero mean and root-mean-square $\sigma = (g/\Delta t)^{1/2}$. The dependence of σ on Δt accounts for the discreteness of time in numerics. A few authors [73, 7] make use of a uniformly distributed random force in the range $[-3\sigma : 3\sigma]$. Both, the normally and uniformly distributed force ensure equilibrium at a temperature T.

Although the integration of the equations of motion with Euler's algorithm accounts for the correct equilibrium properties of the system, it does not handle conservative forces in an appropriate manner. The Verlet class of algorithms are

more suitable for this purpose. In particular, one can integrate the equations of motion with the velocity-verlet algorithm [2] which in its original form reads:

$$\vec{r_i}(t+\Delta t) = r(t) + \vec{v_i}(t)\Delta t + \frac{\Delta t^2}{2M}\vec{F_i}(t), \tag{2.34}$$

$$\vec{v_i}(t+\Delta t) = \vec{v_i}(t) + \frac{\Delta t}{2M}\left[\vec{F_i}(t) + \vec{F_i}(t+\Delta t)\right]. \tag{2.35}$$

Note that the dissipative force on particle i is linearly dependent on its velocity such that, both left and right hand side of Eqn. 2.35 depend on $\vec{v_i}(t+\Delta t)$. Therefore to solve Langevin's equation of motion with the velocity-verlet algorithm, one has to solve Eqn. 2.35 for $\vec{v_i}(t+\Delta t)$. The modified algorithm reads:

$$\vec{r_i}(t+\Delta t) = r(t) + \vec{v_i}(t)\Delta t + \frac{(\Delta t)^2}{2M}\vec{F_i}(t), \tag{2.36}$$

$$\vec{v_i}(t+\Delta t) = \vec{v_i}(t) + \frac{\Delta t}{2M}\left[\vec{F_i}(t) + \vec{F_i^r}(t+\Delta t) + \vec{F_i^c}(t+\Delta t)\right]\left(1 + \frac{\gamma\Delta t}{2M}\right)^{-1}.$$

This algorithm is very easily implemented and will be used in chapters 3 and 4 of this thesis. Other methods have also been used to perform Langevin molecular dynamics [13, 49, 103, 30].

2.5.2 Constant Pressure Simulations

Langevin dynamics is appropriate to describe systems in thermal equilibrium – i.e. for the canonical ensemble. However many situations, especially in soft-matter, require simulations to be performed at both constant temperature and constant pressure. While temperature is computed from the microscopic degrees of freedom using the equipartition theorem, $m\langle\vec{v^2}\rangle/2 = 3k_bT/2$, pressure is obtained from the

virial theorem [42]:

$$\frac{1}{3}\langle\sum_{i=1}^{N}\vec{r_i}\cdot\vec{F_i}\rangle = -Nk_bT, \tag{2.37}$$

where F_i is the sum of intermolecular forces F_i^{int} and external forces F_i^{ext}. The latter are related [2] to the pressure P applied by the walls of a container where the particles are evolving in time: $d\vec{F_i} = -P\hat{n}dA$. When integrated over the area of the container:

$$\sum_{i=1}^{N}\vec{r_i}\cdot\vec{F_i^{ext}} = P\int\hat{n}\cdot\vec{r}dA = 3PV, \tag{2.38}$$

such that the pressure can be written in terms of the microscopic degrees of freedom as:

$$P = Nk_bT + \langle\sum_{i=1}^{N}\vec{r_i}\cdot\vec{F_i^{int}}\rangle. \tag{2.39}$$

Several simulational methods exist to perform molecular dynamics at constant pressure. They consist in constraining the atomic positions inside a simulation box whose volume V vary in order to target a constant pressure. In chapter 5 we use the extended method proposed by Andersen [3]. It assumes that V is an additional variable which contributes with PV to the potential energy of the system and with $\frac{1}{2}Q\dot{V}^2$ to its kinetic energy – where Q has the units of mass and can be interpreted as the mass of the container.

For simulational purpose, it is convenient to describe the position of particles in terms of reduced variables: $\vec{s_i} = \vec{r}/L$ where $L = V^{1/3}$. In terms of these variables, the Lagrangian reads:

$$\mathcal{L} = \sum_i\frac{L^2}{2}m_i\dot{\vec{s}_i}^2 - \sum_{i<j}\mathcal{V}(L,\vec{s}_{ij}) + \frac{Q}{2}\dot{V}^2 - PV, \tag{2.40}$$

where m_i is the mass of ith particle, $\mathcal{V}$ is the interaction potential between particles of the system. The Hamiltonian is obtained via a Legendre transformation,

$$\mathcal{H} = \sum_i \pi_i \dot{\vec{s}}_i + \Pi_V \dot{V} - \mathcal{L}, \tag{2.41}$$

where $\pi_i = \partial\mathcal{L}/\partial\dot{\vec{s}}_i$ and $\Pi_V = \partial\mathcal{L}/\partial\dot{V}$ are the conjugate momenta. The equations of motion read:

$$\begin{aligned} \dot{\vec{s}}_i &= \frac{1}{L^2 m_i}\vec{\pi}_i, \\ \dot{\vec{\pi}}_i &= L\vec{f}_i, \\ \dot{V} &= \frac{1}{Q}\Pi_V, \\ \dot{\Pi}_V &= \mathcal{P} - P, \end{aligned} \tag{2.42}$$

where $\vec{f}_i = -(\partial\mathcal{V}/\partial\vec{r}_i)$ is the force acting on the ith particle, and $\mathcal{P}$ is the instantaneous pressure. $\mathcal{P}$ can cast in terms of the reduced variables:

$$\mathcal{P} = \frac{L}{dV}\sum_{i<j}\vec{f}_{ij}\vec{s}_{ij} + \frac{1}{dL^2V}\sum_i \frac{1}{m_i}\vec{\pi}_i^2, \tag{2.43}$$

where $\vec{f}_{ij} = -(\partial\mathcal{V}/\partial\vec{r}_{ij})$ and $\vec{s}_{ij} = (\vec{s}_j - \vec{s}_i)$ are the force and distance between particles i and j, and d is the number of dimensions of the system.

Therefore Eqs. 2.42 correspond to the equations of motion of N particles evolving in a constant pressure environment. Those equations can be solved numerically using the velocity verlet algorithm, i.e. Eqs. 2.34 and 2.35. Note that apart from the kinetic energy of the piston, the Hamiltonian $\mathcal{H}$ corresponds the enthalpy of the system.

2.5.3 Constant Temperature and Constant Pressure

To account for constant temperature and constant pressure simulations, Feller et al. [34] replaced the canonical equations of motion, Eqs. 2.42, by a stochastic Langevin process:

$$\begin{aligned}\dot{\vec{\pi}}_i &= L\vec{f}_i - \frac{\gamma}{L^2 m_i}\vec{\pi}_i + \sqrt{k_b T \gamma}\vec{\eta}_i(t), \\ \dot{\Pi}_V &= \mathcal{P} - P - \frac{\gamma_V}{Q}\Pi_V + \sqrt{k_b T \gamma_V}\eta_V(t).\end{aligned} \tag{2.44}$$

The random forces are defined according to the fluctuation dissipation theorem:

$$\langle \eta_i^\alpha \rangle = \langle \eta_V \rangle = 0, \tag{2.45}$$

$$\langle \eta_i^\alpha(t)\eta_j^\beta(t') \rangle = 2\delta_{ij}\delta_{\alpha\beta}\delta(t-t'), \tag{2.46}$$

$$\langle \eta_V(t)\eta_V(t') \rangle = 2\delta(t-t'), \tag{2.47}$$

$$\langle \eta_i^\alpha(t)\eta_V(t') \rangle = 0, \tag{2.48}$$

where α and β refer to the Cartesian directions. Those equations are used in chapter 5 of this thesis to account for constant temperature-pressure ensemble. We use the algorithm of Kolb and Dünweg [54] to solve Eqs. 2.44. In terms of the conventional variables $\vec{r}_i(t) = L(t)\vec{s}_i(t)$ and $\vec{p}_i(t) = L(t)^{-1}\vec{\pi}_i(t)$, this algorithm reads:

1. $\vec{p}_i(t + \Delta t/2) = \vec{p}_i(t) + \vec{F}_i(t)\Delta t/2$;
2. Compute $\mathcal{P}$ using $\vec{r}_i(t)$, $L(t)$, and $\vec{p}_i(t + \Delta t/2)$;
3. $\Pi_V(t + \Delta t/2) = \Pi_V(t) + (\mathcal{P} - P)\Delta t/2 - \gamma_V \frac{\Pi_V}{Q}\Delta t/2 + \sqrt{k_b T \gamma_V \Delta t}\vec{\eta_V}$;
4. $V(t + \Delta t/2) = V(t) + Q^{-1}\Pi_V(t + \Delta t/2)\Delta t/2$;
5. $\vec{r}_i(t + \Delta t/2) = \vec{r}_i(t) + (L^2(t)/L^2(t + \Delta t/2))(\vec{p}_i(t + \Delta t/2)/m_i)\Delta t$;
6. $V(t + \Delta t) = V(t + \Delta t/2) + Q^{-1}\Pi_V(t + \Delta t/2)\Delta t/2$;

7. $\vec{r_i}(t+\Delta t) = \frac{L(t+\Delta t)}{L(t)}\vec{r_i}(t+\Delta t/2)$;
8. $\vec{p_i}(t+\Delta t/2) = \frac{L(t)}{L(t+\Delta t)}\vec{p_i}$;
9. Evaluate $\mathcal{P}$ using $\vec{r_i}(t+\Delta t)$,$L(t+\Delta t)$ and $\vec{p_i}(t+\Delta t/2)$;
10. $\Pi_V(t+\Delta t) = \Pi_V(t+\Delta t/2) + (\mathcal{P}-P)\Delta t/2 - \gamma_V\frac{\Pi_V}{Q}\Delta t/2 + \sqrt{k_bT\gamma_V\Delta t}\vec{\eta_V}$;
11. $\vec{p_i}(t+\Delta t) = \vec{p_i}(t+\Delta t/2) + \vec{F_i}(t+\Delta t)\Delta t/2$,

where $\vec{F_i}\Delta t/2 = \vec{f_i}\Delta t/2 - \gamma\vec{p_i}/m_i\Delta t/2 + \sqrt{k_bT\gamma\Delta t}\vec{\eta_i}$. Note that this algorithm reduces to the velocity-verlet algorithm when $Q \longrightarrow \infty$.

Chapter 3

Scaling in Force Spectroscopy of Macromolecules

3.1 Introduction

Many essential biological processes for life depend on the reaction of various bonds and/or molecules to an applied force. One such example is how leukocytes recognize invading pathogenic organisms in blood vessels [68]. Atomic force microscopy (AFM) [63, 104] and bio-membrane force probe [67, 33, 32] (BFP) are now being used to determine the energy landscape of these complex molecules. In the former experiment, a molecule is attached to the tip of the AFM while a counter-molecule is held at the surface. A specific bond [9] between these two molecules is formed when the tip is brought close to the surface. By removing the tip from the surface at a constant loading rate and recording the most likely force at which the specific bond breaks, a complete spectrum of force versus loading rate is obtained.

Experimentally, the force is seen to increase approximately with the logarithm of velocity for about 3 or 4 decades. This behaviour can be explained in terms of a minimal model [31] in which specific bonds are modeled by a Lennard-Jones potential

while the surrounding environment of the bond is described by a stochastic force. The breaking of the bond then takes place through thermal fluctuations, which depends strongly on the energy barrier [98]. The potential energy of the bond added to the energy associated with the external force (the energy of the cantilever) accounts for the energy landscape [6]. An approximate solution of the model reveals that the energy barrier of the landscape decreases linearly with the applied force. This linear dependence then results in a linear force spectrum.

There is a need to expand the minimal model and to perform a detailed theoretical study of the breakup process [22]. In particular, it was recently shown that the linear dependence of the energy barrier on the applied force is rarely valid in similar cases where an interplay between thermal fluctuations and a time-dependent energy barrier exists [8, 90, 37]. The linear logarithmic behaviour must be replaced by a $\ln^{2/3} T/r$ dependence, where T is temperature and r is the loading rate. Since this behaviour is observed in systems as varied as Josephson junctions [37], friction of an AFM tip on a surface in the creep regime [28, 90], and possibly spin glasses [94], it raises the question of the universality of this result and of its applicability to the breaking of biological molecules.

In this chapter, we first use molecular dynamics to study the breaking rate of chain molecule held at a fixed length. We find that the breakup is a thermally activated process and extract the energy barrier and attempt frequency. Two situations are examined. In the first, the chain is composed of identical Lennard-Jones atoms with periodic boundary conditions. We find that the energy barrier $\Delta E \sim (s_c - s)^3$, where s is the strain of the chain and s_c is the critical strain, the strain at which

breakup would occur at $T = 0$. This differs from the universality classes described above, and arises from specific boundary conditions. A more realistic situation is to consider the chain of atoms attached to an harmonic spring representing the membrane of a BFP setup, or to be composed of links with different forces. In this situation, the energy barrier $\Delta E \sim (s_c - s)^{3/2}$, within the ramped creep universality class. A non-linear vanishing of the energy barrier has been observed in other model biological systems [60, 46], and a similar scaling form has also been proposed by Dudko et al [27].

These results are then used to discuss the dynamical breakup. We use a quasi-static approximation, in which the energy barrier $\Delta E(t) \sim (s_c - s(t))^{\gamma}$. We show that when the tension is increased at a constant rate, the strain s at which the system ruptures scales with the temperature T and the loading rate $r = ds/dt$ as :

$$(s - s_c) \sim T^{1/\gamma}[\ln(\kappa T/r)]^{1/\gamma}, \tag{3.1}$$

where $\gamma = 3$ for identical ring molecule and $\gamma = 3/2$ for BFP experiments. These results are confirmed by a direct molecular dynamics study of breaking on a simplified BFP set-up. Even though the rate of elongation that can be simulated using molecular dynamics is unrealistically high, these results validate the quasi-static approximation. Our results can thus be applied to experimentally useful range of stretching rate. In particular, this scaling is based on physical arguments and is universal, the detailed form of the potential enters only through the numerical constant κ.

