ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

A STATISTICAL MECHANICS APPROACH TO PROTEIN FOLDING AND BINDING

Thesis Submitted for the Degree of:
Doctor Philosophiæ

Candidate:
Serge Cattarinussi

Supervisor:
Dr. Giancarlo Jug

Academic Year 1990/1991

TRIESTE

SISSA - SCUOLA INTERNAZIONALE SUPERIORE I STUDI AVANZATI
TRIESTE
Strada Costiera 11

A STATISTICAL MECHANICS APPROACH TO PROTEIN FOLDING AND BINDING

Thesis Submitted for the Degree of:
Doctor Philosophiæ

Candidate:
Serge Cattarinussi

Supervisor:
Dr. Giancarlo Jug

Academic Year 1990/1991

ACKNOWLEDGEMENTS

I am indebted to my supervisor for giving freely of his time, for his patient support and for his commitment beyond the line of duty during the course of this work. I am grateful to Prof. A.Borsellino for accepting me in his group, thus giving me the opportunity to work in the Biophysics Sector of the International School of Advanced Studies in Trieste. I want to thank Dr. F.Seno and Prof. A.Stella for introducing me to the use of the exact enumeration analysis of walks.

CONTENTS

INTRODUCTION

The present work deals with the problems of *protein folding* and *protein adsorption*, some of the most exciting topics in the present scientific scene in biophysics. The large interest raised by these subjects in the scientific community relies upon three main reasons. First of all, substantial progress can only be obtained through the fruitful collaboration of experts from disciplines as different as mathematics, physics, chemistry and biology. Secondly, the protein folding problem constitutes not only a domain of application for well established techniques in each of the mentioned disciplines, but it also instigates, and often requires, fundamental developments in separate fields. Last but not least comes the fact that the protein folding problem perfectly fits in with the present trend in the evolution of science for which, in agreement with the needs of human society, the social and biological sciences acquire a new, more prominent, position.

Owing to its intrinsic multidisciplinary nature, the protein folding problem presents different possible approaches. In our opinion they can be roughly classified in two groups : the biological and the physical approaches. Each is mainly concerned with one of the two fundamental questions arising when dealing with protein folding. Namely: the structure-function relation is the essential matter of concern for biologists, while the study of folding mechanisms and the characterisation of the folded and unfolded states mostly attract the physicists. Biophysics should therefore be concerned with the underlying, unifying question of the prediction of functional structures. Unfortunately, a complete resolution of

the latter still seems to pertain to the domain of fiction or at least involve some unpredictable fundamental discovery in the treatment of complex systems. Thus the question arises of what point of view should be adopted in a thesis submitted for the PhD degree in biophysics. The answer can be rather personal and controversial, but one should be guided by the observation that many protein functions have not yet received a satisfactory explanation of *how* they are, or can be, performed. This question not only requires a good comprehension of the structure-function relation but also a deep insight in the physico-chemical mechanisms underlying biopolymer functions. Equilibrium thermodynamics and statistical physics can provide part of such basic information, to be further included in more specific calculations and/or computer simulations.We have therefore adopted, in the present work, a theoretical physics approach to the protein folding problem, based, precisely, on classical statistical mechanics.

An underlying aim of the present thesis is not only to support the idea that biological systems are suitable subjects for producing nice theoretical physics, but *also* that theoretical physics studies can bring useful, quite general, information on these biological systems despite their intrinsic specificity. We have tried to keep in mind the scope mentioned above and therefore concentrate our efforts on those particular physical aspects both susceptible of being successfully treated by statistical mechanics methods and of playing an important rôle in understanding mechanisms involved in protein functions. As a corollary, some interesting theoretical questions have not received the accurate treatment they would have deserved in a different context. In particular precise numerical evaluations, based on the models in use, have been considered as being of secondary importance in contrast to overall, general features.

Proteins are the building blocks of the living world, to which they confer an

incredibly high degree of variety and richness. Most of the genetic information is expressed by proteins and, as a consequence, they are the most abundant organic molecules in cells, constituting 50 percent or more of their dry weight. They are found in every part of every cell, since they are fundamental in all aspects of cell structure and function. There are many different kinds of proteins, each specialized for a given biological function. Depending on their conformation, proteins can be placed in two major classes, fibrous and globular. The *fibrous proteins* consist of polypeptide chains arranged in parallel along a single axis, to yield long fibers or sheets. They are insoluble in water and represent the basic structural elements in the connective tissue of higher animals. Examples are *collagen* of tendons and bone matrix, *α-keratin* of hair, horn, skin, and nails, and *elastin* of elastic connective tissue.[1]

More relevant for our purposes are the *globular proteins* , in which the polypeptide chains are tightly folded into compact spherical or globular shapes. They are often soluble in water and they usually have mobile or dynamic functions in the cell, such as enzymatic, hormonal or transport functions. Two or more polypeptide chains often associate by means of weak (non-covalent) interactions to form *oligomeric* proteins as for instance *hemoglobin* which consists of four polypeptide chains fitting together to form a compact globular assembly of considerable stability. [1]

Proteins can be classified according to their function. The *enzymes*, which are virtually all globular proteins, represent the largest class. They have an extraordinary catalytic power and are highly specific in their functions. Each type of enzyme molecule contains an active site to which its specific substrate is bound during the catalytic cycle. The *storage proteins* have the function of storing amino acids as nutrients. The *transport* proteins are capable of binding

and transporting specific types of molecules, via the blood or through tissues or cellular membranes. The *contractile* proteins are the major elements of muscles. Antibodies, which play a fundamental rôle in the immune system, are classified among *protective* proteins while *toxins* play a somewhat opposite rôle. Proteins with *hormonal* functions regulate a large variety of processes in the organisms and *structural* proteins form the major extracellular components[2].

Three-dimensional protein configurations profoundly influence and determine their physiological activity. Therefore, the folding process by which proteins acquire their spatial conformation is worthy of a thorough investigation. The determination of the protein conformation involves physical methods and can hardly be performed "in vivo". Nevertheless, conformational changes are believed to play a fundamental rôle in a large number of poorly understood physiological mechanisms. A quite general example could be the proteolytic degradation which ensures the turnover of almost all the proteins in the organisms. Degradation not only allows the organism to dispose at any time of "fresh" and fully functional proteins but also provides an efficient way of controlling the protein concentration. Responsible for degradation are often proteases, enzymes capable of "cutting" proteins at specific location by catalysing the hydrolysis of particular peptide bonds. Significant correlations seem to exist between degradation rates and susceptibility to thermal unfolding, showing that folded conformations are important, for at least minimizing the rate of degradation. Abnormal proteins, such as those resulting from chemical modifications, are almost invariably found to be rapidly degradated proteolytically in cells [3].

Despite its rather generic biological importance, protein folding enters in the physiological mechanisms with a high degree of specificity. A statistical mechanics approach to the phenomenon *alone* seems therefore of little practical interest, or,

at least, too restrictive. However, the prospects radically change if we consider it in relation with *protein adsorption*. The mutual influence of folding and adsorption can generate elementary and fundamental processes underlying a multitude of highly specific protein functions. Actually, there is a question that immediately arises when considering a protein adsorbing at, let us say, a cell membrane. What happens to the protein conformation when attaching to the membrane? It depends ..., would answer any biologist working in the field, and a list of cases probably would follow, long enough to scare the more courageous amongst the physicists. So, let us formulate the question in a slightly different, and more appropriate way. *Is there any tendency in conformational changes induced by the adsorption process*? This is precisely the question we address in the present work and, thus formulated, it certainly pertains to the domain of statistical physics. Our protein models allow us to give an affirmative answer : *the folded phase is stabilized upon adsorption*. Alternatively, we will say that *adsorption tends to enhance the folding temperature*. Can this important result be used as a pattern in trying to understand a variety of molecular mechanisms involving proteins and membranes? We really hope that it will be the case and we even believe the result to be general enough for being successfully applied to similar problems which do not directly imply adsorption on membranes. Actually, we ascribe the origin of this effect to the reduction in protein's entropy induced by the adsorbing substrate. A similar entropy reduction could occur when ligands are bound to enzymes, thus playing an important rôle in the allosteric regulation of enzyme activity. As pointed out by Dill and Alonso [4], the folding process itself could be related to such a mechanism, even when occuring far from any membrane. Disulfide bonds and other cross-links should stabilize the globular states by reducing the conformal entropy of unfolded states [5].

The thesis is organised as follows : In Chapter One we give an overview

of the mechanisms involved in protein folding and adsorption, both from the biochemical and the statistical mechanics points of view. In Chapter Two we present a model for folding and adsorption especially designed to be treated with real-space renormalisation group methods, while in Chapter Three we describe a less elegant but more powerful model relying on the exact enumeration of configurations and finite series analysis. Chapter Four is concerned with a more realistic model in which a random distribution of hydrophylic and hydrophobic residues is considered. In Chapter Five a global critical review of the main results is presented in which a physical as well as a biochemical point of view is taken. In the appendices, the interested reader will find a short description of the main computer algorithm developed for the purposes of this work. Tables of configuration classification are also given in order to facilitate further developments by other workers.

The most innovating features presented in this thesis have already been the subject of previous publications and concern :

- the complete phase diagram in the renormalisation group parameter space for a Θ-point polymer chain adsorbing to a wall, both in $D = 2$ and $D = 3$ dimensions. [6]

- the evidence for a *surface-induced enhancement of the folding transition temperature* or, in other words, a stabilisation of the folded phase at adsorption. [7]

- the analysis of changes in the phase diagram induced by a variable random distribution of hydrophylic and of hydrophobic residues along the protein chain. [8]

CHAPTER ONE

QUALITATIVE ASPECTS OF PROTEIN FOLDING AND ADSORPTION

This first Chapter is devoted to a general overview of the protein folding and protein adsorption phenomena, with no references to their biological importance (see Chapter 5) but centered on chemical and physical aspects. It articulates in two main Sections. First we will adopt a biochemical, and then a statistical mechanics point of view, with the intention of emphasizing the complementarity of the two approaches. Both have some advantages, but also limitations. Biochemistry allows for a very accurate description of a variety of known proteins (and of their behaviour) but offers poor prediction capability. As an example, we will consider the difficulties encountered with sequence analysis in predicting folded structures. On the other hand, statistical mechanics offers more prediction power, but less quantitative accuracy, which is a limitation in the study of biological systems. Therefore, the combined use of both methods should represent the most fruitful approach.

1.1 THE BIOCHEMICAL APPROACH

The enormous amount and the diversity, as well as the high degree of specificity, of the functions performed by proteins is deeply connected with their composition. Actually proteins consist of one or more polypeptide chains generally

linked together by weak bonds and sometimes by a few disulfide bridges. To synthesize the *unbranched* polypeptide chains, the living cells recognize and use 20 different amino acids. With the exception of proline, all these "natural" amino acids have the same structure:

$$H_2N - CH(R) - CO_2H \tag{1.1}$$

and they only differ in the chemical composition of the *side chain* R. During the synthetization process, many such amino acids (generally from 50 to 1000) are linked together in a determined sequence, according to the information stored in the nucleic acids. The resulting chain

$$H - (NH - CH(R_i) - CO)_n - OH \tag{1.2}$$

presents a backbone consisting of the repeated unit of three atoms: the amide N, the alpha C, and the carbonyl C. In principle rotation may occur about any of the three bonds of the polypeptide backbone. However, the peptide bond (which links the carbonyl C with the amide N of the following backbone unit) appears to have partial double-bonded character. Rotation of this bond is then markedly restricted (see figure 1.1).

By itself, the periodic structure of the polypeptide chain backbone would not allow proteins to exhibit their amazing variety of forms and functions. It is clear that a fundamental contribution in that direction comes from the twenty amino acid side chains which differ in many important aspects. First of all they have different chemical affinities and reactivities. An important example is the

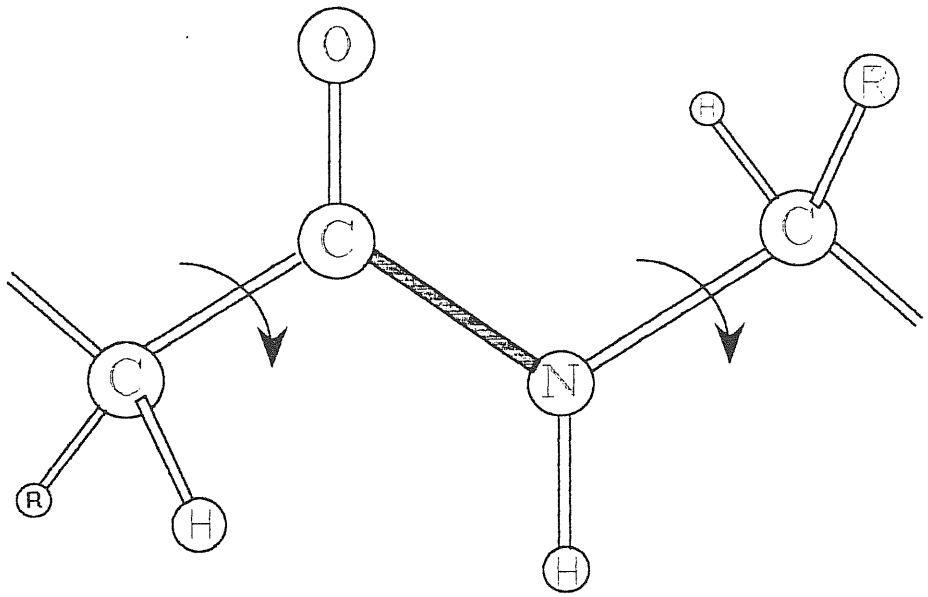

Figure 1.1 The repeating unit of the polypeptide backbone. The arrows indicate the allowed rotations responsible for protein flexibility

strongly reactive thiol group which can form disufide (covalent) bonds between cysteine residues, though playing an important rôle in the stabilisation of some protein's spatial conformation. Side chains also have different protonation and deprotonation tendencies so that, at a given pH, some of them carry a positive charge while others are neutral or negative. The steric constraints imposed to the molecule vary from amino acid to amino acid and finally, but perhaps most important, the hydrophobic or hydrophilic properties of amino acid side chains induce a diametrically opposite behaviour in relation to the solvent.

The specific sequence of the amino acids in the polypeptide chains forming a protein, as well as the location of all the disulfide and other covalent bonds, is called the *primary structure* of the protein. This primary structure is generally determined when the protein is synthesized, nevertheless some modifications can

occur when the protein has already acquired a three-dimensional structure, as for instance in the case of hydrolysis of one or more peptide bonds.

The regular and periodic structure of the chain backbone gives rise to the *secondary structure* of proteins, consisting of a local arrangement of the backbone units. Such an arrangement largely depends on the rotational freedom allowed by the covalent bond in the chain's backbone. Therefore the double-bonded character of the peptide bond plays a crucial rôle in the formation of the secondary structure. The most common conformations found in proteins are α-helices and β-sheets. Both are stabilized by hydrogen bonds formed between the carbonyl ($C = O$) and NH groups of the backbone and both strongly limit the flexibility of the chain. The side chains can hinder or influence secondary structure formation either by their steric constraints, by their intrinsic charge or because of their ability of forming other particular bonds. Of particular importance is the effect of proline, the special backbone unit of which breaks the α-helix structures.

Along a polypeptide chain, ordered and rather rigid segments alternate with flexible, unordered parts usually referred to as *random coils*. This allows the chain to undergo a further folding to acquire a compact three-dimensional *non-periodic* conformation: the *tertiary structure*. The aperiodicity of this three-dimensional molecular organisation certainly represents the most important feature of the folded, *globular*, proteins. Actually, the ability of binding selectively to specific ligands, which is the fundamental function of most globular proteins, is directly related to their non-periodic spatial organisation. Only a non-periodic structure can give rise to the formation of the highly specific binding sites which are present on the globule's surface and which allow for the amazingly rich enzymatic capabilities of globular proteins.

The overall shape of small folded proteins is roughly spherical, but with a very

irregular surface. However, where a protein consists of more than 200 residues, the structure usually appears to consist of two or three rather spherical units, generally referred to as domains, often linked together through a single segment of polypeptide chain. Within a domain, the course of the polypeptide backbone gives the impression of somewhat stiff chain segments interspersed with relatively tight turns which are almost always on the surface of the protein.

The protein *quaternary structure* appears when two or more polypeptide chains link together through weak bonds. Each subunit is generally folded into an apparently independent globular conformation, which then interacts with other subunits.

1.1.1 Protein Folding

There are no precise general rules for the formation of the globular stucture in proteins. Nevertheless some generalisations are possible:

- Most of the non-polar residues are found inside the globule, far from the acqueous sovent.

- The charged groups are essentially distributed on the surface of the molecule, in contact with the solvent.

- Polar groups are frequently found in the proximity of the binding site for ligands.

- The most usual secondary structure in globular proteins is represented by the α-helix, the presence of which varies from 5 to 75 percent of the entire chain.

These features are connected with the different nature of the forces stabilizing the globule: ionic forces, hydrogen bonds or hydrophobic interactions.

Ionic forces appear between residues that carry a net positive charge such as lysine, arginine and histidine, and those that carry a negative charge such as glutamic and aspartic acid. The *hydrogen bonds* form between proton donors (NH, NH_2, NH_3^+, OH) and acceptors (COO^-, CO, OH, NH_2). Water has a strong tendency to break ionic as well as hydrogen bonds. The ability of a group to make a hydrogen bond depends on its ionisation state and thus on the pH of the solution.

The hydrophobic interactions arise from the fact that when non-polar groups are inserted into water, a new interface is created, which requires the adjacent water molecules to assume a more ordered arrangement producing an unfavourable entropy reduction. As a result, a polypeptide chain will tend to assume a conformation in which the non-polar residues are shielded from exposure to water. Furthermore, the hydrophilic groups that could have been dragged inside the molecule can then form ionic and hydrogen bonds which could not exist in an acqueous environment. It is generally admitted that hydrophobic interactions are the principal stabizing forces for the globular state [5].

Among the experimentally accessible physical characteristics [9] which can serve to differentiate folded and unfolded proteins one can mention hydrodynamic and spectral properties. The former, obviously, resume to a much lower viscosity for the more compact globular conformation [10]. The latter involve light absorption [11], fluorescence [12], circular dichroism [13], optical rotatory dispersion [14], and nuclear magnetic resonance [15].

Chemical properties are also widely modified upon folding but they generally

depend upon the nature of the reagent. This suggests that it is the local concentration of the reagent, determined by its interactions with the neighbouring parts of the protein, that is often the crucial factor.

Many difficult questions arise when one considers the folding of proteins. Some of them are connected with the relation between the primary and the tertiary or quaternary structure.

- To which extent the amino acid sequence determines the three-dimensional conformation of proteins?

- How stable is the tertiary structure with regard to chemical transformations such as the proteolysis of some peptide bonds?

- Can different sequences give rise to similar globular conformations?

- Why does a given protein choose a determined configuration instead of another?

- What is the relationship between structural and functional similarity?

It is quite obvious that the answers to the above questions involve preliminary and more physical aspects:

- How do denaturant factors effectively act to induce drastic conformational changes?

- Is folding really a two-state process or does it involve intermediate states?

- How flexible are globular proteins?

- How well are atoms packed inside a protein?

Most of these questions have received partial empirical answers; few, if any, are well supported by a convincing theoretical framework. Since it is not our purpose to discuss all these fascinating topics in full details we simply and briefly discuss those which are the most pertinent to our work.

Denaturation, the unfolding of native globular proteins, has been shown to be a reversible process. This important fact proves that all the information necessary to determine the globular conformation of a protein is contained in the sequence of the amino acid residues in the polypeptide chain [16]. All conditions which can alter hydrophobic interactions, ionic forces, and hydrogen bonds, are susceptible of inducing denaturation. The most important denaturant factors are extremes of pH, non-polar solvents, temperature enhancement, and a number of chemical compounds such as urea and guanidine hydrochloride. In order to obtain insight into the physical basis of globular conformation stabilization, denaturation experiments have been performed using all of the above denaturants. The predominance of the hydrophobic interactions, and therefore that of the entropic effects, in the stabilization of the globular conformations represents the overall conclusions of these experiments. Actually, even chemical denaturants seem to act indirectly by interacting with water molecules, with a consequent decrease of hydrophobic interactions [2]. The rôle of hydrophobic interaction in temperature-driven denaturation has long represented a controversial point since the strength of the hydrophobic interaction increases with increasing temperature. However, recent quantitative arguments support the conjecture that temperature-induced conformational changes could be principally driven by the gain of conformational entropy of the polypeptide chain [17].

The question of the nature of the folding transition has received considerable attention. The sharpness of the denaturation clearly reveals a cooperative process

involving the majority of the chain residues and, therefore, can be seen as a phase transition. Moreover the virtual independence of the measured transition curves from the experimental methods, suggests that the folding-unfolding transition is a two-state process with only the fully folded and unfolded states present. The idea is supported by calorimetric studies [18,19] at least for small proteins. In contrast, complex proteins present independently folding-domains, and therefore a multistate folding process.

Theoretical studies of protein conformation can be classified in three groups: Molecular dynamics, sequence comparison, and simplified protein model analysis.

In *molecular dynamics* calculations, the classical equations of motion for protein atoms are solved using an ad hoc expression for the energy as a function of the conformation. The disavantage of the method is that its application is limited to systems where the nature of the chemical *bonding* is well defined, can be specified at the beginning of a given simulation and does not change during the simulation itself. The principal limitation of the method consists in the virtual impossibility to follow the evolution of the molecule during a long enough period of time to allow large conformational changes. This prevents the method to be used to describe protein folding to a native state from an arbitrary unfolded structure. Moreover it requires that one knows, to a reasonable accuracy, the starting conformation. Therefore, the method can essentially exhibit its usefulness in predicting changes induced by substituting one or a few amino acids in a structurally well known protein [20].

Sequence comparison methods have essentially been applied to secondary structure prediction [21]. The success rates in predicting whether or not a given amino acid belongs to an α-helix reach 80 percent but, drops to approximately 50 percent in predicting whether residues belong to an α-helix, β-strand, reverse

turn, or random coil [22]. These are poor results, especially if we consider the relative "simplicity" of the secondary structure with respect to the tertiary structure in which the interactions between distant residues along the chain play a much more determinant rôle. Nevertheless the method has been improved and *structural profile* comparison, which is essentially a parametric approach to sequence comparison, has made it possible to reveal three-dimensional similarities between proteins that are not demonstrably homologous in their primary structure [23]. Once again the method is more useful for the comparison of slightly mutated sequences to their natural-type counterpart than for tertiary structure prediction, starting from the primary structure.

Unlike the above methods, simplified protein models allow for the investigation of the folding process, but the "picture" of the folded protein they supply is by far not precise enough so as to make structure-function relations possible. As a rule, thermodynamics and statistical mechanics are used to investigate such protein models. The extracted information covers a wide range of interesting effects, from the prediction that short chains do not fold [24] to the suggestion that a significant proportion of all possible protein sequences could have a thermodynamically dominant fold [25], from the affirmation that there exists only a small set of possible folding patterns [26] to the claim that the basis for secondary structure formation in globular proteins may be packing and conformational freedom [27].

As a conclusion to this Section we would say that an understanding of the physico-chemical mechanisms underlying biopolymer folding appears as a necessary step towards the prediction of the secondary and tertiary structure of proteins. Equilibrium thermodynamics and statistical mechanics can provide part of such basic information, to be further included in more specific calculations and

computer simulations.

1.1.2 Protein Adsorption

Strictly speaking adsorption refers to the accumulation of a substance on an impenetrable wall. For a single protein it can be understood as the attachment of a substantial fraction of its residues on the surface of a larger structure. With this proviso, adsorption could seem to play a limited physiological rôle, since proteins most often interact with molecules similar in size or smaller. Nevertheless, the case of cell membranes as the adsorbing substrate is worth being more thoroughly investigated. Membranes contain between 20 and 80 percent protein which are the biochemically active components. The remaining portion is made up by a wide variety of lipids which general structure presents a polar, hydrophilic head and a long non-polar, hydrophobic, tail.

Certainly the major rôle of membrane lipids is to form the bilayer matrix in which the proteins sit. The membrane proteins act as enzymes, transporters, receptors, pores, etc. They are generally viewed as being folded so as to present a non-polar hydrophobic surface compatible with the non-polar portion of the lipid molecules. Polar or charged regions of the protein can adjust to the lipid headgroups at the surface of the bilayer. Many proteins extend through the membrane, others are probably bound exclusively through interactions with membrane-embedded proteins. Figure 1.2 shows the different mechanisms by which proteins are believed to interact with the lipid bilayer membrane [28]. The cases B and D can be readily considered as adsorption phenomena. In the following, these two types of surface interactions will be referred to as *protein adsorption.* It may be used primarily to supplement the membrane-binding properties of proteins which are attached to the membrane by other means, such as through transmembrane "anchors".

