

Collapse transition of SARWs with hydrophobic interaction on a two dimensional lattice

Mathieu Gaudreault

Master of Science

Department of Physics

McGill University

Montreal, Quebec

2007-08-29

A thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of requirements for the degree of
Master of Science

©Mathieu Gaudreault, 2007



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-51269-2
Our file Notre référence
ISBN: 978-0-494-51269-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Acknowledgements

This project would not have been possible without the support of many people. Many thanks to my advisor, Professor Jorge Viñals, for his patience and availability. His knowledge always gives me motivation and creative ideas through my research. Thanks to my mother, girlfriend, and friends, who always offered me support and love at each step of the process.

Abstract

We study the collapse transition of a lattice based protein model including an explicit coarse-grained model of a solvent. This model accounts for explicit hydrophobic interactions, and it is studied by Monte Carlo simulation. The protein is modelled as self-avoiding random walk with nearest neighbor interactions on a two dimensional lattice. Without the solvent, universal quantities of the chain around the collapse transition temperature are well known. Hydrophobicity is then modelled through a lattice of solvent molecules in which each molecule can have Q states depending of an orientation variable. Only one state is energetically favored, when two neighboring solvent molecules are both in the same state of orientation. The monomers are placed in interstitial position of the solvent lattice, and are only allowed to occupy sites surrounded by solvent cells of the same orientation. The potential of mean force between two interstitial solute molecules is calculated, showing a solvent mediated attraction typical of hydrophobic interactions. We then show that this potential increases with the energy of hydrogen bond formation as it appears in the model, while its characteristic range decreases. More importantly, we show that the chain embedded in the solvent undergoes a collapse transition, with the temperature of the transition being shifted relative to that of the chain in isolation. We calculate several critical exponents near the collapse transition, and we observe that their values are not conserved in presence of the explicit solvent.

Abrégé

La transition d'une protéine modélisée avec un modèle explicite de solvant incluant l'effet hydrophobique est étudiée en performant des simulations numériques avec la méthode de Monte Carlo. La protéine est modélisée comme un chemin aléatoire ne pouvant s'intersecter sur lui-même sur un réseau en deux dimensions en utilisant l'algorithme dit de pivot. Sans le solvant, les mesures statistiques universelles de la chaîne autour de la température critique sont exactement connues. L'effet hydrophobique est modélisé comme un réseau carré de molécule de solvant où chaque molécule peut adopter Q états d'orientation spatiale. Seulement un état d'orientation est énergétiquement favorisé, appelé l'orientation spéciale. Les monomères sont placés entre les molécules de solvant sur le réseau et peuvent être accommodés entre deux molécules de solvant seulement si ces deux molécules sont dans leur état d'orientation spéciale. Le potentiel de force moyen entre deux sites pouvant accommoder un monomère est calculé et démontre que l'amplitude de l'attraction augmente en augmentant l'énergie libre de formation de bonds hydrogènes alors que son rayon d'action diminue. Nous démontrons qu'une transition d'un état globulaire à basse température à un état étendu à haute température est incluse dans le modèle lorsque les effets relatifs au solvant sont considérés. La température critique et les exposants universels sont modifiés par les effets hydrophobiques, démontrant que la transition de la chaîne appartient à une classe universelle différente de la transition de phase de la protéine sans solvant.

Table of Contents

1	Introduction	1
2	A coarse grained model of hydrophobicity	7
2.1	Lattice model of hydrophobicity	9
2.2	Reversible work theorem and potential of mean force	11
2.3	Phase transition of the solvent model	15
2.4	Monte Carlo methods and BKL algorithm	16
2.5	Measures analysis of the solvent model	19
2.5.1	P_{11}	22
2.5.2	Heat capacity of the solvent model	24
2.5.3	Order of parameter M	25
2.5.4	Potential of mean force acting on a solute molecule	26
3	Collapse transition of self-avoiding random walks	29
3.1	SARWs with nearest neighbor interaction	30
3.2	SARWs in the BKL scheme using the pivot algorithm	31
3.3	Statistical measures of protein conformation	34
3.3.1	Autocorrelation functions	35
3.3.2	The end to end distance and radius of gyration	36
3.3.3	Derivative of the end to end distance and radius of gyration	37
3.3.4	Specific heat of the chain	38
3.3.5	Partition function of the protein model	39
3.3.6	Structure factor	41
3.4	Tricritical temperature and critical exponents	42
3.4.1	Specific heat of the chain and the tricritical temperature	42
3.4.2	The exponent ν	43
3.4.3	The exponent γ	45
3.4.4	The crossover exponent ϕ	45

3.4.5	Structure factor and tricritical temperature	47
4	Collapse transition with hydrophobic interaction	49
4.1	Thermodynamics of the combined system	49
4.2	Algorithm for chain conformation with solvent effect	54
4.3	Collapse transition with hydrophobic interaction	57
4.3.1	Solvent model with chemical potential	62
4.3.2	Thermodynamical work	63
4.3.3	Distance functions and their derivatives	64
4.3.4	Heat capacity and fluctuations of the shell molecule number	66
4.4	Critical exponents with hydrophobic interaction	70
4.4.1	The specific heat and fluctuations of the shell molecule number	70
4.4.2	The exponent ν^* with hydrophobic interaction	72
4.4.3	The exponent γ^* with hydrophobic interaction	73
4.4.4	Crossover exponent ϕ^* with explicit solvent	74
4.5	Interpretation of the results	75
5	Conclusions	77
A	Exact potential of mean force in one dimension	81
B	Mapping between the Ising and the solvent model	89
C	Q-Potts state model in the Bragg-William approximation	93
	<i>Bibliography</i>	<i>97</i>

List of Figures

1. Introduction	1
1.1 Classification of amino acids based.	3
1.2 Illustration of the model in three dimension.	4
2. A coarse grained model of hydrophobicity	7
2.1 Schematic depiction of the orientation state of a molecule.	11
2.2 Examples of square lattice of solvent molecule for many temperatures.	21
2.3 Temperature and parameter dependence of P_{11}	22
2.4 Concentration of solvent molecules in the special state c versus T.	23
2.5 Heat capacity of the solvent C_S as a function of temperature.	24
2.6 Order of parameter M/N as a function of temperature.	25
2.7 Potential of mean force $W(r)$ as a function of the distance r	27
3. Collapse transition of self-avoiding random walks	29
3.1 Protein conformations for many temperatures.	34
3.2 Temperature dependence of $\langle R_e^2 \rangle$ and $\langle R_g^2 \rangle$ for $L = 10$	36
3.3 Temperature dependence of the detivative of $\langle R_e^2 \rangle$ and $\langle R_g^2 \rangle$	37
3.4 Specific heat of the protein C_P as a function of temperature T	39
3.5 Normalized distribution $P_L(n)$ as a function of n	40
3.6 Linear fit of the temperature of the specific heat maximum T_P vs $1/\sqrt{L}$	42
3.7 Log-log plot of $\langle R_e^2 \rangle$ and $\langle R_g^2 \rangle$ for various protein length L	43
3.8 Exponent γ around the tricritical temperature.	46
3.9 Linear fit of $\langle R'_e \rangle$ and $\langle R'_g \rangle$ for various protein length L	47
3.10 Structure factor $S(q)$ as a function of q	48

4. Collapse transition with hydrophobic interaction	49
4.1 Parameter dependence of the warm denaturation.	50
4.2 Protein conformations together with its shell of solvent molecule.	61
4.3 Probability P_{11} as a function of the temperature T with chemical potential. .	62
4.4 Thermodynamical work required to immerse the chain as a function of T . . .	63
4.5 $\langle R_e^2 \rangle$ and $\langle R_g^2 \rangle$ versus T for $L = 4$ and for many μ	64
4.6 $\langle R_e' \rangle$ and $\langle R_g' \rangle$ versus T for $L = 4$ and for many μ	65
4.7 C_P , C_ν , C_{tot} , and $\langle \Delta N_\nu^2 \rangle$ versus T	68
4.8 Correlation function $C(E_P, E_\nu)$ and $C(E_P, N_\nu)$ versus T	69
4.9 Specific heat and fluctuations of shell molecule number for many lengths. . .	71
4.10 Linear fit of the temperature of the specific heat maximum versus $1/\sqrt{L}$. . .	71
4.11 Radius of gyration versus T for many lengths.	72
Appendix	79
A. Exact potential of mean force in one dimension	81
A.1 Exact potential of mean force in one dimension.	86
B. Mapping between the Ising and the solvent model	89
C. Q-Potts state model in the Bragg-William approximation	93
C.1 Free energy as a function of m for the Q-Potts state model.	95

List of Tables

1. Introduction	1
2. A coarse grained model of hydrophobicity	7
2.1 Definition of the class of a solvent molecule.	17
2.2 Flow chart for solvent configuration updates.	20
3. Collapse transition of self-avoiding random walks	29
3.1 Flow chart for chain conformation updates.	33
3.2 Exponent ν for various temperatures.	44
3.3 Best estimates of the critical exponent ν	45
3.4 Critical exponent ν evaluated from $S(\vec{k})$ close to T_t	46
3.5 Comparison of the exact exponents to the measured one.	48
4. Collapse transition with hydrophobic interaction	49
4.1 Flow chart for solvent and hydrophobic system updates	60
4.2 Exponent ν^* for various temperatures with hydrophobic interaction.	73
4.3 Exponent γ^* for various temperatures with hydrophobic interaction.	74
4.4 Exponent ϕ^* for various temperatures with hydrophobic interaction.	75

Chapter 1

Introduction

Proteins are polymer of amino acids linked together by peptide bonds. The central dogma of biology states that information flows from DNA to messenger RNA (mRNA) by a process called transcription. Information encoded in mRNAs then flows to proteins by the process called translation. During translation, ribonucleotides of the mRNA are read by a molecular machinery formed by the ribosome and transfer RNA (tRNA), in a sequence of three nucleotides called codons. Each codon codes for a specific amino acid which is added one by one by the ribosome to the polypeptide chain. The assembled protein folds in solution with a resulting structure that depends on the linear sequence of its amino acids. This unique three dimensional structure is called the native structure. It has been observed that denaturation of proteins due to thermal change, change in the pH of the solvent or presence of some chemical agents as urea, lead to malfunction of the cell. Hence, proper folding of protein into the native structure is essential for proteins to accomplish their function.

Proteins are constructed by the polymerization of 20 amino acids [1] into a linear chain. There exist 20 amino acids, all of which have a structure consisting of a central α carbon atoms bonded to four different chemical groups, amino group (NH_2), carboxyl group (COOH), hydrogen atoms (H) or a variable group called side chains group (R). Amino acids with polar side chains are hydrophilic and tend to be on the surface of the protein while

amino acids with non polar side chains are hydrophobic and tend to aggregate to form a hydrophobic core. A classification of amino acids based on their affinity with water is shown in figure 1.1. In order to give an idea of the magnitude of the folding problem, we note that any protein consisting of n amino acids could in principle folds approximately into 8^n conformations under the simplification that only 8 bonds angles are allowed in the polypeptide backbone. Despite this large number of possible conformations, any protein folds to a unique conformation, the native structure. Obtaining the conformation in question as a function of the sequence, and the folding path from a linear conformation to the native structure is the protein folding problem.

Experimental methods to resolve protein structure rely mostly on protein crystallization. Proteins are crystallized and exposed to X-ray or electron/neutron scattering experiments to obtain a diffraction pattern of scattered particles. The solvent structure is removed from the experimental diffraction pattern and a protein representation is obtained. However, protein crystallization is a difficult process. Samples must be sufficiently large for a scattering experiment, regular in structure and without impurities; each protein reacts differently to crystallization process and there are in general no specific guidelines to perform the crystallization [2].

Theoretical methods attempt the prediction of the native structure of a protein given its linear sequence of amino acids by using approximate physical or empirical force fields and numerical simulation. It is now believed that hydration forces are the dominant interaction in the folding process [3, 4, 5], being responsible for the formation of a hydrophobic core of amino acids. The network of hydrogen bonds between solvent molecules is also essential to maintain the overall stability of the protein. Although many methods have been developed to include solvent effects in the folding problem, the interaction between the solvent and the protein is not well understood yet. To extend the understanding of the folding problem, we have studied the collapse transition of a coarse-grained protein with explicit hydrophobic

Twenty standard Amino Acids

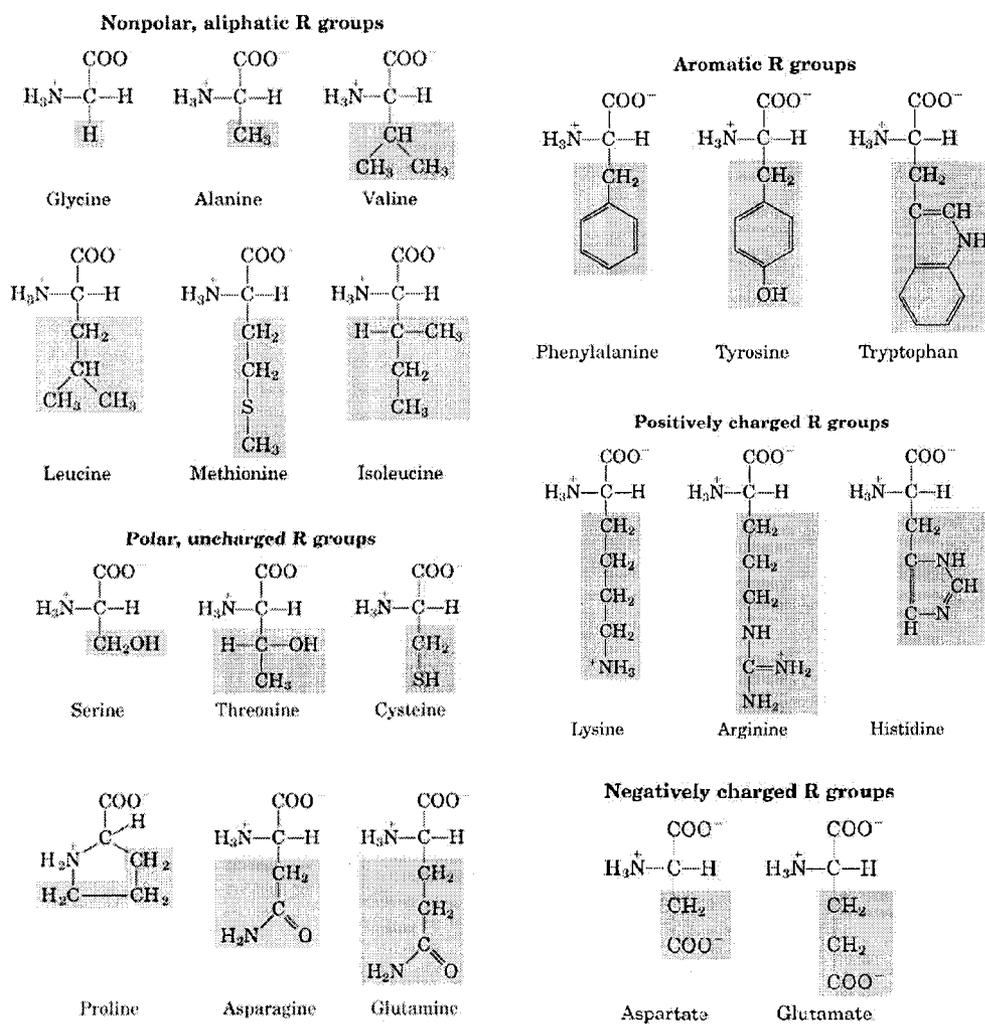


Figure 1.1: Classification of the 20 amino acids.