3.2 Model Description and Simulation

We model the molecule being stretched by a one dimensional chain of N atoms, described by the position x_i $(i = 1, N)$. The atoms interact together through first-neighbor interactions only. For specificity, we choose the Lennard-Jones potential, (although the exact form of the interaction is not essential:

$$U(\Delta x) = \epsilon\left[\left(\frac{a}{\Delta x}\right)^{12} - 2\left(\frac{a}{\Delta x}\right)^{6}\right], \tag{3.2}$$

where ϵ is the binding energy, a is the inter-atomic spacing, and Δx is the distance between the atoms. The dynamics of this system is obtained by simple Langevin equations. The atoms are immersed in a solvent which acts as a friction force ($-\gamma\dot{x}_i$) and a random force $\xi_i(t)$ on each atom i. The intensity of the random force is given by the fluctuation-dissipation theorem

$$\langle \xi_i(t)\xi_j(t+\tau)\rangle = \delta_{i,j}\delta(\tau)2M\gamma k_B T, \tag{3.3}$$

where M and k_B are the mass and Boltzmann's constant, respectively, and the angular brackets denote an average. The equation of motion of atom i can now be written as

$$M\frac{d^2x_i}{dt^2} + M\gamma\dot{x}_i - F(x_i - x_{i-1}) + F(x_{i+1} - x_i) = \xi_i(t), \tag{3.4}$$

where $F(x)$ is the force computed from the potential. Two different situations can then be considered.

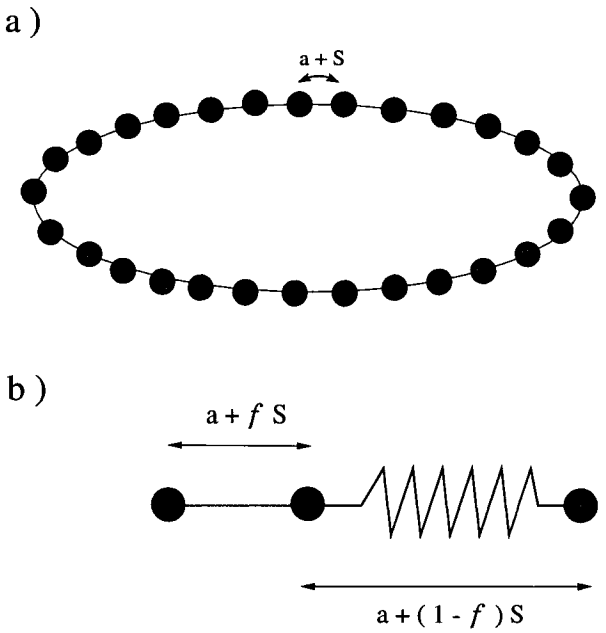

Figure 3–1: Schematic representation of stretched chain models. In (a), a periodic chain with similar atoms is presented in its local equilibrium configuration: all atoms stretch by the same amount S. In (b) we show a system made by a weak bond and a spring. The spring mimics the membrane of a BFP while the bond models the relevant weak bond of a complex molecule. At metastable equilibrium, a fraction f of the stretched length S of the system extends the weak bond.

1) Periodic chain: The simplest system to study is a periodic chain with N similar atoms, as depicted in Fig. 3–1(a). Periodic boundary conditions are imposed with the minimum image convention [2] so that the probability of rupture is equal at each point of the chain. The length of the box is fixed at $L = N(a + S)$ and initially all bonds are strained by an equal amount S. This setup corresponds to a metastable configuration since the total force on each atom is zero. Therefore, the average strain on each bond stays the same along the simulation until rupture occurs. In that case the broken bond stretches gradually to $a + NS$ while the strain on the others $N - 1$

bonds vanishes such that the sum of all bond length is conserved at all times. A similar system, at a fixed length, was already studied by Oliveira [73]. It was found that rupture usually occurred at a single bond through thermal fluctuations.

2) Attached chain: A more realistic situation is when different bonds (corresponding to different values of ϵ) are present in the molecule, with one end of the chain (atom $i = N$) strongly attached to a surface and the other (atoms $i = 1$) to the tip of an AFM or the membrane of BFP. The cantilever of the AFM is modelled by introducing a new atom x_0, linked to the first atom of the chain by an harmonic spring (with spring constant k_h), leading to an additional term in the potential energy:

$$U_h(x_1 - x_0) = \frac{1}{2}k_h(x_1 - x_0 - a)^2. \tag{3.5}$$

Of particular interest for BFP is when one of the bonds (between atoms i^* and $i^* + 1$) is much weaker than the others, and the spring constant less stiffer than the weak bond. Defining $\epsilon(i)$ as the binding energy between atoms i and $i + 1$, and representing the weak bond by $\epsilon = \epsilon(i^*)$, the specific situation that we consider is thus

$$\begin{aligned} \epsilon(i)/\epsilon &\gg 1 \qquad \text{For } i = 1, i^* - 1 \\ \epsilon(i^*) &= \epsilon \\ \epsilon(i)/\epsilon &\gg 1 \qquad \text{For } i = i^* + 1, N \end{aligned} \tag{3.6}$$

In this case, only the weak bond and the spring plays a role in rupture (the 2 subchains formed by the strong links are essentially undisturbed). Therefore to

speed up simulational time we ignore these strong bonds (see Fig. 3–1b). The atoms at the extremity of the setup are held fixed in place.

Dynamical breakup of the chains: The ultimate goal of a molecular dynamic simulation would be to compute the mean breaking force for chains subjected to strain rates spanning several orders of magnitude (typical experimental time scales range from a millisecond to a minute). However, due to prohibitively long computational times [96], molecular dynamics can only probe breaking for a very small range of loading rates. We thus proceed in two steps, we first calculate the energy barrier for a fixed chain length and show that this result can be used to understand the dynamical breakup. The strain s of the chain is thus fixed (through the boundary conditions) and the time $\tau(s)$ necessary for breakup is obtained.

The dynamics of atoms is given by solving the set of Eq. 3.4 using the velocity Verlet algorithm [2, 99]. These are solved numerically for a fixed length of the system (fixed strain) until a bond breaks irreversibly. In order to determine whether rupture has occurred irreversibly, we compute the characteristic time required for the largest bond length of the chain to reach different values d. This is shown in Fig. 3–2. Two distinct regimes are clearly distinguished. The first regime occurs when the length d of the largest bond is smaller than $d_c \sim 1.56$ (in units of a). In this case the time required to produce an increase of d is appreciable. The underlying physics of this regime is the competition between thermal fluctuation, which is responsible for increasing the length of the largest bond, and the elastic restoring force on this bond. The second regime occurs when the length of the largest bond is greater than d_c. Here, τ has reached a plateau: almost no time is required to increase the length of

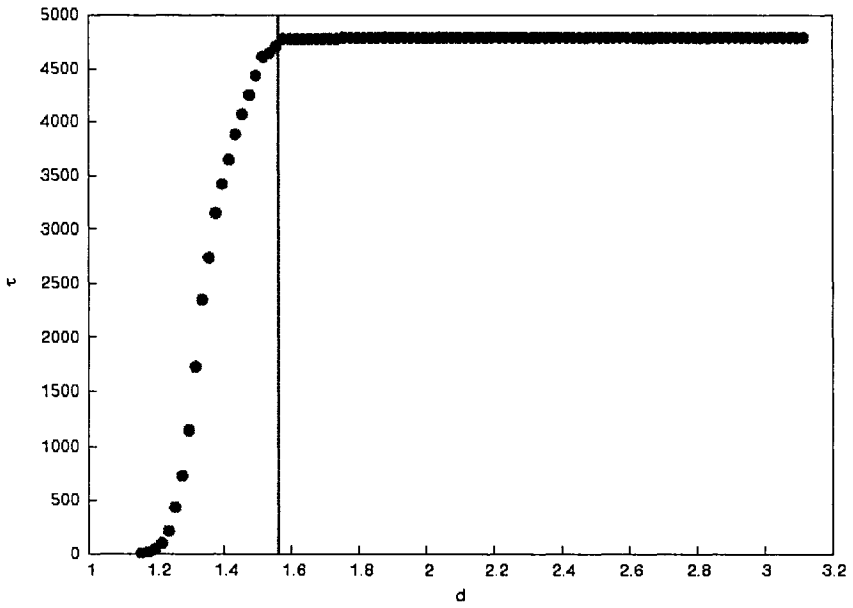

Figure 3–2: Dependence of the characteristic time of rupture τ on the length of the largest bond of the chain. The vertical line separates the two regimes of rupture. This simulation was performed at T=0.019 and S=0.035.

the largest bond. This occurs as the unstable point in the potential energy landscape is crossed when the elastic force changes sign and becomes a breaking force for the largest bond; the bond can not restore its equilibrium length and we can conclude that irreversible breaking has occurred. The average point of rupture d_c is therefore well defined.

The characteristic time of rupture τ can be computed by knowing the number of chains $N(t)$ that have not ruptured at time t. If there are initially N_0 chains, the time evolution is $N(t) = N_o \exp(-t/\tau)$. Unless otherwise stated, we use $N_o = 800$

in this paper. If breakup is thermally activated, the time for breakup follows an Arrhenius form

$$\tau(s) = \Omega^{-1}(s) e^{\Delta E(s)/k_B T}, \tag{3.7}$$

where the inverse of the prefactor, $\Omega(s)$, is the attempted breakup frequency and $\Delta E(s)$ is the *effective* potential energy barrier (*i.e. the amount of the energy the heat bath has to supply in order for rupture to occur*). The rate of breaking of the chain $R(s) = 1/\tau(s)$.

Simulations are performed in reduced units. Energy is written in terms of ϵ, length is given in terms or a and time is given in units of the smallest period of phonon oscillations $\tau_o = \frac{2\pi}{12\sqrt{(2\epsilon/Ma^2)}}$. For simplicity, the mass of each atom is chosen to be one while the friction constant is tuned to $\gamma = 0.25(2\pi/\tau_o)$. Below we present results from simulation. We obtain $\tau(s)$ for various temperatures, from which both the energy barrier $\Delta E(s)$ and the oscillation frequency $\Omega(s)$ can be obtained. The behaviour of $\Delta E(s)$ as a function of the strain is particularly important since it appears as an exponental factor in the thermal rate formula (Eqn. 3.7). At low strain, we expect linear behaviour, ie., $\Delta E(s)/\Delta E(0) = 1 + \mathcal{O}(s)$. At some critical strain value s_c, the energy barrier vanishes and the chain is naturally unstable. Close to s_c, we expect power-law behaviour $\Delta E(s) \sim (s_c - s)^\gamma$.

3.2.1 Numerical simulations: attached chain

This model corresponds to Fig. 3–1(b). The weak bond is defined by $\epsilon = 1$ and $a = 1$ - such that its stiffness (computed from the Lennard-Jones potential) is $k = 72\epsilon/a^2$. The equilibrium length of the harmonic spring is a and its stiffness is $k_h = 10^{-4}k$. Figure 3–3(a) shows the dependence of the logarithm of τ on the

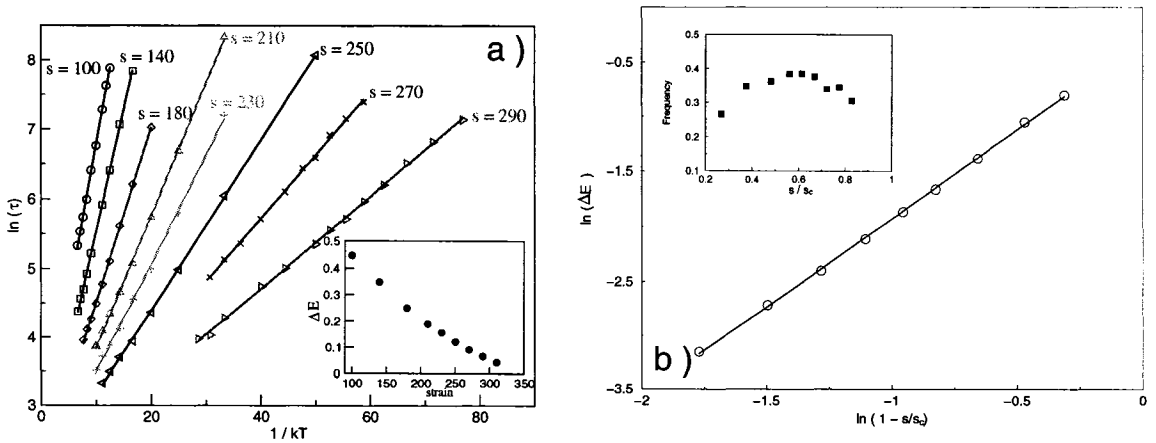

Figure 3–3: Simulational results for the weak + harmonic chain. In (a) the dependence of breakup time on the temperature and strain is presented. The energy barrier as a function of the strain is shown in the inset. In (b) we show the behavior of the energy barrier and attempt frequency (in the inset) on the strain s. The power-law dependence of the barrier is evident.

inverse of temperature for several values of strain. In agreement with Eq. 3.7, this dependence is linear, showing that breaking is a thermally activated process, mostly determined by the binding energy ϵ and the spring constant k_h. The energy barriers as a function of the strain s are shown in the inset of Fig. 3–3(a) and demonstrate the existence of a critical value of the strain s_c. Close to s_c, the energy barrier scales as $\Delta E \sim (s - s_c)^\gamma$, as shown in Fig. 3–3(b). The best fit yields values of $s_c = 370 \pm 10$ and $\gamma \approx 1.5 \pm 0.1$; the large uncertainty comes from having to determine two parameters from the fit, together with the constraint that power law behaviours is observed only for strains sufficiently close to the critical strain. The

prefactor is given in the inset of Fig. 3–3(b). It increases with strain until it reaches a maximum and eventually decreases as s_c is approached.

3.2.2 Numerical simulations: periodic chain

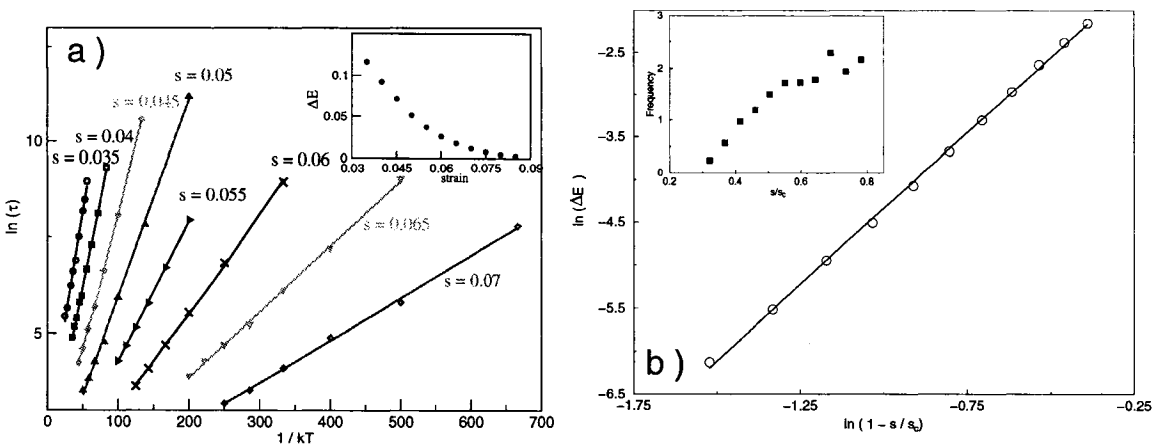

Figure 3–4: Simulational results for the periodic chain. Different colors correspond to different applied strains. In (a) the dependence of breakup time on temperature and strain is presented. The energy barrier as a function of strain is shown in the inset. In (b) we show the behavior of the energy barrier and attempt frequency (in the inset) on the strain s. The power-law dependence of the barrier is evident.