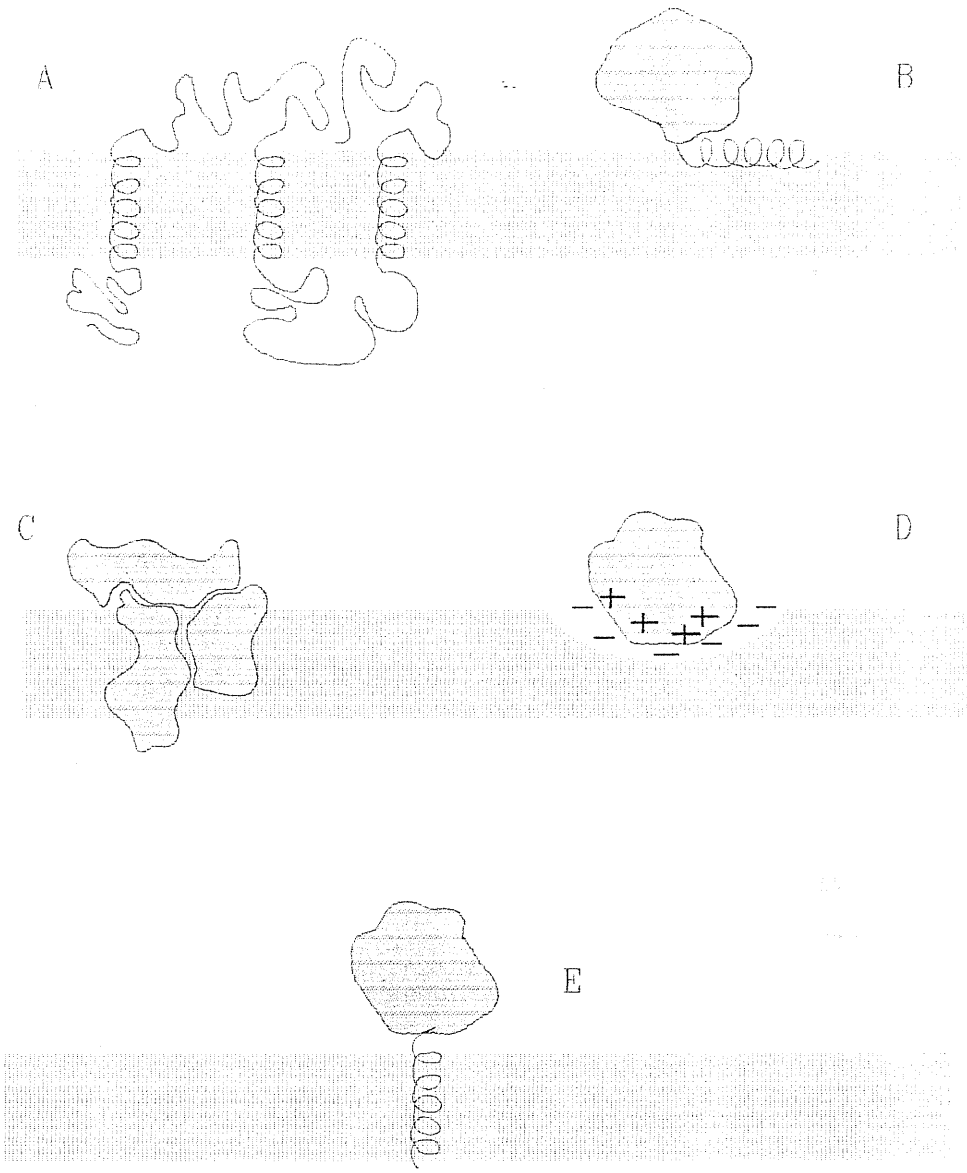

Figure 1.2 A: Polypeptide chain with multiple transmembrane segments (e.g. glycophorin and lactose permease). B: Hydrophobic binding but without penetration (e.g. E. pyruvate oxydase). C: Protein bound to other proteins (e.g. succinate dehydrogenase). D: Electrostatic binding (e.g. myelin basic protein). E: Binding by a terminal segment (e.g. cytochrome b5).

Protein adsorption is certainly involved in a large number of poorly understood physiological processes and we think that the results anticipated in the Introduction, namely the coupling between the folding and the adsorption transitions, should contribute to clarify such mechanisms. Protein translocation and insertion, the processes by which a protein can traverse or can be inserted into a membrane, appear to be possible examples. The post-translational *translocation* is now accepted to occur not only in chloroplast and mitochondrial membranes but also across the bacterial membrane and the endoplasmic reticulum of eukaryotes [29–32]. This ribosome-independent transport phenomenon appears to involve a preliminary adsorption of the protein to the membrane. Moreover important conformational changes are expected to concern the adsorbed protein [28] before translocation. Spontaneous post-translational *insertion* of proteins into the membrane have also been observed. As for translocation, adsorption and conformational changes again appear to be important steps [33].

Structural information about membrane protein is difficult to obtain. Once removed from their native lipid environment the proteins tend to promote non-ordered aggregations unsuitable for X-ray diffraction studies. Moreover the detergents, which are required for the purification step, usually interact with membrane proteins and alter their conformation, as well as their functions, in a way which is hard to control. The determination of conformational changes during a given physiological process is still more difficult. Often the more or less globular conformation of a protein is roughly estimated from the rate of proteolitic degradation during the different steps of the process investigated.

As a conclusion we can say that adsorption is certainly worthy of being carefully considered, especially when related to conformational change since both phenomena seem to occur in a somewhat related way in important physiological

processes. Moreover, as it will be discussed in Chapter 5, protein-protein interactions and protein adsorption present some common features so that our main results on adsorption can be applied to both interactions.

1.2 THE STATISTICAL MECHANICS APPROACH

In this Section we will ignore most of the protein properties that have been described in Section 1.1. Only some physically relevant features will be retained. First of all we must note the relative *large size* of a single protein (typically 200 peptide units), which can therefore be regarded as a macroscopic system. Its description in terms of *classical* statistical mechanics seems thus appropriate. Secondly we will consider proteins as *flexible chains* made up of several hundred linked monomers. This appears as a natural modelisation for unbranched linear macromolecules. A crucial point in modeling proteins is represented by the diversity of the 20 amino acids which, as already discussed, is responsible for their nonperiodic spatial structure and for their highly specific interactions. The procedure adopted to tackle this delicate problem consists of two steps. We will first ignore the differences between amino acids, thus working with polymeric *homogenous* chains. Then we will improve on the model in order to include a random distribution of hydrophobic and hydrophilic monomers to form what will be referred to as the *random chain.* This means that instead of considering a set of 20 amino acids differing on many aspects we will limit ourselves to a "black and white" description of proteins. Future elaborations should consider the introduction of a specific sequence of hydrophobic and hydrophilic monomers (instead of a random one) as well as the use of more parameters for the characterization of monomers. The random chain will be treated in Chapter 4, whilst in the following we will focus on homogeneous chains (homopolymers).

A long chain, made up of several thousand linked monomers, can be characterized by its microscopic states, defined through the mutual positions of the monomer-monomer bonds within the molecule, and by a number of accessible parameters such as its size, usually represented by the *end-to-end distance* or by the *radius of gyration*, its shape, defined by the spatial distribution of the monomers, and many other physical, optical, and chemical properties. Under a progressive change of environmental conditions these parameters can either vary smoothly or undergo a sharp modification which reveals a cooperative process referred to as a *phase transition.*

Clearly, the drastic simplifications introduced in our models will prevent us from obtaining any useful information on the specific local behaviour of proteins. We therefore focus our attention on global phase transitions since it seems very likely that such highly cooperative processes will overcome local modification and deeply influence the general and specific behaviour of proteins. The location of phase transitions, that is the determination of the phase diagram, represents the heart of the present thesis. The biological relevance of the results thus obtained will be discussed in Chapter 5, whereas in the rest of this Chapter we will introduce the concepts which form the basis of our treatment in Chapters 2, 3, and 4.

1.2.1 Basic Concepts in Polymer Physics

The polymeric chains present a particular feature that increases the difficulties encountered in their theoretical analysis. Owing to chain flexibility, each monomer can interact with any other, even if they are located very far along the monomer sequence. Depending on whether these interactions are neglected or not, the physical models are said to deal with *ideal* or *real* chains, respectively. Polymer models also divide into *continuous* and *lattice* models. The former allow for the

polymeric chain to have access at any point in a continuous space, whereas the latter constrain the chain monomers on a regular lattice of points.

A. *The Ideal Chain.*

Most of the calculations on the ideal chain can be performed exactly. This is directly connected with the fact that the monomer distribution is analytically accessible, as an ideal chain is well represented by a *random walk.* The probability for a random walk to be at a distance R from its origin after n steps is known to have, for $n \longrightarrow \infty$, a gaussian distribution [34] which can be written as:

$$P_n(R) = R^{d-1} \left(\frac{d}{2\pi n} \right)^{\frac{d}{2}} \exp\left(-\frac{d}{2} \frac{R^2}{n a^2} \right) \tag{1.3}$$

where d is the dimensionality of space and a is the mean length of the random steps. Using the above distribution we can easily determine the asymptotic dependence of the radius of gyration on the number n of monomers. By definition, the radius of gyration R_G is given by:

$$R_G \equiv \langle (R - \langle R \rangle)^2 \rangle^{\frac{1}{2}} = \left\{ \int_0^{+\infty} R^2 P_n(R) dR \right\}^{\frac{1}{2}} \sim n^{\nu} \tag{1.4}$$

where spatial isotropy has been assumed so that $\langle R \rangle = 0$. Direct integration of equation (1.4) shows that the size exponent is $\nu = \frac{1}{2}$ for any value of the space dimensionality $d \geq 1$.

B. *The Real Chain.*

A particular case of monomer-monomer interaction is represented by the *excluded volume* effect which accounts for the repulsive potential between

monomers on close approach (steric repulsion). When attractive forces between monomers are present, excluded volume interactions must be taken into account to avoid an unphysical collapse of the chain at a single point. Moreover excluded volume interactions have important effects on the size exponent ν of the macromolecule even when no other interaction between monomers is present. The excluded volume implies that less space is available to the molecule so that it spreads over a larger volume and the radius of gyration is increased. *Self-avoiding walks* (SAWs) are natural representations of polymeric chains when excluded volume effects are considered. Depending on whether other monomer-monomer interactions are present or not, we will talk about interacting or noninteracting SAW models. In order to tackle the conceptual and technical problems raised by the theoretical study of real chains, two techniques have been mostly used, namely the *mean field* and the *scaling* approaches.

C. *The Mean Field Approach.*

The *mean field theory* represents a basic tool of statistical mechanics. Essentially, it consists in the application of a variational principle. The hamiltonian H of the system to be described is written in the form

$$H(\lambda) = H_0 + \lambda H_1 \tag{1.5}$$

where H_0 is the hamiltonian of a solvable model which can be used as an approximation for the system at hand. The term H_1 represents the difference between the effective and approximate models. By permitting λ to vary from 0 to 1 we can smoothly interpolate between the solvable model system $H(0) = H_0$ and the system of interest $H = H(1)$. Then the Bogoliubov inequality holds for the Helmoltz potential F

$$F \leq F_0 + \langle H_1 \rangle_0 \tag{1.6}$$

where $\langle H_1 \rangle_0$ is the average value of the perturbation as calculated in the unperturbed system. The above inequality establishes an upper bound for F. Therefore, to find the optimum unperturbed model one has to minimize $F_0 + \langle H_1 \rangle_0$ over a parametrized family of solvable models [35]. This type of procedure has been widely used to study systems of interacting particles exploiting the known solution of the related non-interacting particle models. The particle-particle interactions are then replaced by a fictitious external field the intensity of which has to be chosen in such a way as to minimize the right hand side of the Bogoliubov inequality. In polymer physics, the ideal chain constitutes a solvable model to be used as an approximation for the real chain, whereas excluded volume and other monomer-monomer interactions are considered as perturbations. The mean field approach has been very fruitfully exploited by Flory as we will see in Sections 1.2.2A and 1.2.3A.

D. *The Scaling Approach.*

The theory of phase transitions is very much concerned with the properties of macroscopic systems near their *critical point.* Many unexpected properties have been observed in that region so that criticality is now considered as a particular state of matter. What essentially characterises critical systems is the emergence of divergent thermal fluctuations [36]. Correlation lengths also diverge and, therefore, the critical behaviour does not reflect any more the full atomic complexity of the system. As it happens, a critical system is fully described by a set of *critical exponents* which govern the divergent behaviour of the susceptibilities and the growth of the order parameter. The values of the critical exponents depend only on the dimensionality of the system and on that of the order parameter,

as well as on the symmetries of the hamiltonian. Consequently, the values of the critical exponents are not completely independent. This inter-relation is most economically stated through some postulated *scaling relations* from which it is possible to derive the complete critical behaviour of the properties of the system. The critical exponents are said to be *universal*; since all the systems with the same dimensionalities of order parameter and of space and same hamiltonian symmetry must have identical critical exponents, such systems are considered as belonging to the same *universality class*. The effective use of scaling relations can be summarized as follows. Let a system be characterized by a physical quantity A depending on, say, two variables a and b : $A = F(a, b)$. Suppose that we can rescale the system by a factor λ and that we know the transformation laws for the above variables:

$$a \longrightarrow \lambda^{\alpha} a \quad \text{and} \quad b \longrightarrow \lambda^{\beta} b \tag{1.7}$$

In general, under transformation (1.7), the quantity A will change as

$$A \longrightarrow \lambda^{x} A \tag{1.8}$$

The parameter x depends on the nature of A and can be inferred via physical arguments. It follows from equations (1.7) and (1.8) that

$$F(\lambda^{\alpha} a, \lambda^{\beta} b) = \lambda^{x} F(a, b) \tag{1.9}$$

Equation (1.9) must hold for any λ. It represents a generic scaling relation from which, in particular cases, the form of $F(a,b)$ can be deduced. Scaling relations are not easily determined. The only available physical theory allowing to deduce or prove scaling relations is the *renormalisation group theory*. In the field of polymer physics, considerable progress has been achieved in the scaling approach to the real chain problem when the SAW model has been recognised to be equivalent to the zero-component $O(n)$ model of magnetic systems, as discussed below.

E. *Polymer Statistics and Magnetic Critical Points.*

From a microscopic point of view, ferromagnets can be seen as a collection of particles the dipolar moments of which, also called "classical spins", can have different orientations. In the absence of an external magnetic field and at high temperatures the spin orientations are uncorrelated and no magnetisation would be measured. At the Curie temperature T_c the spin-correlation length will diverge giving rise to a non-zero magnetisation. For temperatures slightly above T_c, that is for small positive values of $\varepsilon = T - T_c$, the correlation length ξ obeys a power-law of the form

$$\xi \cong a|\varepsilon|^{-\nu} \qquad (\varepsilon \longrightarrow 0) \tag{1.10}$$

where a represents the distance between neighbouring atoms. Owing to the fact that $\xi \gg a$ all lattice details become irrelevant and the exponent ν presents a universal character. The only relevant quantities are the dimensionality d of space and that of the order parameter n (i.e. the number of components of the magnetization vector). Therefore, the Curie temperature T_c corresponds to a magnetic critical point.

A first striking analogy between polymer statistics and magnetic systems appears when comparing the relation (1.10) with the scaling form for the growth of the polymer's characteristic size (represented for instance by the radius of gyration R_G)

$$R_G \cong a n^{\nu} \qquad (n \longrightarrow \infty) \tag{1.11}$$

where a is the monomer length. The correspondence

$$\varepsilon \longrightarrow 0 \Longleftrightarrow n \longrightarrow \infty \tag{1.12}$$

denotes the critical behaviour of a polymeric chain in the limit $n \longrightarrow \infty$. The precise form of this correspondence can be expressed through the following theorem :

$$\langle S_i^{\alpha} S_j^{\alpha} \rangle \big|_{n=0} = \sum_n C_n(ij) \left(\frac{K}{T} \right)^n \tag{1.13}$$

where the left hand side represents the correlation between the components α of the spins located at sites i and j, and where the formal limit to zero spin-components ($n = 0$) has been taken. On the right hand side, $C_n(ij)$ is the number of SAWs of n steps linking sites i and j whereas K is the coupling constant for neighbouring spins.

The demonstration of theorem (1.13) was first given by de Gennes [37] and reproposed several times after in different forms [38–41]. Other correspondences

between the critical behaviour of magnets and polymers are easily derived from equation (1.13) (see ref.[42]).

The direct consequences of theorem (1.13) for polymer physics are of two types. On the one hand, many results obtained for magnetic systems can be "translated" and applied to polymers. On the other hand, the study of the limit $n \longrightarrow \infty$ is considerably simplified since this limit can be replaced by the requirement that the system exhibits critical behaviour. Such a requirement receives a precise formulation in the renormalisation group theory as explained below.

F. *Real Space Renormalisation.*

As already discussed, the comparison of the effect of scaling on different physical quantities can give important information on the relative dependence of these quantities. Nevertheless, the above effect of a rescaling transformation is hard to control. The main idea of *renormalisation group theory* consists of repeating a scaling transformation an infinite number of times until all the parameters characterizing the system will become scale-invariant. In other words, the scaling transformation must be repeated until a *fixed point* in the parameter space is reached. All the parameters with zero invariant value are said to be *irrelevant.* If all the parameters are irrelevant the fixed point is called "trivial". Non-trivial fixed points must correspond to systems where the characteristic length diverges. Therefore, they can be associated with the critical points of the system. Let μ be a vector representing all the parameters we need in order to describe the physical system and R_b the transformation under which the system is rescaled by a factor b. We then express the *renormalisation group transformation* by

$$\mu' = R_b(\mu) \tag{1.14}$$

and the fixed point condition by

$$\mu^* = R_b(\mu^*) \tag{1.15}$$

Assuming the existence of the fixed point μ^* we can linearize the transformation (1.14) near this point by:

$$\mu' = \mu^* + J_{R_b}(\mu^*)(\mu - \mu^*) \tag{1.16}$$

where $J_{R_b}(\mu^*)$ is the jacobian matrix of R_b evaluated at μ^*. If we first scale our system by a factor b, and then by a factor b', the two following properties must hold:

$$\begin{aligned} R_b \circ R_{b'} &= R_{b \cdot b'} \qquad \text{Semi-group} \\ [R_b, R_{b'}] &= 0 \qquad \text{Commutativity} \end{aligned} \tag{1.17}$$

Using the properties (1.17) we can easily show that the eigenvalues of the jacobian matrix J_{R_b} *must* have the form $\lambda_i(b) = b^{y_i}$, where the y_i's characterize the critical behaviour of the corresponding relevant parameters, and are usually referred to as the *critical exponents*. Summarizing, we would say that the renormalisation group theory allows one to *calculate* the critical exponents through the fundamental formula:

$$y_i = \frac{\ln \lambda_i}{\ln b} \tag{1.18}$$

1.2.2 The Polymer Chain in a Good Solvent

The concept of *good solvent* for polymers is used to describe a situation in which the *effective* monomer-monomer interactions are null or repulsive. Actually, solvent-solvent, monomer-solvent, and monomer-monomer interactions can be reduced to an overall effective interaction between monomers. We must nevertheless emphasize that passing from a good to a bad solvent does not necessarily imply changing the solvent: temperature changes can alter the relative strength of the interaction components so as to produce a change of sign in the effective interaction strength.

A. *The Flory Approximation.*

A single polymer chain in a good solvent is well modelled by a *noninteracting* SAW . In such a way the excluded volume effect is considered and, owing to their universality, critical exponents can be carefully estimated in the limit $n \longrightarrow \infty$. The size exponent ν is certainly the most interesting and most accessible to experiments. A surprisingly accurate estimate of ν has been obtained by Flory [43] using a mean field-inspired approach. First of all, Flory assumes that the growth of a single polymer chain obeys an asymptotic law of the form

$$R \propto n^{\nu} \tag{1.19}$$

where R is a characteristic length of the molecule such as, for instance, the end-to-end distance or the radius of gyration defined in equation (1.4), and n is the number of monomers. The Helmoltz free energy $F(R) = U(R) - TS(R)$ is

then calculated and minimized with respect to R. The repulsive energy, $U(R)$, due to the excluded volume interactions, is supposed to arise exclusively from two-body contacts between monomers. Neglecting the spatial correlation between monomers, we have $U(R) \propto R^d \rho^2$ with $\rho \propto \frac{n}{R^d}$ the monomer concentration and d the dimensionality of space. The entropy $S(R)$ is estimated from the distribution of R for an ideal chain (1.3) and, neglecting the logarithmic term, is proportional to $\frac{R^2}{n}$. Minimizing $F(R)$ yields the well-known Flory estimate for the size exponent ν

$$R \propto n^{\frac{3}{2+d}} \iff \nu = \frac{3}{2+d} \tag{1.20}$$

which gives the exact values for $d = 1, 2, 4,$ and a very accurate result for $d = 3$. For $d > 4$ the above result cannot be considered since it leads to $\nu(\text{real chain}) < \nu(\text{ideal chain}),$ which is inconceivable without attractive interactions. The space dimensionality $d = 4$ thus appears as the *upper critical dimension* at, and above which, the excluded volume interaction does not influence the asymptotic behaviour of a polymer chain. For a discussion of the amazing accuracy of the method despite the drastic approximations used, see refs. [42] and [44]. It is worth noting that Flory's simple approach, briefly sketched above, is not a real mean field theory in that no mean field is calculated. Actually it is a "local" theory which neglects, like the mean field approach, the local fluctuations and correlations. It should also be noted that, unlike the size exponent ν, other quantities, such as for example the free energy, turn out to be inadequately predicted [45].

B. *The RSRG and the Non-Interacting SAWs.*

In order to apply the real space renormalisation group (RSRG) technique,

the underlying principles of which have been presented in Section 1.2.1, we need a more detailed chain model. Since in the critical region the microscopic details of the chain are irrelevant, the use of lattice models appears to be very convenient. Therefore we first consider, for simplicity, a square two-dimensional lattice on which we "draw" a SAW with a fixed origin. Our purpose is to study the statistics of such a system in the limit of an infinite number of walk steps. The Grand Canonical partition function can be written in the form

$$Z_{SAW}(k) = \sum_{n=1}^{\infty} C_n \cdot k^n \tag{1.21}$$

where C_n is the number of SAW configurations with n steps, and $k = e^{\mu}$ is the weight associated with each step, also called the *step fugacity*. The generating function (1.21) can be shown to become singular when the fugacity reaches a critical value $k_c = \frac{1}{\mu_c}$. The quantity μ_c is the effective *connectivity* of the lattice and corresponds to the mean number of different possible orientations one can choose for the nth step (in the limit $n \longrightarrow \infty$) without violating the self-avoiding requirement. The scale invariance of $Z_{SAW}(k)$ can therefore be used to study criticality. Now let $Z'_{SAW}(k')$ be the generating function for the rescaled system. By requiring

$$Z_{SAW}(k) = Z'_{SAW}(k') \tag{1.22}$$

we will define implicitly a renormalisation transformation

$$k' = R_b(k) \tag{1.23}$$

from which we can proceed as already explained to calculate the critical exponent ν (defined through $R_G \sim |k - k_c|^{-\nu}$ for $n \longrightarrow \infty$).

The equation (1.22) involves infinite sums which must be truncated for practical calculations. Such a truncation can be performed in many different ways and always requires some arbitrary rule to be adopted. The cell-to-cell renormalisation scheme is certainly the most widely used. To implement this scheme we have to define (e.g. in $d = 2$) a finite $n \times m$-lattice cell. Then we exactly enumerate all the SAWs which span the cell in order to determine the coefficients C_n in equation (1.21). The cell is then renormalized by a factor b to obtain a $\frac{n}{b} \times \frac{m}{b}$-cell which spanning SAWs are generated as well. The invariance condition (1.22) provides the approximate renormalisation transformation.

Studies of SAW critical properties by means of real space renormalisation have often been reported [46–49]. Particularly accurate results have been obtained in ref. [50] where the $d = 2$ value $\nu = 0.7503$ has been obtained through the powerful transfer matrix method, often used to calculate correlation lengths in two-dimensional spin systems.

The cell-spanning requirement for SAWs represents an ad hoc "connectivity rule" introduced to obtain a sufficiently good correspondence between bare and renormalized walks. On square lattices the corner rule (see ref.[51]) gives to the spanning condition a more precise form. In Chapter 2 we will define a suitable rule for the SAWs on a trianglar lattice. The use of the renormalisation method with a continuous model for polymer chains in a good solvent has the advantage of permitting an analytic treatment based on field-theoretic arguments (see for instance the paper by des Cloiseaux [52]). Nervertheless, it is difficult to tackle analytically, using continuous models, the problem of the simultaneous polymer collapse adsorption on an impenetrable wall. Therefore, the mathematical

treatment being much simpler for lattice models, we will limit ourselves in this thesis to the use and description of the latter.

1.2.3 Polymer Collapse

When switching from a "good" to a "poor" solvent, either by changing the solvent's nature or temperature, the effective monomer-monomer interactions become attractive. The monomers will tend to aggregate and a single chain will adopt a dense, roughly spherical conformation, usually referred to as a *globule* . The process will be counteracted by the conformational entropy of the chain, which favours the unfolded (*coil*) configurations. Thus infinite chains present two distinct regimes. If conformational entropic effects dominate the energetic preference for globules, then the chain will be in a coil state, otherwise it will be found in a folded, globular, state. The transition point is usually referred to as the Θ-*point* with the corresponding Θ-temperature: T_Θ. This represents an oversimplified description of protein folding.

A. *The de Gennes-Flory Theory.*

In Section 1.2.2A we have presented the Flory approximation for the size exponent of a real chain in a good solvent, which turns out to give surprisingly accurate results. Furthermore this simple method is capable of showing the special rôle of the space dimensionality $d = 4$. This interesting aspect justifies the attempt to generalize it to the case of a single polymer chain in a poor solvent. The entropy is calculated from the distribution (1.3) of the size for a gaussian chain, but, unlike in Section 1.2.2, the logarithmic term must not be neglected. We obtain for a n -monomer chain

$$S_n(R) = (1-d)\ln R + \frac{dR^2}{2na^2} \tag{1.24}$$

The chain energy is calculated using a virial expansion

$$U(R,T) = n(\rho U_1(T) + \rho^2 U_2(T) + \rho^3 U_3(T) + \ldots) \tag{1.25}$$

where $\rho = \frac{n}{R^d}$ is the monomer density. Minimizing the free energy with respect to R one can show that for $d = 3$ all the virial coefficients, except U_1, become irrelevant in the limit $n \longrightarrow \infty$. Conversely, in two dimensions all the virial coefficients are relevant. For good solvents the theory predicts the Flory exponents. For bad solvents it correctly predicts $\nu = \frac{1}{d}$. The case of a Θ-point solvent is obtained for $U_1 \longrightarrow 0$, which corresponds to a compensation of the excluded volume repulsive interactions by the attractive effective forces between monomers. In three dimensions this appears to be a good choice since the theory then predicts $\nu_\Theta = \frac{1}{2}$ and only logarithmic corrections ($\propto \ln n$) are expected to hold. The situation is less favourable in two dimensions, where the value $\nu_\Theta = \frac{2}{3}$ is not well supported by experimental ($\nu_\Theta = 0.56$ [53]) and numerical methods ($\nu_\Theta = 0.55$ [54], $\nu_\Theta = 0.535$ [84], $\nu_\Theta = 0.567$ [55]). Recently-developed conformal invariance methods [56], and a conjectured correspondence between the statistics of polymer rings at the Θ-point and that of the hull of a percolation cluster at threshold [57], yields the value $\nu = \frac{4}{7}$. This value is also well supported by the particularly accurate Monte Carlo studies of reference [58].