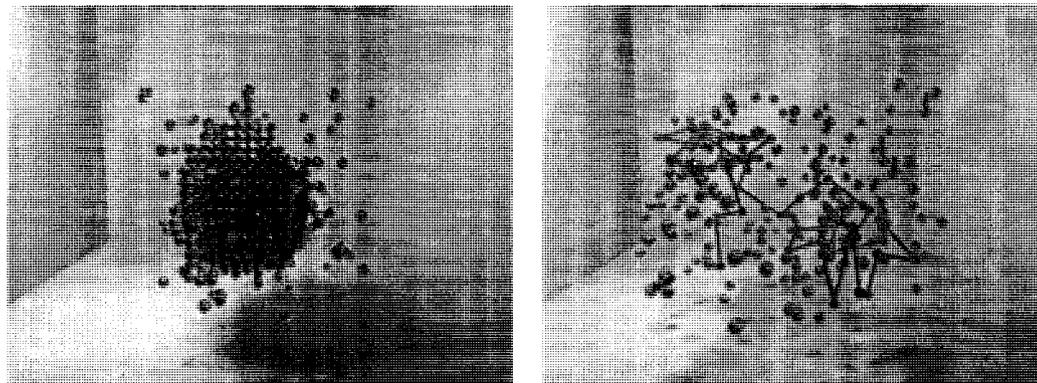


Figure 1.2: Illustration of the coarse grained protein model with explicit solvent in three dimension. A globular conformation is shown in the left figure, whereas a coil conformation is shown in the right figure.

interactions in two dimensions. An illustration of the model in three dimension is shown in figure 1.2.

We report in this thesis the temperature dependence of several statistical quantities and their universal scaling close to the collapse transition for a protein model with explicit solvent interactions. The latter follow by embedding a two dimensional protein model surrounded by a thick shell of vapor molecule into a square lattice of solvent molecules. The specific solvent model used was developed and studied by Widom and coworkers [6, 7]. The results obtained from the solvent model as well as the implementation of the dynamics are presented in chapter 2. This simplified coarse-grained model leads to an attractive potential of mean force acting on solute molecules and hence a possible model of hydrophobicity. A two dimensional protein is then modelled as a self-avoiding random walk with weakly attractive nearest-neighbors interactions. This model was introduced by Lal [9], and extensively studied by Madras and Sokal [10]. The details of the implementation of the Monte Carlo method are presented in chapter 3. The chain undergoes a collapse transition from a globular state at low temperature to a coil state at high temperature. The collapse temperature is called the θ temperature, and statistical measures on the protein model follow universal scaling laws

close to the θ temperature. Our calculated critical exponents agree with their known exact value in two dimensions. The protein model is then embedded in the solvent model. We present in chapter 4 the details of the implementation of this new model, as well as the results obtained. It is shown that the temperature of collapse transition increases with the amplitude of the hydrophobic interaction and that the critical exponents are affected by the presence of the solvent. Density fluctuations of vapor molecules drive the collapse transition with hydrophobic interactions, the chain being in a collapsed state at low temperature due to the entropic cost to add vapor molecules to the layer surrounding the protein and in an extended state at high temperature, as the shell can exchange at random vapor molecules with the bulk solvent.



Chapter 2

A coarse grained model of hydrophobicity

Solvent mediated interactions play a crucial role in the native structure and function of proteins. Proteins fold in water and the associated mediated hydration forces are largely responsible for the 3 dimensional native conformation of proteins. In fact, repulsion between polar water molecules and hydrophobic side chains leads to protein collapse so as to have the latter shielded from water by forming a tightly packed core that contains more than 80 % of the non polar side chains of a typical protein [3]. Water is essential for protein structure, not only with regards to the collapse of the hydrophobic core, but also to maintain the structure and the stability of the chain. Changes in the network of hydrogen bonds between water molecules and the protein influence strongly its stability. Increasing the ordering of water by decreasing temperature can result in protein denaturation, called cold denaturation. On the other hand, some water molecules bind to specific locations in the protein as can be seen by using crystallography as being an intrinsic part of the protein structure. The inclusion of solvent effects is therefore essential to obtain a realistic and accurate description of the native structure of proteins. There are two basics ways to model solvent mediated interactions, which we describe next.

The so called explicit solvent models simulate directly all the solvent molecules, usually by using Molecular Dynamics or Monte Carlo methods. There exist many explicit water models including the ST2 model of Stilliger [11] that involves a pair potential based on a rigid four point charge model (in which the negative charge is placed on a virtual atom placed near the oxygen atom) for all water molecules, the SPC/E model of Berendsen [12] which is a variant of simple point charge models, the TIP3P, TIP4P, and TIP5P models of Jorgensen's [13], in which a rigid water monomer is represented by 3, 4 or 5 interaction sites, and the TIP4P-Ew of Horn [14], in which the Ewald summation is included to treat long-range electrostatics. All of these models treat water as almost a rigid molecule, although some include bond stretching and bond-angle bending, polarization effects, or many body interactions. They are usually parameterized at the single temperature of 298 K, and therefore do not capture the temperature dependence of some properties, like the solvent density or diffusion coefficients. Explicit solvent models are in general highly demanding computationally, as they account for all degrees of freedom in the system. A lot of effort has been devoted to reduce the computational demands by treating the solvent as a continuous medium having the average property of a real solvent. This latter class of models are called implicit solvent models. They include the Generalized Born Surface Area model (GB/SA) [15], which is an approximation to the Poisson-Boltzmann equation for a solute immersed in a dielectric solvent. They also include solvent accessible surface area models [16] that are based on the assumption that once an atom is buried, it doesn't contribute to the free energy of hydration anymore, and that the potential of mean force between pairs of atoms is zero if the two atoms are separated by a distance greater than the sum of their Van der Waals radii plus the thickness of one hydration shell. There has been a great deal of effort devoted recently to improve implicit solvent models as there still exist differences in the results given by implicit and explicit solvent models [17, 18, 19]. It is therefore still a matter of debate the degree to which an implicit solvent model captures the properties of water.

Since explicit solvent models require a large amount of computational time, simplified models with some degree of coarse-graining continue to be explored to study protein folding. The coarse grained solvent model which will consider is defined on a lattice. It has been introduced recently by Widom and coworkers [6], to study thermodynamic properties of hydrophobicity [20]. In particular, this models yields an attractive solvent mediated potential of mean force which is independent of the solute, with a magnitude that increases with temperature while its range decreases, a signature of hydrophobicity. We first describe the model, and present the results we have obtained from Monte Carlo numerical simulations.

2.1 Lattice model of hydrophobicity

The model is defined on a two dimensional square lattice of size $N = n \times n$ with periodic boundary conditions. Each site is occupied by only one solvent molecule. Each molecule can have Q states of orientation with $S_i = \{1, 2, 3, \dots, Q\}$, where S_i is the orientation state of the i^{th} molecule and $i = \{1, 2, \dots, N\}$. Solvent molecules only interact with their nearest neighbors (Figure 2.1). Only one orientation is energetically favored, say $Q = 1$, called the special sate. A hydrophobic solute molecule is then accommodated at interstitial sites between two solvent molecules only if they are both in their special state. A pair of solvent molecule which are both in their special states is said to correspond to hydrogen bonds state between them. The energy of interaction $E_{i,j}$ between a pair of neighboring solvent molecules, one in state S_i and the other in the state S_j is,

$$E_{i,j} = \begin{cases} \omega & \text{if } S_i = S_j = 1 \\ \nu & \text{otherwise} \end{cases}, \quad (2.1)$$

with $(\nu - \omega) > 0$. Without loss of generality, we have set $\nu = 0$. The dynamics of the solvent model follow from the Hamiltonian H_S ,

$$H_S = w \sum_{\langle i,j \rangle} \delta_{S_{i,1}} \delta_{S_{j,1}} \quad , \quad (2.2)$$

where $\omega < 0$ and $\delta_{S_{\mu,1}} = 1$ if the state of the μ^{th} molecule is its special state, and 0 otherwise. The sum is over all the nearest neighbors of the i^{th} molecule. The partition function Z_S for the solvent is given by,

$$Z_S = \sum_C e^{-\beta H_S} = \sum_C \exp \left(-\beta \omega \sum_{\langle i,j \rangle} \delta_{S_{i,1}} \delta_{S_{j,1}} \right) \quad , \quad (2.3)$$

where the sum is over all configurations for the lattice of solvent molecules and $\beta = 1/k_B T$. The quantity $(\nu - \omega)$ represents the favorable energy of hydrogen bonds formation, while $k_B \ln[Q - 1]$, where k_B is the Boltzmann constant, represents the unfavorable entropy involved in constraining a solvent molecule to its special state. At low temperature, free energy is minimized by maximizing the number of hydrogen bonds between solvent molecules. At high temperature, minimization of the free energy is achieved by entropy maximization, hence by increasing the number of molecules in an orientation state other than the special one. Water molecules form short lived tetrahedral structures stabilized by hydrogen bonds, which is a type of short-range order. These structures have an open cage like space between them and hydrophobic molecules can be accommodated in this location without breaking any of the hydrogen bonds. This is the minimum energy configuration for the solvent molecules in the presence of a hydrophobic solute molecule, and is modelled by a pair of neighboring solvent molecule both in their special state. Hydrophobic effects can be understood in the following way. The accommodation of a solute molecule between two solvent molecules is energetically favorable, but entropically unfavorable as we reduce the number of accessible states for this

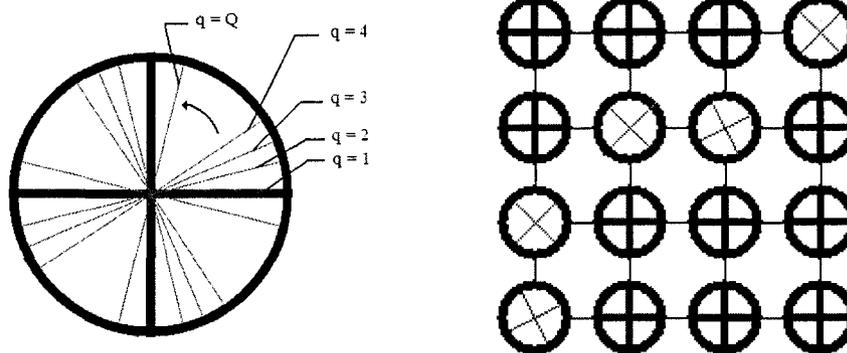


Figure 2.1: Schematic depiction of the orientation state of a solvent molecule (Left). The horizontal and vertical lines correspond to the special orientation state while the other lines correspond to the other $(Q - 1)$ orientation states. Example of a two dimensional lattice of solvent molecules with $n = 16$ (Right).

configuration. This increase is smaller when solute molecules are close together rather than widely separated. Note that this effect is only dependent on the dynamics of the solvent not the solute and can be captured through the calculation of the solvent mediated part of the potential of mean force acting on solute molecules.

2.2 Reversible work theorem and potential of mean force

The calculation of the solvent-mediated part of the potential of mean force acting on solute molecules is based on the potential distribution theorem. Following Widom [21], consider N molecules in a volume V at a temperature T and introduce W_N , the interaction energy of N molecules as a function of position, and ψ as the potential energy of interaction for one molecule with the remaining $N - 1$ molecules as a function of position. The partition

function Z_N is given by,

$$\begin{aligned}
 Z_N &= \int_V dr_1 \dots \int_V dr_N e^{-\beta W_N} \\
 &= \int_V dr_1 \dots \int_V dr_N e^{-\beta \psi} e^{-\beta W_{N-1}} \\
 &= V Z_{N-1} \langle e^{-\beta \psi} \rangle \quad ,
 \end{aligned}
 \tag{2.4}$$

where the average $\langle \dots \rangle$ refers to the canonical ensemble average. In order to understand the potential distribution theorem imagine that the molecules fluctuate in orientation around an equilibrium value. Now for one such state, fix time and introduce one new molecule into the system. Let the enlarged system evolve and measure at each point in space the interaction energy ψ with the new molecule. The radial distribution function $g(r)$ is defined as the probability to find a second molecule in dr given that there is a molecule at r ,

$$g(r) = \frac{V^2}{Z_N} \int_V dr_3 \dots \int_V dr_N e^{-\beta W_N} \quad .
 \tag{2.5}$$

Introduce z as the thermodynamic activity of the fluid molecules, defined to become asymptotic to $N/V = n$ as $n \rightarrow 0$, where n the number density,

$$z = N \frac{Z_{N-1}}{Z_N} \quad .
 \tag{2.6}$$

Combining equations 2.4 and 2.6, we have

$$\left(\frac{n}{z} \right) = \langle e^{-\beta \psi} \rangle \quad .
 \tag{2.7}$$

Define $\phi(r)$ as the intermolecular potential energy between two molecules. Then,

$$\begin{aligned}
g(r)e^{-\beta\phi(r)} &= \frac{V^2}{Z_N} \int_V dr_3 \cdots \int_V dr_N e^{-\beta W_N} e^{-\beta\phi(r)} \\
&= \left(\frac{z}{n}\right)^2 \frac{1}{Z_{N-2}} \int_V dr_3 \cdots \int_V dr_N e^{-\beta W_{N-2}} e^{-\beta\phi(r)} \\
&= \left(\frac{z}{n}\right)^2 \langle e^{-\beta\psi(\vec{r}_1)} e^{-\beta(\psi(\vec{r}_2)+\phi(r))} \rangle \\
&= \left(\frac{z}{n}\right)^2 \langle e^{-\beta\psi(\vec{r}_1)} e^{-\beta(\psi(\vec{r}_2)+\phi(r))} \rangle \\
&= \frac{\langle e^{-\beta\psi(\vec{r}')} e^{-\beta\psi(\vec{r}'+r)} \rangle}{\langle e^{-\beta\psi(\vec{r}')} \rangle^2} .
\end{aligned} \tag{2.8}$$

To apply these results to the solvent model being discussed here, define $\phi(r)$, with r measured in units of the lattice spacing, as the direct interaction of two neighboring solvent molecules, and let $g(r)$ be the solute-solute pair distribution function already defined. Define also P_{11} as the probability that two neighboring solvent molecules be both in their special state, irrespectively of the state of the other solvent molecules, and let $P(r)$ to be the probability that two pairs of consecutive molecules at $(n, n+1)$ and $(m, m+1)$, with $n-m=r$, be all in their special state. Let α be the solute-solvent energy of interaction. To solve for the solvent mediated part of the potential of mean force, we first consider the reversible work theorem [22] that relates the radial distribution function $g(r)$ to the potential of mean force $W(r)$,

$$g(r) = e^{-\beta W(r)} \tag{2.9}$$

The potential of mean force is the reversible work for a process in which two tagged particle are moved from an infinite separation to a separation R . By using the reversible work theorem, the total potential of mean force between pairs of solute molecules is $-k_B T \ln[g(r)]$, in which the solvent mediated part is $W(r) = -k_B T \ln[g(r)] - \phi(r) = -k_B T \ln[g(r)e^{-\beta\phi(r)}]$.