In this case, shown in Fig. 3–1(a), all the bonds of the periodic chains have the same binding energy $\epsilon = 1$. Figure 3–4(a) shows the dependence of the logarithm of τ on the inverse of temperature for several values of strain and several temperatures. Again, the straight lines shows that breakup is a thermally activated process. The inset of Fig. 3–4(a) shows $\Delta E(s)$ vs s. Again, there exists a critical strain at which the barriers disappears.

We show in Fig. 3–4(b) that the energy barrier varies as $\Delta E \sim (s - s_c)^{\alpha}$ close to s_c. A best fit of these simulational data gives $s_c = 0.115 + 0.005$ and $\alpha \approx 3.5 \pm 0.1$. The behaviour of the prefactor is more complicated [88]. Initially it increases and reaches a maximum as shown in the inset of Figure 3–4(b).

For both the periodic and weak bond chain, the behaviour close to s_c is very different from the linear dependence usually expected and on which the model of Ref. [31] is built. We now present mean field models that explain these results.

3.3 Mean-Field Analysis

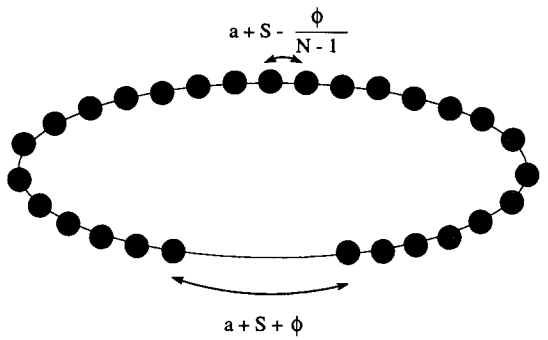

Figure 3–5: Mean-field description of the breakup of a periodic chain of atoms.

The breakup of a chain can be approached through an effective potential that describes rupture [73, 72, 80]. The potential is built around the idea that the molecule breaks locally at a *single* specific bond [73]. During breakup, one of the bonds is stretched by an extra amount ϕ while the other bonds relax according to the prescribed value of the chain length. The complete breakup process can then be described with the potential $U_{eff}(s, \phi)$, where s is the prescribed strain of the chain and ϕ represents the increase in strain for the breaking bond.

This potential typically has a metastable character. There is one minimum $\phi_{\min}$, located near $\phi = 0$ corresponding to a chain which is not broken. This state is separated from a continuum of values of ϕ (corresponding to a broken chain) by an energy barrier, with maximum located at $\phi_{\max}$. These points are found by the condition

$$\left.\frac{\partial U_{eff}(s,\phi)}{\partial \phi}\right|_{\phi=\phi_{\min},\phi_{\max}} = 0 \tag{3.8}$$

The energy barrier for a given strain s is then

$$\Delta E(s) = U_{eff}(s,\phi_{\max}) - U_{eff}(s,\phi_{\min}) \tag{3.9}$$

A zero temperature, the transition to a broken chain occurs when the energy barrier disappear. At this critical strain s_c, the energy barrier goes to an inflexion point, defined by $\phi_{\min} = \phi_{\max} \equiv \phi_c$ and the condition

$$\frac{\partial^2 U_{eff}(s_c,\phi_c)}{\partial^2 \phi} = 0. \tag{3.10}$$

At finite temperature, the transition to a broken chain takes place by thermal activation over the energy barrier. The time τ required for such transition is given by Kramer's reaction rate theory, Eq. 3.7, with the attempt frequency:

$$\Omega(s) = \frac{\omega_o(s)\omega_b(s)}{2\pi\gamma}, \tag{3.11}$$

where ω_o and ω_b are the vibration frequency around the position of local minima and maximum of energy. Eq. 3.11 is the overdamped approximation of Kramer's result for the attempt frequency [43].

If the load on the chain increases at a constant rate, the energy barrier decreases continuously. Due to the exponential character of the activation rate, breakup will occur for strains s close to the critical strain s_c, and this even at finite temperature. In the critical region $(s_c - s)/s_c \ll 1$, it is possible to obtain easily an accurate approximation for the effective potential, and hence the activation rate. We define $\tilde{s} = 1 - (s/s_c)$, and expand the effective potential $U(s, \phi)$ in deviations $\delta\phi = \phi - \phi_c$ and the small parameter $\tilde{s}$, from which the energy barrier can thereafter be calculated. Analytic expression for the attempt frequency on $\tilde{s}$ can also be computed by expanding the second derivative to first order on $\tilde{s}$. We now present the mean-field potential appropriates to both cases.

Attached chain: This case is reprensented in Fig. 3–1(b). The chain is subjected to the total strain s, which is carried by both the weak bond between atoms i^* and i^*+1 and the cantilever between atoms $i^* + 1$ and the first atom of the setup i_o. Without fluctuations, the total elongation of the two active bonds is s, of which a fraction f is taken by the AFM-chain bond. An approximate estimate of f is obtained by linearizing the Lennard-Jones potential, yielding $f = k_h a^2/72\epsilon$. The path of rupture is parametrised by the quantity ϕ, describing the extent by which the weak bond length deviates from $a + fs$. Equivalently, ϕ corresponds to the deviation of the spring from $a + (1 - f)s$, where $(1 - f)s$ is the strain of the spring in the harmonic approximation. The potential energy along ϕ reads:

$$U^h_{eff}(\phi) = \frac{k_h}{2}\Big[\Big(1 - f\Big)s - \phi\Big]^2 + U\Big(a + fs + \phi\Big). \tag{3.12}$$

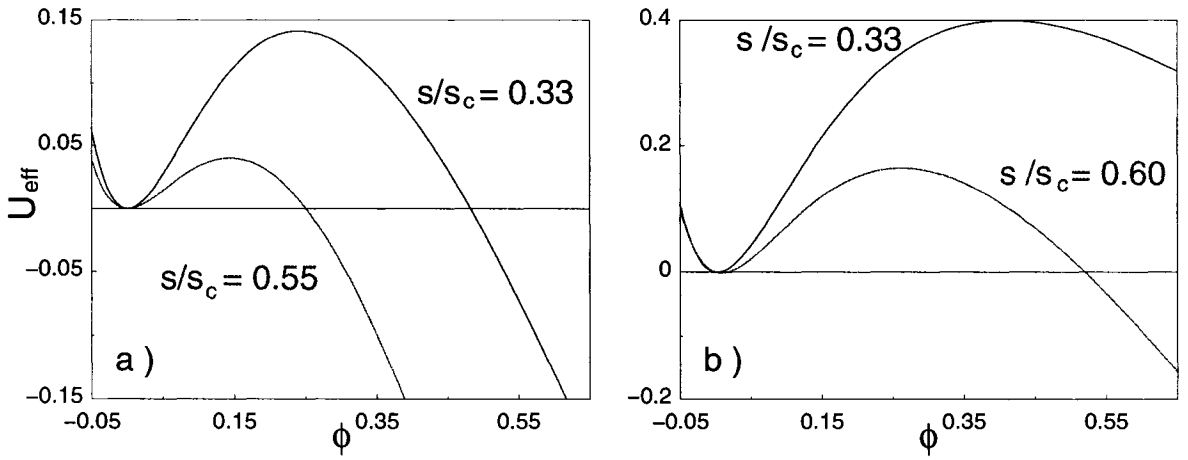

Figure 3–6: Dependence of the effective energy on ϕ. Black and red lines correspond to low and high tension respectively. a) Periodic chain. b) Chain composed of a weak + harmonic bond.

Figures 3–6(a)-(b) show how the effective potential for rupture changes when ϕ is varied under two different tensions. Low tension is shown in black while high tension is given in red.

Numerical resolutions of Eq. 3.8 and Eq. 3.10 show that the critical values $s_c/a = 373.7$ and $\phi_c/a = 0.071$. An expansion around these values then leads to an energy barrier

$$\Delta E^h(\tilde{s}) = E_0 \tilde{s}^{3/2} \tag{3.13a}$$

$$\Omega^h(\tilde{s}) = \Omega_0 \tilde{s}^{1/2}, \tag{3.13b}$$

where E_0 and Ω_0 are given in the footnote [1] . The value of s_c and the $\gamma = 3/2$ exponent for the the barrier height are in good agreement with the values obtained by numerical simulation.

Periodic chain: This case is characterised by the constraint of fixed chain length (see Figure 3–5) and the energetics of this system is described by the effective potential

$$U_{eff}^{pc}(\phi) = U(a+s+\phi) + (N-1)U\Big(a+s-\frac{\phi}{N-1}\Big). \tag{3.14}$$

In this case, $\phi_{\min} = 0$ is a minimum for all values of the strain. Since this minimum is fixed, it implies that $\phi_c = 0$, and the critical strain is simply related to the point at which the single-atom potential becomes unstable

$$\frac{d^2U(a+s_c)}{ds^2} = 0. \tag{3.15}$$

In the specific case of the Lennard-Jones potential, Eq. 3.2, this gives $s_c = a(13/7)^{1/6}$.

Due to the particular form of the effective energy of the periodic chain, $\partial^{1+n}U_{eff}^{pc}/\partial\phi\partial s^n = 0$ for all n. Therefore higher order Taylor's expansion of equation 3.8 has to be carried out and the resulting scaling for the periodic chain becomes:

$$\Delta E^{pc}(\tilde{s}) = E_0\tilde{s}^3 \tag{3.16a}$$

$$\Omega^{pc}(\tilde{s}) = \Omega_0\tilde{s}, \tag{3.16b}$$

[1] For the harmonic+weak link chain: $E_0 = \frac{4}{3}s_c^{3/2}[|2(\frac{\partial^2U_{eff}}{\partial s\partial\phi})/(\frac{\partial^3U_{eff}}{\partial\phi^3})|]^{1/2}(\frac{\partial^2U_{eff}}{\partial s\partial\phi})$, $\Omega_0 = s_c^{1/2}[|2(\frac{\partial^2U_{eff}}{\partial\phi\partial s})/(\frac{\partial^3U_{eff}}{\partial\phi^3})|]^{1/2}(\frac{\partial^3U_{eff}}{\partial\phi^3})/2\pi M\gamma$ and the derivatives are computed at s_c and ϕ_c.

where E_0 and Ω_0 can be simply expressed [2] . The exponent in Eq. 3.16a is similar to that estimated in the simulation (namely 3.54), and likewise for the value of s_c.

Crossover scaling: A crossover scaling function can account for both limits. Consider the parameter $\psi \equiv \partial^2 U_{eff}/\partial\tilde{s}\partial\phi$ which characterizes symmetry: $\psi = 0$ represents the periodic (symmetric) chain, while finite ψ corresponds to the assymetric (weak + harmonic) case. Expanding Eqs. 3.8 and 3.9 to second order in $\tilde{s}$ and $\tilde{\phi}$ gives the scaling form for the energy barrier in the limit of small $\tilde{s}$ and small ψ:

$$E_b(\psi, \tilde{s}) = \tilde{s}^3 g(\psi/\tilde{s}^3), \tag{3.17}$$

where the scaling function obeys $g(\psi^* \to 0) \propto 1$, and $g(\psi^* \to \infty) \propto (\psi^*)^{1/2}$, in terms of the scaled variable $\psi^* = \psi/\tilde{s}^3$. The proportionality constants and the form of f are dependent on the derivative of the energy computed at s_c and ϕ_c. This recovers the previously obtained symmetric and non-symmetric cases in the appropriate limit. Although Eq. 3.17 is more general than Eqs. 3.13 and 3.16, it is preferable to work with these two scalings separately (this is done in the next section).

Notice that by adding a different bond type to the period chain (symmetric case), $\partial^{1+n} U^{pc}_{eff}/\partial\phi\partial s^n \neq 0$ and we regain the scaling of Eq. 3.13. In force spectroscopy a different bond (usually harmonic) is required to probe the molecule. Therefore one can say that the scaling of the periodic chain made of equal bonds is not appropriate to describe these experiments.

[2] For the periodic chain: $E_0 = \frac{2}{3} s_c^3 (\frac{\partial^3 U_{eff}}{\partial s \partial \phi^2})^3 / (\frac{\partial^3 U_{eff}}{\partial \phi^3})^2$, $\Omega_0 = s_c (\frac{\partial^3 U}{\partial s \partial \phi^2})/2\pi M\gamma$ and the derivatives are computed at s_c and ϕ_c.

Finally, we note that Eq. 3.13 and 3.16 are independent on the type of potential used for the interaction between particles.

3.4 Dynamics of breaking

The situation of interest for force spectroscopy experiments is when the length of the system is increased at a constant rate r (*i.e.*, $s(t) = r\,t$). In this case, the chain ruptures at a relative strain s with some probability. Assuming that the breaking of the chain itself takes place on a time scale shorter than any other in the problem, the probability that the chain has not broken at time t is

$$W(s(t)) = \exp\Big(-\int_{t_0}^{t} \tau^{-1}[s(t')]dt'\Big), \tag{3.18}$$

where $\tau(s)$ is Kramer's rate (Eq. 3.7) for the *static* strain s. The probability distribution of breakup is simply $P(s) = -dW/ds$ and can be used to obtain the average strain $\langle s\rangle$ at which breakup occurs. This average strain is given [38] by

$$\langle \tilde{s}\rangle = \Big(\frac{kT}{E_0}\Big)^{1/3} \ln^{1/3}\left[\frac{1}{3}\left(\frac{kT^{2/3}}{E_0}\right)\frac{s_c\Omega_0}{r}\right], \tag{3.19}$$

for the periodic chain. E_0 and Ω_0 are defined in Eq. 3.16. For the chain attached to an harmonic spring and containing a single weak bond, it is more relevant to express our results in terms of the force F, related to the stain s by

$$F(s) = k_h(1 - k_h a^2/72\epsilon)s. \tag{3.20}$$

This is a reasonable approximation for the force since the spring is much softer than the non-linear bond. If we define $\tilde{f} = 1 - F/F_c$, the probability distribution of

rupture force reads:

$$P(\tilde{f}) = \Omega^* \tilde{f}^{1/2} exp\Big[- E_0^* \tilde{f}^{3/2} - \frac{2}{3}\frac{\Omega^* F_c}{E_0^*} e^{-E_0^* \tilde{f}^{3/2}}\Big], \tag{3.21}$$

where $F_c \equiv F(s_c)$, $\Omega^* \equiv \Omega_0/r$, $r \equiv dF/dt$ and $E_0^* \equiv E_0/kT$. From this distribution, average breakup is given by [38]:

$$\langle \tilde{f} \rangle = \left(\frac{kT}{E_0}\right)^{2/3} \ln^{2/3}\left[\frac{2}{3}\left(\frac{kT}{E_0}\right)\frac{F_c \Omega_0}{r}\right], \tag{3.22}$$

where E_0 and Ω_0 are defined in Eq. 3.13.