The failure, in two dimensions, of the above mean-field approach by de Gennes and Flory, must be ascribed to the mentioned relevance of all the virial coefficients, which precludes the use of variational principles. The almost ideal-chain behaviour

of Θ-point polymers in three dimensions shows that the upper critical dimension, which was $d = 4$ for SAWs, becomes $d = 3$ at the Θ-point . The logarithmic corrections are due to three-body interactions [59].

B. *Scaling Approach to the Θ-point .*

The polymer Θ-point has been recognized to be a *tricritical point* [60]. The term "tricritical" has been introduced by Griffiths [61] to describe the end point of a line of three-phase coexistence. When dealing with magnetic systems this definition has a straightforward physical meaning [62]. In the Θ-point polymer problem it can be difficult to identify a three-phase coexistence line. Nevertheless, tricritical points can be regarded in a slightly different, but equivalent way [63]. Let us consider a system exhibiting a simple phase transition at a given temperature T_c. Introducing one new parameter influencing the system, we can enlarge the dimensionality of the phase diagram and turn the critical point into a transition line. As we move along this transition line, we may reach special points at which the properties of the transition change abruptly. Such points represent tricritical points. This directly applies to polymer statistics if we consider the limit $n \longrightarrow \infty$ as a second-order phase transition through which a chain acquires a more symmetric state. Formally, this corresponds to $k \longrightarrow k_c$ in equation (1.21). The monomer-monomer interaction parameter $\omega = -\frac{\varepsilon}{k_B T}$ is the variable giving rise to a transition line $k_c(\omega)$ in the phase diagram. To each point on the transition line one can associate a size exponent which abruptly changes for $\omega(T_\Theta) = -\frac{\varepsilon}{k_B T_\Theta}$. Indeed we have, on the transition line (or, equivalently, in the limit $n \longrightarrow \infty$), the following asymptotic behaviour for the characteristic size R of a polymer:

$$\langle R^2 \rangle \sim n^{2\nu} \tag{1.26}$$

where

$$\nu = \begin{cases} \nu_{SAW}, & \text{if } -\infty < \omega < \omega(T_\Theta) \\ \nu_\Theta, & \text{if } \omega = \omega(T_\Theta) \\ \frac{1}{d}, & \text{if } \omega > \omega(T_\Theta) \end{cases} \tag{1.27}$$

de Gennes' argument [60] for the tricritical nature of the Θ-point is somewhat indirect. The evidence is based on the observation that at the Θ-point the third virial coefficient of the monomer interaction cannot be ignored. This has the theoretical implication that in a Landau free energy expansion near the Θ-point,

$$F = F_0 + \frac{1}{2}F_2M^2 + \frac{1}{4}F_4M^4 + \frac{1}{6}F_6M^6 + \ldots \tag{1.28}$$

where M is the order parameter, the terms of sixth order have to be retained because the coefficient of the sixth-order term is proportional to the third virial coefficient [60]. It is this form of the free energy that is capable of yielding a tricritical point [63].

Belonging to the large family of tricritical points, the polymer Θ-point must obey some typical scaling relations. Let us focus on the Gibbs free energy which, near criticality, must scale as [64]:

$$G(b^{d\lambda_1}\tau, b^{d\lambda_2}H) = b^dG(\tau, H) \tag{1.29}$$

where b is the rescaling factor, $\tau = \frac{T-T_c}{T_c}$ is the reduced temperature, and H is an ordering field. Assuming for the polymer size R a similar scaling law and making use of the correspondence (1.12), we can write :

$$R(\frac{1}{n}, \Delta\omega) = bR(\frac{1}{n}b^{\lambda_1}, \Delta\omega b^{\lambda_2}) \tag{1.30}$$

where n is the polymerisation index and $\Delta\omega = \omega - \omega(T_\Theta)$. Equation (1.30) must be satisfied for all values of b. Thus let us choose it in such a way that $b^{\lambda_1} = n$. Equation (1.30) becomes

$$R(n^{-1}, \Delta\omega) = n^{\nu_\Theta} f(n^{\phi}\Delta\omega) \tag{1.31}$$

with $\nu_\Theta = \lambda_1^{-1}$ and $\phi = \frac{\lambda_2}{\lambda_1}$. Exactly at the tricritical point $\Delta\omega \longrightarrow 0$ and we obtain $R \sim n^{\nu_\Theta}$, which corroborates our affirmations (1.26) and (1.27). The *crossover exponent* ϕ characterizes the smooth transition between two different universality classes.

C. *Real Space Renormalisation Approach to the* Θ*-point*

In order to apply the real space renormalisation technique to the problem of polymer collapse, it is necessary to introduce suitable lattice models allowing to include monomer-monomer interactions. The good models are not very numerous and we are going to present two of them which will be developed in the next Chapters. The most widely used is certainly the *standard* model where *interacting self avoiding walks* (ISAW) are considered. As shown in figure 1.3 for the particular case of a two-dimensional square lattice, a monomer interacts only with its nearest neighbours.

An attractive energy ε is associated with each interaction and the partition function is given by:

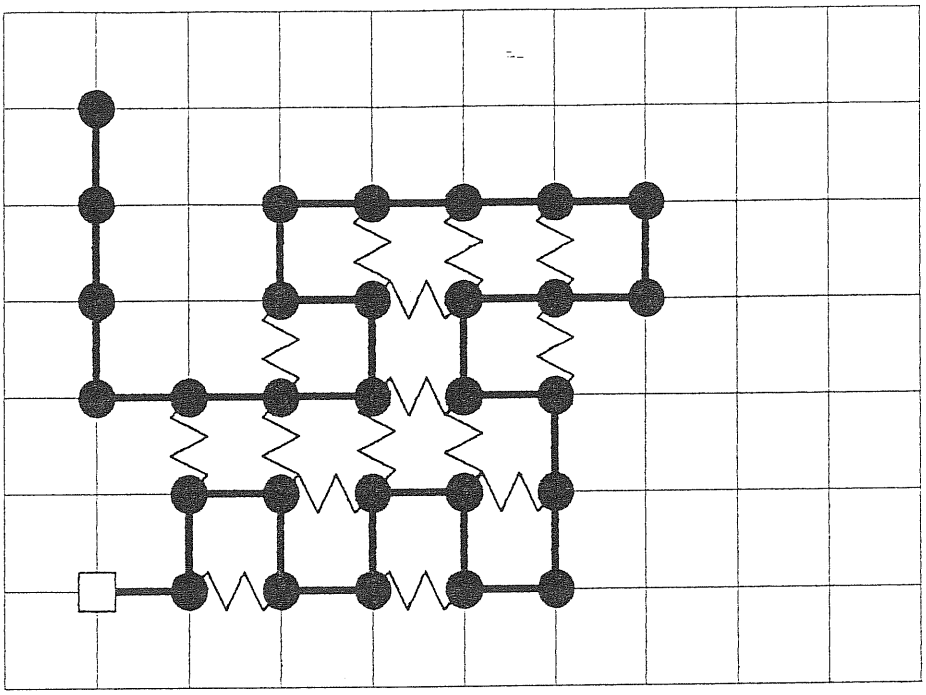

Figure 1.3 An interacting self avoiding walk on a two-dimensional square lattice. Wiggly lines show the nearest neighbour interactions.

$$Z = \sum_{i=0}^{i_{max}} \sum_{\{W(n,i)\}} e^{i\omega} \tag{1.32}$$

where $\omega = -\frac{\varepsilon}{k_B T}$ and $\{W(n,i)\}$ represents the ensemble of the SAWs with n steps and i nearest-neighbour interactions. A cell-to-cell renormalisation scheme has been proposed in ref. [65]. The size exponent has been found to take the following values: $\nu_{SAW} = 0.69$ and 0.69, and $\nu_\Theta = 0.62$ and 0.60, in two and in three dimensions, respectively. Considering that the expected values are $\nu_{SAW} = 0.75$ and 0.59, and $\nu_\Theta = 0.57$ and 0.50 we immediately see that the method is not very precise. However the qualitative agreement with the accepted values can be considered a success considering the simplicity of the

method and the complexity of the problem.

The second model we present has been introduced in ref. [66] and requires a lattice with a connectivity of six or more. A two-dimensional triangular lattice may therefore be used. Two typical configurations are shown in figure 1.4. It can be seen that the walk is allowed to return once but *only once* on an occupied site. This requirement prevents the chain to collapse to unphysically high densities, since it constitutes a geometrical way of taking into account three- and higher-body repulsions which are believed to be always present when collapse takes place.

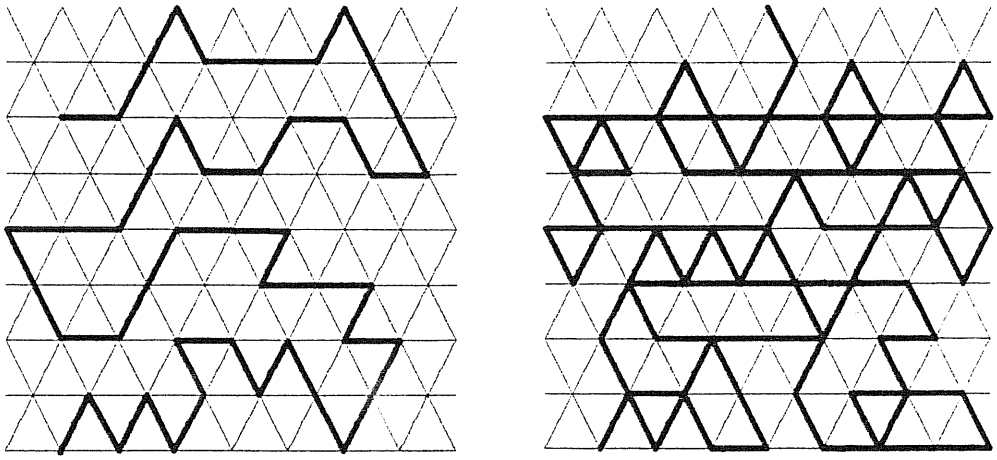

Figure 1.4 Two typical chain configurations: A coil or self avoiding walk (a), and a globule (b).

Let us call this type of walks *self-avoiding-self-attracting walks*(SASAW). To solve the model we can introduce a cell-to-cell renormalistion scheme as in figure 1.5. The essential feature is the introduction of a probability factor f as a weight for each monomer-monomer contact.

The f-factor can represent temperature changes in the solvent via the plausible

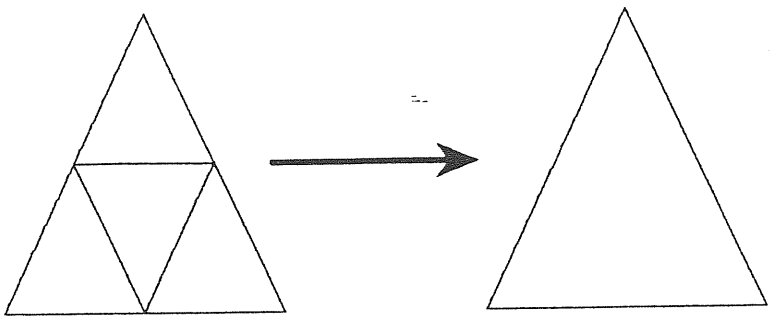

Figure 1.5 Bare and renormalized cells for implementing a real space renormalisation scheme with the SASAW model

functional dependence $f = 1 - \exp(\frac{\varepsilon}{k_B T}) = 1 - \omega^{-1}$, $\varepsilon < 0$ representing the effective solvent-mediated attraction energy. Since non-interacting walks must be weighted by a factor $(1 - f)$ for each contact avoided, the cell generating function is given by

$$Z_\Theta(k, f) = (1 - f)^{n_{max}} Z_0(k) + \sum_{m=1}^{n_{max}} f^m Z_m(k) \tag{1.33}$$

where Z_0 and Z_m enumerate the walks with no (m=0 or SAW) or m contacts in the cell, respectively. They are given by

$$Z_m = \sum_{Spanning Walks} c_m(n) k^n, \tag{1.34}$$

n being the number of random walk steps and $c_m(n)$ the number of walks with n steps and having m contacts which can be constructed in a lattice cell. The spanning condition requires that a walk starting at the origin simply touches one of the two opposite cell corners in a chosen direction. No conditions

are imposed on the position of the ending monomer. n_{max} in equation (1.33) is the maximum number of conctact sites available in the chosen lattice cell.

For the renormalized-cell partition function a modified construction has been used

$$Z'_{\Theta} = \sum_{m=0}^{n'_{max}} f'^{m} Z'_{m} \tag{1.35}$$

since we must allow for the possibility of globule-like walks ($m > 0$) to be renormalized into coil-like ($m = 0$) walks when $f = f' = 1$. As explained in Section 1.2.1 a recursion relation can be obtained by equating the bare and renormalized-cell partition functions

$$Z'_{\Theta}(k', f') = Z_{\Theta}(k, f) \tag{1.36}$$

However, the model has two parameters which must be be renormalized. A second recursion relation is therefore required. For this purpose, the renormalized contact probability is defined in terms of the fraction of weighted walks containing at least one contact in the bare cell, giving rise to the following second recursion relation

$$f' = \frac{\sum_{m=1}^{n_{max}} f^m (1-f)^{n_{max}-m} Z_m}{(1-f)^{n_{max}} Z_0 + \sum_{m=1}^{n_{max}} f^m (1-f)^{n_{max}-m} Z_m} \tag{1.37}$$

The recursion relations implicitly defined by equations (1.36) and (1.37) present three fixed points for $f^* = 0$, $f^* = 1$ and $f^* = f_{\Theta} = 0.66$. It is natural

to associate these fixed points and the relative exponents with the coil, globule, and Θ phases, respectively. The corresponding size exponents (from ref. [7]) are $\nu_{SAW} = 0.78$, $\nu_{Globule} = 0.56$, and $\nu_{\Theta} = 0.54$. As for the standard model, the numerical accuracy of the method is not very good but all the physical features are qualitatively reproduced. Results for $d = 3$ will be discussed in Chapter 2.

1.2.4 Polymer Adsorption

Long molecules in solution can be adsorbed to a solid surface or to a limiting surface of the solution (e.g. water-air or water-oil interfaces) exhibiting a short range attractive potential for the macromolecule's monomers. In the following we will only consider the case of a single chain weakly adsorbed. When the binding is weak, the adsorption layer is thick and a macroscopic description, independent of the detailed structure of the polymer chain and of the interface, is possible.

The most interesting aspect of the phenomenon is the existence of a discontinuity in the adsorption behaviour at some critical temperature T_a. For $T < T_a$ the macromolecule is "condensed" on the surface. For $T > T_a$ no such condensation occurs, although the molecule is still constrained near the surface. The process is somewhat similar to the collapse transition treated in Section 1.2.3 in the sense that when a chain portion sticks to the surface it has much less orientational freedom than when it is free. At high temperature the above entropic effect dominates and the overall interaction of the chain with the surface is repulsive. At low temperature the sign reverses. At T_a the two effects balance exactly. To show the *phase transition* character of the adsorption, we will use a Schrödinger-like equation for the end-to-end distribution of a real

chain following an approach due to de Gennes [67].

A. *A Differential Equation for the Free Chain End-to-End Distribution*

Let us consider a chain configuration where all the n monomers have a prescribed position:

$$\mathbf{r}, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_{n-1}, \mathbf{r}' ; \tag{1.38}$$

Introducing a slowly-varying external potential $V(\mathbf{r})$, the statistical weight associated with the configuration (1.38) is given by

$$P(\{\mathbf{r_i}\}) = f(r_{12}) f(r_{23}) \ldots f(r_{n-1,n}) e^{-\beta V(\mathbf{r}_1)} e^{-\beta V(\mathbf{r}_2)} \ldots e^{-\beta V(\mathbf{r}_n)} \tag{1.39}$$

where $f(r_{ij})$ is a normalized factor ensuring that $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ is equal to the step length a :

$$f(r_{ij}) = \left(\frac{1}{4\pi a^2} \right) \delta(r_{ij} - a) \tag{1.40}$$

Then, the end-to-end distribution function $G(n, \mathbf{r}, \mathbf{r}', V(\mathbf{r}))$ is given by

$$G(n, \mathbf{r}, \mathbf{r}', V(\mathbf{r})) = \int d\mathbf{r}_1 \ldots d\mathbf{r}_n \delta(\mathbf{r}_1 - \mathbf{r}) \delta(\mathbf{r}_n - \mathbf{r}') P(\{\mathbf{r_i}\}) \tag{1.41}$$

An integral equation for G can be derived by adding one monomer:

$$G(n+1,\mathbf{r},\mathbf{r}',V(\mathbf{r})) = \int ds f(\mathbf{r}-\mathbf{s}) e^{-\beta V(\mathbf{r})} G(n,\mathbf{s},\mathbf{r}',V(\mathbf{r})). \tag{1.42}$$

For $n \gg 1$, G will be a smooth function of n ; hence n can be treated as a continuous variable. Expanding G and assuming slow variations on the atomic scale, we arrive at (for details see ref. [68]):

$$-\frac{\partial G(n,\mathbf{r},\mathbf{r}',V(\mathbf{r}))}{\partial n} = -\frac{a^2}{6}\nabla^2 G(n,\mathbf{r},\mathbf{r}',V(\mathbf{r})) + \frac{V(\mathbf{r})}{T} G(n,\mathbf{r},\mathbf{r}',V(\mathbf{r})). \tag{1.43}$$

In order to apply equation (1.43) to polymer adsorption we suppress the external potential $V(\mathbf{r})$ and we introduce the following boundary condition in order to take into account the surface potential:

$$\frac{1}{G}\frac{\partial G}{\partial z}\bigg|_{z=0} = -k(T). \tag{1.44}$$

where $k(T)$ is a phenomenological constant which is considered to be linear near T_a, with a negative slope, and to represent the temperature-dependent surface effect which becomes zero at T_a. Equation (1.43) establishes a direct analogy between the distribution function G and the propagator of the Schrödinger equation. Pursuing the analogy, the eigenfunction $u_p(\mathbf{r})$ of the problem defined by equations (1.43) (with V=0) and (1.44) are the solutions of

$$\begin{aligned} -\frac{a^2}{6}\nabla^2 u_p(\mathbf{r}) &= \epsilon_p u_p(\mathbf{r}) \\ \frac{1}{u_p}\frac{\mathrm{d}u_p}{\mathrm{d}z}\bigg|_{z=0} &= -k(T) \end{aligned} \tag{1.45}$$

Their structure depends on the sign of $k(T)$. If $k < 0$ ($T > T_a$), all the eigenvalues are positive and the eigenfunctions have a finite amplitude on the whole volume of the solution. If $k > 0$ ($T < T_a$), there is a class of bound states, with negative eigenvalues, localized near the surface. Moreover, for large n, the development of G in terms of the eigenfunctions

$$G(n, \mathbf{r}, \mathbf{r}') = \sum_p u_p^*(\mathbf{r}') u_p(\mathbf{r}) e^{-n\epsilon_p} \tag{1.46}$$

is dominated by the lowest bound state corresponding to the eigenvalue $\epsilon_0 = -k^2 \frac{a^2}{6}$. We have

$$G(n, \mathbf{r}, \mathbf{r}') \sim k e^{-k(z+z')} e^{n|\epsilon_0|} \tag{1.47}$$

and we see that, for $T < T_a$, if a monomer is on the surface any other monomer has a finite probability of being on the surface, even when their distance along it becomes infinite. This defines a certain type of long range order present only below T_a.

B. *Random Walks Near a Wall*

Qualitatively interesting information on the behaviour of linear macromolecules can be obtained through the statistical study of random walks in the presence of an impenetrable wall (i.e. a walk that attempts to penetrate the wall is eliminated from the statistical sum). The main problem is to calculate the probability distribution for a walker starting near the wall and ending at a distance R from its origin after n steps. The solution can be found using the fact that we know the corresponding distribution for the isotropic case (without wall). For

simplicity, let us consider a random walk in one dimension which corresponds, in two dimensions, to *directed* SAW configurations where the walk is never allowed to turn back on itself along the direction of the wall (see figure 1.6).

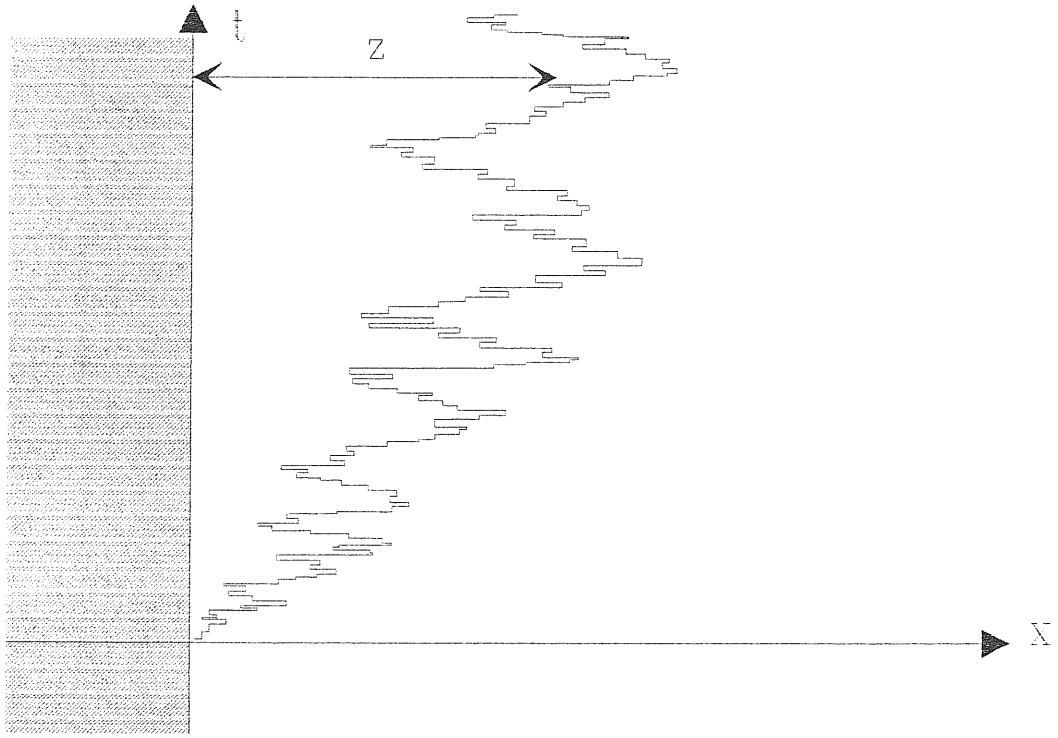

Figure 1.6 A typical random walk in one dimension (x). If the time coordinate t is considered as a spatial dimension we obtain a directed-SAW configuration.

The problem is solved by the method of images [69] which states that the wall probability distribution P_n^W is given in terms of the isotropic distribution by

$$P_n^W(x, x_0) = P_n(x, x_0) - P_n(x, -x_0) \tag{1.48}$$

where x_0 and x represent the positions at step zero and n, respectively. The surface is assumed to be at $x = 0$ and the allowed space corresponds to $x > 0$. Using the gaussian distribution (1.3) where the space dimensionality is $d = 1$ and the step length $a = 1$ as well, we obtain for $n \longrightarrow \infty$

$$P_n^W(x) = \left(\frac{\pi}{2}\right)^{-\frac{1}{2}} \frac{x_0 x}{n^{\frac{3}{2}}} e^{-\frac{x^2}{2n}} \tag{1.49}$$

The total partition function for n -step walks near a wall is thus given by:

$$Z_n^W = \int dx P_n^W(x) \sim n^{-\frac{1}{2}}. \tag{1.50}$$

The *free energy per step at the* nth step is given by:

$$\beta f_n = -\ln\left(\frac{Z_{n+1}^W}{Z_n^W}\right) = \text{Const.} + \frac{1}{2}\ln(1 + n^{-1}) \tag{1.51}$$

Since the most probable distance from the wall is given by $z = n^{\frac{1}{2}}$, the free energy f_n varies with z as

$$\beta f(z) = \text{Const.} + \frac{1}{2}z^{-2} + O(z^{-4}) \tag{1.52}$$

The constant term simply represents the reduced free energy per step in the limit $z \longrightarrow \infty$. The following terms tell us that, due to the presence of the surface, a (directed) two-dimensional SAW feels a *repulsive potential per step*

$$W(z) = \frac{1}{2} k_B T z^{-2} \tag{1.53}$$

The physical origin of this effective potential lies in the *loss of entropy* a walk experiences when approching the wall (steric repulsion). This important fact provides a basis for a possible interpretation of our finding of a surface-induced Θ-temperature enhancement, as discussed in Chapter 3. We expect a long-ranged potential of the form (1.53) to be also induced by the wall in the case of real isotropic interacting polymers.

C. *Scaling and Polymer Adsorption.*

The formal correspondence between the statistics of magnetic and polymer systems can be extended to the case of adsorption [70]. From such a correspondence many scaling relations for polymers at interfaces can be inferred on the basis of the well-known behaviour of magnetic systems [71]. Thus, the scaling analysis of numerical simulations allows one to assign numerical values to the universal critical exponents [72]. It is worth noting that computer simulations are much easier with polymers than with magnets so that exponents are better determined in the context of polymer systems. Nevertheless, as far as we know, the rigorous mapping of the polymer problem onto a corresponding magnetic one has not been performed in the case where an impenetrable attractive wall is present. The theoretical situation is thus quite different from that concerning the pure bulk problem, where theorem (1.13) establishes such a mapping rigorously.