In the infinite dilution limit, equation 2.8 can be written as

$$g(r)e^{-\beta\phi(r)} = \frac{P(r)e^{-2\beta\alpha}}{(P_{11}e^{-\beta\alpha})^2} = \frac{P(r)}{P_{11}^2}, \quad (2.10)$$

at temperature T . The solvent-mediated part of the potential of mean force $W(r)$ is thus obtained,

$$W(r) = -k_B T \ln [g(r)e^{-\beta\phi(r)}] = -k_B T \ln \left[\frac{P(r)}{P_{11}^2} \right] \quad (2.11)$$

Note the cancellation of the parameter α , as all pairs considered can accommodate a hydrophobic particle. The solvent mediated-part of the potential of mean force is thus independent of the solute-solvent interaction and so the effective attraction acting on solute molecules depends on the properties of the solvent alone. The exact calculation of the potential of mean force requires the calculation of P_{11} and $P(r)$. This was done exactly by Widom [6] in one dimension by using the transfer matrix method. In one dimension, the transfer matrix V is a $(Q \times Q)$ square matrix for which the $(1, 1)$ element is given by $b = \exp(-\beta\omega)$ while all the other elements of the matrix are $a = \exp(-\beta\nu)$. By performing direct matrix multiplication of V over all molecules in the system and by using periodic boundary conditions, the partition function is then expressed as the trace of V^N , where N is the total number of molecules. As the trace is independent of the matrix representation, the matrix V can be express in diagonal form and the partition function is the sum of the eigenvalues λ_i^N of V^N . The largest eigenvalue will completely dominate the partition function and thus we end up with the simple result $Z = \lambda_{max}^N$, where λ_{max} is the largest eigenvalue of the transfer matrix. The calculation of the potential of mean force is then straightforward, given the solution of the partition function. The details of the exact calculation of the potential of mean force for a linear chain of solvent molecule with periodic boundary condition are shown in appendix A. It is shown that the magnitude of the potential increases as temperature is increased while its range decreases. In two dimensions, the transfer matrix V is a $(Q^N \times Q^N)$

square matrix. The partition function is again expressed in terms of matrix multiplication, but now over all rows of the lattice, and is given by the trace of V^N , if periodic boundary conditions are used. The exact calculation has not been done yet. The phase transition of the solvent model is analysed in the next section.

2.3 Phase transition of the solvent model

At $T = 0$, molecules only perform transitions into the special state of orientation, and the equilibrium state has all solvent molecules in their special state. We call this phase the ordered phase. As temperature is increased, solvent molecules that are not in their special state are allowed in the equilibrium configuration. As $T \rightarrow \infty$, almost all the solvent molecules are not in the special state of orientation, and this is called the disordered phase. Here the probability of finding a pair of consecutive sites in their special state is $1/(Q - 1)^2$. A mapping between the ferromagnetic Ising model in an external magnetic field in one dimension and the solvent model discussed can be established. The details of the calculation were done by Widom [8] and are shown in appendix B. The mapping leads to a relation between the parameters of the two models,

$$\begin{aligned} 2J &= \frac{(\nu - \omega)}{2} \\ 2H &= \frac{z}{2}(\nu - \omega) - k_B T \ln[Q - 1] \quad , \end{aligned} \tag{2.12}$$

where J is the energy of interaction between two spins and H is the external magnetic field of the ferromagnetic one dimensional Ising model. The Ising model with an external magnetic field undergoes a first order transition as the external magnetic field is varied. For the solvent model, a critical temperature is obtained by setting $H = 0$,

$$k_B T_c = \frac{z}{2} \frac{(\nu - \omega)}{\ln[Q - 1]} \quad , \tag{2.13}$$

where z is the coordination number of the lattice. It agrees with the numerical results. The Q-Potts state model in one dimension in the Bragg-William approximation has also been considered. The details of the calculation are shown in appendix C. This model undergoes also a first order transition for $Q \geq 3$ at the critical temperature,

$$k_B T_c = \frac{z(Q-2)(\nu-\omega)}{2(Q-1)\ln[Q-1]}, \quad (2.14)$$

which is a good approximation to equation 2.13. Transition from the ordered phase to the disordered phase is determined by the parameters $(\nu - \omega)$ and Q . Increasing $(\nu - \omega)$ or decreasing Q increases the critical temperature of the solvent. Given the mapping between the solvent model and the ferromagnetic Ising model in an external magnetic field, the phase transition between the ordered state at low temperature and the disordered phase at high temperature is first order. The solvent model can also be viewed as a Q-state Potts model as each solvent molecule can have Q orientations (or states). Transition of the Q-Potts state model is first order for $Q \geq 3$ and is another indication that the transition of the solvent model is first order.

2.4 Monte Carlo methods and BKL algorithm

We have implemented the dynamics of the solvent model using Monte Carlo methods through the BKL algorithm. In a standard Monte Carlo method, one would choose at random the orientation state of a solvent molecule chosen also at random, and perform a transition from the current orientation state to the randomly chosen one. According to the Metropolis rule, transitions would be always accepted if the energy change from the new configuration with respect to the old configuration, ΔE , is less or equal to zero, and with probability $P = \exp(-\beta\Delta E)$, if $\Delta E > 0$. As there are Q possible states of orientation, standard Monte Carlo sampling involves transitions among many configurations for which a solvent

Class	Current State	Numbers of n.n. in special state
1	1	4
2	1	3
3	1	2
4	1	1
5	1	0
6	other	4
7	other	3
8	other	2
9	other	1
10	other	0

Table 2.1: Table summarizing the construction of the classes of solvent molecules. Each class number represents the current state of a solvent molecule and the number of nearest neighbors molecules that are in their special state.

molecule is not in the special state, leading to no change in the energy, but increasing the computational time for equilibrium sampling. To increase the speed of computation, we have implemented the dynamics through the BKL algorithm, first introduced by Bortz, Kalos and Lebowitz [23]. In this method, instead of sampling at constant rate until a solvent molecule performs a transition, one takes advantage of the fact that the probability for a solvent molecule to perform a transition to a new orientation state is already known, and directly performs the transition according to its probability of occurrence. The simulation time variable is then updated accordingly.

Solvent molecules have one orientation state, the special one, that is degenerate in energy depending on its $z = 4$ nearest neighbors. The other $(Q - 1)$ orientation states all share the same energy ν . We construct a table in which we attribute a class to all possible transitions for a solvent molecule, describing its current state of orientation and the number of its nearest neighbors that are in the special state. There are 10 of those possible transitions, 5 for which the solvent molecule is originally in its special state and leave it for another state, and 5 for which the solvent molecule is not initially in its special state and enters it. The change in

energy ΔE for a transition of a solvent molecule with orientation state S to the orientation state S' is

$$\Delta E = \omega \Delta N_{11} - k_B T \ln[Q - 1] \Delta N_Q \quad , \quad (2.15)$$

where ΔN_{11} is the change in the number of hydrogen bonds between solvent molecules and ΔN_Q is the change in the number of molecules that are in one of the other $Q - 1$ orientation state. Let $P(S, S')$ be the transition probability from an orientation state S to another orientation state S' be defined as,

$$P(S, S') = \begin{cases} 1 & \text{if } \Delta E \leq 0 \\ e^{-\beta \Delta E} & \text{if } \Delta E > 0 \end{cases} \quad . \quad (2.16)$$

This choice of transition probability is called the Metropolis rule, and ensures that detailed balance is satisfied. For the canonical ensemble, the detailed balance condition demands that

$$\frac{P(S, S')}{P(S', S)} = e^{-\beta \Delta E} \quad . \quad (2.17)$$

From the choice of the transition probabilities, we have for $\Delta E \leq 0$,

$$\frac{P(S, S')}{P(S', S)} = \frac{e^{-\beta \Delta E}}{1} = e^{-\beta \Delta E} \quad , \quad (2.18)$$

and for $\Delta E > 0$, we have

$$\frac{P(S', S)}{P(S, S')} = \frac{1}{e^{\beta \Delta E}} = e^{-\beta \Delta E} \quad , \quad (2.19)$$

which shows that detailed balance is satisfied. The first five class involve transition from the special state to one of the other $(Q - 1)$ other states of orientation and are given transition probability $P = \exp(-\beta \Delta E)$ while the last five classes involve transition from one of the other $Q - 1$ states of orientation to the special one and are given transition probability $P = 1$. The construction of this table is summarized in table 2.1. For each attempted transition,

we calculate ten numbers,

$$Q_i = \sum_{j=0}^i P_j N_j \quad i = 1, 2, \dots, 10 \quad , \quad (2.20)$$

where N_μ is the number of molecules in class μ , P_μ is the probability of the transition of the μ^{th} class, and $Q_0 = 0$. An uniformly distributed random number x is chosen in the interval $I = [0, Q_{10}]$ to identify the class i of the molecule which will undergo the transition so that $Q_{i-1} \leq x < Q_i$. A molecule is chosen at random in this class and a transition for this molecule to the new orientation state is performed by updating its energy and the energy of its nearest neighbors by adjusting their class. After each transition, time is updated by an amount

$$\Delta t = -\frac{\ln(y)}{Q_{10}} \quad , \quad (2.21)$$

where y is an uniformly distributed random number in the interval $I = [0, 1]$. These transitions are iterated many times until a specified simulation time is reached. Statistical measures are then averaged at a given temperature and the process is repeated for many temperatures to study temperature dependence of the model. The flow chart of the algorithm used is shown in table 2.2. Examples of square lattice configurations of solvent molecules obtained with this algorithm are shown for six different temperatures in figure 2.2.

2.5 Measures analysis of the solvent model

We have performed all our numerical simulations with the algorithm described above on a two dimensional square lattice. Anywhere between 20 000 - 100 000 independent configurations are considered to compute equilibrium averages. Without loss of generality, we have set the parameter ν to be zero through all the simulations. In the initial condition, all solvent molecules were in the disordered state (all class 10). We then have thermalized the system up

 Flow chart of the solvent algorithm

1. Initialize the two dimensional lattice of solvent molecules, all in the disorder state (All in class 10).

 2. Thermalize the solvent
 - a. Construct the numbers $Q_i = \sum_{j=0}^i N_j P_j$, with $i = \{1, \dots, 10\}$, where P_j is the probability to perform a transition to the j^{th} class, where N_j is the number of molecules in the j^{th} class and where $Q_0 = 0$.
 - b. Choose a uniformly distributed random number x in the interval $I = [0, 1]$. Identify the class of the molecule which will undergo the transition such that $Q_{i-1} \leq x < Q_i$. Choose a molecule in the i^{th} class at random.
 - c. Perform the transition of the molecule by updating its class, +5 if the molecule is initially in its special state and -5 if the molecule is in another state.
 - d. Update the class of the 4 nearest neighbors of the molecule, +1 if the molecule was originally in its special state (all the 4 nearest neighbors of the molecule lose a neighbor in its special state) and -1 if the molecule was not originally in its special state (all the 4 nearest neighbors of molecule gain a neighbor in its special state).
 - e. Update the number of molecules in each class.
 - f. Update the simulation time by an increment of $\Delta t = -\ln(y)/Q_{10}$, where y is a uniformly distributed random number in the interval $I = [0, 1]$.

 3. Repeat step a - f up to a maximal simulation time providing an equilibrium lattice.

 4. For many independent configurations average,
 5. Repeat Step a - f up to a maximal simulation time set to provide uncorrelated measures.
 6. Take measures of P_{11} , c , C_S , M and $W(r)$.

 7. Display the mean of the quantities of interest.
-

Table 2.2: Flow chart for solvent configuration updates.