To derive Eqs. 3.19 and 3.22, quasi-static equilibrium was assumed: the mean time of rupture was assumed to obey Kramer's equation at any time. This is justified [96] since experimentally a significant increase in the length of the chain occurs at a macroscopic time-scale (determined by experiments and ranging from milliseconds to minutes) which is much larger than the correlation time of molecules in a liquid ($\sim 10^{-9}$ second). Therefore, at any elongation of the chain, the molecule vibrates several times ensuring equilibrium.

Now we present additional simulations relating the scaling given in Eq. 3.22 to a typical BFP setup. Here, the length of the chain composed of weak + harmonic bond is increased at a constant velocity v: the last atom of the chain is kept fixed while the position of the atom representing the tip of the AFM (see Figure 3–1b) is given by $x_0(t) = x_0(0) + vt$. The dynamics is determined by Newton's law until the chain ruptures. A typical simulational result is presented in Fig. 3–7(a). Here the force on the spring is shown while the molecule is being stretched at $v = 2.7 \times 10^{-3}$ (in units of a/τ_o) and and $T = 0.02$ (in units of ϵ). The force increases until it drops to zero,

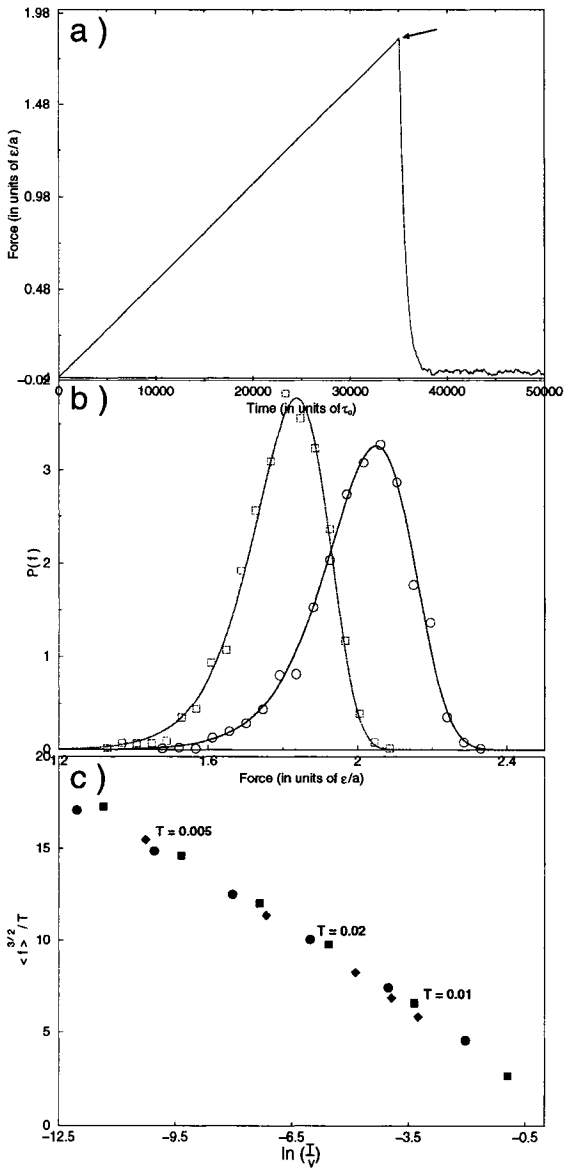

Figure 3–7: a) Typical force on the spring while the molecule is being stretched at $v = 2.7 \times 10^{-3}$ (in units of a/τ_o) and $T = 0.02$ (in units of ϵ). The arrow indicates rupture. b) Normalized probability distribution of breaking force. Simulational results are indicated by symbols. These simulations were performed at $T = 0.02$ and $v = 0.01$ (squares, red), $v = 0.08$ (circles, blue). Curves correspond to analytical result. c) Validation of the scaling form, Eq. 3.22.

indicating rupture. The cusp occurring right at rupture is a typical experimental results of AFM pulling experiments [46] or other stick-slip phenomena [90, 28]. Note that it is not due to any kind of singularity in the inter-atomic potentials but only to the fact that rupture itself is a very fast process compared to the elongation of the chain.

The distribution of breakup force is illustrated in Figure 3–7(b) for an ensemble of 1900 chains. These simulations were carried out at $T = 0.02$ and two different velocities $v = 0.01$ (squares, shown in red) and $v = 0.08$ (circles, blue). Lines in this Figure correspond to the analytic result (Eq. 3.21) and symbols are results obtained from simulation. A good match between simulations and analytical results is observed. We see that the mean force $\langle \tilde{f} \rangle$ of rupture depends on v.

To illustrate this dependence, we perform some simulations using ensembles of 1000 chains at three temperatures and various velocities. The mean force $\langle \tilde{f} \rangle$ of rupture is extracted from these simulations. We note that Eq 3.22 provides a temperature independent scaling by plotting $\langle \tilde{f} \rangle^{3/2}/T$ in terms of $\ln(T/v)$. Within this choice of axes, data computed at different temperatures collapses into a single line — as shown in Figure 3–7(c). The quality of the collapse in this Figure for the wide range of temperatures and velocities used in the simulation validates the proposed scaling.

In the simulation performed at the slowest stretching rate (the point at the left side of Figure 3–7b), each atom of the chain executed on average $\sim 3.3 \times 10^7$ oscillations around its equilibrium position before the chain ruptured. If we consider that each vibration is executed in 10^{-9} seconds (the vibration period of a molecule in

a liquid), than this simulation lasted approximately 33 milliseconds, which is within the time range of experiments. At the other end of the spectrum, only 640 vibrations (or 0.64 microseconds) were executed before rupture. While this is outside the range of experiments, it is interesting to note that quasi-static approximation remains valid at such a high loading rate.

3.5 Discussion and Conclusion

The typical bond used for force spectroscopy experiments is the specific bond formed between a biotin and a streptavidin molecule [67]. These experiments are usually performed at room temperature ($k_BT = 4.1\ pN.nm^{-1}$) and the spring constant used to characterize the membrane of a BFP lies in the range $0.1 - 3\ pN.nm^{-1}$. The binding energy of the specific bond is $\epsilon \sim 50 k_BT = 205\ pN.nm$ and its bond length is assumed to be of the order of nanometer ($a \sim 1\ nm$). If this bond were Lennard-Jones like, its stiffness would be $72\epsilon/a^2 = 14760\ pN.nm^{-1}$. In this manner, the stiffness of the membrane is 10000 times smaller than the stiffness of the specific bond. This set of parameters correspond to the blue circles in Figure 3–7 - which falls in the ramped creep universality class discussed in this paper. Dudko *et al.* have also proposed the same scaling for force spectroscopy [27]. In their work, they use Morse potential to perform simulations and validate their scaling (which is equivalent to our Eq. 3.22). This provides additional support for our universality argument: the proposed scaling is independent of bond type.

In summary we modeled the rupture of a specific bond which is being stretched at a constant rate r and temperature T. We showed that if there is just one type of bond being stretched, the strain at which the molecule ruptures scales as: $s \sim$

$T^{1/3}[\ln(T^{1/3}/r)]^{1/3}$; when at least two different bonds are stretched, the molecule ruptures according to: $f \sim const - T^{2/3}[\ln(r/T)]^{2/3}$. Since in force spectroscopy experiments there are at least two types of bonds being stretched (bonds of the molecule and the spring of the BFP), the last scaling should be used to describe these experiments and to extract the relevant parameters from them. In order to test our results experimentally, it would be valuable to probe rupture in an extended range of temperatures.

Chapter 4

Designable structures are easy to unfold

4.1 Introduction

While the number of different proteins exceeds 10^5, when classified in terms of structures, only of order of 10^3 families of protein folds exist [11, 71]. These structural templates for amino-acid sequences can be understood [57, 59, 69] in terms of minimal microscopic models. In these models, the positions of amino acids are restricted to lattice sites, and interaction energies between residues are described by a coarse-grained model. Emergent structures are classified by their designability, the number of different amino acid sequences that design the same structure – see Fig. 4–1. A few structures are highly designable, and correspond to an enormous number of sequences. These are thereby stable to amino acid mutation, a desirable and natural feature for evolution. As well, highly designable structures are thermodynamically stable [57, 105], and have protein-like symmetry [57, 59, 102].

In this chapter, we investigate the dynamical behavior of designable structures [23, 20]. Some calculations suggest [66, 12] that sequences of amino acids which are

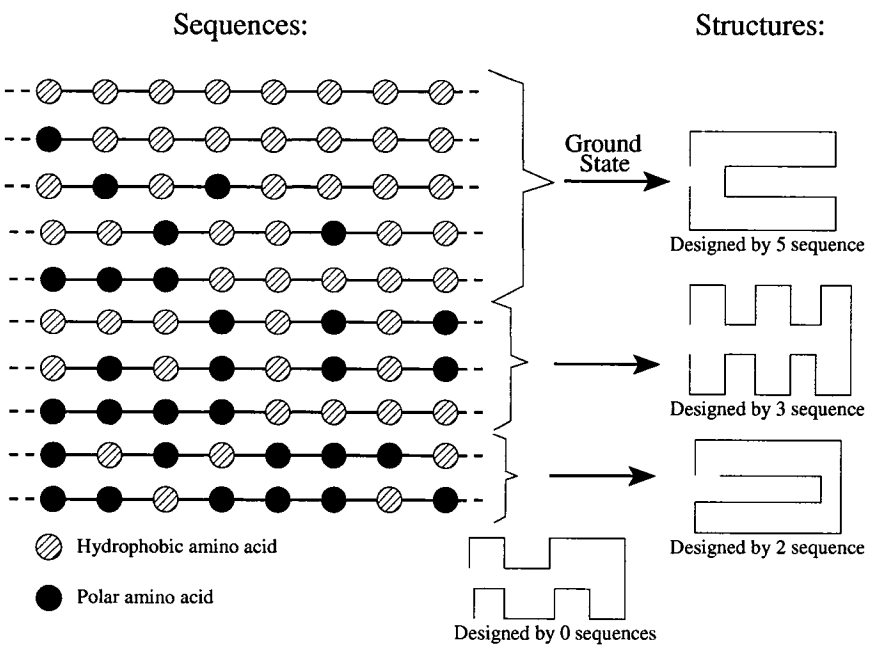

Figure 4–1: Schematic representation of designability. Sequences that have the same structure as ground state are inside the same bracket. A structure that is not designed by sequences is also shown.

thermodynamically stable and whose ground state are highly designable, fold faster than random sequences. Another important aspect of proteins is their reaction to forces [16]. We study proteins under shear and find a dependence of their phase diagram on designability. This diagram reveals that highly designable structures are weaker: they are easier to unfold than low designable structures. This result is a consequence of how strong covalent bonds in the backbone and weak bonds are distributed in designable structures. We expect this to be related to specific function, and in particular to protein flexibility.

4.2 Model

Designable structures are qualitatively different from other structures [59]. Topologically, a large number of α helices and a lack of β sheet secondary structures [17], seems to account for the peculiar geometry of highly designable structures. Since the type of secondary structures determines how the backbone connects surface monomers and this affects the dynamics of unfolding, one would expect that highly designable structures respond differently to forces than other structures. To investigate this, we consider a hydrophobic-polar (HP) model where a protein is a chain made up of polar (P) and hydrophobic (H) amino acids. The model incorporates hydrophobicity, the main driving force for folding [51, 58]. The energy of a structural sequence is given by the short-range contact interaction:

$$\mathcal{H} = \sum_{i<j} \epsilon_{i,j} \Big[\delta(|\vec{r}_i - \vec{r}_j| - \sigma) - \delta_{j-1,i} \Big] \tag{4.1}$$

where the N monomers located at spatial positions $\vec{r}_i$ are labeled by indices i and j on a two-dimensional triangular lattice, as described below. The first delta function

allows only nearest-neighbors interactions at a distance σ, while the second excludes interactions between residues which are adjacent along the backbone. The interaction energy $\epsilon_{i,j}$ between monomers i and j can have three values depending on the type of monomers being binded: H-H, H-P, or P-P. These values are chosen to minimize energy when H-like (P-like) amino acids are within (on the surface of) the protein, namely, $\epsilon_{PP} > \epsilon_{HP} > \epsilon_{HH}$. To account for the segregation of different types of amino acids an additional condition is imposed: $2\epsilon_{HP} > \epsilon_{PP} + \epsilon_{HH}$. Since compact shapes have maximum contact and the lowest energy states, they are the only shapes considered for representing proteins. With this simplification, the interaction energies can be shifted without changing the relative energies of a sequence when folded into different conformations. Following Li *et al* [57], we use: $\epsilon_{HH} = -2.3$, $\epsilon_{HP} = -1$ and $\epsilon_{PP} = 0$. For studying unfolding, the triangular "lattice" is created by assigning an energy for each structure through two potentials: adjacent monomers along the backbone protein interact through harmonic potentials, others by a Lennard-Jones potential:

$$V(r_{i,j}) = \sum_{i=1}^{N-1} \frac{k}{2}\left(r_{i,i+1} - \sigma\right)^2 + \frac{1}{2}\sum_{\substack{j\neq i\pm 1 \\ j\neq i}} \epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{4.2}$$

where $r_{i,j} \equiv |\vec{r}_i - \vec{r}_j|$. The harmonic bonds, with spring constant k and equilibrium length σ, ensure that the backbone of the protein is preserved during the simulation. The monomers are bound by the Lennard-Jones potential, characterized by energy ϵ and the same equilibrium length σ. These Lennard-Jones bonds can be driven apart, changing the structure of the protein. A cut-off distance of 2.5σ is used. Since the equilibrium length of harmonic and Lennard-Jones potentials are the same, the

minimal energy structure of the model in two dimensions is a triangular lattice, up to small corrections due to surface effects. Conveniently, then, the equilibrium states can be studied by simply assuming all monomers sit on the positions of a triangular lattice.