The above-mentioned analogy is with a semi-infinite spin system with modified couplings ($J \longrightarrow J_s$) at the surface. The critical behaviour of such a system is summarized in figure 1.7. It is essentially determined by the parameter c which is connected to the relative values of J_s and J. For $c > 0$ the transition from a paramagnetic (disordered) phase (P) to a ferromagnetic (ordered) phase (F)

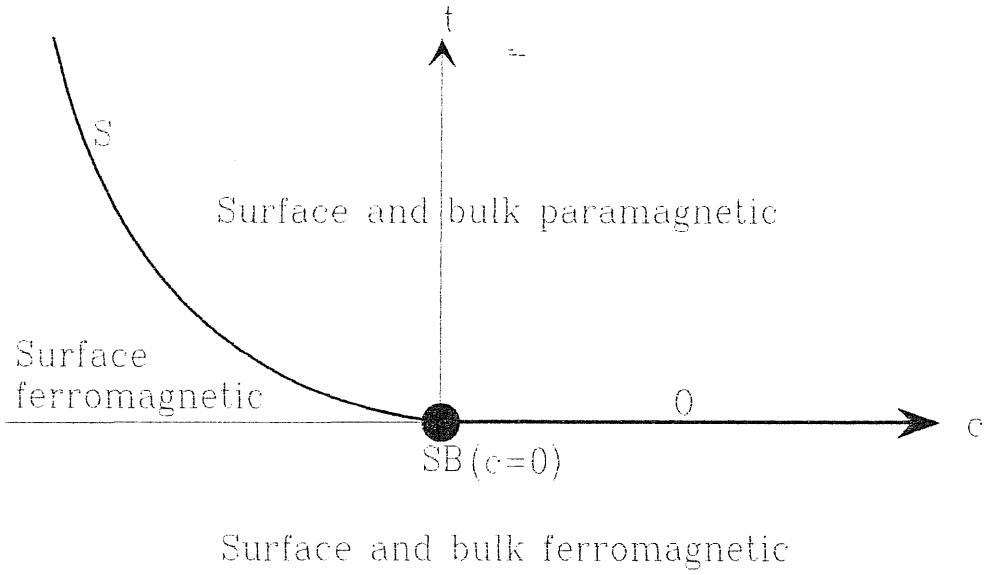

Figure 1.7 Phase diagram of a spin system with a free surface (see text)

occurs at the same critical temperature T_c for the bulk and for surface systems. For $c < 0$, when lowering the temperature $t = T - T_c$, the surface orders before the bulk which transition still occurs at $t = 0$. In the renormalisation group language the line O in figure 1.7 is mapped onto the *ordinary* fixed point O^* ($t = 0$ and $c = \infty$) and the line S is mapped onto the *surface* fixed point S^* ($t = c^2$ and $c = -\infty$). For $c = t = 0$ the two above fixed points coincide giving rise to a multicritical point SB which corresponds to a simultaneous surface and bulk ordering transition. A crossover exponent ϕ must therefore govern the transition from a bulk to a surface behaviour.

For a polymer chain, the line O corresponds to the limit $n \longrightarrow \infty$ while the line S represents the adsorption transition already discussed in this Section. Thus we have the following correspondences

$$\begin{aligned} t - c^2 \longrightarrow 0 &\iff n \longrightarrow \infty \\ c \longrightarrow 0 &\iff T \longrightarrow T_a \end{aligned} \tag{1.54}$$

Therefore, if the chain's monomers experience an attractive short-ranged surface potential, we will have effectively a repulsive, neutral or attractive surface-chain interaction for $c > 0$, $c = 0$ and $c < 0$, respectively. More precisely, we have $c \propto \frac{T}{T_a} - 1$.

A similar correspondence is expected to hold for an ideal as well as for a real chain [72]. Let us consider first the case of a real chain in a good solvent. The analysis by Eisenriegler *et al.* [72] shows that, in the limit $n \longrightarrow \infty$, the partition function at adsorption (point SB^* in figure 1.7) must obey the following scaling relation

$$Z_n = n^{\gamma_1 - 1} \phi(cn^{\varphi}) \tag{1.55}$$

with $\phi(x)$ linear for small values of x. The following scaling laws are found to hold for the characteristic sizes of the molecule. In the direction perpendicular to the surface

$$\langle R^2 \rangle_{\perp}^{1/2} = n^{\nu} \phi_{\perp}(cn^{\varphi}) \tag{1.56}$$

with the limiting behaviour, as $n \longrightarrow \infty$,

$$\langle R^2 \rangle_{\perp}^{1/2} \sim \begin{cases} |c|^{-\frac{\nu}{\varphi}}, & \text{for} \quad c < 0 \ ; \\ n^{\nu} & \text{for} \quad c = 0 \ ; \\ n^{\nu} & \text{for} \quad c > 0 \end{cases} \tag{1.57}$$

In the direction parallel to the surface

$$\langle R^2 \rangle_{\|}^{1/2} = n^{\nu} \phi_{\|}(cn^{\varphi}) \tag{1.58}$$

with the limit for $n \longrightarrow \infty$

$$\langle R^2 \rangle_{\|}^{1/2} \sim \begin{cases} n^{\nu_{d-1}} |c|^{-\frac{\nu_{d-1}-\nu}{\varphi}}, & \text{for} \quad c < 0 \ ; \\ n^{\nu} & \text{for} \quad c = 0 \ ; \\ n^{\nu} & \text{for} \quad c > 0 \end{cases} \tag{1.59}$$

In the above relations, ν and ν_{d-1} refer to the usual bulk size exponent and to the bulk size exponent in $d-1$ dimensions, respectively. In the next Chapters we will refer to the above exponent ν as ν_b and we will use the definition $\nu_s = \frac{\nu_b}{\varphi}$. It is also important to note that the perpendicular size of the adsorbed molecule is independent of the polymerization index n (see equation (1.57)). Similar scaling relations are expected to hold for a real chain in a Θ or poor solvent, but with other values for the exponents.

While the exponent ν has been dicusses previously, the crossover exponent φ is worthy of being briefly considered. The mean field estimate of φ has been shown to fail by de Gennes [70], who proposed $\varphi \approx 0.41$. Actually, the latter value is not well supported by numerical studies, which predict $\varphi \approx 0.6$ for chains in a good solvent. In two dimensions and for $c > 0$, the γ_1 exponent in equation (1.55) is known exactly from conformal-invariance arguments extended to the half plane geometry. Its value is $\gamma_1 = \frac{61}{64}$ (in the bulk we have $\gamma = \frac{43}{32}$). These values are confirmed by transfer-matrix calculations [73] which have also

provided, in two dimensions and for $c = 0$, $\varphi = 0.501$ and $\gamma_1 = 1.454$.

D. *Adsorption in a Θ-point Solvent*

We have seen that a polymer chain at the collapse transition (Θ-point) exhibits a multicritical behaviour. The same has been found to hold for a polymer at adsorption. Thus two multicritical points may coincide when the adsorption of a Θ-polymer occurs. This problem of the simultaneous collapse and adsorption of a polymer chain has only recently received the deserved attention. Considering the fact that Θ-polymers are nearly gaussian in three dimensions but deviate strongly from ideality in two dimensions, a model has been proposed in which the adsorbed (essentially two-dimensional) chain is modelled by SAWs whereas the free (three-dimensional) chain is represented by random walks. Both numerical [74] and field-theoretic studies [75] of this model have been carried out. Both approaches arrive at the same important conclusions. If it is true that at adsorption the bulk governs the exponents, as suggested by equations (1.57) and (1.59), then important logarithmic corrections to the ideal behaviour of three-dimensional Θ-polymers must be expected at (and near ?) adsorption. In other words, at adsorption, the bulk behaviour is strongly perturbed by the surface.

Two other studies concerning adsorption in Θ-point solvents must be mentioned. Both treat two-dimensional chains adsorbing on a line. In one case the model considers SAWs on a fractal structure [76] where exact renormalisation recursions can be written, whereas in ref. [77] a powerful transfer-matrix method has been used to analyse both directed and isotropic SAWs. The essential result was in both cases the demonstration of the existence of a multicritical point for simultaneous collapse and adsorption. The structure of the phase diagram was found for the fractal-SAW, and for the directed-SAW models, while for the isotropic model numerical difficulties have hindered the complete determination of

the phase diagram. The structure of the obtained phase diagrams is very similar and, somewhat in contradiction with our resuts, the Θ-line (that is the line formed by the Θ-points at different surface interactions) appears to be independent of the surface interaction parameter. For the directed-SAW model this feature has received a final confirmation by the exact calculations of ref. [78].

E. *Real-Space Renormalisation Approach to Polymer Adsorption.*

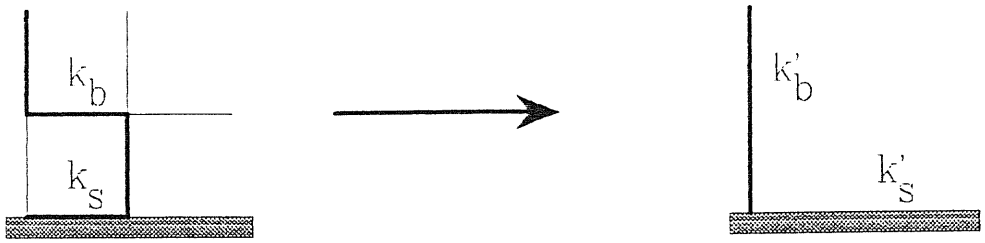

Figure 1.8 Bare and renormalized smallest cells on the square two-dimensional lattice. The "surface" is represented by the hashed area. A typical SAW is represented in the 2 x 2-cell and the bond fugacities are indicated.

To conclude this Chapter we now present the real-space renormalisation scheme introduced in ref. [79] to describe the adsorption of a polymer chain in a good solvent. The model is the classical SAW on a square lattice, but the chain is confined to a semi-infinite plane. The starting point or origin of the walks is on the limiting edge. As usual, a fugacity k_b is associated with each monomer-monomer bond in the bulk, whereas, in order to take into account the attractive nature of the wall, the bonds on the limiting edge will be attributed a different fugacity k_s. Figure 1.8 shows the bare and renormalized lattice cells for the simplest case in two dimensions. Two cell partition functions are then generated:

$$Z_b(k_b) = \sum_{n_b} c(n_b) k_b^{n_b}$$
$$Z^{s(x)}(k_b, k_s) = \sum_{n_b, n_s} c^{(x)}(n_b, n_s) k_b^{n_b} k_s^{n_s} \qquad (1.60)$$

The bulk cell generating function Z_b is constructed without considering the presence of the surface and $c(n_b)$ represents the number of SAW spanning the considered cell and having n_b steps. In the surface partition function $Z^{s(x)}$, the coefficients $c^{(x)}(n_b, n_s)$ represent the number of SAWs with n_b steps in the bulk and n_s steps on the surface (edge). The upper index x represents the minimum allowed fraction of surface bonds in the partition sum (that is, we sum only over walks spanning in the direction of the surface and having $\frac{n_s}{n_b+n_s} \geq x$). The recursion relations are obtained by equating the generating functions (1.60) evaluated for the bare cell, with those evaluated for the renormalized cell. The phase diagram given by the renormalisation flux [79] is sketched in figure 1.9. Its structure supports the correspondence previously discussed between the "special" critical point (in magnetic language) and the adsorption of a coil polymer chain. The best values of the critical exponents are obtained for renormalisation factors approaching 1 and for $x \longrightarrow 0$.

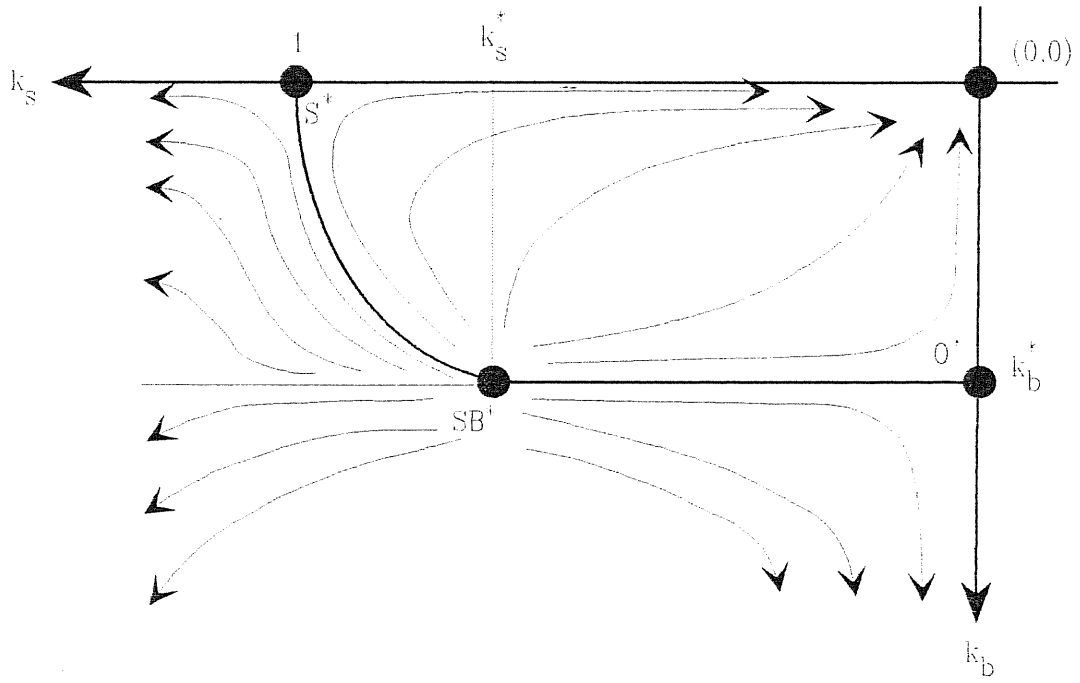

Figure 1.9 The renormalisation flux with the three non-trivial fixed points at $k_s=0$, k_s^*, 1 which must be associated with a free (nonadsorbed) chain, with a chain at adsorption, and with an adsorbed chain, respectively.

CHAPTER TWO

A REAL-SPACE RENORMALISATION-GROUP APPROACH TO POLYMER COLLAPSE AND ADSORPTION

2.1 THE SELF-AVOIDING-SELF-ATTRACTING-WALK MODEL

In the previous Chapter we have presented two separate real-space renormalisation approaches to the problem of polymer collapse [66] and polymer adsorption [79] (see Sections 1.2.3C and 1.2.4E). Both calculations have been performed in two dimensions. Elaborating on these original treatments, we have been able to propose a model capable of describing the collapse and adsorption transitions simultaneously. As it will be shown, the results of refs. [66] and [79] can be recovered as particular cases of our model (see also ref. [6]). Moreover, three-dimensional polymer chains adsorbing on a two-dimensional impenetrable wall have also been successfully studied.

In order to extend the SASAW-model to the adsorption problem in $d = 2$, we construct a semi-infinite two-dimensional triangular lattice as shown in figure 2.1. A bond fugacity k_b is associated with each monomer-monomer bond in the bulk, whereas in order to take into account the attractive nature of the wall the bonds on the limiting edge will be attributed a different fugacity k_s. A third parameter f,

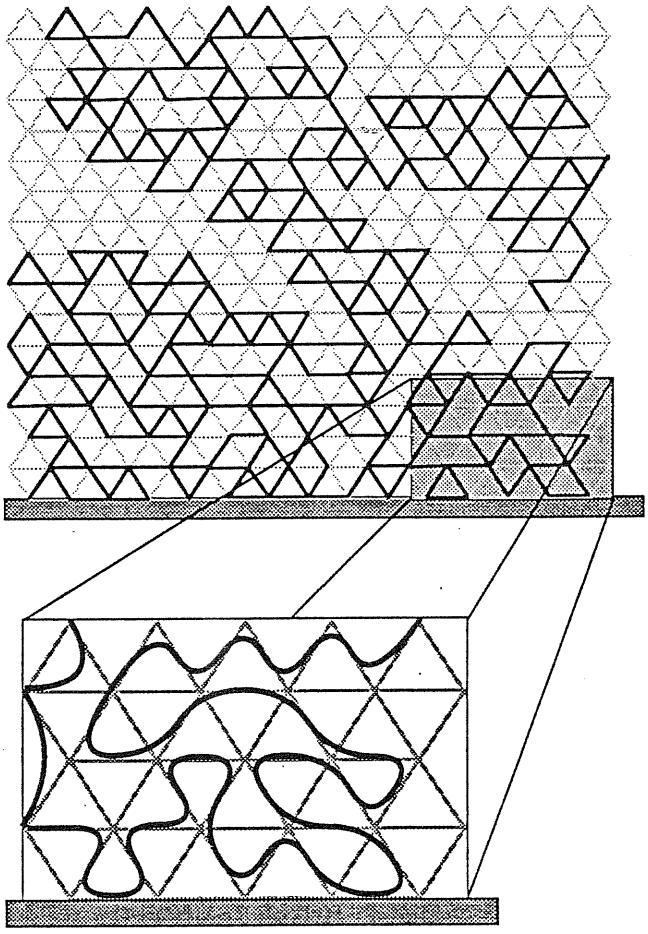

Figure 2.1 Schematic representation of the SASAW-model of adsorption of a collapsing macromolecule. In the window, a more realistic, continuum-space representation of a section of the polymer chain is given.

as defined in Section 1.2.3C, is then introduced to weight the monomer-monomer contacts. To deal with three-dimensional molecule configurations, we consider a "stack" of the above two-dimensional semi-infinite lattices. Thus, adsorption occurs on an infinite surface.

2.2 THE RENORMALIZATION PROCEDURE

As explained in Section 1.2.2B, real-space renormalisation-group calculations are implemented through the enumeration of the walks compatible with the model on small bare and renormalized lattice cells. All the different renormalisations that have been performed are shown in figure 2.2. In $d = 2$, three renormalisation factors have been considered: $b = 2$, $\frac{3}{2}$, and 3. In $d = 3$ we have performed two renormalisations, both with $b = 2$ but with different non-equivalent locations of the wall.

Unlike in refs.[66] and [79], our SASAW model requires three separate recursion relations to determine the three parameters k_b, k_s, and f. They are given by:

$$Z_{\Theta}^{\prime b}(k_b', f') = Z_{\Theta}^{b}(k_b, f) \tag{2.1}$$

$$Z_{\Theta}^{\prime s(x)}(k_b', k_s', f') = Z_{\Theta}^{s(x)}(k_b, k_s, f) \tag{2.2}$$

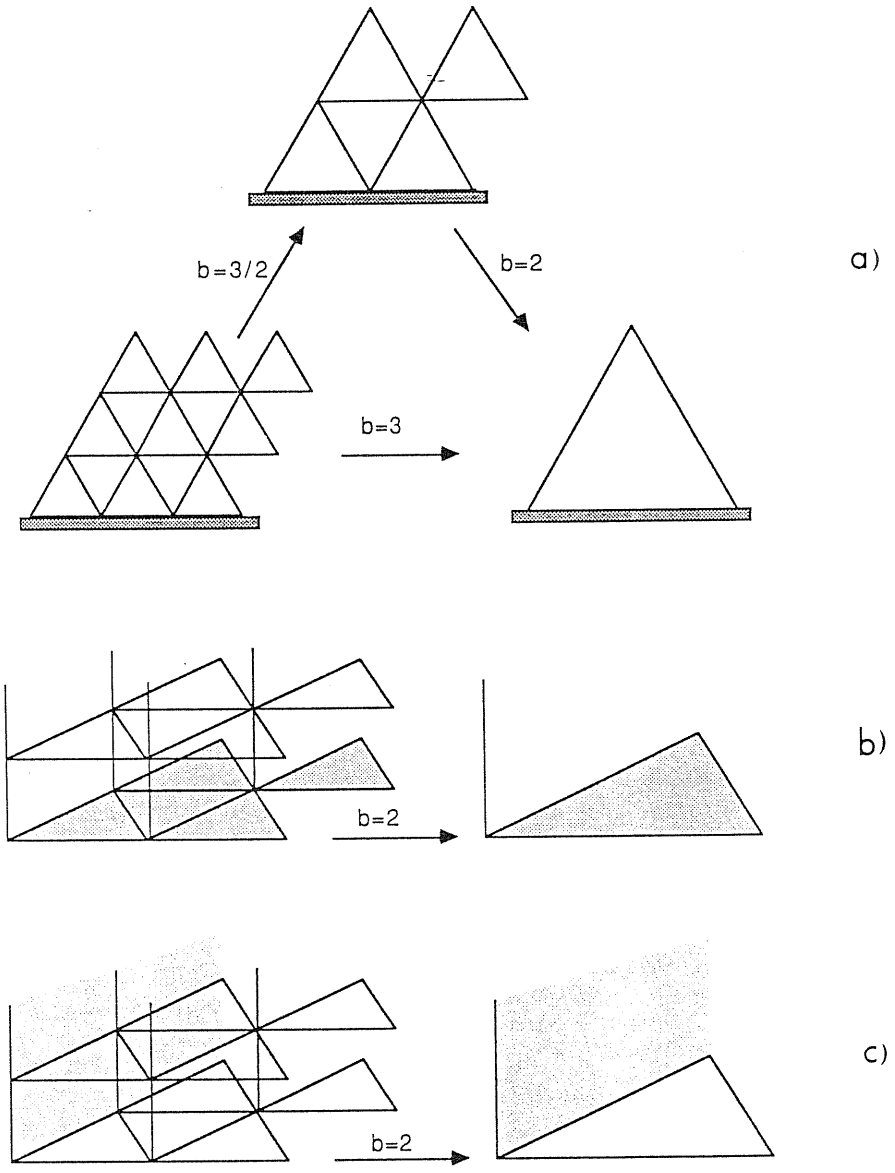

Figure 2.2 (a): Bare and renormalized cells for the three renormalizations performed in the two-dimensional space. (b) and (c): As in (a), but for the three-dimensional case. The shaded areas show the two possible positions for the surface.

$$f' = \frac{\sum_{m=1}^{n_{max}} f^m (1-f)^{n_{max}-m} Z_m^{s(x_{min})}}{(1-f)^{n_{max}} Z_0^{s(x_{min})} + \sum_{m=1}^{n_{max}} f^m (1-f)^{n_{max}-m} Z_m^{s(x_{min})}} \tag{2.3}$$

The cell partion functions are given by:

$$\begin{aligned} Z_\Theta^b(k_b, f) &= (1-f)^{n_{max}} Z_0^b(k_b) + \sum_{m=1}^{n_{max}} f^m Z_m^b(k_b) \\ Z_\Theta^{s(x)}(k_b, k_s, f) &= (1-f)^{n_{max}} Z_0^{s(x)}(k_b, k_s) + \sum_{m=1}^{n_{max}} f^m Z_m^{s(x)}(k_b, k_s) \\ Z'_\Theta &= \sum_{m=0}^{n'_{max}} f'^m Z'_m \end{aligned} \tag{2.4}$$

where the functions Z_m and Z'_m enumerate the walks with $m = 0, 1, \ldots, n_{max}$ contacts. They can be written as

$$Z_m = \sum_{Spanning Walks} c_m(n_b, n_s) k_b^{n_b} k_s^{n_s} \tag{2.5}$$

and similarly for Z'_m. Each of the above equations has a corresponding one in Section 1.2.3C where all the necessary explanations have been given. The meaning of the superscript x can be found in Section 1.2.4E. In order to find the fixed points we follow the standard procedure presented in Section 1.2.1F. The only practical difficulty is represented by the fact that the renormalization transformation is not known explicitly. Defining $\vec{\mu} = (k_b, k_s, f)$ we can represent

the three equations (2.1),(2.2), and (2.3) in the form $\vec{Z}'(\vec{\mu}') = \vec{Z}(\vec{\mu})$. For $\vec{\mu_0}$ sufficiently near to a fixed point we can write

$$\begin{aligned} \vec{R}(\vec{\mu}_0) &= \vec{Z}'(\vec{\mu}_0) - \vec{Z}(\vec{\mu}_0) \\ &\cong \nabla \vec{R}(\vec{\mu}_0) \cdot \delta\vec{\mu} \\ &= 0 \end{aligned} \tag{2.6}$$

Solving for $\delta\vec{\mu}$ we obtain a new guess $\vec{\mu}_1 = \vec{\mu}_0 + \delta\vec{\mu}$. Repeating the procedure, we can find a fixed point defined as $\vec{\mu}^* \equiv \vec{\mu}_{i+1}$ if $|\vec{\mu}_{i+1} - \vec{\mu}_i| < \varepsilon$, ε being the precision required. To determine the critical exponents for a given fixed point we derive the implicit recursions with respect to $\vec{\mu}$

$$\frac{\partial \vec{Z}'}{\partial \vec{\mu}'} \cdot \left.\frac{\partial \vec{\mu}'}{\partial \vec{\mu}}\right|_{\vec{\mu}^*} = \left.\frac{\partial \vec{Z}}{\partial \vec{\mu}}\right|_{\vec{\mu}^*} \tag{2.7}$$

Solving for the unknowns $\left.\frac{\partial \vec{\mu}'}{\partial \vec{\mu}}\right|_{\vec{\mu}^*}$, we obtain the jacobian matrix which eigenvalues give the critical exponents via equation (1.18).

2.3 RESULTS OBTAINED AND COMMENTS

2.3.1 Fixed Points and Critical Exponents

A sketch of the location of all the fixed points in the three-dimensional parameters space is presented in figure 2.3.