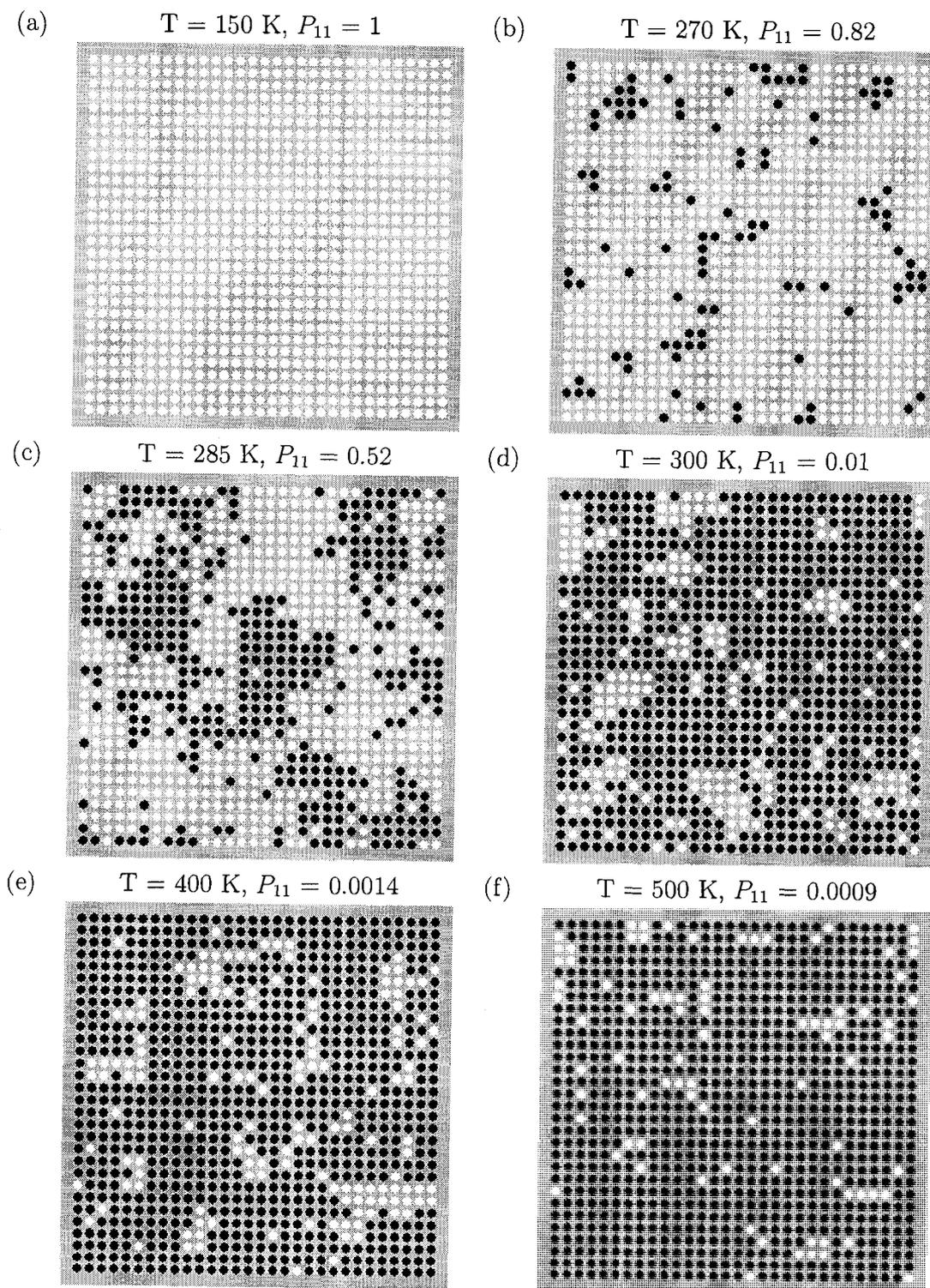


Figure 2.2: Examples of the square lattice of solvent molecule for a unique configuration with the parameters fixed at $\omega = -421 \text{ K}$ and $Q = 20$, for $N = 900$ solvent molecules. White circles represent molecules that are in the special state of orientation while the black circles represent the ones which are in one of the other $(Q - 1)$ states. For those parameters, the critical temperature of the solvent is $T_c^S \approx 286 \text{ K}$.

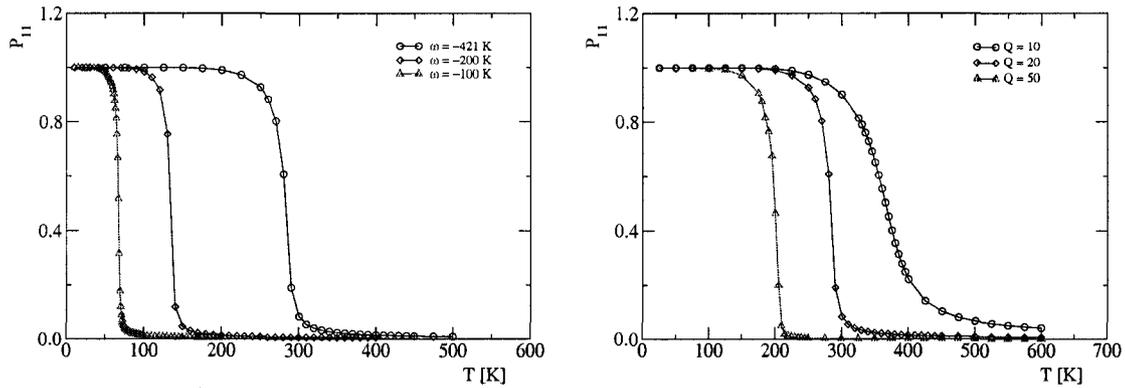


Figure 2.3: Probability that a pair of neighboring solvent molecules be both in their special state P_{11} as a function of temperature T [K] for different values of the energy of interaction ω and fixed number of accessible states $Q = 20$ (Left), and for different value of Q with $\omega = -421$ [K] fixed (Right) for a lattice of $N = 10\,000$ solvent molecules. At low temperatures, most of the orientations are on the preferred state, whereas at high temperature, the orientation of the spin variable is random.

to a maximal simulation time providing an equilibrium configuration at a given temperature T and measured the probability that two neighboring solvent molecules are in their special state P_{11} , the heat capacity C_S , the order of parameter M , and the potential of mean force $W(r)$ for a simulation time which provides uncorrelated measurements. Since we work in the infinite dilution limit, all the numerical simulations were implemented with periodic boundary condition to reduce finite size effects. We present in this section the results obtained for the different measures.

2.5.1 P_{11}

The quantity P_{11} is the probability that a pair of neighboring solvent molecules be in their special state, meaning the probability of hydrogen bonds formation between two solvent molecules. At low temperature, all molecules are in their special state (class 1) and $P_{11} = 1$. Increasing temperature allows molecule to adopt an orientation state different than the special one in the equilibrium configuration, and there is a phase transition between the

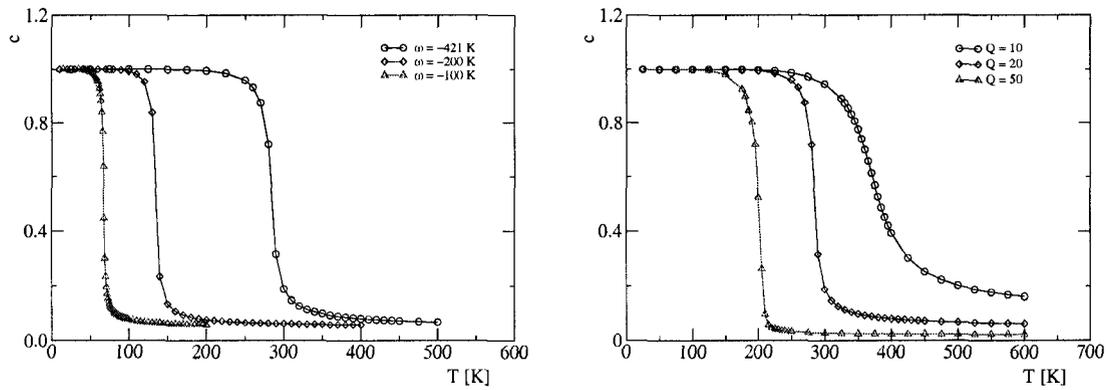


Figure 2.4: Concentration of molecules in the special state of orientation c as a function of temperature T [K] for a lattice of $N = 10\,000$ solvent molecules. The number of accessible states is $Q = 20$ and the energy of interaction ω is varied (Left). Same concentration for various values of Q with $\omega = -421$ K (Right). Note that the concentration of molecule in their special state as $T \rightarrow \infty$ is $c = 1/(Q - 1)$.

ordered and disordered phase, which is fairly sharp over the relevant range of parameters. As $T \rightarrow \infty$, the system fluctuates at random and is characterized by a lattice in the disordered phase in which small fluctuations of molecules in their special state are allowed and $P_{11} = 1/(Q - 1)^2$. Our results for the measure P_{11} as a function of the temperature for many energy of interaction ω as the number of accessible state is fixed and vice and versa are shown in figure 2.3. The calculation of P_{11} was done by counting directly the numbers of pairs of solvent molecules in their special state with respect to the total number of pair ($2N$), and averaged for many independent configurations. The concentration of molecules in the special state of orientation, $c = N_1/N$, where N_1 is the number of molecule in the special state, was calculated the same way than P_{11} , the results are shown in figure 2.4. At low temperature, all the solvent molecules are in the special state and $c = 1$ while at high temperature, small fluctuation of molecule in the special state is allowed and $c = 1/(Q - 1)$.

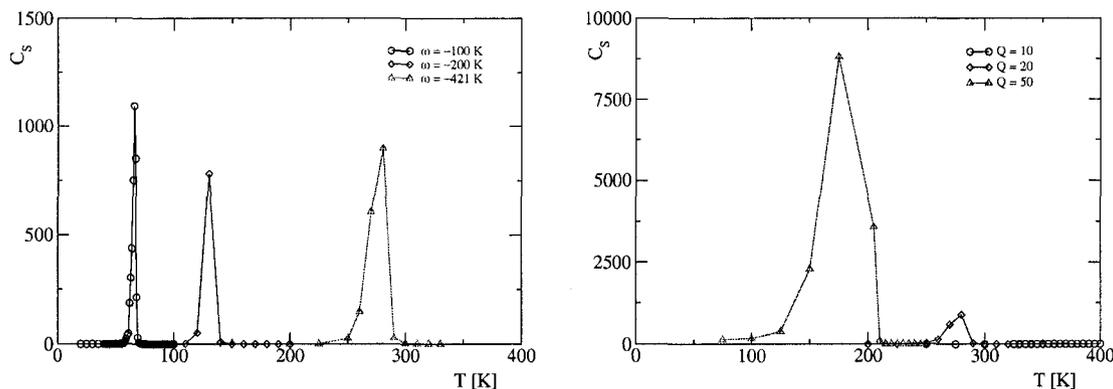


Figure 2.5: Heat capacity of the solvent C_S as a function of temperature T [K] for different values of the energy of interaction ω and fixed $Q = 20$ (Left) and for different values of Q for fixed $\omega = -421$ K (Right) for a lattice of $N = 10\,000$ solvent molecules. Increasing ω or decreasing Q leads to an increase in the location of the critical temperature and in the magnitude of the heat capacity.

2.5.2 Heat capacity of the solvent model

Define the heat capacity of the solvent C_S as

$$C_S = \frac{-1}{k_B T^2} \frac{\partial \langle E_S \rangle}{\partial \beta} = \frac{\langle \Delta E_S^2 \rangle}{k_B T^2} = \frac{(\langle E_S^2 \rangle - \langle E_S \rangle^2)}{k_B T^2}, \quad (2.22)$$

where E_S is the energy of the solvent configuration. The heat capacity is a measure of the fluctuations in the energy of the solvent molecules. Figure 2.5 shows the heat capacity of the lattice of solvent molecules as a function of the temperature as ω is fixed and Q is varied, and vice versa. As expected from the shifts in the transition temperature, increasing $-\omega$ with Q fixed or decreasing Q with $-\omega$ fixed increases the temperature of the heat capacity maximum as well as its magnitude. At low temperature, almost all the solvent molecules are in their special state and fluctuations in energy are small. Close to the critical temperature, the fluctuations are expected to diverge as the chemical potential of hydrogen bonds formation between two neighboring solvent molecule balances the chemical potential of neighboring solvent molecule in which one is in a state different than the special one. As $T > T_c$, the

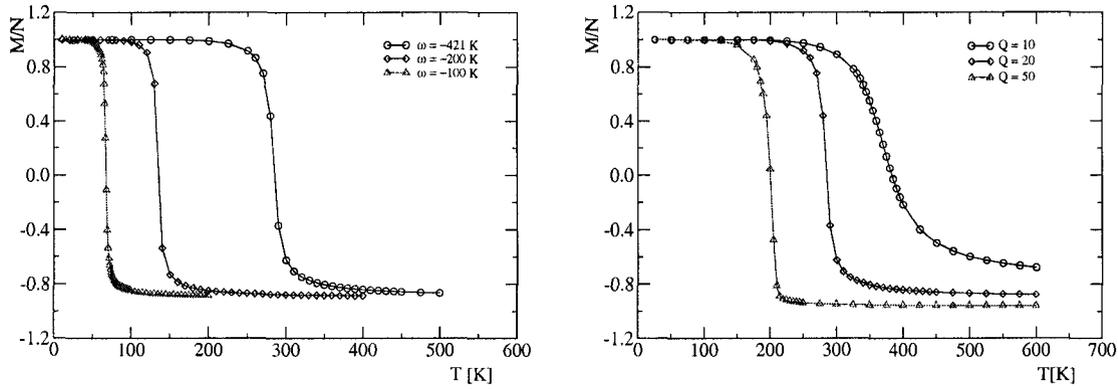


Figure 2.6: Order of parameter M/N as a function of temperature T [K] on a lattice of $N = 10\,000$ solvent molecules for various values of ω and Q . The number of accessible states is $Q = 20$ and ω is changed (Left) and where Q changes for $\omega = -421$ K (Right). Temperature dependence of the order of parameter M is similar to the magnetization of the ferromagnetic Ising model with an external magnetic field.

fluctuation in energy is constant, and the heat capacity goes to zero, as $C \sim 1/k_B T^2$. The heat capacity was calculated using equation 2.22, by directly averaging for many independent configurations the total energy of the solvent molecules.

2.5.3 Order of parameter M

Define the order of parameter M as the difference between the number of molecules in their special state N_1 and the number of molecules that are not in their special state N_Q ,

$$M = N_1 - N_Q \quad , \quad (2.23)$$

where N_1 is the number of solvent molecules that are in their special states and $N_Q = N - N_1$. This order parameter can be seen as the equivalent of the magnetization in the Ising model. In the low temperature regime, the solvent is in its ordered phase and almost all the molecules are in their special state, the order of parameter M , shown in figure 2.6, is positive and of the order of N . As temperature is increased, some disordered molecules are present in the

configuration and the order of parameter decreases. For very large temperature, the solvent model is in a disordered phase and the probability to find a molecule in its special state is $1/(Q - 1)$. Around the critical temperature, the favorable energy of hydrogen bonds formation balances the unfavorable entropy, leading to $M \sim 0$.