To study dynamics, we use a Langevin approach where friction and a random force act on each monomer. The intensity of the random force is given by a fluctuation-dissipation theorem. The friction force on each monomer is proportional to the relative velocity of the monomer with respect to a prescribed velocity field, which can apply a shear [5]. If the ith monomer is located at $\vec{r_i} = x_i\ \hat{x} + y_i\ \hat{y}$, the prescribed velocity is $\vec{v}_{fluid}(\vec{r_i}) = Sy_i\ \hat{x}$, where S is the shear rate. The equation of motion inside the shear flow is

$$M\frac{d^2\vec{r_i}}{dt^2} = \sum_j \vec{F}(r_{ij}) - M\gamma\Big[\frac{d\vec{r_i}}{dt} - \vec{v}_{fluid}(r_i)\Big] + \vec{f_i}(t), \tag{4.3}$$

where the sum is over all atoms inside the cut-off. Here, M is the mass of a monomer, and $\vec{F}$ is the force computed from the interacting potential. For simplicity, σ, ϵ and M are chosen to be unity. The spring is chosen to be five times stiffer than the Lennard-Jones potential: $k = 5(72\epsilon/\sigma^2)$. Simulations are carried out in units of the fastest atomic vibration time $\tau_0 = 2\pi\sqrt{k/M}$, and the friction constant is given a value $\gamma = (\tau_0/4)^{-1}$.

First we consider and review equilibrium structure. We study chains of 25 amino acids. Possible structures are restricted to compact self-avoiding walks on a 5×5 triangular lattice, implying 352,375 independent structures. The ground state of all the 2^{25} sequences is computed and we count the number of sequences that fold

uniquely into a structure. This number corresponds to the designability of a given structure. We find that 135,216 ($\sim$ 38%) are non-degenerate ground states of at least one sequence.

4.3 Results

The distribution of designability for those 135,216 structures is given in Fig. 4–2(a). A small number of highly designable folds accommodate more than 500 sequences. These are thermodynamically stable [57]. The stability of a structure can be quantified as the difference between the energy of its ground state and first excited state, E_{gap}. In Fig. 4–2(b) we illustrate the relation between stability of a structure and designability. In this figure, E_{gap} is averaged over a given range of designabilities and plotted versus designability. Highly designable structures are seen to be more stable thermodynamically than other structures. Therefore, one can conclude that those rare structures which are highly designable, and thus stable against mutation, are also thermodynamically stable. These structures have a large number of bonds connecting surface monomers to core monomers [102, 45, 92, 93]. This is shown in Fig. 4–2(c), where the number of bonds connecting surface to core, averaged over structures of a given range of designability, is plotted against designability. A systematic increase of surface-to-core bonds with designability is observed. A particular example of such a highly designable structure is shown in Fig. 4–2(d).

Now we will quantitatively evaluate how structures with differing designabilities react to an applied shear and thermal fluctuations. Rather than simulate all 135,216 structures, we sample as follows. We study all the 1,500 structures with highest

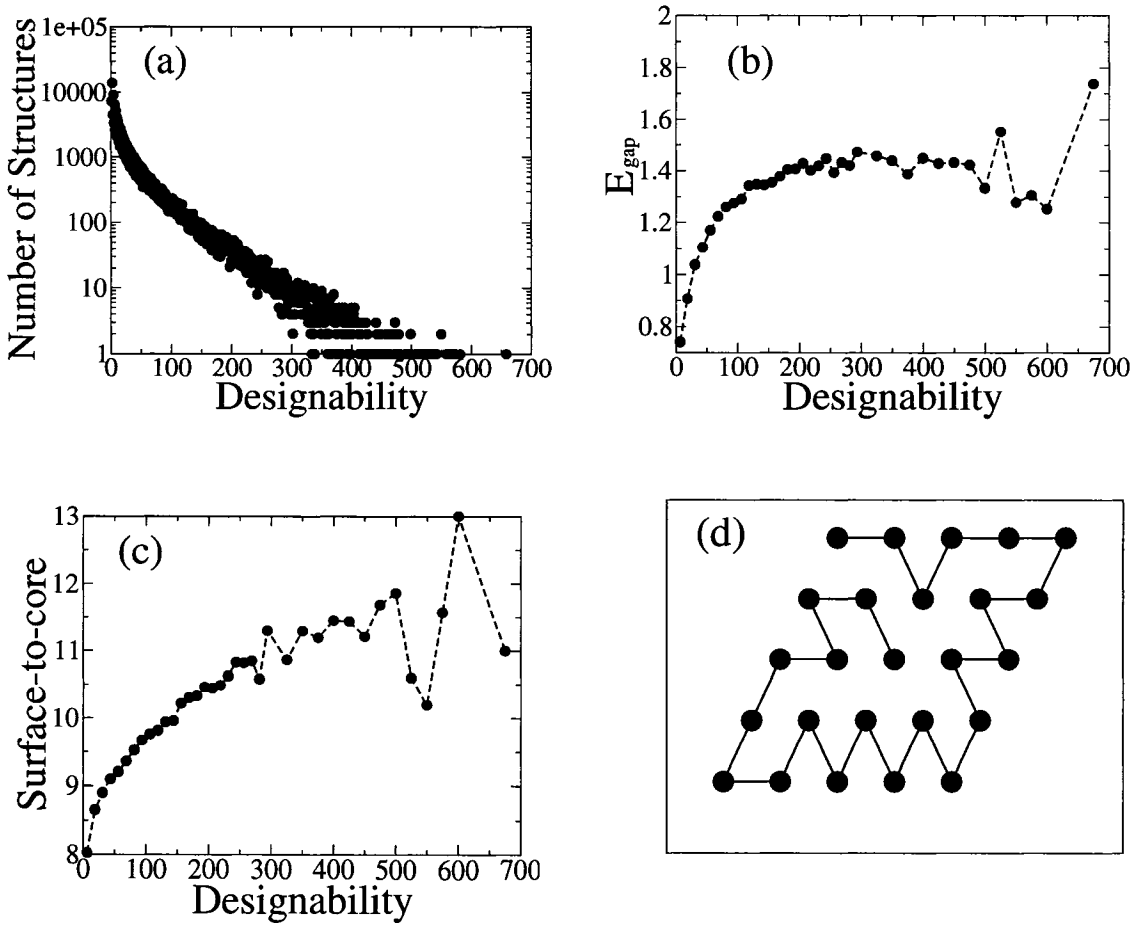

Figure 4–2: a) Histogram of designability. b) Energy gap, averaged over structures of a given range of designability, versus designability. c) Number of bonds connecting surface to core residues versus designability. d) Fifth most designable structure.

designability, ranging from 200 to 700. For the more numerous structures which are less designable, we consider eight randomly-chosen structures for each designability. This ensemble of 3,100 structures is representative of the diversity of folds.

At zero temperature a structure only unfolds if the shear rate is greater than a critical value S_c. This critical value is a measure of structural stability to an applied force: the larger S_c is, the more stable the structure. To determine the relation between S_c and designability, we probe each structure at varying shears and different simulation times. A structure is considered to be unfolded whenever 5 or more bonds have broken. The quantitative dependence of S_c on designability is illustrated in Fig. 4–3. In the Figure, the ensemble of 3,100 structures was divided into 12 bins, each containing structures with the same number (4 to 15) of surface-to-core bonds. The average designability and the average S_c of each bin is plotted in the figure. A clear correlation between designability and S_c exists [1] . Structures which are highly designable are easier to unfold by a shear force — that is, more unstable to a shear force — than low designable structures.

The other extreme condition for unfolding is zero shear and high temperatures. In this case, thermal fluctuations are the mechanism responsible for unfolding. We

[1] The critical shear rate of high and low designable structures differs by up to 10% in Fig. 3. This is a large effect because thermal unfolding of structures depend exponentially upon those shear rates. When the rate is smaller than S_c, a structure can still unfold via thermal fluctuations: the time for unfolding $\tau(S) \sim \exp\left(E_b(S)/k_BT\right)$, where E_b is the energy barrier and k_B is Boltzmann's constant. Close to the critical flow, the energy landscape is dominated by the fixed point, which is a point of inflection. To lowest order in $(1 - S/S_c)$ the energy barrier is $E_b(S) \propto (1 - S/S_c)^{3/2}$.

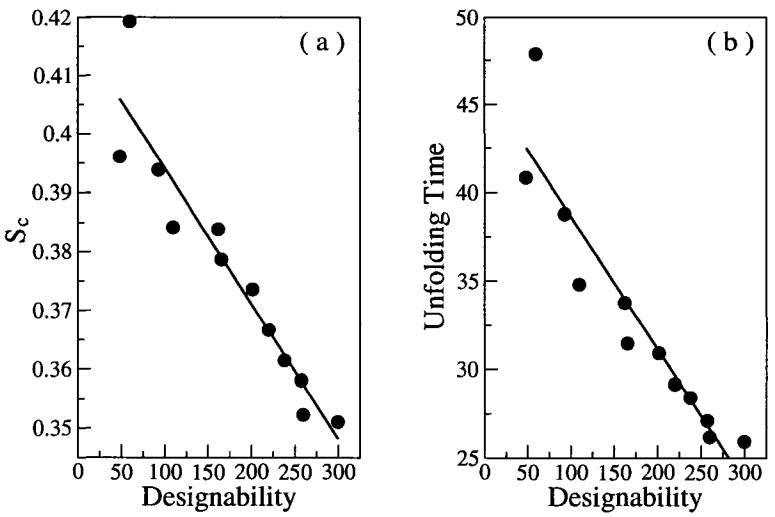

Figure 4–3: (a) Dependence of S_c on designability. (b) Time required to unfold designable structures at zero shear and T=0.5 (in units of ϵ). Lines are just a guide to the eye.

study how the time required to unfold a structure depends on its designability at a temperature of 0.50 (in units of ϵ). In our simulations, the unfolding time τ is computed by tracking the population of folded chains. The number of chains that unfold at time t (dN/dt) is proportional to the population of folded chains $N(t)$. In this case, $N(t) = N_0 exp(-Rt)$ where R is the rate of unfolding and the characteristic unfolding time is given by the inverse of the rate $\tau = 1/R$. We use 1,000 copies (*i.e.* $N_0 = 1,000$) of each structure in the simulation. The larger the unfolding time of a structure, the more stable it is to thermal fluctuations. Results are shown in Fig. 4–3(b). Again, each point in this Figure corresponds to the ensemble of structures having the same number of surface-to-core bonds. A clear downward trend shows

that highly designable structures are less robust to thermal fluctuation: they unfold faster.

To investigate the dependence of highly designable structures on simultaneous applied shear and thermal fluctuations, the phase-diagram was estimated. This diagram is constructed by computing the applied shear rate required to unfold a structure in $5,000$ units of time at different temperatures. This shear rate is then averaged over structures having the same number of surface-to-core bonds. Notice that the computed shear delimits two regions of the diagram: folded structures are found below this shear and unfolded structures above it. In Fig. 4–4 the phase-diagram is shown for structures having 4 and 15 surface-to-core bonds. These two sets of structures have an average designability of 60 and 300 respectively. At any temperature, the set of structures with lower designability is more robust and require a higher shear rate to unfold. *One can therefore state that high designable structures are easier to unfold than low designable ones.*

4.4 Discussion

It is constructive at this point to visualize the protein while it is unfolding – see Fig. 4–5. The upper (lower) panels of this Figure correspond to the unfolding of a low (highly) designable protein fold. These simulations were performed at a temperature of 0.70 (in units of ϵ) and zero shear. Low designable folds have few surface to core bonds. As a result, many weak bonds are aligned forming sub-structures where monomers are correlated over long distances. For those folds, the time of unfolding is dominated by the slow unbinding of the largest sub-structures. In contrast, high designable folds are formed by many small sub-structures which are approximately of

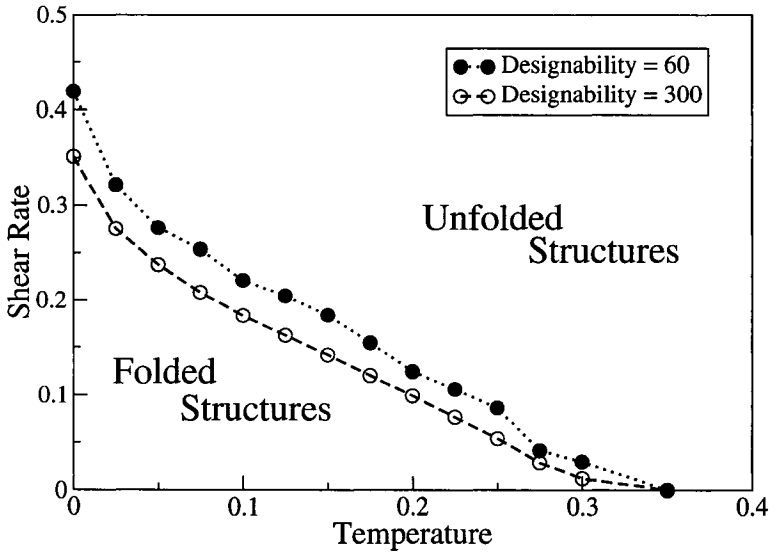

Figure 4–4: Phase diagram of low and highly designable structures – filled and open circles, respectively.

the same size. Hence it is easy to separate these sub-structures: only a few bonds need to rupture. This is illustrated in Fig. 4–5: for the low designable folds, the largest sub-structures is still preserved after 30 τ_o (panel c) while for the high designable folds all the small sub-structures have been destroyed (panel f).

This is in marked contrast to the relationship of designability to thermodynamic stability, namely that highly designable structures are *more* stable than low designable structures. The implication is that, although highly designable structures are more stable in the folded region of the phase diagram, they require less force/perturbation to unfold. We speculate this to be related to protein flexibility [62]: many globular proteins are stable to thermal fluctuations but undergo conformational changes (and are said to be flexible) when performing their functions.

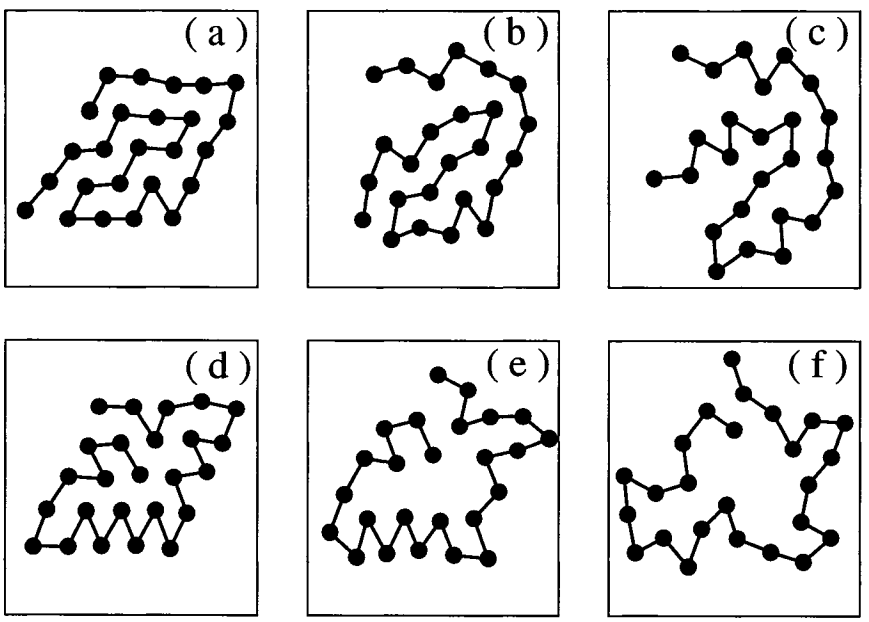

Figure 4–5: Snapshot of a low (panels *a*,*b* and *c*) and a highly (panels *d*,*e* and *f*) designable structure during thermally induced unfolding ($S = 0$ and $T = 0.70$). Panels *a* and *d* show beads position at time τ_o and a time interval of 15 τ_o has elapsed between each panel.