For a more detailed description of the parameter space see ref. [6] where two-dimensional sections are presented and the renormalisation flux is given by

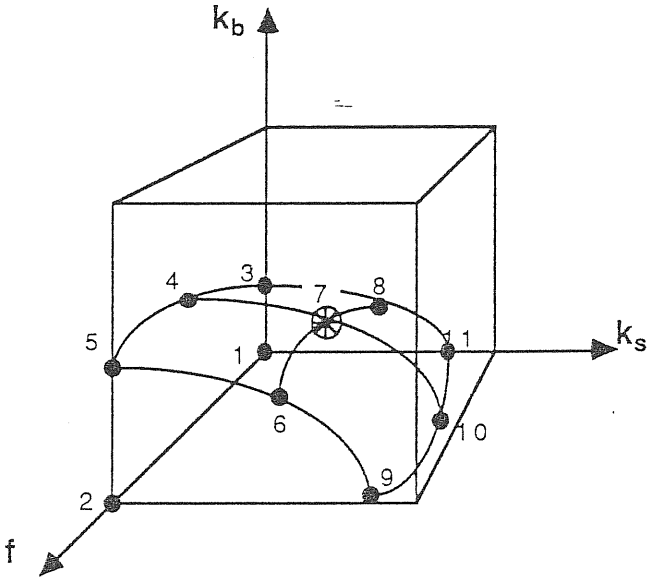

Figure 2.3 A schematic representation (for the d=3 problem) of the fixed points location in the three-dimensional parameter space. The exact coordinates of the adsorption-desorption fixed points 6, 7, and 8, are given in table 2.1 (for d=2) and table 2.2 (for d=3).

projected and normalized arrows. Tables 2.1 and 2.2 give the values and the relative exponents for the relevant fixed points corresponding to the different renormalizations presented in figure 2.2. In each table a study as a function of the minimum adsorbed fraction x is reported in one case.

2.3.2 Discussion of the Results

Looking at tables 2.1 and 2.2 we can see that the size exponents obtained with our SASAW model are close to the expected values, reported in brackets when known. As for the new exponents, we point out that they take reasonable values. In particular, the ν_s exponent decreases when going from SAW- to Θ-

			SAW	Θ	Globule
2 x 2 → 1 x 1	Bulk		k_b=.316 ν_b=.782 (3/4) f =0	k_b=.498 ν_b=.542 (4/7) f =.663 ν_f=1.945	k_b=.451 ν_b=.560 (1/2) f =1
	Surface	x = 1/4	k_b=.316 ν_b=.782 (3/4) k_s=.432 ν_s=1.465 (3/2) f =0	k_b=.518 ν_b=.580 k_s=1.005 ν_s=1.060 f =.480 ν_f=2.466	k_b=.451 ν_b=.560 (1/2) k_s=.862 ν_s=1.000 (1) f =1
		x = 1/5	k_b=.316 ν_b=.782 (3/4) k_s=.384 ν_s=1.613 (3/2) f =0	k_b=.516 ν_b=.572 k_s=.795 ν_s=1.241 f =.510 ν_f=2.521	k_b=.451 ν_b=.560 (1/2) k_s=.687 ν_s=1.155 (1) f =1
		x = 1/6	k_b=.316 ν_b=.782 (3/4) k_s=.377 ν_s=1.630 (3/2) f =0	k_b=.514 ν_b=.564 k_s=.615 ν_s=1.507 f =.534 ν_f=2.563	k_b=.451 ν_b=.560 (1/2) k_s=.553 ν_s=1.365 (1) f =1
3x3 → 1x1	Bulk		k_b=.298 ν_b=.772 (3/4) f =0	k_b=.376 ν_b=.615 (4/7) f =.535 ν_f=3.056	k_b=.338 ν_b=.632 (1/2) f =1
	Surface x = 1/6		k_b=.298 ν_b=.772 (3/4) k_s=.413 ν_s=1.547 (3/2) f =0	k_b=.392 ν_b=.618 k_s=.617 ν_s=1.271 f =.404 ν_f=6.063	k_b=.338 ν_b=.632 (1/2) k_s=.592 ν_s=1.241 (1) f =1
3x3 → 2x2	Bulk		k_b=.284 ν_b=.757 (3/4) f =0	k_b=.449 ν_b=.552 (4/7) f =.256 ν_f=1.622	k_b=.356 ν_b=.487 (1/2) f =1
	Surface x = 1/6		k_b=.284 ν_b=.757 (3/4) k_s=.453 ν_s=1.490 (3/2) f =0	k_b=.435 ν_b=* * * k_s=.847 ν_s=1.109 f = 180 ν_f=* * *	k_b=.356 ν_b=.487 (1/2) k_s=.675 ν_s=.999 (1) f =1

Table 2.1. Fixed point characterization for the recursion relation in d=2. Reported are the *SAW*, Θ and *G* fixed point values both with and without surface interactions. Results refer to 2 × 2 ⇒ 1 × 1, 3 × 3 ⇒ 1 × 1 and 3 × 3 ⇒ 2 × 2 cell renormalizations. A study as a function of the minimum adsorbed fraction *x* is reported in one case. Stars denote complex eigenvalues; in brackets are the expected exponent values, when known.

and globular-phases, as expected for increasingly more compact structures. Taking into account the relative small size of the cells considered, we can conclude that the model gives satisfactory quantitative results even if it does not allow us to

2x2x2 → 1x1x1			SAW	Θ	Globule
	Bulk		k_b= .222 ν_b= .635 (.588) f =0	k_b= .319 ν_b= .410 (1/2) f = .524 ν_f= 2.383	k_b= .281 ν_b= .425 (1/3) f =1
	Surface (1)	x = 1/4	k_b= .222 ν_b= .633 (.588) k_s= .256 ν_s= .911 f =0	k_b= .341 ν_b= .454 k_s= .379 ν_s= .688 f = .268 ν_f= 3.616	k_b= .281 ν_b= .425 (1/3) k_s= .306 ν_s= .643 f =1
		x = 1/8	k_b= .222 ν_b= .635 (.588) k_s= .245 ν_s= .972 f =0	k_b= .335 ν_b= .418 k_s= .348 ν_s= .741 f = .372 ν_f= 6.173	k_b= .281 ν_b= .426 (1/3) k_s= .291 ν_s= .711 f =1
		x =1/12	k_b= .222 ν_b= .635 (.588) k_s= .245 ν_s= .972 f =0	k_b= .333 ν_b= .415 k_s= .342 ν_s= .753 f = .394 ν_f= 5.894	k_b= .281 ν_b= .425 (1/3) k_s= .288 ν_s= .733 f =1
	Surface (2) x = 1/12		k_b= .255 ν_b= .537 (.588) k_s= .292 ν_s= .953 f =0	k_b= .370 ν_b= .344 k_s= .390 ν_s= .735 f = .498 ν_f= 2.226	k_b= .327 ν_b= .352 (1/3) k_s= .344 ν_s= .717 f =1

(1) Surface as in fig. **2.2 b)**
(2) Surface as in fig. **2.2 c)**

Table 2.2. As in table 2.1, but for d=3 and a 2 × 2 × 2 ⇒ 1 × 1 × 1 cell renormalization.

make a particular claim for any new precise exponent.

A. *Internal Consistency of the SASAW Model.*

An important check for the internal consistency of the model can be obtained by a careful analysis of the phase diagram in figure 2.3. This can be done by considering some particular cases. Let us look at the plane defined by $f = 0$, which must correspond to the SAW-phase. We note that the structure of this two-dimensional phase diagram exactly corresponds to that of figure 1.9 obtained by Kremer in ref. [79], with the three non-trivial fixed points (3), (8) and (11) that can be identified with the ordinary (O^*), special (SB^*) and surface (S^*) fixed points, respectively. Note that in $d = 2$ the k_s coordinate of (3) is $k_s^* = 1$, while in $d = 3$ we have $k_s^* = 0.316$. This can be easily understood since for $k_b = 0$ the chain is constrained to the ($d-1$)-dimensional surface. If $d = 2$ the

chain occupies entirely the available space and the critical fugacity must be equal to one. If $d = 3$ the chain is constrained to the two-dimensional surface and we recover the bulk two-dimensional case, k_s replacing k_b . Hence, relaxing the condition $f = 0$ but keeping $k_b = 0$, we recover for the $d = 3$ problem the two-dimensional collapse phase diagram obtained in ref. [66], with the variables (k_s, f) replacing (k_b, f) . Let us point out that the phase diagram sketched in figure 2.3 concerns the three-dimensional chain. Since, for the two-dimensional problem, the bound (or adsorbed) chain becomes one-dimensional, the collapse can no longer take place and the fixed points (9) and (10) must disappear, as we indeed observe.

B. *Enhancement of the Coil-Globule Transition Temperature*

An important new feature in the behaviour of a collapsing polymeric chain when a wall is present is suggested by the analysis of our results in tables 2.1 and 2.2. We have found that the value of f_Θ^* is always significantly *depressed* by the presence of the attractive surface [6], corresponding (for example) to an enhancement of the Θ-point transition temperature (collapse temperature). Alternatively, we could say that a collapse transition can be induced by adsorption, or that desorption is induced by a globule to coil transition in a bound macromolecule.

Since a finite-size-scaling analysis is beyond the capabilities of our small-cell renormalization method, one could always argue that the observed shift is due to a finite size effect. This would be the case in a magnetic problem, where the modified surface exchange cannot modify the bulk transition temperature in the thermodynamic limit. To assess this argument for the polymer problem at hand we have performed the following calculation. In the bulk 3×3 to 2×2 renormalization scheme, we have not allowed the chain to have any bond on one of the edges of

both the bare and renormalized cell. We obtain in this way a Θ-fixed-point with $f_{\Theta}^{bulk} = 0.240$, instead of $f_{\Theta}^{bulk} = 0.256$ (see table 2.1). Since in the case of a semi-infinite lattice the fraction of sites on the edge vanishes in the thermodynamic limit, we expect this shift to disappear in that limit. It is therefore plausible to consider this shift as an approximate measure of the finite-size effect. But we observe from table 2.1, for the related renormalization scheme, that this manifestly finite-size shift in the value of f_{Θ} is considerably smaller than the surface-induced decrease of the same parameter ($f_{\Theta}^{surface} = 0.180$), which, therefore we feel, cannot be easily attributed to a finite-size effect alone. Nevertheless, real-space cell-renormalization is and remains a poorly controlled method of characterizing the asymptotic behaviour of a lattice model. This fact and the great importance that the above result may have for surface macromolecular science in general, and for protein biochemistry in particular, motivated us in using a more powerful technique to investigate the simultaneous adsorption and collapse of polymer chains, as presented in the next Chapter.

CHAPTER THREE

A FINITE-SERIES APPROACH TO POLYMER COLLAPSE AND ADSORPTION

In this Chapter we consider the standard model, which we have presented in Section 1.2.3C (see figure 1.3) and which has been used by Kremer in ref. [79] to study the adsorption of polymer chains in a good solvent. Here, unlike in ref. [79] and instead of a small-cell real-space renormalisation technique, we use an exact enumeration of lattice walk configurations to tackle the problem of simultaneous collapse and adsorption. The following study has been essentially motivated by the interesting, but not sufficiently reliable, results obtained with the SASAW model and presented in Chapter 2.

3.1 THE STANDARD MODEL FOR POLYMER Θ-POINT COLLAPSE AND ADSORPTION

The method of finite series analysis [80–84] has provided a reliable technique for the determination of the Θ-point and related bulk critical exponents characterizing chains at collapse [83,84]. The adsorption of coil-like chains has also been examined with success by this method [80–82]. The enumeration of configurations by Monte Carlo [58] allows one to generate much longer chains and to give more accurate results than exact enumerations, but is much less versatile and, as pointed out below, is less suitable for simultaneous investigation of collapse,

adsorption and (see Chapter 4) disorder effects. For simplicity's sake, use has been made of exact enumerations of SAW-configurations on simple *square* and *cubic* lattices for two- and three-dimensional macromolecules, respectively. In two dimensions, three situations have been considered. We have first introduced a surface (or impenetrable wall) by disallowing the SAWs to visit one half of the space. The origin of each walk was fixed on the surface. Except for the self-avoiding and surface constraints, the walk is free to expand in all the lattice directions. In this case, chains of up to 22 steps in length have been considered. The second type of enumerations we have performed concern *directed*-SAW configurations, that is where the walk is never allowed to turn back on itself along the direction of the surface. The reduced lattice connectivity has allowed us to consider chains of up to 24 steps in length. For both isotropic and directed SAWs, energies ε_b and ε_s have been attributed to nearest-neighbour monomer-monomer interactions and to bond-surface interactions, respectively. As a third possibility, we have considered a system with an interfacial, instead of surface, potential consisting of an attractive line crossing the two dimensional space. The walk is allowed to visit both the semi-infinite planes defined in this way, but each monomer-monomer bond on the delimiting line contributes a term ε_s to the energy of the system. Before analysing the results, let us describe the method we have used to extract useful information from the computer enumeration of the configurations.

3.2 FINITE-SERIES TREATMENT FOR THE STANDARD MODEL

We define the squared radius of gyration of a given lattice configuration with n steps, i non-adjacent nearest neighbour interactions and j bonds on the surface, $R^2(W(n,i,j))$, as the variance of the chain monomers' positions. The statistical average over the configuration set $\{W(n,i,j)\}$ gives the mean squared radius of gyration for an n-step chain:

$$\langle R_n^2(\omega_b,\omega_s)\rangle = \frac{\sum_{i=0}^{i_{max}}\sum_{j=0}^{j_{max}}\sum_{\{W(n,i,j)\}} R^2(W(n,i,j))e^{i\omega_b}e^{j\omega_s}}{\sum_{i=0}^{i_{max}}\sum_{j=0}^{j_{max}}\sum_{\{W(n,i,j)\}} e^{i\omega_b}e^{j\omega_s}} \tag{3.1}$$

where $\omega_b = -\frac{\varepsilon_b}{k_B T}$ and $\omega_s = -\frac{\varepsilon_s}{k_B T}$. We also define $\omega_b(T_\Theta) = -\frac{\varepsilon_b}{k_B T_\Theta}$ and $\omega_s(T_a) = -\frac{\varepsilon_s}{k_B T_a}$ as the bulk and surface interactions at collapse and adsorption, respectively. Following Privman [84] we form the effective exponent estimates:

$$2\nu_{nk}(\omega_b,\omega_s) = \frac{\ln\left(\frac{\langle R_n^2(\omega_b,\omega_s)\rangle}{\langle R_{n-k}^2(\omega_b,\omega_s)\rangle}\right)}{\ln\left(\frac{n}{n-k}\right)} \tag{3.2}$$

Inserting in equation (3.2) the scaling relation (1.31), keeping ω_s constant and noting that $\Delta\omega$ in equation (1.31) corresponds to $\omega_b(T)-\omega_b(T_\Theta)$, we can easily see that for $T \longrightarrow T_\Theta$, and for n large enough for equation (1.31) to hold, the curves $\nu_{nk}(\omega_b,\omega_s = \text{const.})$ should intersect at $(\omega_b(T_\Theta),\nu_\Theta)$, independently of n and k. Similarly, keeping ω_b constant and using the scaling relations (1.56) or (1.58) we can conclude that the curves $\nu_{nk}(\omega_b = \text{const.},\omega_s)$ should intersect at $(\omega_s(T_a),\nu_{T_a})$, independently of n and k. Searching for the intersection of curves $\nu_{nk}(\omega_b)$ and of that for curves $\nu_{nk}(\omega_s)$ thus constitutes a practical way of finding the collapse and adsorption points, respectively. This, obviously, works perfectly only for infinite chains.

However, for relatively short chains a natural procedure is to consider the center of intersection region as an approximation for the Θ- and adsorption points. Also, when dealing with a square lattice the curves $\nu_{nk}(\omega_b)$ and $\nu_{nk}(\omega_s)$ must

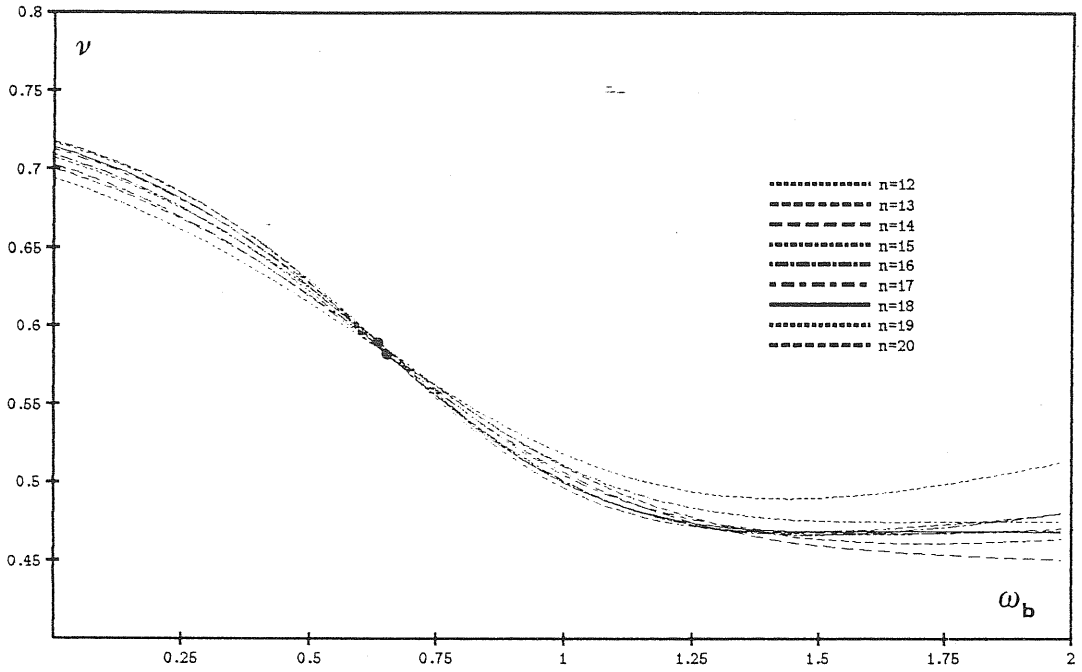

Figure 3.1 The curves $\nu_{n2}(\omega_b,\omega_s=0)$ with the intersection region clearly visible. It is also possible to distinguish the slightly separate intersection regions for even and odd values of n.

be separated in two groups, corresponding to even and odd values of n, having slightly different intersection regions (see figure 3.1).

To minimize the effects of these odd-even oscillations inherent to the square lattice, and as proposed in ref. [58], we choose $k = 2$ in equation (3.2) and determine, separately, the centre of each intersection region defined as the point corresponding to a minimal dispersion of the curves $\nu_{n2}(\omega_b)$ with n varying from a certain n_{max} to, say, $n_{max} - 5$. The final estimate of the multicritical points is then simply the arithmetic mean of the two values obtained in the way above (for even and odd values of n).

In principle, by varying n_{max}, one could obtain a set of intersection point locations and then extrapolate to $n_{max} = \infty$. Unfortunately, the values of

n_{max} accessible through exact enumerations give rise to results for the chain behaviour not systematic enough to allow for a meaningful extrapolation of trends. Nevertheless the precise location of transition points and relative exponents is beyond the scope of this work. Yet, the method described appears flexible enough to tackle complex problems arising when dealing with biological systems, such as the mutual relation between folding and adsorption subjected to a preliminary investigation in Chapter 2.

Let us now return to equation (3.1) and perform the sums over the configuration set $\{W(n,i,j)\}$. We are led to define the following coefficients:

$$\begin{aligned} D_n(i,j) &= \sum_{\{W(n,i,j)\}} R^2(W(n,i,j)) \\ C_n(i,j) &= \sum_{\{W(n,i,j)\}} 1 \end{aligned} \tag{3.3}$$

which are calculated *once and for all* during the computer generation of the set of configurations. Here we see the advantage of using exact enumerations with respect to the (more asymptotic) Monte Carlo sampling method, for which a small but consistent fraction of the configurations must be randomly generated for *each* values of k_b and k_s.

3.3 RESULTS AND COMMENTS

The complete phase diagram will be discussed in the next Chapter, where a chain consisting of a random sequence of hydrophilic and hydrophobic monomers will be considered. Here we focus our attention on the part of the parameter space characterized by positive values of ω_b and ω_s, that is for negative ε_b and

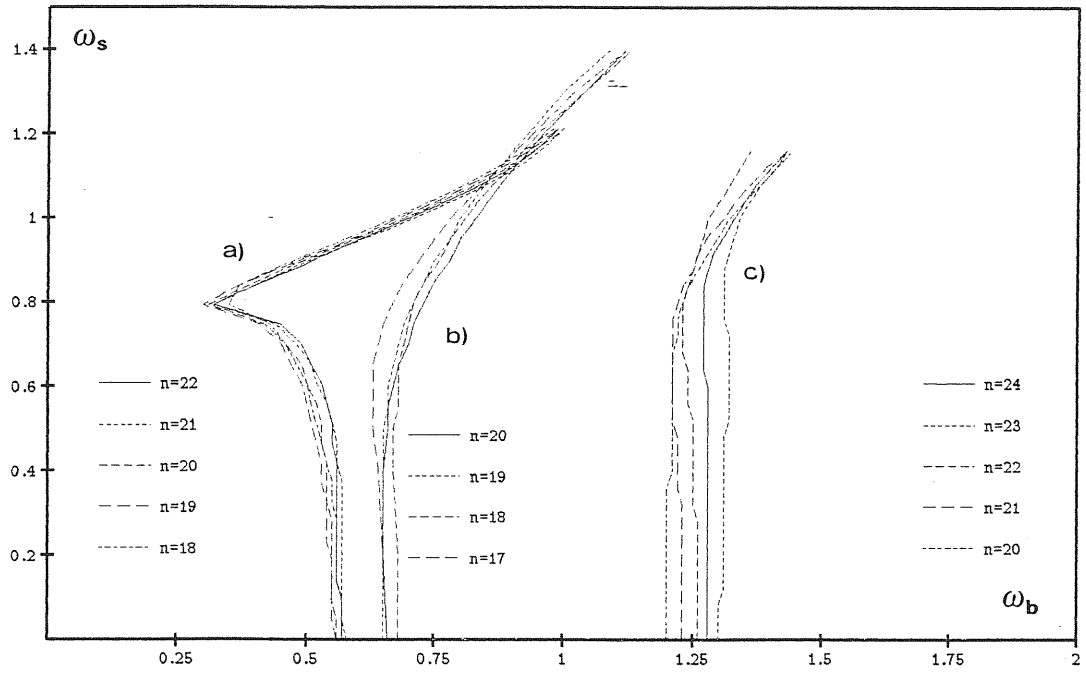

Figure 3.2 Part of the phase diagram for the following two-dimensional systems: (a) An isotropic chain near a surface. (b) An isotropic chain near an interface. (c) A directed chain near a surface. Transition lines are given for various chain lengths as indicated.

ε_s corresponding to a poor solvent and an attractive surface. The main results for the three different cases we have considered are summarized in figure 3.2.

The collapse lines are shown entirely for different values of n_{max} as indicated. For clarity's sake, only the right hand side branches of the adsorption lines are shown. In two cases, (b) and (c), we have a smooth connection between collapse (vertical) lines and adsorption (oblique) lines. The multicritical point corresponding to the coexistence of adsorption and collapse transitions can be roughly located at the above connection point. The situation is radically different for the case (a) where, as ω_s increases, the collapse line bends to the left hand side before reaching the adsorption line at the coexistence point. This corresponds to a progressive enhancement of the collapse temperature on increasing monomer-

surface attraction.

Before discussing more thoroughly the latter (unexpected?) effect, let us note that for the case (c) our results are in perfect qualitative agreement with the exact solution proposed by Foster in ref. [78]. More surprisingly, even quantitatively the results are rather good since, without any extrapolation to infinite chains, the transition coexistence point can be located on figure 3.2 (using $n = 24$) at $(\omega_b^*, \omega_s^*) = (1.25, 0.8)$ while the exact values are $(\omega_b^*, \omega_s^*) = (1.218, 0.784)$. For the case (b) and at $\omega_s = 0$, that is for the purely bulk case, we have $\omega_b(T_\Theta) = 0.65$ if we consider the transition line for $n = 20$. This should be compared with $\omega_b(T_\Theta) = 0.64$ proposed in ref. [58] and obtained with much more asymptotic Monte Carlo configurations. We can conclude that the method seems to give qualitatively good results. Therefore, the surprising feature of case (a) is worthy of being carefully considered.

3.3.1 The Surface-Induced Enhancement of the Collapse Temperature

The fact that the collapse transition of a chain with an end fixed on an impenetrable wall should occur at a higher temperature than if the chain were free is not by itself very much surprising. Such a behaviour was already conjectured by Dill and Alonso for protein folding [4] on the basis of the entropic arguments we are going to develop in the following. However, it is generally admitted that in the thermodynamic limit of infinite systems surface details cannot, for obvious reasons, affect bulk properties. This is true, for instance, for semi-infinite magnetic systems where a modified surface exchange cannot induce any change in the bulk phase transition. Accepting the above considerations we should ascribe the enhancement of the collapse temperature to a very strong finite-size effect which

certainly *must hold* for real, finite molecules such as polymers, poly-peptides, and proteins. However, there are some intriguing points which should be clarified:

1.- Our real-space renormalisation treatment also exhibits the same shift in the Θ -temperature and a test was performed to show that its amplitude is greater than that expected for a finite-size-induced shift.

2.- The case of a chain near an interface (b) and that of a directed chain attached to a surface (c) undergo the same finite-size limitations, but no shift has been observed for these situations.

A convincing physical interpretation of our results should clearly explain the observed differences between the three systems studied. A possible solution of the problem can be found by considering the reduction of entropy entailed by a chain attached to a wall. Dill and Alonso [4] argue that such a reduction is by far more drastic for coil-like than for globule-like chains. Thus, by constraining near a wall a chain which is just above its collapse temperature we can induce it to favour a globular conformation. In order to test the above affirmation we have performed the following calculation. We first define the free energy per step at the nth. step by

$$f_n = \frac{1}{2}(F_{n+1} - F_{n-1}) \tag{3.4}$$

where F_n represents the total free energy for an n-step chain. Similarly we define s_n and u_n, the entropy and the energy per step at the nth. step. Our purpose is to calculate $s_n = \beta(u_n - f_n)$. Therefore we must first obtain the energy U_n and the free energy F_n by making use of our exact enumerations of the system's configurations. This is readily done since we have

$$\begin{aligned}
\beta U_n &= -\frac{1}{Z_n}\sum_{i,j}(i\omega_b + j\omega_s)C_n(i,j)e^{i\omega_b+j\omega_s} \\
\beta F_n &= -\ln Z_n \\
Z_n &= \sum_{i,j} C_n(i,j)e^{i\omega_b+j\omega_s}
\end{aligned} \tag{3.5}$$

We then need an evaluation of the changes in s_n produced by adsorption, for both the coil and the globule. Choosing for ω_b and ω_s typical values of the coil and collapsed phases, and of the free and adsorbed phases, respectively, we determine $\delta^{ads}s_n(coil)$ and $\delta^{ads}s_n(globule)$. The results, as a function of the number of steps n, are shown in figure 3.3a for both the systems with an interface and with a surface. It can be seen that, as naively expected, adsorption is always accompanied by an entropy reduction (negative values of $\delta^{ads}s_n$). The conjecture of Dill an Alonso appears substantially correct since it is clear that this reduction is much greater for coil-like chains than for globules. In figure 3.3b we have reported $\Delta s_n = \delta^{ads}s_n(globule) - \delta^{ads}s_n(coil)$, again for both of the surface and interface cases. Δs_n is invoked in ref. [4] as the globule-stabilizing entropic-term. Here we see that $\Delta s_n \neq 0$ even for a chain constrained near an interface, but we know (see figure 3.2 b) that no globule stabilization occurs in that case; hence, for the interface, this Δs_n is due to adsorption.