2.5.4 Potential of mean force acting on a solute molecule

By using the reversible work theorem, the solvent mediated part of the potential of mean force is obtained by using equation 2.11. The results obtained are shown in figure 2.7. In the ordered phase, all molecules are in their special state and thus $P_{11}^2 = P(r)$, independent of distance r , leading to a vanishing potential of mean force. A negative potential denotes an effective solvent mediated attraction between two solute particles, an intrinsic feature of hydrophobicity. For temperatures higher than the critical temperature, the magnitude of the potential of mean force increases as the temperature is increased while its range decreases, an observation that agrees with the calculation of Widom [6, 7] in one and two dimensions. This leads over this temperature range to an effective solvent mediated potential of mean force which is attractive. This is a signature of hydrophobicity, favoring solute molecules to cluster. The potential of mean force was calculated in a straightforward way by using equation 2.11 where the quantities P_{11} and $P(r)$ were obtained by numerical simulations. The probabilities P_{11} and $P(r)$ were measured directly by counting the appropriate number of pairs in which two neighboring molecules are in their special state. A solute molecule can only occupy the interstitial site in between two solvent molecules, and this point was taken as origin for the distance r separating this pairs to another one in its special state. By the geometry of the lattice, the allowed value of r taken from this origin are $r = \{\sqrt{1/2}, 1, \sqrt{2}, \sqrt{5/2}, 2, \sqrt{17/4}, \sqrt{9/2}, 5, \dots\}$. As the temperature is further increased, the solvent will fluctuates at random and $P_{11}^2 = P(r) = 1/(Q - 1)^4$, leading to a null potential of mean force. We have presented in this section the results obtained by nu-

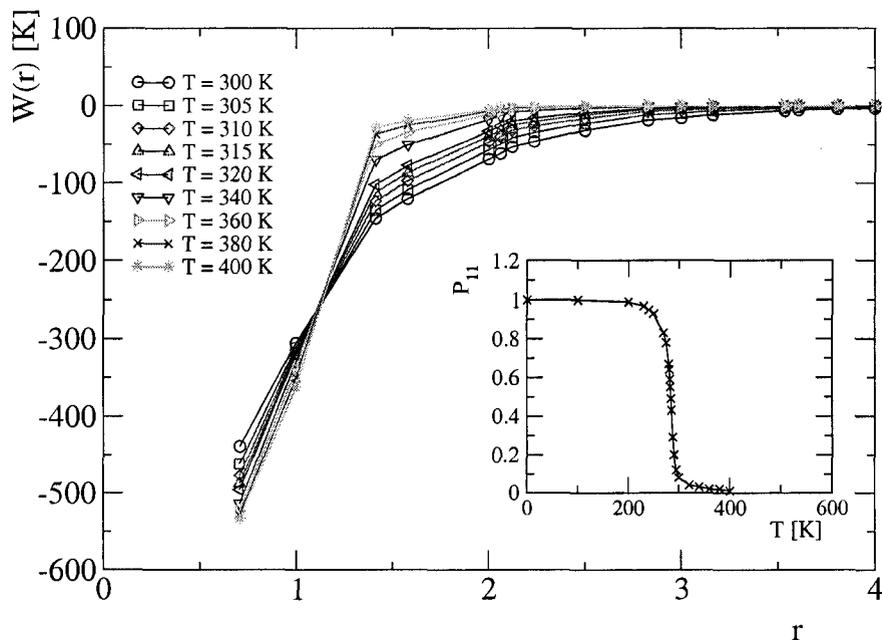


Figure 2.7: Potential of mean force $W(r)$ [K] as a function of the distance r in units of the lattice spacing. The results shown correspond to $\omega = -421$ K and $Q = 20$ on a lattice of $N = 250\,000$ solvent molecules. 20 000 independent configurations were averaged for all temperatures. The magnitude of the potential of mean force increases with temperature while its range decreases slightly. The inset shows the function P_{11} as a function of temperature T [K] as obtained with the same values of the parameters.

merical simulation of an explicit coarse grained solvent model that incorporates some basic features of hydrophobicity. A first order transition is included by the model from an ordered phase at low temperature, where almost all the solvent molecules are in their special state, to a disordered phase at high temperature, where almost all the solvent molecules have an orientation state different than the special one. The two parameters of the model, the energy of interaction between two consecutive sites ω and the number of accessible states Q , determine the critical temperature of the transition between the two phases of the solvent. The potential of mean force between two pairs of neighboring solvent molecules that are in their special state has been calculated. This potential is attractive, and its magnitude increases as temperature is increased while its range decreases, both known features of hydrophobicity.

1

2

3

Chapter 3

Collapse transition of self-avoiding random walks

To achieve the goal of this thesis, the solvent model presented in the last chapter will be used to study the collapse transition of a model protein immersed in it. We will use a simple model of a protein on a two-dimensional lattice, and compare its collapse transition with hydrophobic interactions to that of the protein alone. Of course, protein modelling can be thought of as a part of polymer science, a field widely developed [25]. There are many simplified models of a protein, which can be classified mainly as on and off lattice models, depending on whether the model includes the use of a lattice or not. We focus here on a specific model in which the chain is a random walk in space, with the special condition that chain segments don't intersect themselves (self-avoiding random walks or SARWs) [10, 24]. With nearest-neighbors interaction, SARWs undergo a phase transition at some critical temperature $T = \theta$ [26], the Flory temperature, from a globular phase at low temperature to a coil phase at high temperature. Around the critical temperature, some measures of protein conformation follow power laws with the chain length, in which the exponents are universal and independent of the type of lattice used. Their values are known exactly in two dimensions. We present in this section the theoretical methods and results

obtained by Monte Carlo methods of a SARW in 2 dimensions.

3.1 SARWs with nearest neighbor interaction

Consider a SARW of L monomers on a two dimensional square lattice of size $N = n \times n$, with periodic boundary conditions. Monomers can only interact with their nearest neighbors with interaction energy $\epsilon = -|\epsilon|$. The Hamiltonian H_P describing this chain is,

$$H_P = \epsilon \sum_{\langle i,j \rangle} \delta_i \delta_j \quad , \quad (3.1)$$

where the sum is over the nearest neighbors of the i^{th} monomer, $\epsilon < 0$ to simulate weakly attractive monomers, and $\delta_\mu = 1$ if the μ^{th} lattice site is occupied by a monomer, 0 otherwise. At low temperature, energy minimization dominates, and low energy states (or collapsed states) result. Increasing the temperature allows more conformations in the sampling by increasing the probability for transitions that increase the energy, and the protein undergoes a transition from a globular phase to a coil phase at high temperature. As $T \rightarrow \infty$, the protein conformation fluctuates randomly in space and behaves like a SARWs with no interaction. It has been shown by de Gennes [27, 28] that the θ temperature of the phase transition is in fact a tricritical temperature with an upper critical dimension of three. The collapse transition depends on the interaction parameter ϵ between monomers. Close to the θ temperature, statistical measures of chain conformation such as the mean squared end-to-end distance $\langle R_e^2 \rangle$, the mean squared radius of gyration $\langle R_g^2 \rangle$, the partition function Z_P of the chain, the derivative of the mean squared end to end distance $\langle R_e' \rangle$, and the derivative of the mean squared radius of gyration $\langle R_g' \rangle$, are known to follow universal scaling laws in two dimensions [29],

$$\langle R_e^2 \rangle, \langle R_g^2 \rangle \sim L^{2\nu}, \quad \nu = \frac{4}{7} = 0.5714... \quad , \quad (3.2)$$

$$Z_P \sim \mu^L L^{\gamma-1}, \quad \gamma = \frac{8}{7} = 1.1428... \quad , \quad (3.3)$$

$$\langle R'_e \rangle, \langle R'_g \rangle \sim L^\phi, \quad \phi = \frac{3}{7} = 0.4285\dots, \quad (3.4)$$

where L is the number of monomers. We have first performed numerical simulations to reproduce the collapse transition of the chain as well as the universal scaling of the measures in absence of solvent. Details of the implementation of the chain dynamics are presented in the next section.

3.2 SARWs in the BKL scheme using the pivot algorithm

Lattice polymer models can be considered as coarse-grained models of continuous polymers, different from the off lattice class in that they have fixed bond angles. To be able to investigate later the dynamics of the protein and the solvent together, we have implemented an on lattice algorithm to sample chain conformation. There exist many on lattice algorithms for a protein e.g., the Kink-Jump and Crankshaft algorithm [30], involving random monomer displacements from an old position to a new one and a crankshaft-like rotation. This algorithm gives a good description of the Rouse model. The reptation regime has been simulated by a different algorithm [31, 32] in which one chooses one end of the chain at random, and transfers a monomer from this end to the other repeatedly, or with the Growth and Scanning algorithm [33, 34], which is an improvement of the classical Rosenbluth technique for generating polymers chains. We have chosen the pivot algorithm to model the protein, a technique which has been extensively studied by Madras and Sokal [10, 35]. The pivot algorithm was devised by Lal [9] and is a Monte Carlo algorithm which generates SARWs in the canonical ensemble (the length of the protein is fixed) with free end-points. A monomer is chosen at random and acts as a pivot for rotation of the right or left segment of the protein with respect to the pivot. Conformations violating excluded volume condition are rejected. This algorithm is ergodic, meaning that all microstates are equally probable over a long period of

time, and satisfies detailed balance. In two dimensions, the symmetry operation for segment displacement consists of the dihedral group D_4 , which has 8 elements, the identity (1), $\pm\pi/2$ rotations (2), π rotation(1), axis reflections (2) and diagonal reflections (2). Madras and Sokal [10] have shown that a sufficient condition for ergodicity is a non vanishing probability for either $\pm\pi/2$ rotations or both diagonal reflections and either π rotation or both axis reflections. We have chosen $\pm\pi/2$ rotations and π symmetry to model the dynamics of the protein. The dynamics of the system were implemented through Monte Carlo methods. The initial condition is a linear chain, thermalized at a given temperature before measures of independent configurations for the chain are averaged. An attempted move starts by choosing a monomer at random (the pivot), and then performing the transition of the protein to a new conformation according to one of the symmetry operations described, also chosen at random. The energy change between the old and the new conformation is calculated and the new conformation is accepted if $\Delta E \leq 0$ with probability $P = 1$ and with probability $P = \exp(-\beta\Delta E)$ if $\Delta E > 0$, where $\Delta E = \epsilon\Delta N_P$ and N_P is the number of contacts between monomers. The resulting conformation becomes the new starting point for another transition of the protein. These transition probabilities are the Metropolis transition probabilities and ensure that detailed balance is satisfied, just in the same way as in chapter 2. A conformation for which excluded volume is violated is automatically rejected and a new attempted transition is performed. According to Madras and Sokal, this algorithm is one of the most efficient algorithms to simulate the dynamics of proteins and converges to equilibrium in a reasonable amount of computational time [10]. The flow chart of the algorithm is shown in table 3.1 and examples of chain conformation generated with this algorithm at various temperatures are shown in figure 3.1.

 Flow chart for chain conformation updates

1. Initialize the protein as a linear chain on the lattice by assigning (x,y) coordinates and energy parameter to all monomers.

2. Thermalize the protein.
 - a. Choose a monomer at random.
 - b. Choose one of the symmetry operations ($\pm\pi/2$, π rotation) for the transition to be perform at random.
 - c. Perform the transition by updating (x,y) coordinates and energy of the monomers concerned by the chosen symmetry operation.
 - d. Check for exclude volume condition. If it is violated, go directly to a.
 - e. Calculate the energy change from the old to the new conformation, $\Delta E = \epsilon\Delta N_P$, where N_P is the number of contacts between neighboring monomers.
If $\Delta E \leq 0$, accept the new conformation with probability $P = 1$.
Otherwise, accept the new conformation with probability $P = \exp(-\beta\Delta E)$
 - f. Update the (x,y) coordinates and energy of all the monomers in the new conformation

3. Repeat Step a - f up to a maximal simulation time set to provide an equilibrium configuration of the chain.

4. For many independent conformation averages
 5. Repeat Step a - f up to a maximal simulation time providing that the autocorrelation function of the measures is 0.
 6. Compute $\langle R_e^2 \rangle$, $\langle R_g^2 \rangle$, $\langle R'_e \rangle$, $\langle R'_g \rangle$, C_P , $P_L(n)$, $S(\vec{k})$

7. Display the mean of the quantities of interest.

 Table 3.1: Flow chart for chain conformation updates.

Chain conformations for many temperatures

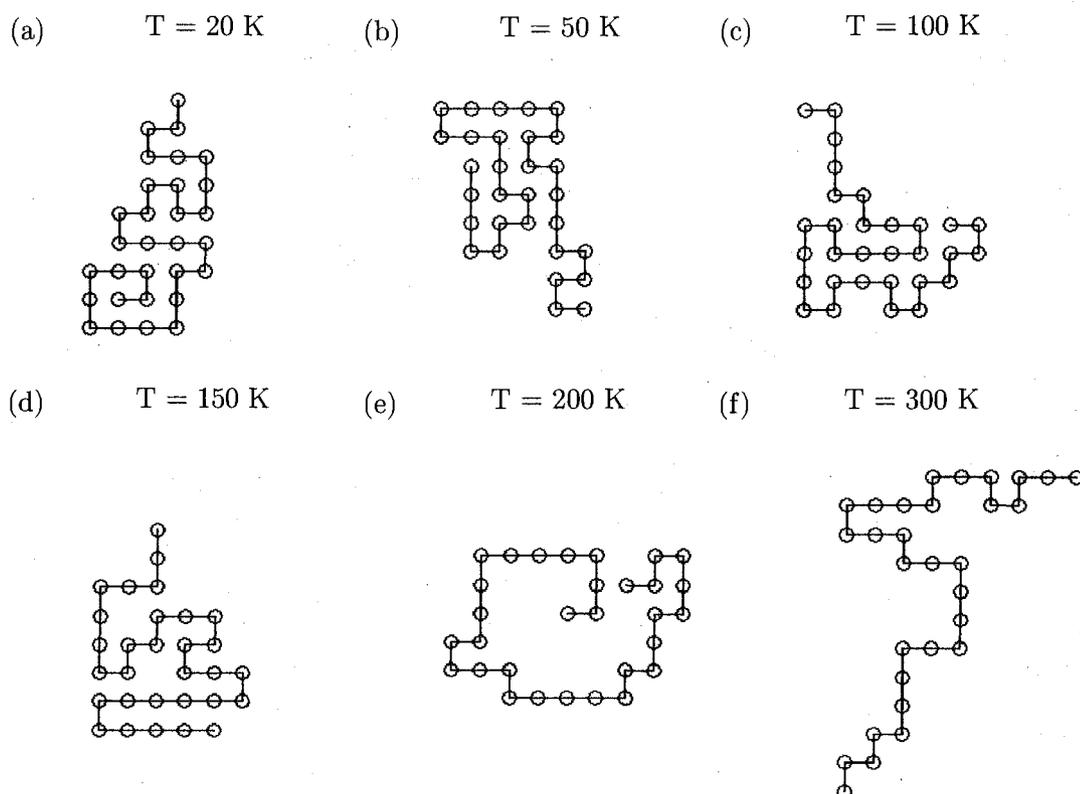


Figure 3.1: Examples of chain conformation for many temperatures with parameters $\epsilon = -100$ K and $L = 30$. With those parameters, the chain undergoes a phase transition from a globular phase to a coil phase around $T \sim 152$ K.