The phenomenology of this is as follows. Highly designable structures are weaker due to the large number of surface-to-core bonds they contain: as a result of this feature, protein folds contain many small sub-structures. These are easy to unfold since only a few bonds need to rupture in order to separate the sub-structures. Also, the presence in large number of surface-to-core bonds makes it difficult to transform highly designable structures into other distinct compact shapes through local rearrangements of the backbone [59]. Such a transformation would require the partial unfolding of the structure, which is unlikely in the region of the phase diagram where folded structures are at equilibrium, followed by folding into the new shape. Therefore, the presence of surface-to-core bonds might explain why high designable structures are thermodynamically stable but easier to unfolding. Finally, we expect interesting insights to be obtained by expanding the model to three dimensions and including hydrodynamics effects (*i.e.* modeling the solvent explicitly).

Chapter 5

Microscopic mechanism for cold denaturation

5.1 Introduction

Under physiological conditions, proteins adopt a unique three-dimensional (3D) structure [4]. It is maximally stable at about 17°C and becomes unstable thus denaturing the protein at both high ($\sim$ 60°C) and low (~ -20°C) temperatures [78, 79, 55]. The latter phenomenon is called *cold denaturation*, where the protein unfolds and thereby increases its entropy, which in turn is accompanied by a decrease in the entropy of the entire system. This counter-intuitive behavior has been experimentally verified [79, 82] but has remained a subject of controversy [78, 55], since a satisfactory microscopic explanation for this phenomenon has not yet emerged. Resolving cold denaturation microscopically would facilitate understanding the forces responsible for the structure of proteins, and in particular the role of the complex hydrophobic effect

In the case of diluted proteins, hydrophobicity is considered as the main driving force for folding and unfolding [24]. Consequently, different classes of models

describing hydrophobicity through varying explicit models have been used to study cold denaturation [14, 75, 86, 18]. One such class [14, 86] considers the different energetic states of shell water, i.e. water molecules neighboring the protein, in a lattice. A more realistic water model [75] supports this view, as water-water hydrogen bonding among shell water has been found to increase at low temperatures and to correlate with cold denaturation. Meanwhile, another class of models suggests that the density fluctuations of water are responsible for cold denaturation [64, 15]. Despite the lack of consensus in the explanation of cold denaturation, the solvent is widely accepted as the key player. This is also supported by the fact that denaturation also takes place under pressure [47, 65]. By focusing on the transfer of water molecules to the protein interior, pressure denaturation has been explained through the destabilization of hydrophobic contacts in favor of solvent separated configurations [47]. This destabilization has been verified using different water models [39].

In the present chapter, we examine the microscopic physical mechanism behind cold denaturation [21]. To this end, we consider the two-dimensional Mercedes-Benz (MB) model to describe water molecules in the solvent and a simple bead-spring model for the protein. The MB model reproduces many of the properties of water [95], including the temperature dependent behavior of the hydrophobic effect [25]. Our molecular dynamics (MD) simulations of the MB model provide a simple microscopic picture for cold denaturation in terms of changes in hydration: at low temperatures water molecules infiltrate the folded protein in order to passivate the "dangling" water-water hydrogen bonds (H-bonds) found in shell water. At the same time, hydrophobic contacts are destabilized and an ordered layer of water molecules

forms around the protein monomers such that they become separated by a layer of solvent in the cold denatured state. Hence, increasing pressure and decreasing temperature destabilize hydrophobic contacts in favor of similar solvent separated configurations. We expect that this aggravated destabilization of hydrophobic contacts at high pressure explains why the transition temperature for cold denaturation increases with increasing pressure [55]. Here, we study cold denaturation at the equivalent of ambient pressure.

5.2 Model

As in water, the interaction between the MB molecules is given by a sum of hydrogen bonds and van der Waals bonds. The directionality of H-bonds is accounted for by three arms separated by an angle of 120°. This interaction has maximal strength when arms of neighboring molecules are aligned. If $\vec{r}_{ij}$ is the distance vector between the center of mass of molecules i and j, and $\vec{r}_{i\alpha}$ is the distance vector between the center of molecule i and the extremity of arm α, then the interaction energy is given by:

$$V_h(\vec{r}_{ij}, \{\vec{r}_{i\alpha}\}, \{\vec{r}_{j\beta}\}) = \epsilon_h \exp\left(-\frac{(r_{ij} - R_h)^2}{2\sigma_R^2}\right) \times \left[\sum_{\alpha=1}^{3} \exp\left(-\left(\frac{\vec{r}_{i\alpha}\cdot\vec{r}_{ij}}{r_{ij}r_{i\alpha}} - 1\right)^2 \frac{1}{2\sigma_\theta^2}\right)\right] \times \left[\sum_{\beta=1}^{3} \exp\left(-\left(\frac{\vec{r}_{j\beta}\cdot\vec{r}_{ij}}{r_{ij}r_{j\beta}} - 1\right)^2 \frac{1}{2\sigma_\theta^2}\right)\right], \quad (5.1)$$

where ϵ_h and R_h are the binding energy and the equilibrium (reference) length of the bond, respectively. The constants σ_R and σ_θ are attenuation parameters of the

interaction. Equation (1) favors configurations where the distance between molecules i and j is R_h, one arm of molecule i is aligned with the line joining the two centers of mass, and the same for one arm of molecule j. The van der Waals interaction is described by a Lennard–Jones (LJ) potential V_{ww} with binding energy ϵ_{ww} and equilibrium length R_{ww}:

$$V_{ww}(r_{ij}) = 4\epsilon_{ww}\left[\left(\frac{R_{ww}}{r_{ij}}\right)^{12} - \left(\frac{R_{ww}}{r_{ij}}\right)^{6}\right]. \tag{5.2}$$

The LJ potentials are shifted so that the force becomes zero at the cut-off distance $R_c = 2.5R_h$ [2]. We use the parameter set which has been studied extensively by Silverstein *et al.* [95]: $\epsilon_h = 1.0$, $R_h = 1.0$, $\sigma = \sigma_R = \sigma_\theta = 0.085$, $\epsilon_{ww} = 0.1$, and $R_{ww} = 0.7$. The total interaction energy $V_{i,j}$ between two water molecules is given by the sum of Eqs. 5.1 and 5.2.

Here, we set $M_w = 1$ for water. To mimic the distribution of mass in water, 1/10 of the total mass of a water molecule is located at each arm's extremity and the extremity of an arm is located at a distance $R_{\text{arm}} = 0.36R_h$ from the center of mass [81]. This defines the angular momentum of the water molecule.

The force on a MB molecule can now be computed: on the extremity of the arm the force is given by $\vec{F}_{i\alpha} = \left(\partial V/\partial \vec{r}_{i\alpha}\right)$, and on the center of mass by $\vec{F}_i =$

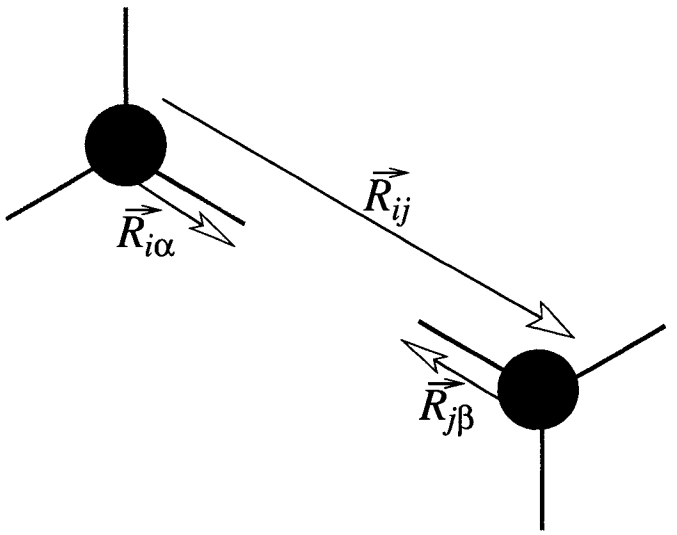

Figure 5–1: Schematic representation of two MB molecules.

$(\partial V/\partial \vec{r}_i) + \sum_\alpha \vec{F}_{i\alpha}$. Explicitly, these forces read:

$$\begin{aligned}
\vec{F}_i &= -\epsilon_H \Big[\frac{1}{\sigma^2}(r_{ij} - R_h) G_r \sum_{\alpha\beta} G_\alpha G_\beta \hat{r}_{ij} + \\
&\frac{1}{\sigma^2} G_R \sum_{\alpha,\beta} (cos(\theta_\alpha) - 1) G_\alpha G_\beta \left[-\frac{\hat{r}_{i\alpha}}{r_{ij}} - \left(\frac{\vec{r}_{i\alpha} \cdot \vec{r}_{ji}}{\vec{r}_{ij} \cdot \vec{r}_{ij}} \right) \frac{\hat{r}_{ij}}{R_{arm}} \right] + \\
&\frac{1}{\sigma^2} G_R \sum_{\alpha,\beta} (cos(\theta_\beta) - 1) G_\alpha G_\beta \left[-\frac{\hat{r}_{i\beta}}{r_{ij}} - \left(\frac{\vec{r}_{i\beta} \cdot \vec{r}_{ij}}{\vec{r}_{ij} \cdot \vec{r}_{ij}} \right) \frac{\hat{r}_{ij}}{R_{arm}} \right] \Big] + \\
&\frac{\partial V_{ww}}{\partial \vec{r}_i},
\end{aligned}$$

and,

$$\vec{F}_{i\alpha} = -\epsilon_H \left[\frac{1}{\sigma^2} (cos(\theta_\alpha) - 1) G_R G_\alpha \left(-\frac{\hat{r}_{ij}}{R_{arm}} - \left(\frac{\vec{r}_{i\alpha} \cdot \vec{r}_{ji}}{\vec{r}_{i\alpha} \cdot \vec{r}_{i\alpha}} \right) \frac{\hat{r}_{i\alpha}}{R_{ij}} \right) \sum_\beta G_\beta \right],$$

where $G_R \equiv \exp\left(-\frac{(r_{ij}-R_h)^2}{2\sigma^2}\right)$, $G_\alpha \equiv \exp\left(-\left(\frac{\vec{r}_{i\alpha}\cdot\vec{r}_{ij}}{r_{ij}r_{i\alpha}} - 1\right)^2 \frac{1}{2\sigma^2}\right)$ and a similarly definition applies for G_β. Vectors used in those equations are represented in Fig. 5–1.

Energies, distances, and time are given in units of ϵ_h, R_h, and $\tau_o = \sqrt{\epsilon_{ww}/M_w R_{ww}^2}$, respectively. To model the protein, we use a bead-spring model: monomers which are adjacent along the backbone of the protein are connected to each other by springs, and non-adjacent monomers are connected by a shifted LJ potential. The LJ potential is described by a binding energy $\epsilon_{mm} = 0.375$ and distance R_{mm}. The equilibrium length and stiffness of the spring are R_{spring} and $K_{\text{spring}} = 2(456\epsilon_{mm}/R_{mm}^2)$. This corresponds to twice the stiffness of the LJ potential. Monomers are set to be ten times heavier than water molecules. The interaction between monomers and water molecules is given by a shifted LJ potential with binding energy $\epsilon_{wm} = \epsilon_{ww}$ and equilibrium length R_{wm}.

When the side-chain of a hydrophobic amino acid is exposed to the solvent, the liquid surrounding the side-chain assumes a cage-like configuration [10] in order to minimize the amount of broken H-bonds of water molecules. This configuration has a low entropy and proteins minimize their free-energy by burying these hydrophobic amino-acids in their interior. To reproduce this, we choose $R_{\text{spring}} = 2.0$ and $R_{wm} = 0.9$ such that monomers can be surrounded by a layer of water molecules when exposed to the solvent. To allow for the formation of a dry protein core, we use $R_{mm}R_{wm}$, though $2R_{mm} > R_{\text{spring}}$ to avoid the backbone from intersecting itself. Taking these restrictions into account, we choose $R_{mm} = 1.1$.

Having defined the interaction between the different particles we now perform MD in the isothermal-isobaric ensemble. Constant pressure is achieved using the Andersen extended method [3]. To suppress oscillations of the simulation box, the canonical equations of motion are replaced by a Langevin stochastic process [34]

implemented using the simplectic algorithm [54]. For the mass Q and the friction constant γ_V of the piston acting on the simulation box, we use $Q = 0.054/R_{ww}^4$ and $\gamma_V = 0.5$. A parallelogram with equal sides and defined by an angle of 120° is used for the simulation box. This geometry retains the periodicity of a crystal made of water molecules through the boundaries. Periodic boundary conditions are implemented using the minimum image convention. For the Langevin equations describing the motion of particles, we use the friction constant $\gamma^{-1} = 0.93\tau_o$. The noise term in the Langevin equations of motion is given by the fluctuation-dissipation theorem. Pressure is set to 0.2 in units of ϵ_h/R_h^2. At this pressure, the MB model reproduces water-like anomalies seen at ambient pressure [25] and hydrates non-polar molecules in a realistic manner [97]. The simulation box is packed with 512 molecules comprised of a 10-monomer long protein and 502 water molecules. To represent the solvent in its liquid state, we use temperatures ranging from 0.145 to 0.25 in units of ϵ_H.

The system was initially equilibrated at a temperature of 0.25 for 5000 time steps, followed by a data collection period of 50000 time steps. The temperature was then lowered and the equilibration-collection cycle was repeated. This cooling procedure was repeated until the lowest temperature was reached. Four samples with different initial conditions were prepared using this protocol and the distribution of the protein's radius of gyration R_G was computed. To obtain equilibrium properties, the final configuration at each temperature was used to extend the simulation time until the distribution of R_G of the four samples converged within a root-mean-square value of 0.02.