On the basis of the above calculation (which certainly needs to be further refined, especially for what concerns the dependence of the results on the chosen values of ω_b and ω_s), we must improve on the conjecture of Dill and Alonso, at least as presented by the authors. Actually, an important feature of figure 3.3b must be pointed out. The difference $\Delta(\Delta s_n)$ between the two (surface and interface) curves is a growing function of the number of steps n.

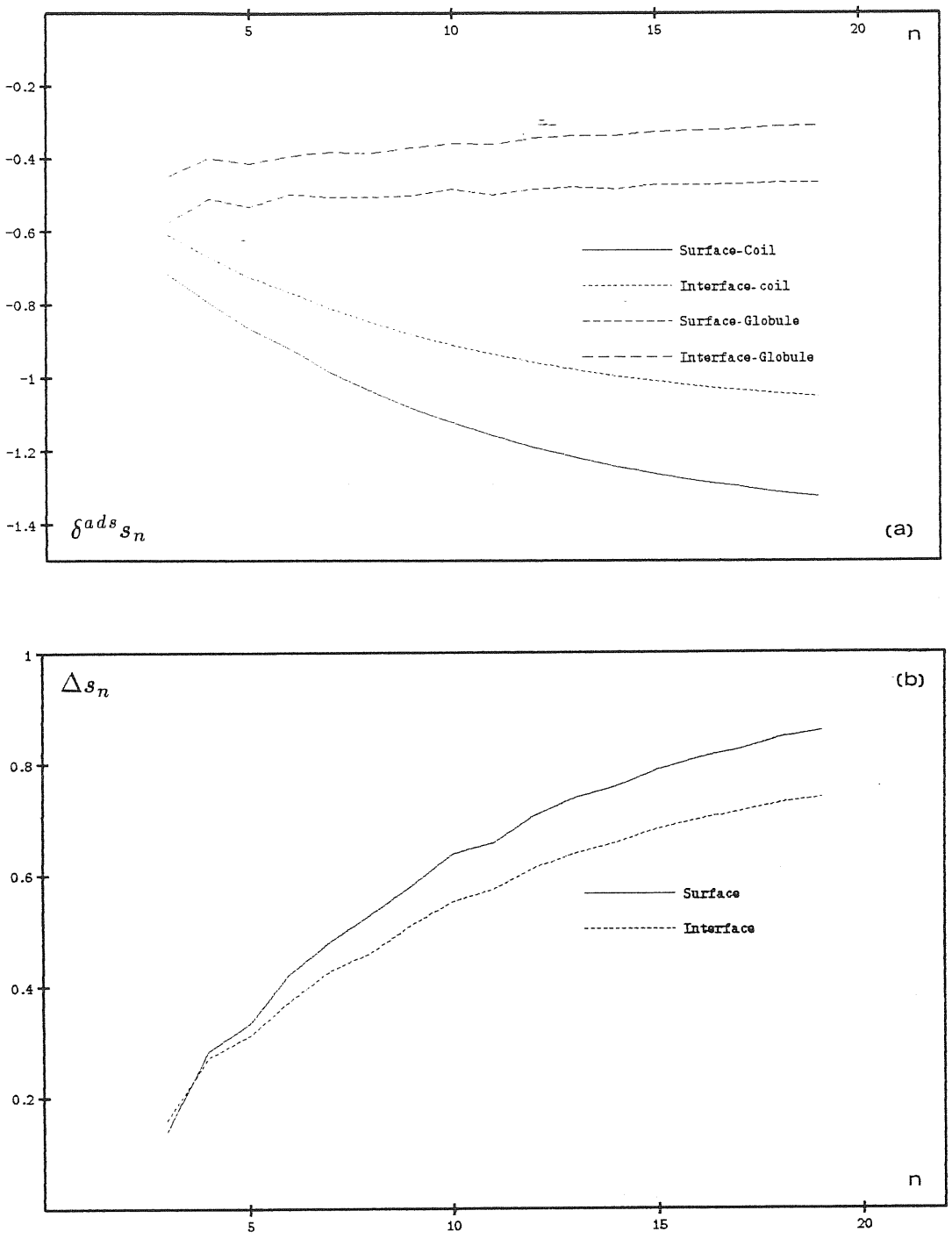

Figure 3.3 (a) The change (reduction) of entropy per step at the nth. step due to adsorption for the indicated cases and as a function of n . (b) The difference in entropy reduction between coil-like and globule-like chains (see also text).

This difference refers to one monomer only, and probably does not diverge in the thermodynamic limit. However the corresponding difference referred to the whole chain $(\Delta(\Delta S_n))$ does certainly diverge. We therefore believe that the observed shift in the Θ-temperature should be in some way connected with the above considerations.

In order to give a physical interpretation for the observed differences between the surface and interface cases, we can return to our discussion of Section 1.2.4B, where a one-dimensional random walk or an equivalent two-dimensional directed-SAW were shown to experience a long-range potential $W(z)$ arising from the steric repulsion induced by an impenetrable wall (see equation (1.53)). Such a repulsion is generally attributed to the entropy reduction produced by the wall. If the potential $W(z)$ is effectively long-ranged, particularly for an isotropic SAW we must expect its effect to be ω_s-dependent. Furthermore, coil- and globule-like chains are likely to experience this effective potential differently. The features of figure 3.2a thus follow. The absence of the temperature shift for the case where an interface is present is obvious since no entropy reduction occurs. For a directed chain, we note that the collapse occurs only in the direction parallel to the surface so that collapsed and coil chains will feel the same potential and no shift must be observed.

To examine the effect of the above repulsive potential we have calculated the ratio of the perpendicular to the parallel component of the radius of gyration as a function of ω_b and for different values of ω_s . The results are presented in figure 3.4 and concern a chain near a wall. Values of the ratio greater than 1, which occur for $\omega_s < 0.3$, indicate a chain slightly elongated in the direction perpendicular to the surface and reveal the presence of a repulsive potential. When the surface attraction increases the chain tends to flatten. It is interesting to note

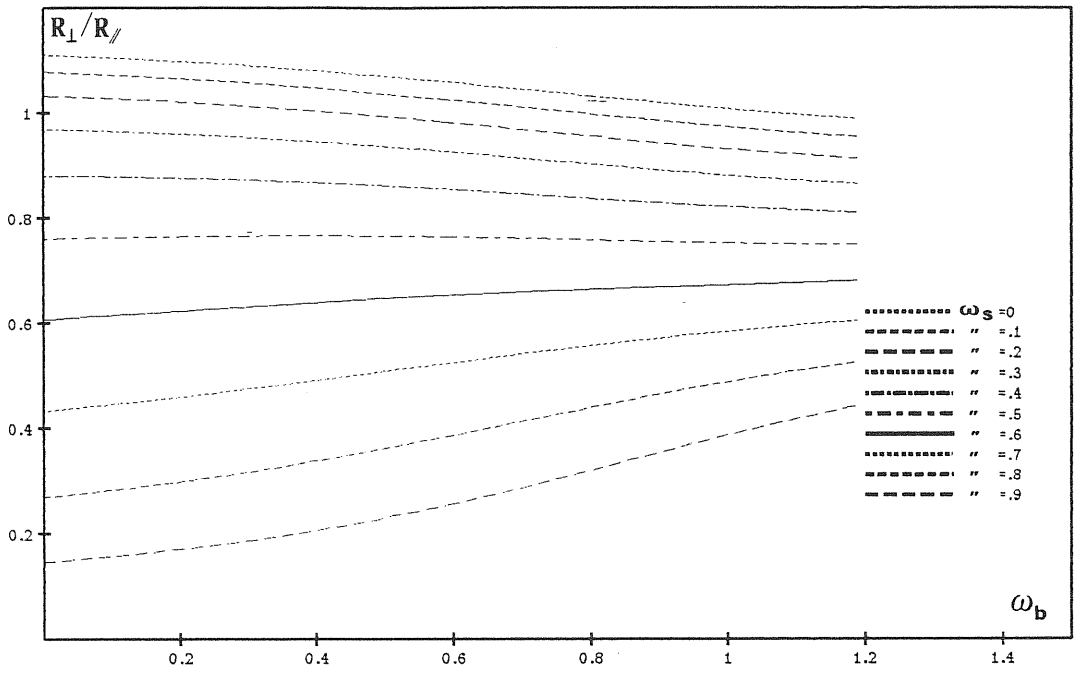

Figure 3.4 The ratio $\frac{R_\perp}{R_\parallel}$ of the perpendicular to parallel component of the radius of gyration as a function of ω_b and for various values of ω_s

that globular chains (high values of ω_b) resist much better to deformations than coils do.

3.3.2 The Three-dimensional Phase Diagram

To conclude this Chapter we present in figure 3.5 the results of a calculation for three-dimensional chains adsorbing to a two-dimensional impenetrable wall. Once adsorbed and essentially two-dimensional, macromolecules can collapse and all of the four phases (coil-free, coil-adsorbed, globule-free, globule-adsorbed) are now present. We note that three-dimensional collapse occurs at higher temperature than two-dimensional collapse, the location at $\omega_b = 0.7$ of which is in qualitative agreement with the two-dimensional bulk Θ-point which is expected to occur at

$\omega_b = 0.64$. Note that, due to the high connectivity of the lattice, we have been constrained to limit our enumerations to $n = 13$. Very surprising, indeed, appears the presence of a collapse and adsorption coexistence line where, according to our real space renormalisation, a single coexistence point should be expected. In our opinion, the results obtained with the exact enumeration analysis are more reliable than those obtained with the real-space renormalisation techniques presented in Chapter 2. To conclude, let us point out that the shift in the Θ-temperture is always present.

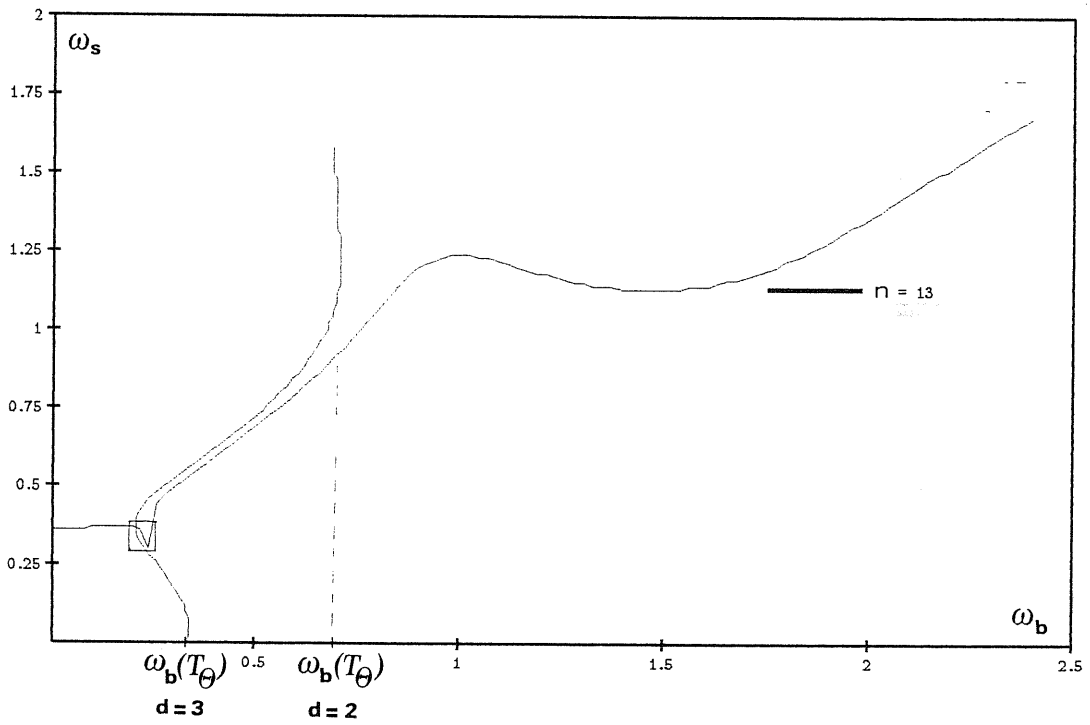

Figure 3.5 The phase diagram for a three-dimensional chain adsorbing on a two-dimensional surface.

CHAPTER FOUR

THE DISORDERED CHAIN

As already discussed in Chapter 1, a principal driving force for the folding of globular proteins is the solvophobic interaction, defined as the transfer process of non-polar residues between polar and non-polar environments. Disregarding their specific origin, these interactions are usually supposed to arise from the contact of non-polar amino-acid side chains and to be dominated by short-ranged attractive forces [17]. It is therefore of fundamental importance, when modelling folding processes, to account for the *random* distribution of non-polar (hydrophobic) and polar (hydrophilic) residues along a generic polypeptide chain. In this Chapter we present the first attempt to deal with the inherent disorder of proteins and with its implications for the folding transition, the latter having been modelled hitherto only through the oversimplified Θ-point collapse of a homopolymer in bad solvent. In a preliminary work by Obukhov [85], the problem of polymers with a fixed number of monomers of each kind, but with different possible ordering of these monomers, is considered. It is shown that for a finite chain near the Θ-point a substantial dispersion of the the polymer size must occur. The problem we intend to tackle here is different, since it concerns the dependence of the collapse and the adsorption temperatures on the average concentration of hydrophilic and hydrophobic monomers.

Randomness and disorder have been the subject of long-standing interest in the statistical mechanics of magnetic spin systems, such as for instance dilute magnets [86] and spin glasses [87]. Owing to the existence of a well-known

formal connection between the statistical mechanics of magnetic and polymeric systems (see Section 1.2.1E), one could think of a straightforward extension of the methods of theoretical analysis from one field to the other. In the case of proteins some attempt towards this direction has indeed appeared in the literature [88,89]. However, spin-glass type properties appear to dominate protein behaviour only well below biologically relevant temperatures, so that the frustration aspect of the problem loses part of its attraction. On the other hand, the renormalisation methods extensively and most fruitfully used to study standard disordered magnetic systems near criticality typically involve the determination of scale-invariant distributions [90], a procedure which we have found extremely delicate to implement when applied to inhomogeneous polymer chain models.

We have tried, indeed, to modify the SASAW model presented in Chapters 1 and 2 by replacing the contact probability f by a random variable F_i associated with each site where a monomer-monomer contact is possible. We choose the following form for the distribution of F_i

$$D_{F_i}(p, f_0) = p^2\delta(F_i) + 2p(1-p)\delta(F_i - f_0) + (1-p)^2\delta(F_i - 1) \qquad (4.1)$$

with p the fraction of hydrophilic monomers in the chain. This distribution corresponds to a weight $f = 0,\ f_0,\ 1$ for the contact between two hydrophilic, one hydrophilic and one hydrophobic, and two hydrophobic monomers, respectively. The recursion relations (1.36) and (1.37) then define, for the renormalized contact probability F', a distribution $D_{F'}(p, f_0')$ containing 3^k delta functions, where k represents the number of possible contacts in the considered bare cell. In order to determine a fixed distribution it is necessary to approximate $D_{F'}(p, f_0')$

by a simpler distribution with only three delta functions. The latter, unavoidable approximation has turned out to be fatal for the internal consistency of the model, which had to be abandoned. Fortunately, the finite-series analysis of the standard model presented in Chapter 3 has revealed to be more suitable for an extention to non-homogeneous polymer chains.

4.1 THE EXTENDED STANDARD MODEL

We consider a chain consisting of a *random* sequence of hydrophilic and hydrophobic residues, and we construct the following interacting SAW model, which should apply to a protein molecule in water (see also figure 4.1):

- a bond between two neighbouring hydrophobic residues will be attractive, thus weighted with a Boltzmann factor e^{ω_b}
- a bond between two hydrophilic residues is taken as repulsive, thus weighted with a factor $e^{-\omega_b}$
- a bond between an hydrophilic and an hydrophobic residues will be considered as neutral, thus corresponding to a unit weighting factor

An impenetrable wall is then introduced which interacts with the monomer-monomer *bonds*. In the case of a *non-polar* surface we have the following interactions:

- Two successive hydrophobic residues lying on the surface will be attracted and thus weighted with a Boltzmann factor e^{ω_s}
- a bond between two hydrophilic residues is repelled by the surface and thus weighted with a factor $e^{-\omega_s}$

- a bond between an hydrophilic and an hydrophobic residues will not interact with the surface, thus corresponding to a unit weighting factor

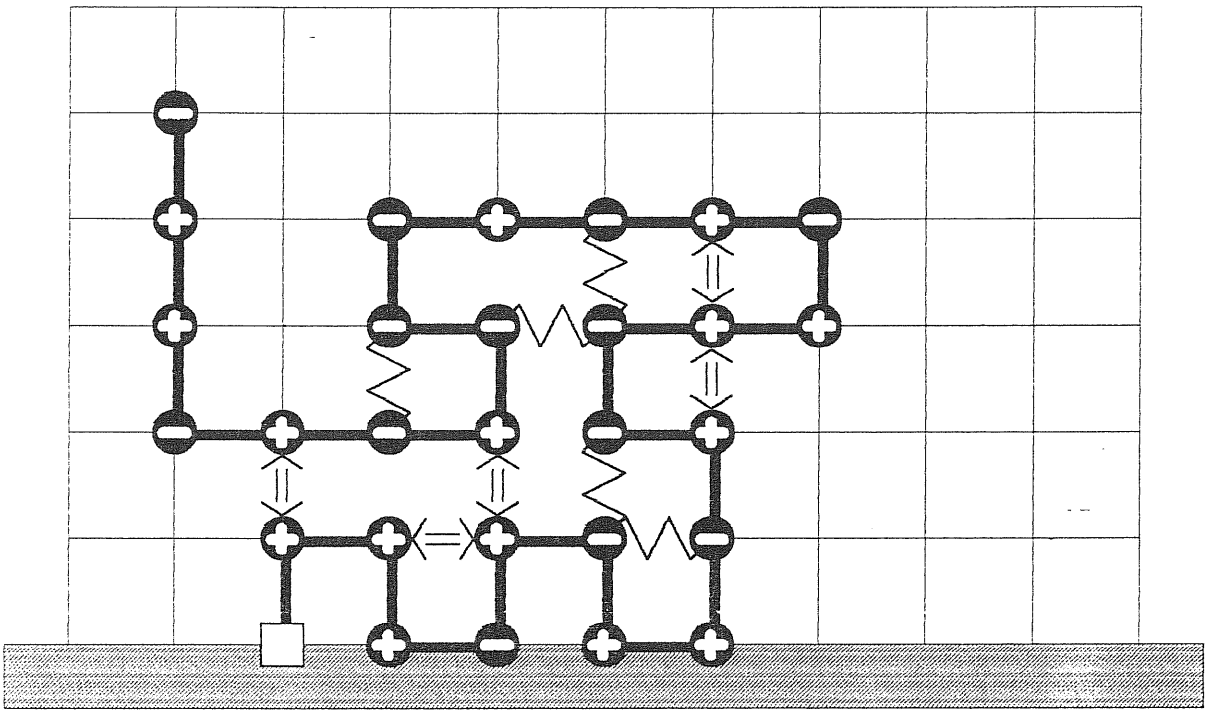

Figure 4.1 Example of a SAW configuration used in the present model. Pluses represent hydrophilic residues, minuses hydrophobic ones, while the interactions are represented by different symbols according to whether these are repulsive (plus-plus) or attractive (minus-minus). All configurations begin with one end on the surface, while the sign of the bond-surface interaction depends on the type of residues attached to the bond (see text).

In order to implement the finite series analysis presented in Chapter 3 we need to calculate the mean squared radius of gyration. Using the same notation as in equation (3.1)we have:

$$\langle R^2_{nk}(\omega_b,\omega_s)\rangle = \frac{\sum_{i,j=0}^{i_{max}\,j_{max}} \sum_{\{W(n,i,j)\}} R^2(W(n,i,j)) \sum_{\{Q(n,k)\}} e^{\sum_{b=1}^{i}\sigma_b\omega_b+\sum_{s=1}^{j}\sigma_s\omega_s}}{\sum_{i,j=0}^{i_{max}\,j_{max}} \sum_{\{W(n,i,j)\}} \sum_{\{Q(n,k)\}} e^{\sum_{b=1}^{i}\sigma_b\omega_b+\sum_{s=1}^{j}\sigma_s\omega_s}} \qquad (4.2)$$

where $\sigma_b = \sigma_b(\ W(n,i,j);Q(n,k)\ )$ and $\sigma_s = \sigma_s(\ W(n,i,j);Q(n,k)\ )$ take the values 1, 0 or -1 depending on the nature of the nearest-neighbour or bond-surface interaction as defined above.

Sums are now performed over the set $\{Q(n,k)\}$ of all possible combinations of k hydrophilic residues and $(n-k)$ hydrophobic residues along a n-step chain. Such an approach has the severe disadvantage that σ_b and σ_s depend on the spatial configuration $W(n,i,j)$ as well as on the sequence $Q(n,k)$, so that the calculation of the coefficients (3.3) cannot be performed separately.

A slightly different, statistical, approach nevertheless makes the problem numerically tractable. We first define the average fraction $f = \frac{k}{n}$ of hydrophilic residues and then assign to each contact a value of σ_b with a probability $p(\sigma_b)$ depending on f in the following way:

$$p(\sigma_b) = \begin{cases} f^2, & \text{if } \sigma_b = -1\ , \text{ hydrophilic-hydrophilic residue contact.} \\ 2f(1-f), & \text{if } \sigma_b = 0\ , \text{ hydrophobic-hydrophilic residue contact.} \\ (1-f)^2, & \text{if } \sigma_b = 1\ , \text{ hydrophobic-hydrophobic residue contact.} \end{cases} \qquad (4.3)$$

Similarly, to each bond-surface contact we assign a value of σ_s with a probability $p(\sigma_s)$ depending on f through:

$$p(\sigma_s) = \begin{cases} f^2, & \text{if } \sigma_s = -1\ , \text{ hydrophilic-hydrophilic bond-surface contact.} \\ 2f(1-f), & \text{if } \sigma_s = 0\ , \text{ hydrophobic-hydrophilic bond-surface contact.} \\ (1-f)^2, & \text{if } \sigma_s = 1\ , \text{ hydrophobic-hydrophobic bond-surface contact.} \end{cases} \tag{4.4}$$

The mean squared radius of gyration, $\langle R_n^2(\omega_b, \omega_s, f)\rangle$, will now be given by :

$$\frac{\sum_{i,j=0}^{\substack{i_{max}\\ j_{max}}} \sum_{\{W\}} R^2(W) \sum_{\{c_b(i)\},\{c_s(j)\}} P(c_b(i), c_s(j), f) e^{\sum_{b=1}^{i} \sigma_b\omega_b + \sum_{s=1}^{j} \sigma_s\omega_s}}{\sum_{i,j=0}^{\substack{i_{max}\\ j_{max}}} \sum_{\{W\}} \sum_{\{c_b(i)\},\{c_s(j)\}} P(c_b(i), c_s(j), f) e^{\sum_{b=1}^{i} \sigma_b\omega_b + \sum_{s=1}^{j} \sigma_s\omega_s}} \tag{4.5}$$

where $c_b(i) = (\sigma_1, \sigma_2, \ldots, \sigma_i)_b$ and $c_s(j) = (\sigma_1, \sigma_2, \ldots, \sigma_j)_s$ are the combinations of i nearest-neighbour and j bond-surface interactions, respectively. $P(c_b(i), c_s(j), f)$ is the probability of the corresponding combinations and is given by:

$$\begin{aligned} P(c_b(i), c_s(j), f) &= P(c_b(i), f) \cdot P(c_s(j), f) \\ &= \prod_{b=1}^{i} p(\sigma_b) \cdot \prod_{s=1}^{j} p(\sigma_s) \end{aligned} \tag{4.6}$$

The sum over the sets $\{c_b(i)\}$ and $\{c_s(i)\}$ can now be factorized out, and, using the following definitions:

$$H_b(i,\omega_b,f) = \sum_{\{c_b(i)\}} P(c_b(i),f) e^{\sum_{b=1}^{i} \sigma_b \omega_b}$$
$$H_s(i,\omega_s,f) = \sum_{\{c_s(j)\}} P(c_s(j),f) e^{\sum_{s=1}^{j} \sigma_b \omega_s} \tag{4.7}$$

we can write

$$\langle R_n^2(\omega_b,\omega_s,f)\rangle = \frac{\sum_{\substack{i=0 \\ j=0}}^{\substack{i_{max} \\ j_{max}}} H_b(i,\omega_b,f) H_s(j,\omega_s,f) D_n(i,j)}{\sum_{\substack{i=0 \\ j=0}}^{\substack{i_{max} \\ j_{max}}} H_b(i,\omega_b,f) H_s(j,\omega_s,f) C_n(i,j)} \tag{4.8}$$

The coefficients $C_n(i,j)$ and $D_n(i,j)$ are those defined in equation (3.3). A further simplification occurs if equation (4.7) is rewritten in the following way. Let a, b and c be the number of repulsive, neutral and attractive bonds, respectively. The sums over the set of combinations $\{c_b(i)\}$ and $\{c_s(j)\}$ become:

$$H_b(i,\omega_b,f) = \sum_{\substack{a,\ b,\ c \\ a+b+c=i}} \frac{i!}{a!\ b!\ c!} p^a(-1) p^b(0) p^c(1) e^{-a\omega_b + c\omega_b}$$
$$H_s(j,\omega_s,f) = \sum_{\substack{a,\ b,\ c \\ a+b+c=j}} \frac{j!}{a!\ b!\ c!} p^a(-1) p^b(0) p^c(1) e^{-a\omega_s + c\omega_s} \tag{4.9}$$

which can be more conveniently calculated than equation (4.7).