3.3 Statistical measures of protein conformation

We present in this section the results obtained for the different measures introduced above by performing Monte Carlo numerical simulations. The energy of interaction has been set to $\epsilon = -100$ K and about 20000 – 50000 independent conformations of the chain were averaged over about 2×10^6 total conformations generated. The number of transitions for thermalization as well as the number of transitions between two measures required for

equilibrium sampling vary as temperature is varied. They have been set to satisfy the condition that the autocorrelation function for all measures be around zero at equilibrium. Simulations were performed for all temperatures for a given length to present the temperature dependence of the quantities of interest, and around the θ temperature for many lengths to study universal scaling of the quantities of interest.

3.3.1 Autocorrelation functions

With interaction, the convergence of SARWs towards the equilibrium state requires more computational time, especially close to the tricritical temperature where the correlation length goes to infinity. We are interested in equilibrium properties of proteins for all temperatures and for convergence reasons, it is desirable that all measure contributing to the ensemble average be uncorrelated from each other. Correlation information can be obtained by calculating the autocorrelation function $\rho(X)$ of a given measure X on the chain,

$$\rho(X) = \frac{\langle X(t_0)X(t+t_0) \rangle - \langle X^2 \rangle^2}{\langle X^4 \rangle - \langle X^2 \rangle^2}, \quad (3.5)$$

where t_0 is a reference time and t is the actual time. An uncorrelated measure X on the chain is thus defined as a measure for which $\rho(X) \sim 0$. This condition has been satisfied through all numerical simulations for the chain. The autocorrelation function was computed for all measures and the increment of time iteration between two measures was set to have uncorrelated quantities, meaning $\rho(X) \sim 0 \forall X$. This leads to an autocorrelation time of τ of the order of $\sim [L^2, L^4]$.

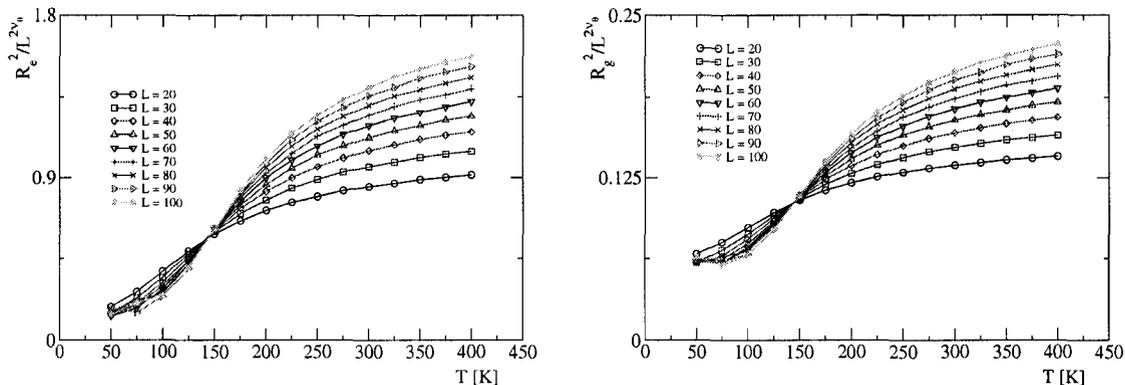


Figure 3.2: Mean squared end to end distance $\langle R_e^2 \rangle$ (left) and mean squared radius of gyration $\langle R_g^2 \rangle$ (Right) scaled by $L^{2\nu_\theta}$ as a function of temperature T [K] for various chain length and lattice constant $a = 1$. At low temperature, favorable energy of interaction dominates and the chain is in a collapsed state. Increasing the temperature allows more conformational energy and the protein undergoes a phase transition to a coil phase at high temperature. The exponent is evaluated at its universal value, $\nu_\theta = 4/7$. The transition occurs close to the tricritical temperature $T_\theta \approx 152$ K, which is the temperature where the distance functions intersect each other.

3.3.2 The end to end distance and radius of gyration

We define the mean squared end to end distance $\langle R_e^2 \rangle$ as,

$$\langle R_e^2 \rangle = \langle (r_L - r_1)^2 \rangle \quad , \quad (3.6)$$

where r_i denote the position of the i^{th} monomer and where the average $\langle \dots \rangle$ is over many independent conformations for the protein, measuring the squared distance between the first and the last monomer along the chain. The mean squared radius of gyration is defined as,

$$\langle R_g^2 \rangle = \frac{1}{(L+1)^2} \sum_i \sum_j \langle (r_i - r_j)^2 \rangle \quad , \quad (3.7)$$

where the average is taken the same way as the mean squared end to end distance. The radius of gyration measures the squared distance between all possible pairs of monomers

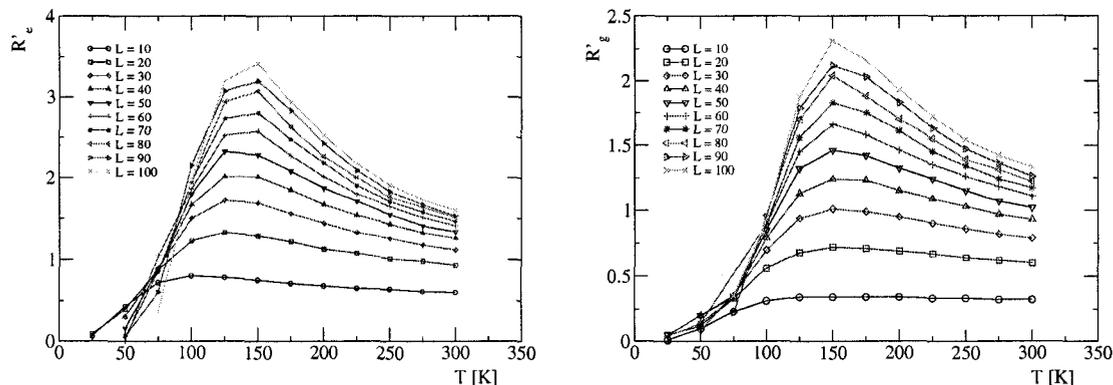


Figure 3.3: Derivative of $\langle R_e^2 \rangle$ (Left) and $\langle R_g^2 \rangle$ (Right) as a function of temperature T [K]. The derivative $\langle R'_e \rangle$ and $\langle R'_g \rangle$ are the correlation function of $\langle R_e^2 \rangle$ and $\langle R_g^2 \rangle$ with the conformational energy of the chain (equation 3.8).

along the chain, and is therefore a more accurate measure on the geometrical conformation of the chain. At low temperature, energy of interaction dominates and the protein is in its globular phase represented by small end to end distance and radius of gyration. The chain undergoes a phase transition by increasing the temperature and the distance functions increase. At large temperature, the chain fluctuates randomly in space, behaving as a SAWS without nearest neighbors interaction, as shown in figure 3.2. The calculation of those two distances was done directly by measuring the appropriate distances between the involved monomers in the calculation.

3.3.3 Derivative of the end to end distance and radius of gyration

The derivative of the mean squared end to end distance and radius of gyration give useful information about a cross over exponent ϕ . By evaluating the derivative of the distance

functions,

$$\begin{aligned}
 \langle R'_\alpha \rangle &= \frac{-1}{\langle R_\alpha^2 \rangle} \left(\frac{\partial \langle R_\alpha^2 \rangle}{\partial \beta} \right) = \frac{-1}{\langle R_\alpha^2 \rangle} \frac{\partial}{\partial \beta} \left[\frac{1}{Z_P} \sum_C R_\alpha^2 e^{-\beta E_P} \right] \\
 &= \frac{1}{\langle R_\alpha^2 \rangle} \left[\left(\frac{1}{Z_P} \sum_C R_\alpha^2 E_P e^{-\beta E_P} \right) - \left(\frac{1}{Z_P} \sum_C R_\alpha^2 e^{-\beta E_P} \right) \left(\frac{1}{Z_P} \sum_C E_P e^{-\beta E_P} \right) \right] \\
 &= \frac{1}{\langle R_\alpha^2 \rangle} [\langle R_\alpha^2 E_P \rangle - \langle R_\alpha^2 \rangle \langle E_P \rangle] \quad ,
 \end{aligned}
 \tag{3.8}$$

with $\alpha = \{e, g\}$, where the index e stands for the end to end distance and g for the radius of gyration, and where E_P is the chain conformational energy. In those terms, the derivative of the mean squared end to end distance and mean squared radius of gyration is the correlation function of the end to end distance or the radius of gyration and the conformational energy of the chain. The temperature dependence of the two quantities is shown in figure 3.3. As $T = 0$, the derivative of the mean squared end to end distance and the radius of gyration is zero meaning that a measure of one of the two distance functions is uncorrelated from a measure of the conformational energy of the protein. Increasing the temperature increases the derivative of the two distance function, indicating that a measure of the distance function is now correlated with the conformational energy of the chain. Those measures were also calculated directly by taking the appropriate average of the distance function and the conformational energy of the protein for all temperatures.

3.3.4 Specific heat of the chain

The specific heat of the protein C_P is given by,

$$C_P = \frac{1}{k_B L T^2} (\langle E_P^2 \rangle - \langle E_P \rangle^2) \quad .
 \tag{3.9}$$

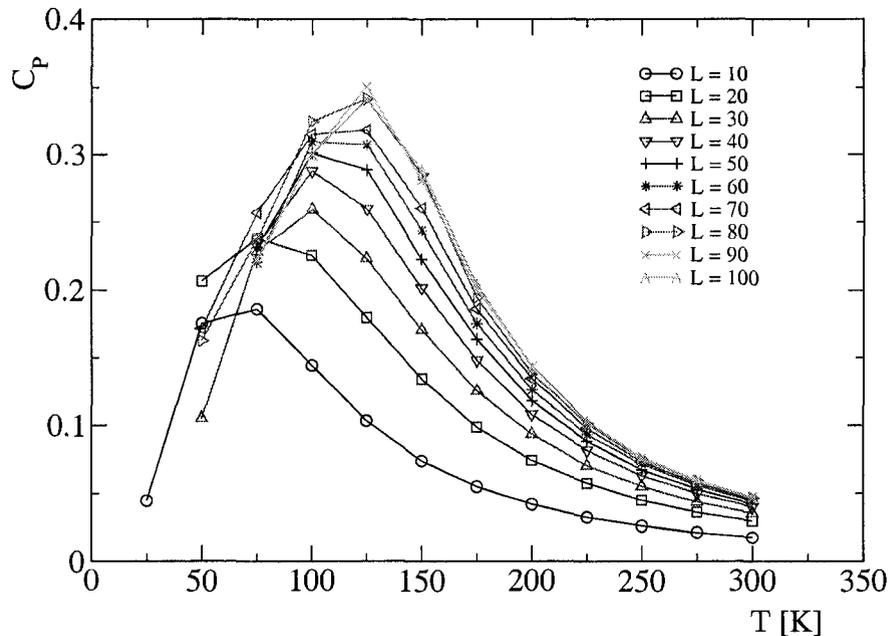


Figure 3.4: Specific heat of the protein C_P as a function of temperature T [K] for various lengths L of the protein. The temperature of the maximum of the specific heat extrapolates to the tricritical temperature in the limit of $L \rightarrow \infty$. The magnitude of the maximum as well as the temperature at which the maximum occurs increases with the length of the protein.

The specific heat is a measure of the fluctuations in energy of the protein, normalized by the length of the chain. At low temperature, the fluctuations are small as the chain is in a collapsed state, and $C_P \rightarrow 0$. Close to the θ temperature, fluctuations in energy diverge as the correlation length is infinite. Since we are working with finite polymer lengths, this divergence is instead a finite maximum in the specific heat, for which the temperature of the maximum as well as its amplitude increase as the length of the polymer is increased, as shown in figure 3.4. At large temperatures, fluctuation of the conformational chain energy is constant and the specific heat decays to zero as $C_P \sim 1/k_B T^2$.