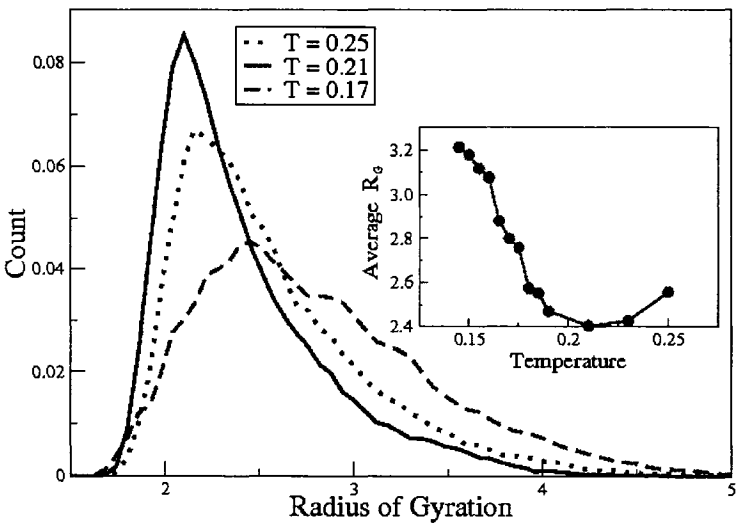

Figure 5–2: Normalized distribution of the radius of gyration R_G at three temperatures: $T = 0.25$, $T = 0.21$ and $T = 0.17$. Inset: The temperature dependence of R_G of the protein.

5.3 Results

Let us now move on to describe the results. In Fig. 5–2, we show the equilibrium distribution of R_G averaged over the four samples at three different temperatures. An initial decrease in temperature, from 0.25 to 0.21, shifts the peak of the distribution to a lower value. Therefore, in *hot water*, proteins favor more compact configurations when the temperature of the system is lowered. However, a further decrease of temperature results in completely different behavior: as the temperature decreases from 0.21 to 0.17, the peak shifts to a larger value indicating that *in cold water* proteins become less compact for decreasing temperature. This behavior in hot and

cold water is shown systematically in the inset of Figure 5–2, which depicts the temperature dependence of R_G. The radius of gyration decreases as temperature decreases towards 0.21 – in hot water the protein folds as temperature decreases. Below that temperature, R_G increases monotonically as temperature decreases – in cold water the protein unfolds as temperature decreases. These two types of behavior are characteristic of heat and cold denaturation of real proteins and are in line with previous studies [55, 86, 75].

The parabolic-like shape of R_G (see the inset of Fig. 5–2) cannot be mapped into a model with local monomer-monomer interactions only [14]. To study the role of water, we show in Fig. 5–3 the average H-bond energy per water molecule for shell and bulk water. The energy of shell water averaged over the different configurations is higher than the energy of bulk water at high temperatures. This changes gradually as temperature decreases such that the creation of shell water becomes energetically favorable at low temperatures. Therefore, when a protein is immersed in cold water it releases heat to form the shell, while in hot water it absorbs heat. These features are again characteristic of cold and heat denaturation of real proteins [78]. In the inset of Fig. 5–3 we show the energy absorbed by the system to create the shell around the protein. The absorbed energy is defined as the difference in H-bond energy between shell and bulk water multiplied by the average number of molecules forming the shell. The absorbed energy decreases monotonically with decreasing temperature and becomes negative below some T indicating heat release.

Characteristic configurations of the protein at different temperatures are shown in Fig. 5–4. In cold water (upper panels), the solvent forms a cage around each

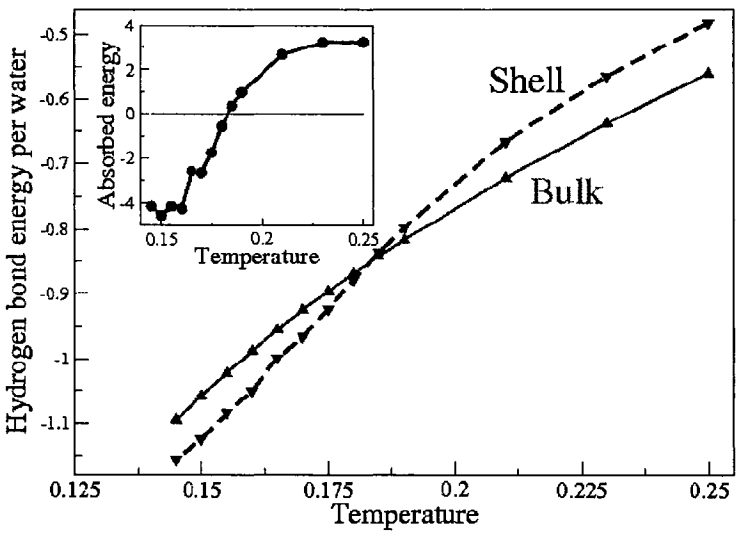

Figure 5–3: Hydrogen bond energy per water molecule for shell and bulk water. Inset: Absorbed energy to accommodate the protein at different temperatures. The shell is defined by water molecules whose distance to the protein is less than 2.5 in units of R_h.

monomer of the protein, i.e. monomers are surrounded by an ordered layer of water molecules. Molecules forming the cage are strongly H-bonded to each other and therefore have a low energy. At $T = 0.21$, the protein favors compact configurations. Water molecules close to the protein have at least one non-saturated H-bond which is pointing towards the protein. When the temperature is increased to $T = 0.25$, most monomers are in contact with the solvent. The solvent forms either incomplete cages around monomers, i.e. cages which do not surround monomers from all sides, or they correspond to particles which are weakly bonded to the other solvent particles and are thus energetic. The crossover behavior of shell water shown in Fig. 5–3 is therefore

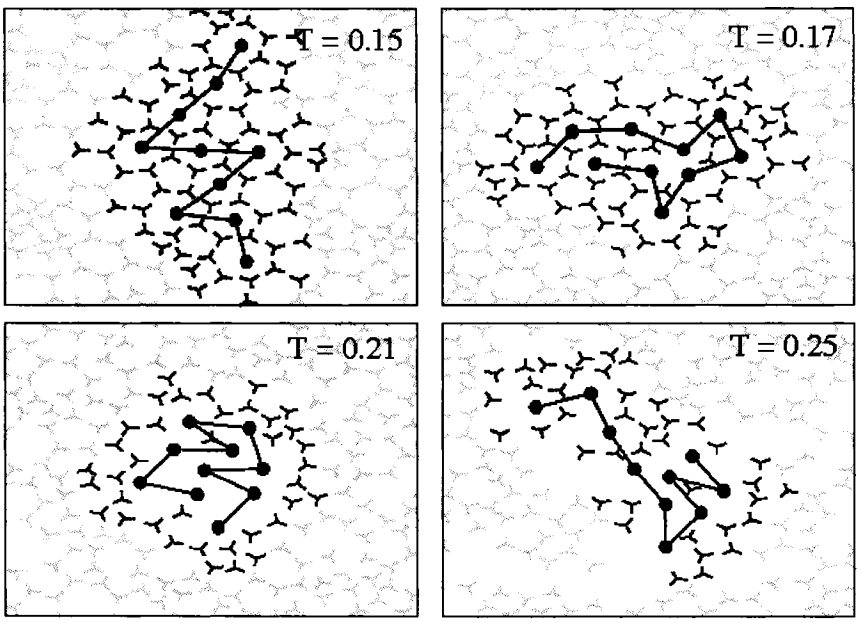

Figure 5–4: Characteristic configurations of a protein in cold water ($T = 0.15$ and $T = 0.17$), at an intermediate temperature ($T = 0.21$), and in hot water ($T = 0.25$). The distance of highlighted (shell) water molecules to the protein is less than 2.5 in units of R_h. In cold water, the monomers are typically surrounded by clathrate-like cages.

characterized by the formation of cage-like configurations at low temperatures and the presence of dangling H-bonds at high temperatures.

Configurations where monomers are separated by an ordered layer of solvent molecules have also been shown to become more stable, as temperature decreases, in models for the hydrophobic effect of methane-like solutes [74, 97]. Solvent layers around those monomer-pairs are highly ordered such that their formation decreases

the entropy of the system [1] . Unfolding at low temperatures is therefore accompanied by a lowering in the entropy of the *total* system in accordance with experiments [78], shell water molecules becoming more ordered as the protein becomes less ordered. This mechanism explains the counter-intuitive decrease in entropy during cold denaturation. The phenomenology is as follows. When non-polar solutes are transferred into water, the system relaxes by ordering those solvent molecules around the solute. This ordering has an entropic cost which is minimized by clustering non-polar solutes together, as this decreases the amount of surface around solutes. As the temperature decreases below a particular value, the system rebuilds the ordered layer of solvent around non-polar solutes to saturate the dangling H-bonds left on the surface of clustered solutes – minimizing the enthalpy. Although hydrophobicity is not the only force responsible for the stability of proteins, the formation of a hydrophobic core plays the dominant role.

5.4 Conclusion

In conclusion, we find that, at low temperatures, shell water forms hydrogen-bonds better than bulk water. Microscopically this correlates with the presence of solvent-separated-configurations which accounts for the unfolding of the protein at low temperatures. The existence of such low energetic states for shell water at low

[1] According to table 1 of Southall and Dill [97], the entropy term $(-T\Delta S)$ of contact configurations is -0.55 while the same term for solvent separated configuration is 0.18 – for two monomers embedded in 120 water molecules at $T = 0.21$. Therefore the entropy decreases when hydrophobic contacts are destabilized in favor of solvent separated configurations.

T explains why cold denaturation proceeds with heat release as opposed to heat absorption seen during heat denaturation. Although here we studied cold denaturation in two dimensions, solvent-separated-configurations have also been shown to become more favorable as temperature decreases in a 3D model for the hydrophobic effect [74]. Therefore we expect that the results found in this work remain valid in 3D systems. Our results further suggest that cold and pressure denaturation could be studied under a single framework: a transition towards solvent-separated-configurations [47].

Chapter 6

Conclusion

In this thesis we developed and studied minimal models for proteins undergoing structural changes. In particular, we studied bond rupture under external stress, and unfolding resulting from shear flow and temperature changes. To study those non-equilibrium processes, we explored minimal models using Langevin dynamics.

In the first part of the thesis we showed that bond rupture is a thermally activated process. This phenomena depends therefore on the form of the energy landscape and, in particular, on its curvatures and energy barrier. The dependence of those quantities on the applied stress was obtained analytically by studying the landscape around the critical stress where the barrier is zero. This dependence was obtained for two classes of models, and simulations of specific models falling into those classes were performed to confirm analytical results. The experimentally relevant situation, where the stress applied to the protein increases linearly with time, was also studied. We found that the force F, at which the protein ruptures, depends on the pulling rate r and temperature T as $F \sim const - T^{2/3}|\ln(r/T)|^{2/3}$. Experiments, like the biomembrane-force-probe, are expected to fall into this universality

class, and could benefit from the use of this equation to extract relevant microscopic parameters for biomolecules.

In the second part of the thesis we studied how resistance to shear flow and temperature depends on the conformation of a protein model. Those conformations were classified according to their stability to mutation, i.e., designability. This quantity was, then, shown to be negatively correlated with resistance to shear and temperature: highly designable structures are easier to unfold than low designable ones. By counting the number of covalent bonds between surface and core monomers, a phenomenological explanation is provided for this correlation. Although this negative correlation seems to contradict the designability principle, which states that highly designable structures are more stable thermodynamically, we show that this contradiction is only apparent. It would be interesting at this stage to verify if the correlation found in this work holds for more refined models of proteins structure.

In the last part of this thesis we studied the microscopic mechanism of cold denaturation. This phenomena corresponds to the loss of stability of protein structures at low temperatures. We showed that water molecules, which form clathrate-like cages around hydrophobic solutes, were responsible for cold denaturation. At high temperatures the entropic cost of forming clathrate-like cages is considered to be the driving force for protein folding. At low temperatures, we demonstrated that it is enthalpically favorable to form clathrate-like cages. Therefore proteins unfold at low temperatures to hydrate hydrophobic monomers which are inside the dry protein core. This mechanism explains cold denaturation and we showed that it is consistent with the experimentally observed heat release that occurs during this phenomena.

Since pressure denaturation of proteins is also considered to be driven by the formation of clathrate-like cages, we suggested that pressure and cold denaturation could be studied under a single framework.

In conclusion, we have studied dynamical processes of conformational changes in proteins using minimal models. We obtained: i) an equation that can be used in force spectroscopy to extract parameters of biomolecules; ii) a correlation between structural stability to mutation and stability to external pertubation; and iii) the microscopic mechanism of cold denaturation of proteins. At this point it would be interesting to validate the results obtain in this study with more realistic three-dimensional models.

References

[1] http://www.lsbu.ac.uk/water

[2] M. P. Allen and D. J. Tildesley. *Computer Simulations of Liquids.* Clarendon, Oxford, 1990.

[3] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72:2384, 1980.

[4] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[5] Michal Ciesla Armando M. Maroja, Fernando A. Oliveira and Lech Longa. Polymer fragmentation in extensional flow. *Physical Review E*, 63:61801, 2001.

[6] George I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200:618–627, 1978.

[7] Gerhard Besold, Ilpo Vattulainen, Mikko Karttunen, and James M. Polson. Towards better integrators for dissipative particle dynamics simulations. *Physical Review E*, 62:R7611, 2000.

[8] K. Binder. Nucleation barriers, spinodals, and the Ginzburg criterion. *Physical Review A*, 29:341–349, 1984.

[9] Pierre Bongrand. Ligand-receptor interactions. *Rep. Prog. Phys.*, 62:921–968, 1999.

[10] Daniel T. Bowron, Adriano Filipponi, Mark A. Roberts, and John L. Finney. Hydrophobic hydration and the formation of a clathrate hydrate. *Phys. Rev. Lett.*, 81:4164–4167, 1998.

[11] Steven E. Brenner, Cyrus Chothia, and Tim JP Hubbard. Population statistics of protein structures: lessons from structural classifications. *Current Opinion in Structural Biology*, 7:369–376, 1997.

[12] R. A. Broglia, G. Tiana, H. E. Roman, E. Vigezzi, and E. Shakhnovich. Stability of designed proteins against mutations. 82(23):4727, 1999.

[13] Axel Brunger, Charles L. Brooks III, and Martin Karplus. Stochastic boundary conditions for molecular dynamics simulations of st2 water. *Chemical Physics Letters*, 105:495, 1984.

[14] Pierpaolo Bruscolini and Lapo Casetti. Lattice model for cold and warm swelling of polymers in water. *Phys. Rev. E*, 61:R2208, 2000.

[15] Sergey V. Buldyrev, Pradeep Kumar, and H. Eugene Stanley. A physical mechanism underlying the increase of aqueous solubility of nonpolar compounds and the denaturation of proteins upon cooling. *cond-mat*, page 0701485, 2007.

[16] Siedlecki C.A., Lestini B. J., KottkeMarchant K., Eppell S. J., Wilson D. L., and Marchant R. E. Shear-dependent changes in the three-dimensional structure of human von willebrand factor. *BLOOD*, 88:2939–2950, 1996.

[17] Hu Chen, Xin Zhou, and Zhong-Can Ou-Yang. Secondary-structure-favored hydrophobic polar lattice model of protein folding. 64:041905, 2001.