The determination of the $\omega_b(T_\Theta)$ and $\omega_s(T_a)$ for the collapse and adsorption transitions, respectively, follows the scheme presented in Chapter 3.

The results thus obtained are presented in the following Section.

4.2 RESULTS AND COMMENTS

In figure 4.2 we present the central result of this Chapter, namely the dependence on the fraction f of hydrophilic residues of the collapse and adsorption parameters. The curve $\omega_b(T_\Theta)$ represents the collapse transition when the monomer-surface interaction is absent, ($\omega_s = 0$, bulk problem).

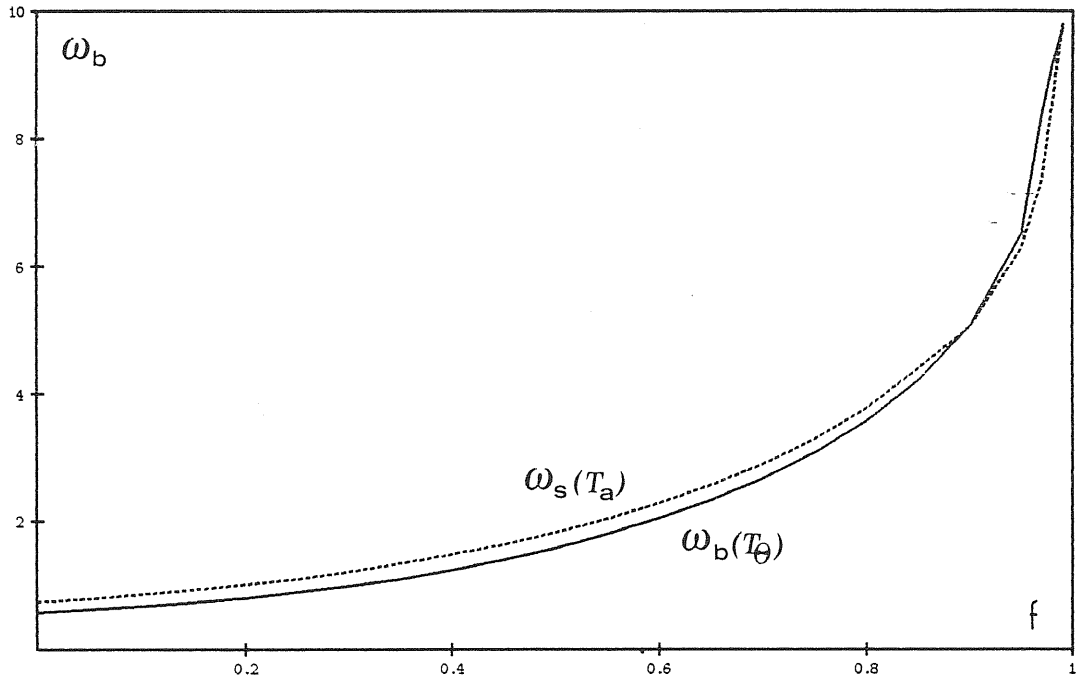

Figure 4.2 Dependence of the collapse temperature $\omega_b(T_\Theta) = -\frac{\varepsilon_b}{k_B T_\Theta}$ and of the adsorption temperature $\omega_s(T_a) = -\frac{\varepsilon_s}{k_B T_a}$ as a function of the hydrophilic concentration f for $\omega_s = 0$ and $\omega_b = 0$, respectively.

When $f \to 0$, $\omega_b(T_\Theta)$ approaches the standard value ≈ 0.65 which holds for an homogeneous chain (hydrophobic residues only). As f is increased,

$\omega_b(T_\Theta)$ increases too, corresponding to a fall in collapse temperature: for $f \to 1$ collapse takes place only at zero temperature. A similar pattern emerges from the curve $\omega_s(T_a)$, where adsorption in presence of a good solvent is considered ($\omega_b = 0$) . The reported curve corresponds to a wall exibiting attractive, neutral and repulsive interactions with hydrophobic-hydrophobic, hydrophobic-hydrophilic and hydrophilic-hydrophilic bonds, respectively. Very similar results are obtained for $\omega_b \neq 0$ also. We note that the above results, which constitute the first quantitative estimates of the effects of randomness on polymer adsorption and collapse, are those one would qualitatively expect.

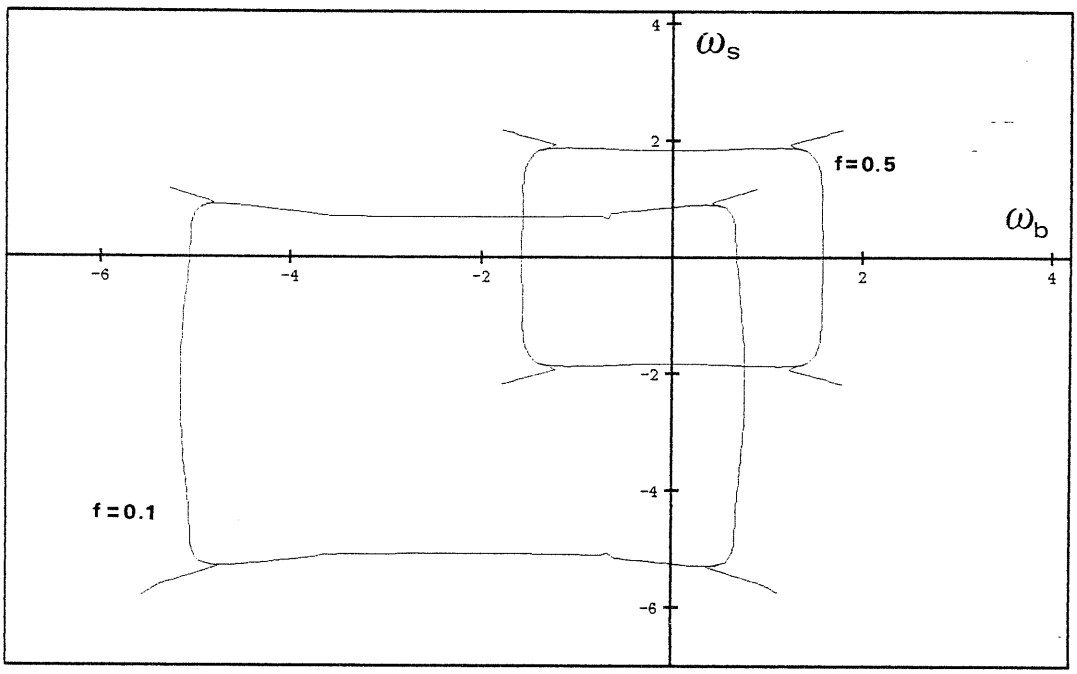

Figure 4.3 Generalised phase diagram for the folding and adsorption transitions of our disordered chain model. (a) is for f=0.5 , when complete symmetry holds, (b) is for f=0.1 .

In figure 4.3 the complete phase diagram is presented for the model introduced and for some values of f . Complete symmetry with respect to the origin is

noticed for the particular choice of $f = 0.5$ owing to the fact that for negative values of ω_b or ω_s there is a precise exchange of the rôle of hydrophilic and hydrophobic residues. Hence the reported diagram covers all of the four possible combinations of a polar or non-polar surface and a polar or non-polar solvent. For values of f other than 0.5 we observe a distortion of the adsorption and collapse boundary lines. The full boundaries for $f = 0.1$ are also reported in figure 4.3 for comparison. Data refer always to $n = 22$ chain steps.

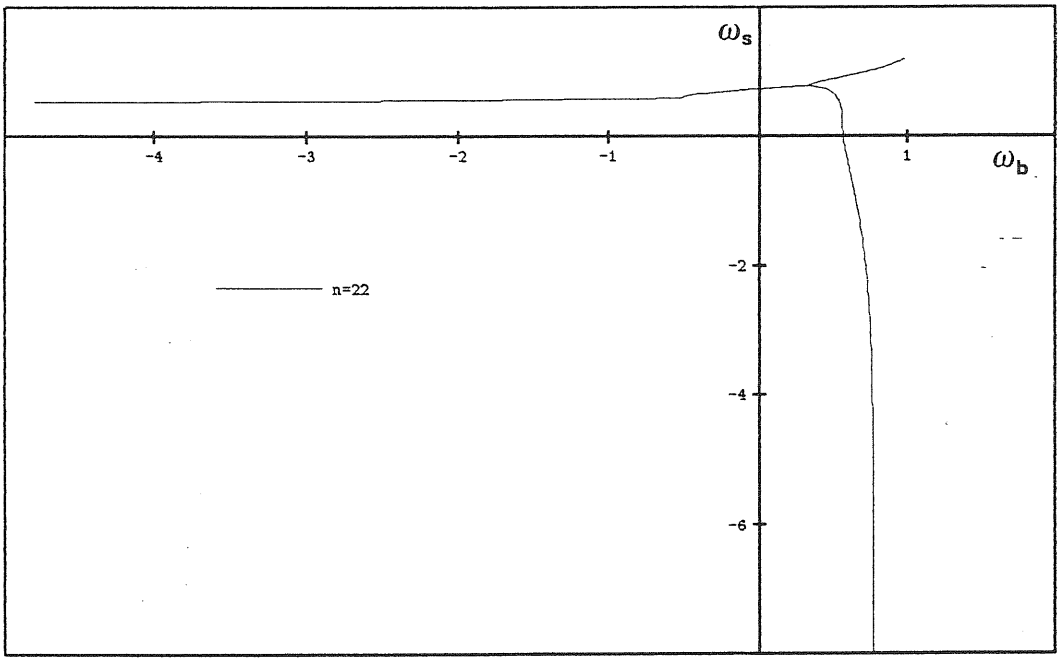

Figure 4.4 When $f \to 0$ only the first quadrant remains and coincides with the homogeneous system's diagram, while the multicritical point in the three other quadrants disappears to infinity.

In figure 4.4 we present the special case $f = 0$, which corresponds to a homogeneous chain. We note that the collapse and the adsorption lines are not present for negative values of ω_b and ω_s since they correspond to a self-repulsive chain and to a repulsive bond-surface interaction. In Chapter 3 we

have compared, for the homogeneous case, the phase diagram of a chain near an impenetrable wall and that of a chain near an attractive interface. In figure 4.5 we repeat the comparison but for an inhomogeneous chain and for different values of f . The same conclusion holds: the shift in the Θ-temperature for a chain near adsorption concerns only the surface case, while in presence of an interface the Θ-line is absolutely vertical.

However, two new features must be pointed out. First, the maximum amplitude of the shift, as also shown in table 4.1, is maximum for $f = 0.5$. This corroborates our hypothesis on the origin of the shift, since this value corresponds to a maximum in the entropy of the chain and, therefore, must also correspond to a maximum in the entropy reduction induced by the presence of the surface. Furthermore, we observe in table 4.1 that within the numerical accuracy allowed by the method $\Delta\omega_b \longrightarrow 0$ as $f \longrightarrow 1$, a demonstration that the shift is due to surface adsorption as in the limit $f \longrightarrow 1$ the chain becomes unbound.

The second feature we would like to point out concerns the collapse line for $f = .9$ in figure 4.5. It can be seen that for $\omega_s \approx 0.2 - 0.3$, the curves for the surface and the interface cases coincide. This can be understood in the following way. Since 90% of the monomers are repelled by the surface, the chain conformations approaching the surface are energetically unfavourable. Therefore the surface-induced entropy reduction plays a much less important rôle. Note that for $\omega_s \approx 0$ and $\omega_s \approx \omega_s(T_a)$ the above argument does not hold and the effect of the surface is again important. A similar explanation can be given to the behaviour of the Θ-line in figure 4.4 for negative values of ω_s.

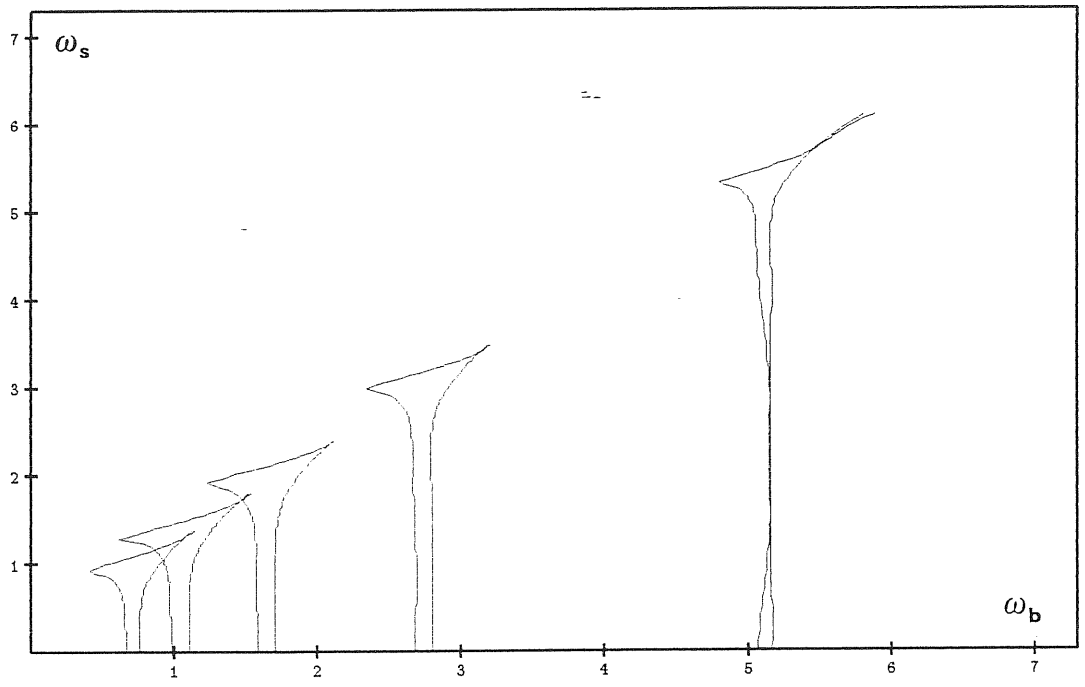

Figure 4.5 Comparison between (part of) the phase diagram for the isotropic disordered polymer chain adsorbing at a surface or at an interface, respectively. Phase boundaries are shown for values of f=0.1,0.3,0.5,0.7,0.9 . The shift is present only for the surface case, and is maximum at f=0.5.

f	$\Delta\omega_b(T_\Theta)$
0	0.20
0.1	0.24
0.3	0.32
0.5	0.37
0.7	0.31
0.9	0.24
0.95	0.22
0.97	0.13
0.98	0.12

Table 4.1. Variation of the (maximum) shift in inverse Θ temperature as a function of f. Observe that the change is roughly symmetric and is maximum for f=0.5, when entropy reduction for the SAW phase is greatest. Values near f=1 carry the highest uncertainty, but are consistent with $\Delta\omega_b(T_\Theta)\rightarrow 0$ as $f \rightarrow 1$.

CHAPTER FIVE

CONCLUSIONS AND OUTLOOK

At this point it remains for us to summarize the results obtained and to comment on their possible biological implications. This is better done by considering the physical aspects separately from the biological ones.

5.1 PHYSICAL ASPECTS

Even though the primary motivation for this work was the attempt to model the folding transition of proteins occuring near an adsorbing substrate, a number of other physically interesting features have emerged from the models studied.

We have first shown how the analysis of a the simple *SASAW-model* by means of a *real space renormalisation* method allows for the determination of the complete *phase diagram* for a complex system such as the *isotropic* flexible chain near an impenetrable wall exhibiting a short-range attractive potential towards the chain monomers. A similar phase diagram has been obtained already only for the simplified (and not too realistic) cases of a directed chain and of a chain embedded in a fractal lattice space. Our model allows us to find all the non-trivial *fixed points* and to give approximate, but consistent, values for the relative critical exponents. It is important to note that calculations have been performed for both two-dimensional and three-dimensional chains. The phase diagram structure reveals the existence of an *enhancement* of the collapse temperature due to partial chain adsorption.

We have then considered the nearest-neighbour interacting SAW model, also referred to as the standard model. A finite series analysis of computer-generated configurations allows us to obtain a more precise determination of the phase diagram with a complete location of the collapse and adsorption transition curves as a function of the monomer-monomer and monomer-surface interaction parameters. For two-dimensional chains, a perfect correspondence between the phase diagrams obtained with real space renormalisation and finite-series analysis, respectively, can be established. In particular, the shift in the collapse temperature near adsorption is seen to be always present. For three-dimensional chains the same shift is present, but finite series analysis strongly suggests the presence of a coexistence *line* where collapse and adsorption take place simultaneously. This improves on the real-space renormalisation result, which allows for a unique coexistence point.

By extending the standard model in order to include a random distributon of hydrophilic and hydrophobic monomers along the chain we have then obtained the first quantitative estimate of the effects of *disorder* on the collapse *and* adsorption transitions. Previous studies of disorder effects on adsorption alone have considered either random ideal chains or homogenous chains in a random environment. Beside its intrinsic relevance, the disorder effect analysis has provided useful information on the nature of the shift in the collapse temperature.

As we will see in the next Section, the shift in the collapse temperature occuring near the adsorption transition represents our most important result for biology. Nevertheless, the difficult *conceptual* problem of its persistence in the limit of infinite chains is not yet fully resolved. The comparison of different systems (isotropic and directed chains adsorbing on an impenetrable wall, isotropic chain localising at an interface) as well as the analysis of the shift as a function of the

fraction of hydrophilic monomers allows us to ascribe the physical nature of the shift to the reduction of the chain conformational entropy induced by the confining surface.

5.2 BIOLOGICAL ASPECTS

The phase diagrams we have calculated are drawn in the space of the parameters $\omega_b = -\beta\varepsilon_b$ and $\omega_s = -\beta\varepsilon_s$ which measure the strength of the monomer-monomer and monomer-surface interactions, respectively. It must be clear that a temperature change affects both factors $\beta = \frac{1}{k_B T}$ and $\varepsilon_{b,s} = \varepsilon_{b,s}(T, \text{solvent})$ whereas solvent changes affect only the effective interaction energies $\varepsilon_{b,s}$. Therefore, our results can be tested experimentally only if we know the temperature and/or solvent dependence of $\varepsilon_{b,s}$. Moreover, the experimental determination of a similar phase diagram requires the ability to vary ω_b and ω_s independently. This suggests that experiments should be more easily performed at constant temperature.

As already mentioned, the most biologically-relevant feature in our phase diagrams is certainly the stabilization of the globular conformations induced by the presence of an adsorbing surface. We must emphasize that the doubts concerning the thermodynamic limit of infinite chains do not affect the biological implications of our findings, since all real molecules have a finite length. Moreover the effect is rather strong and should therefore concern even large proteins. In order to roughly evaluate the shift in the case of a temperature-driven folding transition, we can consider the ratio $\frac{\omega_b(T_{\Theta_1})}{\omega_b(T_{\Theta_2})} \approx \frac{3}{2}$ (see figure 3.2) where T_{Θ_1} and T_{Θ_2} refer to the collapse temperatures far and near adsorption, respectively. Then assuming $\varepsilon_b \propto T^\alpha$ we obtain

$$T_{\Theta_2} = T_{\Theta_1}\left(\frac{2}{3}\right)^{\frac{1}{\alpha-1}}. \qquad (5.1)$$

We see that if (as implicitly supposed until now in this thesis) ε_b is temperature independent, that is $\alpha = 0$, we have $T_{\Theta_2} = \frac{3}{2}T_{\Theta_1}$, a rather strong temperature enhancement indeed. For decreasing values of $\alpha < 0$ the shift decreases as well. More suprisingly, if $\alpha > 1$ our shift corresponds to a drop in the folding transition temperature and thus to a coil stabilisation. It is not our purpose to consider other, more realistic, relations for $\varepsilon_b(T)$. We simply note that, on the basis of this discussion, we consider of fundamental importance, when looking at the protein folding transition, to take into account the effect of surrounding structures such as cell membranes or larger macromolecules.

Membrane proteins with an hydrophobic "anchor" embedded in the lipid bilayer and a long segment moving freely in the vicinity of the membrane are the biological systems which are more closely represented by our models. If the free part weakly interacts with the membrane surface, for example through screened electrostatic or hydrophobic binding without penetration (see also figure 1.2) and if such interactions are modulated by environmental factors such as ionic concentration, then, according to the phase diagram we have calculated, the folding of the free part of the chain can occur as a result of an increased interaction with the surface. A similar behaviour can also be expected for bulk proteins bound to other membrane proteins. Since conformational changes of proteins upon adsorption seems to play a fundamental rôle in some cell processes such as for instance post-translational translocation (see Section 1.1.2) we strongly believe that the importance of the structure of the collapse-adsorption phase diagram cannot be overestimated.

To conclude, the point we would like to raise in these final considerations

concerns the possible extensions of our results to the protein folding itself. The steric constraint due to side chain residues has been recognised to destabilize the folded conformation. Our claim of a surface-induced stabilisation of the folded conformation corresponds to a somewhat opposite phenemenon where the steric constraint is not related to the chain itself but to its environment. Let us now consider the folding process. An inhomogeneous chain is very likely to present some nucleation process with one part of the chain folding before the other. The steric contraints imposed on the rest of the chain by the rather compact and large nucleus can then induce its folding. Such a working hypothesis could then explain the formation of protein domains and our result would be susceptible to form the basis of further theoretical investigations directed, for instance, to the prediction of the protein domain sizes. Folding processes controlled by enzymes could obviously find similar theoretical bases.

APPENDIX A

COMPUTER ALGORITHM

We have generated walks on a triangular lattice for the SASAW-model and on a square lattice for the standard model. The algorithm structure is similar in both cases. The problem consists in generating, and classifying, suitable walks for the considered model. The classification parameters are the total number of bonds, the number of monomer-surface interactions and the number of monomer-monomer contacts.

The enumeration algorithm is based on a labelling of sites that allows for each site to "know" its connected neighbours and to "visit" one of them in a step of the walk, remembering which sites have already been visited. In such way, from visit to visit a walk is generated through the lattice. A visit is only valid if it does not violate the conditions imposed by the model. The walk stops when an "end site" is encountered, that is when a site cannot visit any of its neighbours either because they have all been visited, because the maximum allowed number of steps has been reached or because they are not valid sites. At this point, the currently generated walk, recorded as a sequence of sites from the origin to the ending site, is analysed. This means that the classification parameters characterising the walk are determined, and that the corresponding element of the configuration array (see appendix B for the standard model) is updated. In order to continue the generating process, the ending site is removed and, provided that the new final site is not itself an ending site, the walk is allowed to visit another region of the cell. This procedure is repeated until the origin itself becomes an ending site.

Besides the lattice structure, the main differences between the two models concern the valid sites and the analysis process to which a walk is submitted before being recorded. For the SASAW-model a site is valid, and therefore can be visited, only if it has not been already visited more than once and if it belongs to the considered small lattice cell. A walk is only recorded if it spans the cell, from a given origin (a corner of the cell lying on the given surface), in a direction parallel to the surface. Moreover overlapping walks are considered unphysical and must be eliminated. For the standard model a site is valid if it has never been visited before and if it belongs to the allowed half space. When a walk is recorded, its radius of gyration is calculated and stored as well.

It should be noticed that, for the SASAW-model, our renormalization scheme requires a fixed value for the minimum adsorbed fraction of bonds x for each configuration. However this check can be performed directly on the final configuration tables which are given in ref. [6]. The configuration tables for the standard model are presented in Appendix B for the case of isotropic walks.