3.3.5 Partition function of the protein model

While the conformational energy and the specific heat of the chain can be measured directly from the numerical simulations, the free energy is more difficult to obtain as the number of

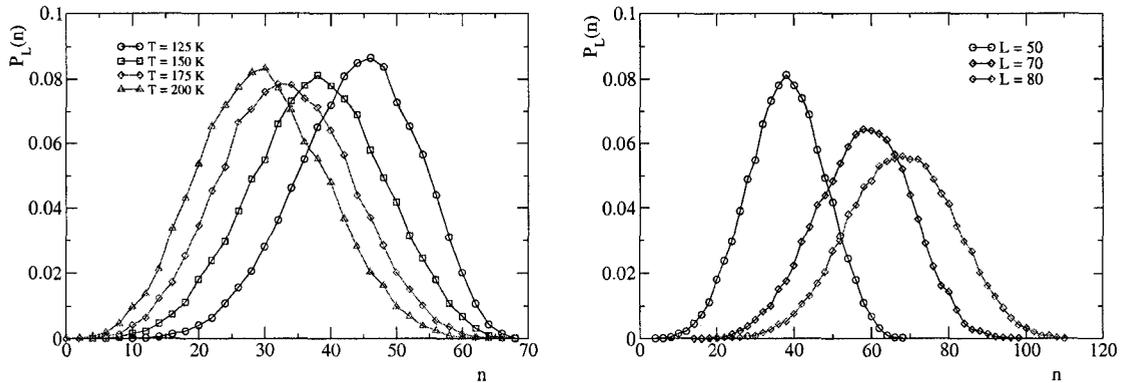


Figure 3.5: Distribution function $P_L(n)$ (normalized such that $\sum_n P_L(n) = 1$) as a function of the number of contacts n for a chain of length L . The left figure shows the distribution function for various temperatures T [K] as the chain length is fixed to $L = 50$ and the right figure shows the distribution for a fixed temperature $T = 150$ K as the chain length is varied.

possible states of a given energy is unknown. We can however estimate the free energy difference ΔF between SARWs and NRRWS (non reversal random walks, or ordinary random walk without the self-avoiding restriction). Following Binder [38], the difference in the free energy is given by

$$-\beta\Delta F(T, L) = \beta [F_{NRRW}(T \rightarrow \infty, L) - F_{SARW}(T, L)] = \ln \left[\frac{Z_{SARW}(T, L)}{Z_{NRRW}(T \rightarrow \infty, L)} \right], \quad (3.10)$$

where Z_{SARW} is the partition function of the generated SARWS and Z_{NRRW} is the partition function of NRRWs. With the equilibrium sampling method, every attempt to generate a NRRWs chain is successful and thus the ratio of the two partition functions is estimated from the fraction of successful SARWs construction. Denoting $P_L(n)$ the fraction of successful SARWs of length L with n nearest neighbor contacts, the ratio of the two partition functions is given by

$$\frac{Z_{SARW}(T, L)}{Z_{NRRW}(T \rightarrow \infty, L)} = \sum_n P_L(n) e^{-\beta n \epsilon}. \quad (3.11)$$

From des Cloizeaux [39] for $L \rightarrow \infty$, the free energy difference can be written as

$$\beta F_{SARW}(T, L) = -L \ln [\mu(T)] - (\gamma - 1) \ln [L] \quad , \quad (3.12)$$

and

$$\beta F_{NRRW}(T, L) = -L \ln [z - 1] \quad z = 4 \quad , \quad (3.13)$$

where μ_t is called the growth parameter and can be interpreted as the effective coordination number of the chain. The probability distribution of the number of contacts of a given chain of length L was directly calculated from the numerical simulation and the exponent γ and the growth parameter were estimated by integrating the distribution function using equation 3.11 and 3.12. The probability distribution of the number of contacts of a chain of length L as calculated by equation 3.11 for many temperatures is shown in figure 3.5.

3.3.6 Structure factor

The structure factor $S(\vec{k})$, measured as a function of the reciprocal lattice vector \vec{k} , is a powerful measure which can be compared to experimental results in three dimensions. It is a description of how a crystallized protein scatters incident radiation. Obviously, no crystallized proteins exist in two dimensions, but the structure factor follows universal scaling laws close to the θ temperature. The structure factor was circularly averaged to eliminate effects due to the orientation of the lattice axes, which agree with the view of chains in a dilute solution[38]. Define $S(\vec{k})$ as,

$$S(\vec{k}) = \left\langle \frac{1}{(L+1)^2} \left| \sum_{i=1}^{L+1} \exp(i\vec{k} \cdot \vec{r}_i) \right|^2 \right\rangle_{|\vec{k}|} \quad . \quad (3.14)$$

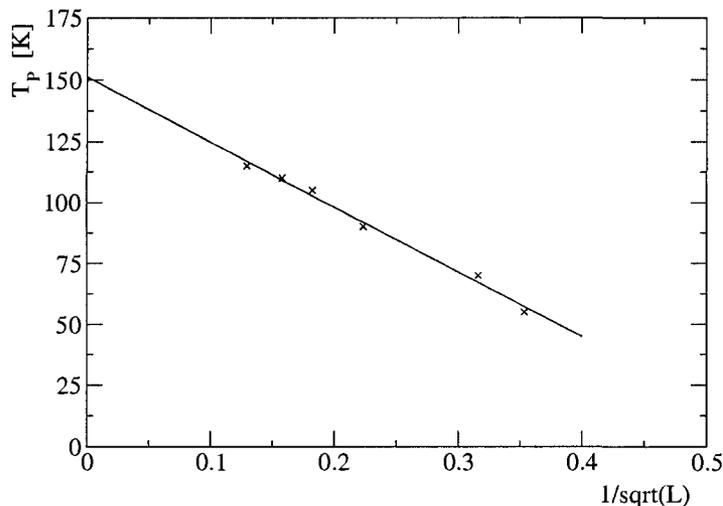


Figure 3.6: Linear fit of the critical temperature of the maximum of the specific heat T_P [K] as a function of the square root of the length of the protein, $1/\sqrt{L}$. The critical temperature of the protein extrapolates to $T_\theta = (151 \pm 3)$ K.

According to Debye [42] and Farnoux [43], the structure factor $S(\vec{k})$, evaluated in its spherical average, follows a universal scaling law around the tricritical temperature given by,

$$S(\vec{k}) \sim k^{-1/\nu}/L \quad \text{with} \quad (2\pi)^2/\langle R^2 \rangle \ll k^2 \ll (2\pi)^2/L^2 \quad , \quad (3.15)$$

$$S(\vec{k}) \sim 1/L \quad \text{with} \quad (2\pi)^2/L^2 \ll k^2$$

Around the tricritical temperature, we have $\nu_t = 4/7$ and therefore $k^2 S(\vec{k})$ should approach a constant, which is another way to identify the θ point.

3.4 Tricritical temperature and critical exponents

3.4.1 Specific heat of the chain and the tricritical temperature

The temperature corresponding to the maximum specific heat extrapolates to the tricritical temperature in the limit of $L \rightarrow \infty$, and is a way to identify the θ temperature of the

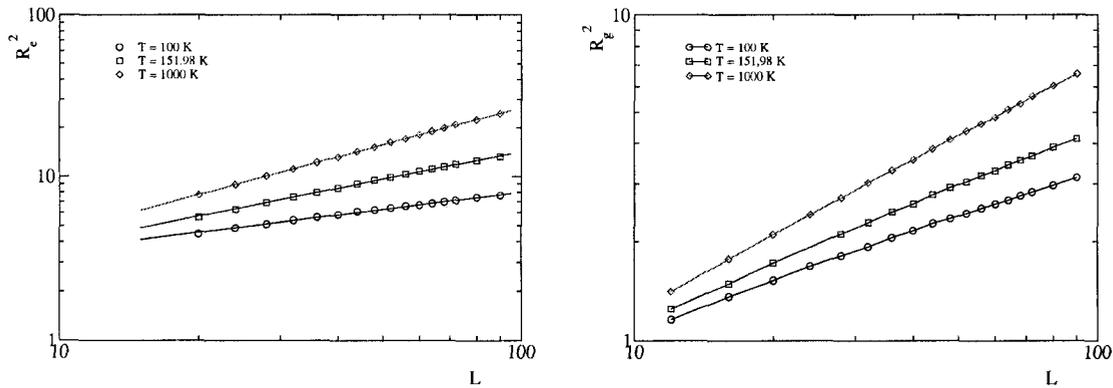


Figure 3.7: (Left) Logarithm of the mean squared end to end distance $\langle R_e^2 \rangle/L$ as a function of the log of the length of the protein L and (Right) logarithm of the mean squared radius of gyration $\langle R_g^2 \rangle/L$ as a function of the log of the chain length. The slope of the power fit corresponds to the exponent ν . Best estimates for ν_c , ν_e and ν_t are shown in table 3.3.

collapse transition of the protein. Let T_P be the temperature at which the specific heat of the protein attains its maximum value. We can estimate T_θ by evaluating the temperature T_P as a function of the inverse square length of the protein, shown in figure 3.6. Our best estimate for the tricritical temperature obtained by the specific heat is $T_\theta = (151 \pm 3)$ K, which agrees with the known value in the literature [40] [41].

3.4.2 The exponent ν

The mean squared end to end distance and the mean squared radius of gyration satisfy the following scaling relation[36],

$$\langle R_e^2 \rangle, \langle R_g^2 \rangle \sim L^{2\nu} \quad . \quad (3.16)$$

As $T \rightarrow \infty$, the polymer behaves as a SARWs and $\nu_e = 3/4$, where the index 'e' stands for extended, while as $T < T_\theta$, the polymer is in a collapsed state and scales as $\nu_c = 1/d$, where the index 'c' denotes the collapsed state and where d is the dimensionality of the system. Around the tricritical temperature, $\nu_t = 4/7$. We can thus identify the θ temperature by looking for temperatures leading to a scaling of $\nu_t = 4/7$ for the end to end distance

T [K]	$\nu_{\langle R_e^2 \rangle}$	$\nu_{\langle R_g^2 \rangle}$
151.6	0.5682	0.5743
151.7	0.5711	0.5741
151.8	0.5705	0.5762
151.9	0.5673	0.5766
151.976	0.5714	0.5708
152.2	0.5687	0.5721
152.3	0.5748	0.5766
152.4	0.5722	0.5733

Table 3.2: Comparison of the exponent ν obtained by the mean squared end to end distance $\langle R_e^2 \rangle$ and the mean squared radius of gyration $\langle R_g^2 \rangle$ for many temperatures, obtained from numerical simulations with chain length $L = [20, 100]$. The best estimates for the exponent ν close to the tricritical temperature is $\nu_t = 0.571 \pm 0.002$ at $T_\theta = 151.976$ K.

and the radius of gyration. An estimate of the θ temperature was obtained by comparing the universal exponent ν for many temperatures around the collapse region. The results are summarized in table 3.2. Our best estimates for the critical temperature is $T_\theta = (151.9 \pm 0.7)$ K, which was the temperature for which the exponent ν is the closest to the exact value, for both the end to end distance and radius of gyration. The universal exponent ν can be calculated from the numerical simulations by fitting equation 3.16 to many protein lengths. Figure 3.7 show a log-log plot of the distance functions as a function of chain length. Our best estimates of the exponent ν at different temperatures are shown in table 3.3. The estimate of the exponent ν for $\langle R_e^2 \rangle$ and $\langle R_g^2 \rangle$ at the tricritical temperature is found to be $\nu_t = 0.571 \pm 0.002$, in excellent agreement with the theoretical value $\nu_t = 4/7 = 0.5714\dots$ We also obtained $\nu_c = 0.50 \pm 0.02$ at $T = 100$ K, which agrees with the theoretical value of $\nu_c = 1/2$, in two dimension. At $T \rightarrow \infty$, we obtained $\nu_e = 0.752 \pm 0.007$, also close to the exact value of $\nu_e = 3/4$ for SARWs without interaction.

T [K]	ν_{exact}	$\nu(\langle R_e^2 \rangle)$	$\nu(\langle R_g^2 \rangle)$	L
100	1/2	0.5 ± 0.02	0.500 ± 0.005	[20,100]
151.976	0.571...	0.571 ± 0.005	0.570 ± 0.003	[40,100]
∞	0.75	0.752 ± 0.007	0.756 ± 0.005	[4,100]

Table 3.3: Best estimates of the critical exponent $\nu(R_\alpha^2)$, where R_α^2 is the mean squared end to end distance $\langle R_e^2 \rangle$ or the mean squared radius of gyration $\langle R_g^2 \rangle$, compared to the theoretical values of ν , ν_{exact} , for 3 different temperatures. The last column indicates the length involved in the estimates of the exponents.

3.4.3 The exponent γ

The probability distribution function $P_L(n)$ was integrated over to obtain the ratio Z_{SARW}/Z_{NRRW} and thus the difference in the free energy between the two walks. The exponent γ was calculated using equation 3.12 around the tricritical point and compared to the exact value of $\gamma = 8/7 = 1.1428\dots$. An example of the linear fit used and a comparison of the calculated exponent γ and the growth parameter μ_t around the tricritical temperature is shown in figure 3.8. Our best estimate of the exponent γ is obtained at $T = 151.976$ K, leading to $\gamma = 1.14 \pm 0.01$ and $\mu_t = 3.24 \pm 0.03$, where chain length between $L = 28$ to $L = 100$ were considered. Those results are in agreement with the exact value of the exponent γ and confirms furthermore the location of the tricritical temperature. The value of the growth parameter agrees with previous calculation performed by Meirovitch and Lim ($\mu_t = 3.213 \pm 0.0013$) [40] and Meirovitch and Chang ($\mu_t = 3.212 \pm 0.007$) [41].

3.4.4 The crossover exponent ϕ

The cross over exponent ϕ can be obtained from the derivative of the mean squared end to end distance and the mean squared radius of gyration. As defined by equation 3.8, the derivative of the distance function corresponds to the correlation function of the distance functions with the conformational energy of the chain. The derivatives of the distance function are

T [K]	γ	μ_t
151.8	1.1370	3.2082
151.9	1.1388	3.1711
151.976	1.1423	3.2429
152.0	1.1635	3.2734
152.1	1.1705	3.1705

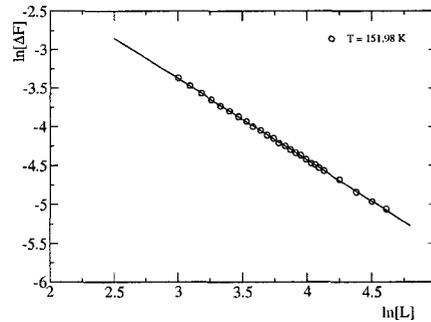


Figure 3.8: Comparison of the exponent γ and the growth parameter μ_t for many temperatures (Left). Our best estimates occurs at $T = 151.976$ K in which we got $\gamma = 1.1423 \pm 0.0108$ and $\mu = 3.2429 \pm 0.0257$. This value is close to the exact value of $\gamma = 8/7 = 1.1428\dots$ and confirms the location of the tricritical temperature. An example at $T = 151.976$ K of the linear fit used is shown on the right figure.

known to scale,

$$\langle R'_\alpha \rangle = \frac{-1}{\langle R_\alpha^2 \rangle} \frac{\partial \langle R_\alpha^2 \rangle}{\partial \beta} = \frac{1}{\langle R_\alpha^2 \rangle} (\langle E_P R_\alpha^2 \rangle - \langle E_P \rangle \langle R_\alpha^2 \rangle) \sim L^{\phi_t} \quad (3.17)$$

where $\alpha = \{e, g\}$. The cross over exponent can be obtained by a power law fit to equation 3.17 around the tricritical temperature, and the results are shown in figure 3.9. Our best estimates is $\phi_t = 0.436 \pm 0.007$ for $\langle R'_e \rangle$ and $\phi_t = 0.43 \pm 0.01$ for $\langle R'_g \rangle$ which average to $\phi_t = 0.436 \pm 0.009$, very close to the exact value of $\phi_t = 3/7 = 0.4285\dots$

L	ν
20	0.574 ± 0.008
30	0.571 ± 0.008
40	0.573 ± 0.003

Table 3.4: Critical exponent ν evaluated from the decay region of the structure factor $S(\vec{k})$ at the tricritical temperature for different protein lengths, scaling as $-1/\nu$. The average leads to $\nu = 0.572 \pm 0.006$.