[18] O. Collet. Warm and cold denaturation in the phase diagram of a protein lattice model. *Europhysics Letters*, 53:93–99, 2001.

[19] Thomas M. Devlin, editor. *Textbook of Biochemistry With Clinical Correlations*. John Wiley, York, 1997.

[20] C. L. Dias and Martin Grant. Unfolding designable structures. *The European Physical Journal B*, 50:265, 2006.

[21] Cristiano L. Dias, Tapio Ala-Nissila, Mikko Karttunen, Ilpo Vattulainen, and Martin Grant. Microscopic mechanism for cold denaturation. *Submitted for publication*, page 00, 2007.

[22] Cristiano L. Dias, Martin Dube, Fernando Oliveira, and Martin Grant. Scaling in force spectroscopy. *Physical Review E*, 72:011918, 2005.

[23] Cristiano L. Dias and Martin Grant. Designable structures are easy to unfold. *Physical Review E*, 74:042902, 2006.

[24] Ken A. Dill. Dominant forces in protein folding. *Biochemistry*, 29:7133, 1990.

[25] Ken A. Dill, Thomas M. Truskett, Vojko Vlachy, and Barbara Hribar-Lee. Modeling water, the hydrophobic effect, and ion solvation. *Annual Review of Biophysics and Biomolecular Structure*, 34:173, 2005.

[26] Yong Duan and Peter A. Kollman. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282:740, 1998.

[27] O.K. Dudko, A.E. Filippov, J. Klafter, and M. Urbakh. A new look at dynamics force spectroscopy of adhesion bonds. *PNAS*, 100:11378, 2003.

[28] T. Gyalog Ch. Loppacher M. Bammerlin E. Meyer E. Gnecco, R. Bennewitz and H.-J. Gntherodt. Velocity dependence of atomic friction. *Phys. Rev. Lett.*, 84:1172, 2000.

[29] John T. Edsall. Apparent molal heat capacities of amino acids and other organic compounds. *Journal of the American Chemical Society*, 54:1506–1507, 1935.

[30] Donald L. Ermak and Helen Buckholz. Numerical integration of the langevin equation: Monte carlo simulation. *Journal of Computational Physics*, 35:169–182, 1980.

[31] E. Evans and R. Merkel. Dynamic strength of molecular adhesion bonds. *Biophysical Journal*, 72:1541–1555, 1997.

[32] Evan Evans. Looking inside molecular bonds at biological interfaces with dynamic force spectroscopy. *Biophysical Chemistry*, 82:83–99, 1999.

[33] Evan Evans, Andrew Leung, Dan Hammer, and Scott Simon. Chemically distinct transition states govern rapid dissociation of single l-selectin bonds under force. *PNAS*, 98(7):3784–3789, 2001.

[34] Scott E. Feller, Yuhong Zhang, Richard W. Pastor, and Bernard R. Brooks. Constant pressure molecular dynamics simulation: The langevin piston method. *The Journal of Chemical Physics*, 103:4613, 1995.

[35] Adriano Filipponi, Daniel T. Bowron, Colin Lobban, and John L. Finney. Structural determination of the hydrophobic hydration shell of Kr. *Phys. Rev. Lett.*, 79:1293–1296, 1997.

[36] Henry S. Frank and Marjorie W. Evans. Free volume and entropy in condensed systems iii. entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *Journal of Chemical Physics*, 13:507, 1945.

[37] T. A. Fulton and L. N. Dunkleberger. Lifetime of the zero-voltage state in Josephson tunnel junctions. *Physical Review B*, 9:4760, 1974.

[38] Anupam Garg. Escape-field distribution for escape from a metastable potential well subject to a steadily increasing bias field. *Physical Review B*, 51:15592, 1995.

[39] Tuhin Ghosh, Angel E. Garca, and Shekhar Garde. Molecular dynamics simulations of pressure effects on hydrophobic interactions. *J. Am. Chem. Soc.*, 123:10997–11003, 2001.

[40] S. J. Gill, S. F. Dec, G. Olofsson, and I. Wadsoe. Anomalous heat capacity of hydrophobic solvation. *J. Phys. Chem.*, 89:3758, 1985.

[41] Nigel Goldenfeld. *Lectures on Phase Transitions and the Renormalization Group.* Perseus Books, 1992.

[42] Herbert Goldstein. *Classical Mechanics.* Addison-Wesley, 1980.

[43] Peter Hanggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after Kramers. *Reviews of Modern Physics*, 62(2):251–, 1990.

[44] S. A. Hawley. Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry*, 10:2436 – 2442, 1971.

[45] Robert Helling, Hao Li, Regis Melin, Jonathan Miller, Ned Wingreen, Chen Zeng, and Chao Tang. The designability of protein structures. *Journal of Molecular Graphics and Modelling*, 19:157–167, 2001.

[46] A. Hummer, G.; Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical Journal*, 85:5, 2003.

[47] G. Hummer, Shekhar Garde, Angel E. Garcia, and Michael E. Paulaitis. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. 95:1552, 1998.

[48] Jacob N. Israelachvili. *Intermolecular and Surface Forces: With Applications to Colloidal and Biological Systems.* Elsevier Science & Technology Books, 1992.

[49] Jess A. Izaguirre, Daniel P. Catarello, Justin M. Wozniak, and Robert D. Skeel. Langevin stabilization of molecular dynamics. *Journal of Chemical Physics*, 114:2090, 2001.

[50] Ruth S. Spolar Jeff R. Livingstone and M. Thomas Record. Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry*, 30:4237, 1991.

[51] W. Kauzmann. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.*, 14:1, 1959.

[52] J. C. Kendrew, G. Bodo, H. M. Dintzis, T.G. Parrish, H. Wyckoff, and D. C. Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181:662, 1958.

[53] J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore. Structure of myoglobin: A three-dimensional fourier synthesis at 2 . resolution. *Nature*, 185:422, 1960.

[54] A. Kolb and B. Dnweg. Optimized constant pressure stochastic dynamics. *The Journal of Chemical Physics*, 111:4453, 1999.

[55] Shigeru Kunugi and Naoki Tanaka. Cold denaturation of proteins under high pressure. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1595:329–344, 2002.

[56] B. Lee and Giuseppe Graziano. A two-state model of hydrophobic hydration that produces compensateion enthalpy and entropy changes. *Journal of the American Chemical Society*, 118(22):5163, 1996.

[57] Hao Li, Robert Helling, Chao Tang, and Ned Wingreen. Emergence of preferred structures in a simple model of protein folding. *Science*, 273:666, 1996.

[58] Hao Li, Chao Tang, and Ned S. Wingreen. Nature of driving force for protein folding: A result from analyzing the statistical potential. 79(4):765, 1997.

[59] Hao Li, Chao Tang, and Ned S. Wingreen. Are protein folds atypical ? 95:4987–4990, 1998.

[60] Pai-Chi Li and Dmitrii E. Makarov. Theoretical studies of the mechanical unfolding of the muscle protein titin: Bridging the time-scale gap between simulation and experiment. *Journal of Chemical Physics*, 119:9260, 2003.

[61] Jie Liang and Ken A. Dill. Are proteins well-packed? *Biophys. J.*, 81:751, 2001.

[62] Dennis R. Livesay and Donald J. Jacobs. Conserved quantitative stability/flexibility relationships (qsfr) in an orthologous rnase h pair. *Proteins: Structure, Function, and Bioinformatics*, 62:130–143, 2006.

[63] Yu-Shiu Lo, Ying-Jie Zhu, and Thomas P. Beebe Jr. Loading-rate dependence of individual ligand-receptor bond-rupture forces studied by atomic force microscopy. *Langmuir*, 17:3741–3748, 2001.

[64] Manuel I. Marques, Jose M. Borreguero, H. Eugene Stanley, and Nikolay V. Dokholyan. Possible mechanism for cold denaturation of proteins at high pressure. *Physical Review Letter*, 91(13):138103, 2003.

[65] Filip Meersman, Christopher M. Dobson, and Karel Heremans. Protein unfolding, amyloid fibril formation and configurational energy landscapes under high pressure conditions. *Chem. Soc. Rev.*, 35:908–917, 2006.

[66] Regis Melin, Hao Li, Ned S. Wingreen, and Chao Tang. Designability, thermodynamic stability, and dynamics in protein folding: A lattice model study. 110(2):1252, 1999.

[67] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, and E. Evans. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature*, 397:50, 1999.

[68] Rudolf Merkel. Force spectroscopy on single passive biomolecules and single biomolecular bonds. *Physics Reports*, 346:343–385, 2001.

[69] Jonathan Miller, Chen Zeng, Ned S. Wingreen, and Chao Tang. Emergence of highly designable protein-backbone conformations in an off-lattice model. *PROTEINS: Structure, Function, and Genetics*, 47:506–512, 2002.

[70] Norbert Muller. Search for a realistic view of hydrophobic effects. *Acc. Chem. Res.*, 23:23–28, 1990.

[71] A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia. Scop: a structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology*, 247:536–540, 1995.

[72] F. A. Oliveira. Transition-state analysis for fracture nucleation in polymers: The lennard-jones chain. *Physical Review B*, 57:10576, 1998.

[73] F. A. Oliveira and P. L. Taylor. Breaking in polymer chains. ii. the lennard-jones chain. *Journal of Chemical Physics*, 101(11):10118, 1994.

[74] Dietmar Paschek. Temperature dependence of the hydrophobic hydration and interaction of simple solutes: An examination of five popular water models. *The Journal of Chemical Physics*, 120:6674, 2004.

[75] Dietmar Paschek, Sascha Nonn, and Alfons Geiger. Low-temperature and high-pressure induced swelling of a hydrophobic polymer-chain in aqueous solution. *Phys. Chem. Chem. Physics*, 7:2780–2786, 2005.

[76] Linus Pauling and Robert B. Corey. Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets. *Proceedings of the National Academy of Sciences*, 37:729, 1951.

[77] Linus Pauling, Robert B. Corey, and HR Branson. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences*, 37:205, 1951.

[78] P. L. Privalov, Yu. V. Griko, and S. Yu. Venyaminov. Cold denaturation of myoglobin. *Journal of Molecular Biology*, 190:487–498, 1986.

[79] Peter L. Privalov. Thermodynamics of protein folding. *J. Chem. Thermodynamics*, 29:447–474, 1997.

[80] Rosabella Puthur and K. L. Sebastian. Theory of polymer breaking under tension. *Physical Review B*, 66:024304, 2002.

[81] Aneesur Rahman and Frank H. Stillinger. Molecular dynamics study of liquid water. *The Journal of Chemical Physics*, 55:3336, 1971.

[82] Revanur Ravindra and Roland Winter. On the temperature-pressure free-energy landscape of proteins. *ChemPhysChem*, 4:359 – 365, 2003.

[83] F. Reif. *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill, 1976.

[84] Timothy R. Rettich, Y. Paul Handa, Rubin Battino, and Emmerich Wilhelm. Solubility of gases in liquids. 13. high-precision determination of henry's constants for methane and ethane in liquid water at 275 to 328 k. *J. Phys. Chem.*, 85:3230, 1981.

[85] F. M. Richards. Areas, volumes, packing, and protein structure. *Annual Review of Biophysics and Bioengineering*, 6:151–176, 1977.

[86] Paolo De Los Rios and Guido Caldarelli. Putting proteins back into water. *Physical Review E*, 62:8449, 2000.

[87] Paolo De Los Rios and Guido Caldarelli. Cold and warm swelling of hydrophobic polymers. *Physical Review E*, 63:031802, 2001.

[88] Anirban Sain, Cristiano L. Dias, and Martin Grant. Rupture of an extended object: A many-body Kramers calculation. *Physical Review E*, 74:046111, 2006.

[89] K. Y. Sanbonmatsu and A. E. Garcia. Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *PROTEINS: Structure, Function, and Genetics*, 46:225, 2002.

[90] Yi Sang, Martin Dubé, and Martin Grant. Thermal effects on atomic friction. *Physical Review Letters*, 87(17):174301–4, 2001.

[91] G. M. Schutz. Exactly solvable models for many-body systems far from equilibrium. *Phase Transitions and Critical Phenomena*, 19:1, 2001.

[92] C. T. Shih, Z. Y. Su, J. F. Gwan, B. L. Hao, C. H. Hsieh, J. L. Lo, and H. C. Lee. Mean-field hp model, designability and alpha-helices in protein structures. 84(2):386, 2000.

[93] C. T. Shih, Z. Y. Su, J. F. Gwan, B. L. Hao, C. H. Hsieh, J. L. Lo, and H. C. Lee. Geometric and statistical properties of the mean-field hydrophobic-polar model, the large-small model, and real protein sequences. 65:41923, 2002.

[94] Scott Sills and Ren M. Overney. Creeping friction dynamics and molecular dissipation mechanisms in glassy polymers. *Phys. Rev. Lett.*, 91:95501, 2003.

[95] Kevin A. T. Silverstein, A. D. J. Haymet, and Ken A. Dill. A simple model of water and the hydrophobic effect. *J. Am. Chem. Soc.*, 120 (13):3166–3175, 1998.

[96] S.Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical Journal*, 72:1568–1581, 1997.

[97] Noel T. Southall and Ken A. Dill. Potential of mean force between two hydrophobic solutes in water. *Biophysical Chemistry*, 101-102:295–307, 2002.

[98] T. Strunz, K. Oroszlan, I. Schumakovitch, H.-J. Guntherodt, and M. Hegner. Model energy landscapes and the force-induced dissociation of ligand-receptor bonds. *Biophysical Journal*, 79:1206–1212, 2000.

[99] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. Computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *Journal of Chemical Physics*, 76:637, January 1982.

[100] J. M. Thornton. *Protein Structure: The End Point of the Folding Pathway. In "Protein folding"*. W. H. Freeman and Company. USA, 1992.

[101] Carl Branden & John Tooze. *Introduction to protein structure.* Garland Publishing, 1998.

[102] Tairan Wang, Jonathan Miller, Ned S. Wingreen, Chao Tang, and Ken A. Dill. Symmetry and designability for lattice protein models. 113(113):8329, 2000.

[103] J. H. Weiner and R. E. Forman. Rate theory for solids. iv. classical brownian-motion model. *Physical Review B*, 10:315, July 1974. Langevin + Euler.

[104] Oscal H. Willemsen, Margot M. E. Snel, Alessandra Cambi, Jan Greve, Bart G. De Grooth, and Carl G. Figdor. Biomolecular interactions measured by atomic force microscopy. *Biophysical Journal*, 79:3267, 2000.

[105] Ned S. Wingreen, Hao Li, and Chao Tang. Designability and thermal stability of protein structures. *Polymer*, 45:699–705, 2004.

[106] Adam Zipp and Walter Kauzmann. Pressure denaturation of metmyoglobin. *Biochemistry*, 12:4217 – 4228, 1973.