APPENDIX

In the following table we give the coefficients $C_n(i,j)$ for two-dimemsional isotropi walks with their origin on an impenetrable wall. A square lattice is considered. Th indices n, i and j refer to the total number of steps, the number of monomer-monome interactions and the number of monomer-surface interactions, respectively.

j = 0

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	818369	1328860	1311994	935356	578332	301160	140712	58810	21104	4274	32				
n=18	1888233	3268254	3406210	2566850	1662736	909496	455796	187662	73908	25702	3182				
n=19	4362273	8007152	8782794	6965936	4709774	2717136	1407770	642448	258506	95354	28236	1526			
n=20	10069649	19560136	22528134	18758858	13215558	7983882	4317750	2089926	902390	344402	124832	27486	1010		
n=21	23268551	47638934	57432964	50069558	36708548	23101962	13062568	6566248	3045182	1235754	455850	151148	32278	172	
n=22	53731117	115745626	145846832	132818662	101140202	66202872	38919450	20427456	9912650	4298732	1688564	595340	175726	18914	

j = 1

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	502334	1033708	1132658	879486	568544	314052	155244	64042	24454	8820	702				
n=18	1229840	2518272	2896794	2369336	1603258	934314	474544	217730	84800	29602	8004	176			
n=19	2831868	6108366	7389948	6347796	4496468	2734332	1476756	710020	302676	113532	41996	7978	28		
n=20	6543766	14806428	18730480	16838822	12439558	7878210	4446526	2216376	1012236	408522	140804	45056	4560		
n=21	15080040	35764136	47302978	44456796	34184362	22567736	13247306	6962708	3033122	1434494	554916	192214	55402	3056	
n=22	34852104	86338872	118940818	116476668	93031680	63735716	38778490	21221028	10540640	4783728	1971728	700476	229142	46430	1236

j = 2

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	260252	451230	504418	402514	279036	156472	85574	36952	14214	6574	1428				
n=18	602816	1106498	1290664	1063006	775184	465616	251706	123018	50812	18864	7402	490			
n=19	1389704	2702374	3291986	2892160	2149224	1357152	762692	388964	179028	67968	27942	9398	448		
n=20	3218158	6586306	8351850	7656450	5909300	3860262	2284692	1182360	580908	242256	90876	34926	6760		
n=21	7423316	15998324	21118486	20167556	16135948	10961022	6731574	3650514	1846828	843480	341104	124828	48366	7308	
n=22	17186696	38802028	53175140	52790984	43664104	30778972	19452402	11037280	5712570	2728914	1187248	446046	164820	47878	3062

j = 3

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	127982	215774	225380	176802	122642	71848	38966	18024	7072	3526	820				
n=18	297868	532352	582648	476218	343350	208908	117126	57558	24544	9758	4454	316			
n=19	688724	1306098	1501700	1286238	957430	614382	350564	183894	87712	35428	14218	5592	448		
n=20	1600888	3199294	3839018	3424018	2629480	1750288	1039118	557538	280180	119182	47598	20362	4382		
n=21	3701626	7799454	9785644	9096872	7232602	4975402	3083964	1708854	888152	420996	175702	66432	27434	6062	
n=22	8596840	18992114	24801478	23924632	19618874	13938288	8906916	5124304	2718790	1347092	590554	235684	93926	29284	3696

j = 4

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	61444	99784	100710	76224	50636	30520	16344	7130	3262	1632	188				
n=18	143852	248202	263064	207834	144964	88676	48300	23870	10440	4124	1802	50			
n=19	334124	613004	684502	565778	412506	258364	151824	78080	36972	15478	7008	2428	76		
n=20	780342	1510792	1764434	1522908	1143324	746754	445184	236584	118982	53086	20292	8818	1732		
n=21	1810624	3702958	4528802	4084156	3167808	2155964	1324990	740732	389660	180858	81708	30544	14082	2124	
n=22	4221130	9064252	11551400	10832986	8673646	6070804	3855542	2313726	1184464	587654	263884	105616	41032	14474	892

j = 5

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	28882	45168	43830	31474	20954	11422	6300	2394	1418	400	12				
n=18	68004	113184	115182	88340	58858	36006	18230	9372	3756	1716	330				
n=19	158980	282000	303796	244116	170880	105482	59634	30640	13486	5894	2796	414			
n=20	373230	699726	789196	660968	484874	304876	182270	92648	46684	20522	7912	3160	328		
n=21	870324	1726482	2046814	1795364	1358856	897110	543288	300008	153528	70768	30728	12532	4890	192	
n=22	2037944	4249822	5254452	4800008	3750448	2571084	1590318	911824	479292	233214	107162	41308	16380	4822	84

j = 6

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	13388	20140	18460	12544	8164	4078	2156	826	512	24					
n=18	31630	50778	49658	35820	23844	13036	6558	3184	1258	406	32				
n=19	74442	127518	131588	101670	68268	41646	21706	11234	4462	2378	634	14			
n=20	175596	318570	346164	280816	197016	121812	68928	33396	17080	6406	2736	562	14		
n=21	411632	791558	905864	767474	565242	360112	214852	113166	57196	25178	10756	4604	784		
n=22	968328	1960342	2347156	2077632	1583698	1048118	633146	353888	177404	86250	36646	13608	5230	634	

j = 7

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	6128	8894	7594	5076	3780	1476	626	224	74						
n=18	14558	22564	20796	14242	9146	4566	2258	988	396	30					
n=19	34484	56944	56244	40778	27104	14694	7548	3604	1298	674	38				
n=20	81528	143112	149124	115706	77752	47012	24616	11706	5546	2086	622	24			
n=21	192320	358360	394120	322164	225950	140950	79256	39236	19908	7414	3660	856	30		
n=22	453844	891822	1030272	879920	648968	414192	245088	129754	61936	30398	11238	4352	934	14	

j = 8

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	2776	3774	3134	1858	984	478	206	38	2						
n=18	6616	9892	8546	5686	3088	1650	674	234	54						
n=19	15780	25188	23364	16122	10242	5198	2496	1050	380	82					
n=20	37538	63634	63428	46224	30594	16434	8296	3810	1504	498	44				
n=21	88976	160194	168812	131506	88512	53198	27752	13308	6126	2142	968	42			
n=22	210680	401776	446296	366996	256918	160272	89074	44074	20788	8922	3144	796	40		

j = 9

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	1254	1596	1198	666	334	112	34	2							
n=18	2976	4156	3504	2062	1090	518	214	38	2						
n=19	7122	10952	9592	6336	3422	1834	722	336	52	2					
n=20	17056	28012	26158	18168	11434	5694	2726	1116	400	58					
n=21	40746	70856	71280	52264	34422	18368	9038	4212	1580	470	102				
n=22	96908	178700	190424	148888	100300	59814	31084	14692	6408	2418	742	48			

j = 10

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	526	740	412	218	118	24	2								
n=18	1336	1752	1306	742	360	114	34	2							
n=19	3182	4558	3900	2286	1202	558	222	38	2						
n=20	7646	12074	10738	7028	3782	2028	770	238	52	2					
n=21	18384	31040	29188	20382	12728	6322	2958	1182	410	56	2				
n=22	44112	78620	79808	58912	38572	20462	10350	4616	1660	484	74				

j = 11

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	224	256	204	72	22	2									
n=18	552	816	446	336	124	24	2								
n=19	1420	1916	1418	824	386	116	34	2							
n=20	3304	4980	4332	2532	1320	598	230	38	2						
n=21	8188	13258	11990	7764	4168	2232	818	240	52	2					
n=22	19764	34278	32466	22768	14132	6992	3192	1248	420	56	2				

j = 12

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	98	80	60	14	2										
n=18	234	272	230	76	22	2									
n=19	578	896	482	254	130	34	2								
n=20	1506	2088	1534	912	412	118	34	2							
n=21	3612	5422	4770	2802	1444	638	238	38	2						
n=22	8748	14504	13354	8546	4580	2446	866	242	52	2					

j = 13

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	50	32	12	2											
n=18	102	84	64	14	2										
n=19	344	388	358	80	32	2									
n=20	604	980	520	272	136	34	2								
n=21	1594	2268	1654	1006	438	120	34	2							
n=22	3836	5884	5244	3098	1574	678	246	38	2						

j = 14

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	10	32	2												
n=18	52	34	12	2											
n=19	106	88	68	14	2										
n=20	254	304	288	84	22	2									
n=21	630	1068	560	390	142	34	2								
n=22	1684	2456	1778	1106	464	122	34	2							

j = 15

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	4	2													
n=18	10	34	2												
n=19	54	36	12	2											
n=20	110	92	72	14	2										
n=21	264	320	320	88	22	2									
n=22	656	1160	602	308	148	24	2								

j = 16

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	2														
n=18	4	2													
n=19	10	36	2												
n=20	56	38	12	2											
n=21	114	96	76	14	2										
n=22	274	336	354	92	32	2									

j = 17

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	2														
n=18	2														
n=19	4	2													
n=20	10	38	2												
n=21	58	40	12	2											
n=22	118	100	80	14	2										

j = 18

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18	2														
n=19	2														
n=20	4	2													
n=21	10	40	2												
n=22	60	42	12	2											

j = 19

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18															
n=19	2														
n=20	2														
n=21	4	2													
n=22	10	42	2												

j = 20

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18															
n=19															
n=20	2														
n=21	2														
n=22	4	2													

j = 21

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18															
n=19															
n=20															
n=21	2														
n=22	2														

j = 22

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18															
n=19															
n=20															
n=21															
n=22	2														

In the the following table we give the coefficients $D_n(i,j)$ which refer to the normal component of the radius of gyration for two-dimemsional isotropic walks with their origin on an impenetrable wall. A square lattice is considered. The indices n, i and j refer to the total number of steps, the number of monomer-monomer interactions and the number of monomer-surface interactions, respectively.

j = 0

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	5239973	7262162	6254517	3844184	2051087	913233	368031	122804	41879	7490	55				
n=18	13195274	19630650	17973493	11765036	6625915	3127066	1354623	485104	169178	50921	5597				
n=19	33093848	52584962	50964406	35387058	20952796	10494942	4740708	1881045	668215	217908	57042	2807			
n=20	82629125	139769966	142993175	104923978	65158323	34401706	16316392	6892464	2821546	889722	286607	55443	1873		
n=21	205630514	368819342	397010850	306838436	199441251	110446410	55016146	24291017	9940006	3571439	1184285	348987	46494	350	
n=22	509846737	967066415	1092812018	867082663	602097416	346661052	181577037	84126342	36135176	13876608	4885744	1550601	408142	39230	

j = 1

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	3458096	5713053	5455411	3680374	2061514	977774	415515	147919	49970	16021	1238				
n=18	8709673	15299301	15466276	11067252	6541223	3298498	1450605	579691	195867	59589	15152	332			
n=19	21804327	40611819	43382289	32827482	20445912	10844240	5104619	2136291	798431	266034	86510	15687	76		
n=20	54441613	107102600	120357940	95915784	62694864	34869235	17272054	7531377	3032652	1074484	328626	95677	9370		
n=21	135265444	280456232	331004268	277173646	189649527	110671050	57312792	26509298	11188431	4257700	1472786	458445	122134	6399	
n=22	335351087	730510139	902464252	791476442	565494809	344510765	188028856	90051773	39581216	15947114	5846373	1859766	550700	105213	2679

j = 2

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	1549686	2270288	2214331	1514909	906171	434208	205158	74687	26010	10681	2196				
n=18	3931319	6156028	6314342	4583915	2859408	1478910	694416	290659	105876	35212	12879	766			
n=19	9897346	16533478	17816903	13654706	8910545	4877622	2393475	1061092	424378	143080	53492	16858	797		
n=20	24854185	44044915	49702017	40054536	27316250	15613735	8110124	3661439	1580719	580141	197112	69531	12629		
n=21	62033436	116437479	137421236	116144747	82564137	49419671	26776108	12736342	5697785	2283008	824909	275549	101098	14175	
n=22	154515704	305839125	376532987	332847575	245986720	153738395	86266405	43180871	19807449	8365124	3248123	1096285	375932	102142	6184

j = 3

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	671126	948002	855295	568509	335893	166340	76547	29899	10989	4848	1009				
n=18	1720834	2606053	2486733	1743806	1081763	558943	271634	112987	42538	15644	6239	406			
n=19	4371432	7082249	7142966	5295102	3422314	1884795	933256	421610	173564	63855	24052	8164	638		
n=20	11071258	19072367	20224702	15753026	10589867	6111263	3154524	1476436	645832	241384	88625	34801	6749		
n=21	27834744	50890002	56710621	46361547	32482239	19527600	10597247	5126841	2347454	968742	368044	129362	48484	9900	
n=22	69819685	134830397	157289089	134407068	97764109	61137888	34449939	17378247	8157117	3557867	1387377	502972	188188	53274	4610

j = 4

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	279154	376135	322011	202146	113572	57172	25210	9521	4173	1711	166				
n=18	725324	1052014	956534	635982	378286	193898	90541	37721	14742	5516	1973	60			
n=19	1863306	2901569	2801328	1971702	1235014	655738	332202	144699	59462	23240	9651	2678	95		
n=20	4769419	7917360	8063456	5990551	3895294	2181313	1123032	515297	225086	88058	32104	12232	2169		
n=21	12097856	21374122	22943784	17959420	12155308	7154290	3815174	1851272	849659	344312	141809	50765	19634	2811	
n=22	30607665	57233387	64467798	52915922	37235247	22716981	12618522	6307187	2970743	1290660	515357	191643	69635	21469	1296

j = 5

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	111993	143023	115496	67749	37474	16824	7640	2650	1334	323	8				
n=18	295213	407996	350854	222134	124557	63341	27393	11545	4557	1770	313				
n=19	769148	1146135	1052247	707928	420299	218182	104387	45260	17816	7158	2930	364			
n=20	1992045	3175487	3086286	2191942	1372263	732170	375459	164217	70737	28334	10505	3388	309		
n=21	5109490	8693802	8944116	6718531	4378672	2470724	1291316	610467	272508	110343	44272	16541	5410	228	
n=22	13052756	23565182	25499721	20142172	13653540	8071476	4332704	2148408	989780	419159	173693	63456	22617	5511	101

j = 6

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	43472	52255	39298	21430	11569	4698	2085	816	330	15					
n=18	116498	152579	123791	72890	40329	18080	7911	3123	1222	330	14				
n=19	308275	437353	379193	241689	136102	69316	30648	13104	5117	2206	489	9			
n=20	808960	1233646	1139459	771748	458159	238360	113794	48078	20748	7492	2824	476	9		
n=21	2100939	3431292	3367061	2411490	1513321	813951	417899	187657	81776	32780	13006	4675	677		
n=22	5426730	9432096	9776360	7393184	4833129	2729454	1424933	679130	300140	126572	49720	17216	5643	585	

j = 7

=	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
ı=17	16336	18400	12659	6694	3062	1316	507	167	32						
ı=18	44633	55060	41571	22845	12238	5019	2166	889	282	18					
ı=19	120262	161522	131927	78042	43224	19382	8579	3406	1244	448	24				
ı=20	320100	464642	405678	259729	146352	74352	32965	13698	5762	2036	486	15			
ı=21	843319	1316913	1226501	836133	496277	260226	124057	53156	23034	8229	3371	669	20		
ı=22	2203000	3675547	3633089	2618127	1643553	886648	452750	204125	86429	36933	13350	4426	788	9	

j = 8

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	5923	6134	3922	1885	806	356	142	18	0						
n=18	16583	19170	13249	7032	3211	1397	524	168	27						
n=19	45588	57653	43739	24228	12901	5347	2296	926	277	34					
n=20	123550	169905	139726	83023	45974	20598	9111	3551	1323	374	27				
n=21	330488	491070	431897	277815	156855	79680	35444	14700	6164	2086	635	25			
n=22	875164	1396669	1310694	898461	532739	280174	132882	56906	23921	9087	3058	638	26		

j = 9

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	2077	1930	1162	471	215	69	16	0							
n=18	5952	6318	4076	1950	842	372	144	18	0						
n=19	16773	19878	13815	7349	3358	1475	539	166	27	0					
n=20	46384	60075	45819	25575	13547	5667	2414	957	282	28					
n=21	126410	177824	147306	87947	48687	21831	9668	3770	1356	362	44				
n=22	339857	516129	457579	295644	167246	84873	37832	15602	6382	2211	546	28			

j = 10

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	692	595	291	117	63	10	0								
n=18	2070	1970	1187	492	220	68	16	0							
n=19	5967	6484	4223	2012	877	385	145	18	0						
n=20	16919	20536	14363	7651	3502	1551	552	164	26	0					
n=21	47055	62359	47832	26897	14183	5980	2523	985	285	28	0				
n=22	128931	185363	154712	92836	51357	23059	10224	3977	1382	362	37				

j = 11

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	218	169	75	27	9	0									
n=18	683	606	293	120	64	10	0								
n=19	3060	2007	1210	512	224	66	16	0							
n=20	5972	6638	4367	2074	910	398	146	18	0						
n=21	17033	21153	14901	7940	3646	1627	563	162	26	0					
n=22	47629	64531	49797	28200	14817	6288	2624	1010	287	27	0				

j = 12

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	64	37	21	4	0										
n=18	214	170	77	27	9	0									
n=19	674	616	294	124	65	10	0								
n=20	2048	2040	1232	533	227	65	16	0							
n=21	5869	6781	4508	2136	941	409	147	17	0						
n=22	17121	21737	15433	8220	3789	1701	573	159	26	0					

j = 13

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	19	7	3	0											
n=18	62	36	22	4	0										
n=19	209	170	80	26	9	0									
n=20	665	626	296	126	65	10	0								
n=21	2036	2071	1252	554	231	64	15	0							
n=22	5961	6915	4646	2198	971	419	148	17	0						

j = 14

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	4	4	0												
n=18	19	7	3	0											
n=19	60	36	22	4	0										
n=20	205	170	82	26	8	0									
n=21	656	635	298	129	66	9	0								
n=22	2022	2100	1271	576	234	63	15	0							

j = 15

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	1	0													
n=18	4	4	0												
n=19	18	7	3	0											
n=20	59	36	22	4	0										
n=21	202	171	85	26	8	0									
n=22	647	644	300	132	66	9	0								

j = 16

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17	0														
n=18	1	0													
n=19	4	4	0												
n=20	18	7	3	0											
n=21	57	35	23	4	0										
n=22	198	171	88	26	8	0									

j = 17

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18	0														
n=19	1	0													
n=20	3	4	0												
n=21	18	7	3	0											
n=22	56	35	23	4	0										

j = 18

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18															
n=19	0														
n=20	1	0													
n=21	3	4	0												
n=22	17	7	3	0											

j = 19

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18															
n=19															
n=20	0														
n=21	1	0													
n=22	3	4	0												

j = 20

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18															
n=19															
n=20															
n=21	0														
n=22	1	0													

j = 21

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18															
n=19															
n=20															
n=21															
n=22	0														

j = 22

i =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n=17															
n=18															
n=19															
n=20															
n=21															
n=22															

REFERENCES

[1] A.L.Lehniger, *Biochemistry* (Worth Publishers Inc., New York 1981)

[2] T.E. Creighton, *Proteins* (W.H.Freemann & Co., New York 1984)

[3] R.Cantor and P.R Schimmel, *Biophysical Chemistry* (W.H.Freemann & Co., San Francisco 1980)

[4] K.A.Dill and D.O.V.Alonso, *Protein Structure and Protein Engineering, Colloquium Mosbach 1988* (Springer-Verlag, Berlin 1988).

[5] W.Kausmann, *Adv. Protein Chem.* **14** (1959) 1

[6] S.Cattarinussi and G.Jug, *J.Phys.II(Paris)* **1** (1991) 397

[7] S.Cattarinussi and G.Jug, *J.Phys. A:Math.Gen* **23** (1990) 2701

[8] S.Cattarinussi and G.Jug, *SISSA Preprint* (1991)

[9] S.N.Timasheff, *The Enzymes*, P.D.Boyer(ed) (Academic Press, New York 1970)

[10] D.C.Teller et al., *Methods Enzymol.* **61** (1979) 103

[11] J.W.Donovan, *Methods Enzymol.* **27** (1973) 497

[12] G.Weber, *Ann.Rev.Biophys.Bioeng.* **1** (1972) 553

[13] I.Z.Steinberg, *Ann.Rev.Biophys.Bioeng.* **7** (1978) 113

[14] J.C.Sutherland and B.Holmquist, *Ann.Rev.Biophys.Bioeng.* **9** (1980) 293

[15] K.Nagayama et al, *Eur.J.Biochem.* **114** (1981) 375

[16] C.B.Anfinsen, *Science* **181** (1973) 223

[17] K.A.Dill et al., *Biochemistry* **28** (1989) 5439

[18] P.L.Privalov and N.N.Khechinashvili, *J.Mol.Biol.* **86** (1974) 665

[19] P.L.Privalov, *Adv.Protein Chem.* **33** (1979) 167

[20] P.Kollman and W.F.Van Gunsteren, *Methods in Enzymology* **154** (1987) 430

[21] P.Y.Chou and G.M.Fasman, *Adv.Enzymol.* **47** (1978) 45

[22] W.Kabsh and C.Sander, *FEBS Letters* **155** (1983) 179

[23] S.Pongor, *Methods in Enzymology* **154** (1987) 450

[24] J.-R.Garel et al., *J.Phys.(Paris)* **50** (1989) 3067

[25] E.I.Shakhnovich and A.M.Gutin, *Nature* **346** (1990) 773

[26] A.V.Finkelstein and B.A.Reva, *Nature* **351** (1991) 497

[27] H.S.Chan and K.A.Dill, *J.Chem.Phys.* **90** (1989) 1

[28] R.B.Gennis, *Biomembranes: Molecular Structure and Function* (Springer Verlag, New York 1989)

[29] G.Schatz, *Nature* **321** (1986) 108

[30] M.G.Watres and G.Blobel, *J.Cell Biol.* **102** (1986) 1543

[31] W.Hansen et al., *Cell* **45** (1986) 397

[32] E.Perara et al., *Science* **232** (1986) 348

[33] A.Kuhn et al., *Nature* **322** (1986) 335

[34] A.Papoulis, *Probability, Random Variables, and Stochastic Processes, 2nd edition* (McGraw-Hill International Editions, 1984)

[35] H.B.Callen, *Thermodynamics and an Introduction to Thermostatistic, 2nd edition* (John Wiley & Sons, New York 1985)

[36] D.J.Amit, *Field Theory, the Renormalisation Group, and Critical Phenomena* (McGraw-Hill, 1978)

[37] P.G.de Gennes, *Phys. Lett.* **A38** (1972) 339

[38] M.Daoud et al., *Macromolecules* **8** (1975) 804

[39] V.Emery, *Phys. Rev.* **B11** (1975) 239

[40] D.Jasnow and M.Fisher, *Phys. Rev.* **B13** (1976) 1112

[41] G.Jug and G.Rickayzen, *J.Phys.A:Math.Gen.* **14** (1981) 1357

[42] P.G.de Gennes, *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca 1979)

[43] P.Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, New York 1971)

[44] S.P.Obukhov, *J.Phys.A:Math.Gen.* **17** (1984) L965

[45] C.Domb et al., *Proc.Phys.Soc.* **85** (1965) 625

[46] H.J.Hilhorst, *Phys.Lett.* **A38** (1976) 339

[47] H.J.Hilhorst, *Phys. Rev.* **B16** (1977) 1253

[48] B.Shapiro, *J.Phys.C:Solid State Phys.* **11** (1978) 2829

[49] A.Coniglio and M.Daoud, *J.Phys.A:Math.Gen.* **12** (1979) L259

[50] B.Derrida, *J.Phys.A:Math.Gen.* **14** (1981) L5

[51] H.E Stanley, P.J.Reynolds, S.Redner, F.Family, *Real Space Renormalisation* (Springer-Verlag, Berlin 1982)

[52] J.des Cloiseaux, *J.Phys.(Paris)* **42** (1981) 635

[53] R.Villanove and F.Rondelez, *Phys.Rev.Lett.* **45** (1980) 1502

[54] B.Derrida and H.Saleur, *J.Phys.A:Math.Gen.* **18** (1985) L1075

[55] K.Kremer and J.W.Lyklema, *J.Phys.A:Math.Gen.* **18** (1985) 1515

[56] J.L.Cardy, in: Phase Transitions and Critical Phenomena **11**, C.Domb and J.L.Lebowitz (editors) (Academic, New York 1987)

[57] A.Coniglio et al., *Phys.Rev.B* **35** (1987) 3617

[58] F.Seno and A.Stella, *J.Phys.(Paris)* **49** (1988) 739

[59] I.C.Sanchez, *Macromolecules* **12** (1979) 980

[60] P.G.de Gennes, *J.Phys.(Paris)* **36** (1975) L55

[61] R.J.Griffiths, *Phys.Rev.* **141** (1973) 545

[62] I.D.Lawrie and S.Sarbach, in: Phase Transitions and Critical Phenomena **9**, C.Domb and J.L.Lebowitz (editors) (Academic, New York 1984)

[63] A.Aharony, in: *Critical Phenomena*, Lecture Notes in Physics **186**, F.J.W.Hahan (editor) (Springer-Verlag, Berlin 1983)

[64] L.P.Kadanoff, in: Phase Transitions and Critical Phenomena **5a**, C.Domb and M.E.Green (editors) (Academic, New York 1976)

[65] A.Maritan et al., *Physica A* **156** (1989) 679, (North-Holland, Amsterdam)

[66] G.Jug, *J.Phys.A:Math.Gen.* **20** (1987) L503

[67] P.G de Gennes, *Rep.Prog.Phys* **32** (1969) 187

[68] G.Jug, *Annals of Physics* **142** (1982) 140

[69] M.E.Fisher, *Journal of Statistical Physics* **34** (1984) 667

[70] P.G.de Gennes, *J.Phys.(Paris)* **37** (1976) 1445

[71] K.Binder, in: Phase Transitions and Critical Phenomena **10**, C.Domb and J.L.Lebowitz (editors) (Academic, New York 1986)

[72] E.Eisenriegler et al., *J.Chem.Phys* **77** (1982) 6296

[73] I.Guim and T.W.Burkhardt, *J.Phys.A:Math.Gen.* **22** (1989) 1131

[74] F.van Dieren and K.Kremer, *Europhysics Letters* **4** (1987) 569

[75] E.Eisenriegler, *Phys.Rev.B* **37** (1988) 5257

[76] E.Bouchaud and J.Vannimenus, *J.Phys.(Paris)* **50** (1989) 2931

[77] A.R.Veal et al., *J.Phys.A:Math.Gen.* **23** (1990) L-109

[78] D.P.Foster, *J.Phys.A:Math.Gen.* **23** (1990) L-1135

[79] K.Kremer, *J.Phys.A:Math.Gen.* **16** (1983) 4333

[80] T.Ishinabe, *J.Chem.Phys.* **76** (1982) 5589

[81] T.Ishinabe, *J.Chem.Phys.* **77** (1982) 3171

[82] T.Ishinabe, *J.Chem.Phys.* **80** (1983) 1318

[83] T.Ishinabe, *J.Phys.A:Math.Gen* **18** (1985) 3181

[84] V.Privman, *J.Phys.A:Math.Gen* **19** (1986) 3287

[85] S.P.Obukhov, *J.Phys.A:Math.Gen* **19** (1986) 3655

[86] R.B.Stinchcombe, in: Phase Transitions and Critical Phenomena **7**, C.Domb and J.L.Lebowitz (editors) (Academic, New York 1983)

[87] K.Binder and A.P.Young, *Rev.Mod.Phys.* **58** (1986) 801

[88] H.Frauenfelder, *Helv.Phys.Acta* **57** (1984) 165

[89] R.Rammal et al., *Rev.Mod.Phys* **58** (1986) 765

[90] D.Andelmann and A.N.Berker, *Phys.Rev.B* **29** (1983) 2630