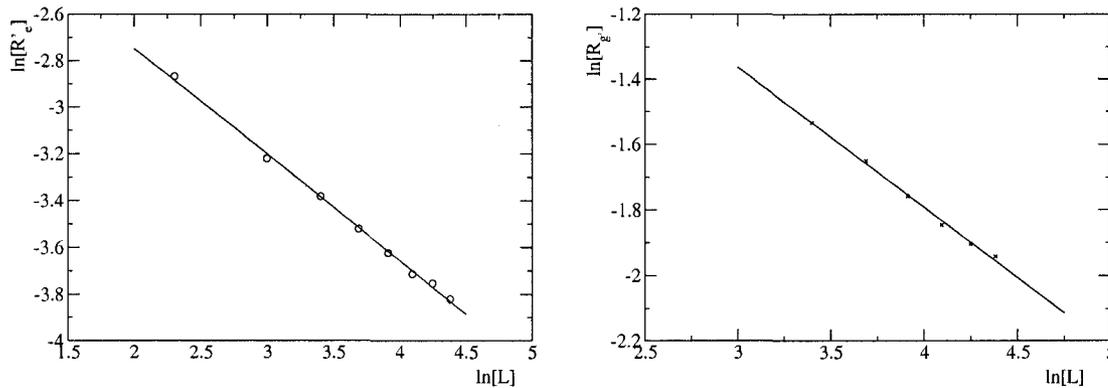


Figure 3.9: Linear fit of the log of the derivative of the mean squared end to end distance $\langle R_e^2 \rangle$ and the mean squared radius of gyration $\langle R_g^2 \rangle$ as a function of the log of the length of the protein. Our best estimates for the cross-over exponent is $\phi_t = 0.436 \pm 0.009$.

3.4.5 Structure factor and tricritical temperature

The structure factor $S(\vec{k})$ is another way to identify the θ temperature and to obtain the exponent ν . Based on equation 3.15, we can obtain ν by evaluating the slope of $S(\vec{k})$ in the decay region, scaling as $-1/\nu$. We have performed numerical simulations for three chain lengths and the average value of the exponent ν is $\nu = 0.572 \pm 0.006$. This estimate agrees with the exact value and the three exponents ν obtained from the three protein's lengths are summarized in table 3.4. An estimate of the θ temperature can be obtained by evaluating $k^2 S(\vec{k})$ around the tricritical temperature, as shown in figure 3.10. In fact, this function should approach a constant value at the tricritical temperature, characterized by having a slope of zero in this region. Our best estimate was obtained by the curve at $T = 151.9$ K, which is close to our other estimates of the tricritical temperature. In summary, our best estimate for the tricritical temperature obtained by the mean squared end to end distance, the mean squared radius of gyration, the specific heat and the structure factor is $T_\theta = (151.9 \pm 0.2)$ K. Table 3.5 presents our value for the exponents ν, γ, ϕ around the tricritical point compared to the exact values in two dimension. We have presented in this section the results obtained by Monte Carlo numerical simulation of a simplified protein

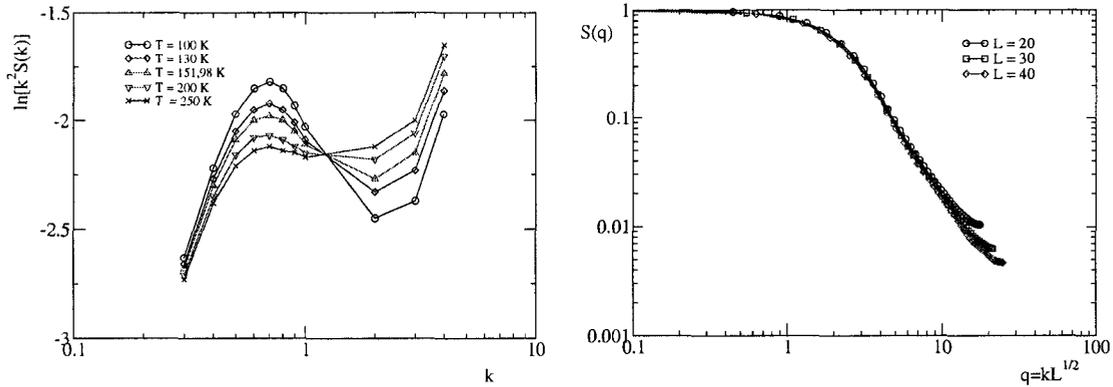


Figure 3.10: (Left) Structure factor $S(q)$ as a function of $q = k\sqrt{L}$ for various polymer lengths. (Right) Structure factor $S(\vec{k})$ as a function of \vec{k} for various temperatures with a chain length of $L = 40$.

model with weakly attractive interaction energy by using the pivot algorithm on a two dimensional lattice. A phase transition is given by the model where the chain is described by a globular phase at low temperature, and by a coil state at high temperature. The tricritical temperature of collapse transition and the universal exponents were measured and are in general very close to the exact values. This analysis validates the protein part of our model. In the next section, we present the results of combining the solvent model and the protein model analysed in this section to study the collapse transition of a chain with hydrophobic interaction.

	Exact	Measured
ν	$4/7 = 0.5714\dots$	0.571(1)
γ	$8/7 = 1.1428\dots$	1.143(5)
ϕ	$3/7 = 0.4285\dots$	0.436(7)

Table 3.5: Comparison of the measured and exact exponent through the analysis. The number in (\dots) in the last column represents the uncertainty on the measures.

Chapter 4

Collapse transition with hydrophobic interaction

The protein chain model introduced in chapter 3 is combined in this section with the solvent model introduced in chapter 2. The aim is to study the collapse transition of SARWs with explicit hydrophobic interactions. We report in this section the results obtained by implementing the combined solvent effects and the dynamics of SARWs by Monte Carlo numerical simulations. The temperature dependence of statistical measures of the protein and the solvent are presented. We show that the collapse transition of the chain is shifted to higher temperatures, in agreement with other numerical simulations with implicit and explicit solvent models [44, 45]. We also calculate the values of the universal exponents close to the collapse temperature, which are modified by the presence of the solvent.

4.1 Thermodynamics of the combined system

The solvent model consists of a square lattice of size $N = n \times n$ solvent molecules with lattice constant $a = 1$. Each molecule can have Q states of orientation, with $S_i = \{1, 2, 3, \dots, Q\}$ and $i = \{1, 2, \dots, N\}$, where S_i corresponds to the state of the i^{th} solvent molecule. Only one orientation state is energetically favored, the special state, $Q = 1$ say. Neighboring

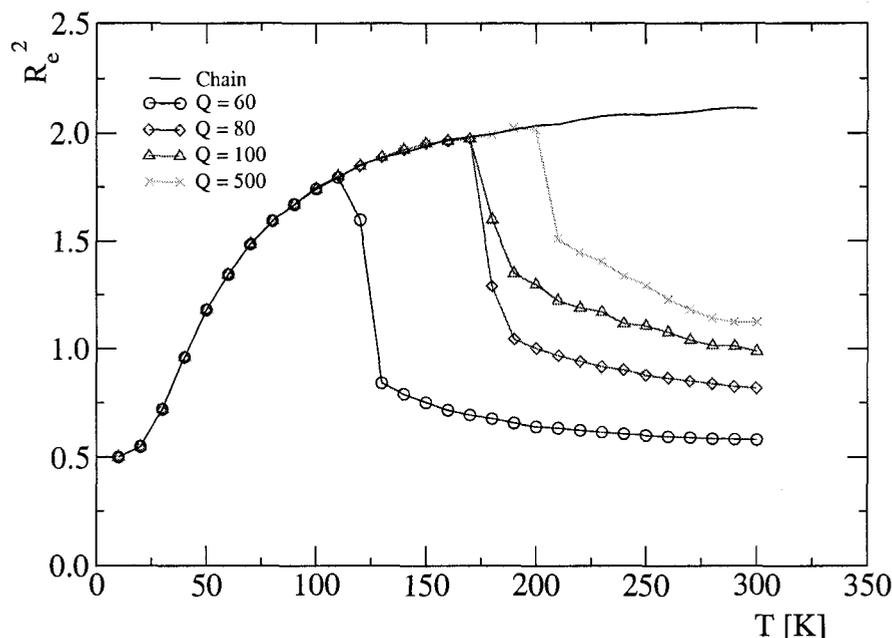


Figure 4.1: Mean squared end to end distance $\langle R_e^2 \rangle$ as a function of the temperature T [K] for $\omega = -200$ K and $\epsilon = -50$ K. In this special implementation, the shell of vapor molecule can not exchange molecule with the solvent. As the temperature is lower than T_c^S , all the bulk solvent molecules are in their special state ; protein's transition are not restricted by the solvent and denaturation of the protein corresponds to collapse transition in vacuum. Above T_c^S , the protein is in a collapsed state due to unfavorable entropy of forcing a bulk solvent molecule to be in its special state. The probability for a solvent molecule to be added in the shell approaches $1/(Q - 1)$ at high temperature. The solid line is the results obtained by numerical simulations of the chain with the same parameters without the solvent.

solvent molecules interact with energy ω ($\omega < 0$) if they are both in their special state and with zero energy otherwise. The accommodation of a hydrophobic particle between two solvent molecules is allowed only if the two molecules are both in their special state. This condition demands that each hydrophobic monomer be always surrounded by a shell of molecules in their special state at all temperatures. The properties of the bulk solvent were studied in chapter 2. At low temperature, the lattice of solvent molecules is ordered, modelling the vapor state. Immersion of the protein model involves no thermodynamical work and equilibrium sampling on the chain is the same as if there was no solvent. At high temperature, the lattice is disordered, modelling liquid state, and immersion of the chain is

entropically unfavored. In this temperature range, free energy minimization is achieved by removing vapor molecules to the shell surrounding the chain, forcing the protein to adopt configurations in a small volume. Increasing the number of accessible states Q increases the entropic cost to have vapor molecule in the equilibrium configuration at high temperature, keeping the protein in a collapsed state. Numerical simulations were performed to illustrate this process, and the end to end distance of the protein model is shown in figure 4.1. This implementation was done by using the same algorithm for the solvent and protein model described respectively in chapters 2 and 3. A collapse transition from a globular state at low temperature to a coil state at large temperature is not observed for temperature greater than the critical temperature of the solvent and the solvent model has to be modified to relax the entropic condition at large temperature. This was done by allowing the exchange of vapor molecules between the shell and the bulk solvent, by introducing a new parameter in the model, the chemical potential μ . The chemical potential corresponds to the change in energy involved in the addition or removing vapor molecules in the bulk solvent. In those terms, the ensemble average for the solvent model is now the grand canonical ensemble. The Hamiltonian of the solvent H_S is given by

$$H_S = \omega \sum_{\langle i,j \rangle} \delta_{S_i,1} \delta_{S_j,1} + \mu \sum_i (1 - \delta_{S_i,1}) \quad , \quad (4.1)$$

where $\omega < 0$, $\mu > 0$ and where $\delta_{S_i,1} = 1$ if $S_i = 1$ and zero otherwise. The chemical potential acts as an external magnetic field in the ferromagnetic Ising model. If $\mu = 0$, there is no energetic cost to add a vapor molecule in the system and the model reduces to the one presented in chapter 2. If μ is small, the energy of interaction dominates the energetic cost to add a vapor molecule in the sampling and a phase transition from an ordered phase at low temperature to a disordered phase at high temperature is included in the system. Measures of the solvent are similar to the results obtained in chapter 2, except that the chemical potential changes the location of the critical temperature, decreasing as μ is increased. In

the regime of high μ , the penalty cost to constrain a solvent molecule in the special state dominates the energy of interaction and no phase transition is given by the model. At low temperature, the lattice is in disordered phase and the probability to find a pair of solvent molecules that are both in their special state, P_{11} , is low. At very large temperature, vapor molecules are added and removed at random, but the entropic cost for constraining a solvent molecule to the special orientation state dominates and $P_{11} = 1/(Q - 1)^2$. The probability of hydrogen bonds formation is thus bounded by $0 \leq P_{11} \leq 1/(Q - 1)^2$ and there is no phase transition in this regime from an ordered phase to a disordered phase by increasing temperature, the bulk solvent is always in the liquid phase.

We model the protein as a SARW on a square lattice with lattice constant $a = 1/\sqrt{2}$, moving in between solvent molecules in their special state. Hydrophobic solute molecules only interact with their nearest neighbors, with energy $\epsilon < 0$. We restrict the model accessible configurations so that a solute molecule can only be accommodated between two solvent molecules that are in their special state. This condition ensures that monomers are always surrounded by vapor molecules, and hence that the adequate entropic penalty for hydrophobicity is captured by the model. The molecules adjacent to solute molecules are different than bulk solvent molecules as they can have only one state of orientation (as opposed to the Q orientation states available to bulk molecules). We refer to molecules that belong to the layer surrounding the protein as shell molecules. Solvent molecules in the shell and bulk interact, and the latter can exchange vapor molecule with the bulk solvent reservoir at chemical potential μ . Immersion of the solute in the solvent consists in the immersion of the protein model between solvent molecules and its shell of vapor molecules. Protein transition is accompanied by a transition of its shell of vapor molecules. Denoting the system chain-shell molecule the hydrophobic system, the Hamiltonian of the hydrophobic system

H_{HS} is,

$$\begin{aligned}
 H_{HS} &= H_P + H_\nu \\
 H_P &= \epsilon \sum_{\langle m,n \rangle} \delta_m \delta_n \\
 H_\nu &= \omega \sum_{\langle k,l \rangle} \delta_k^m \delta_l^m + \mu \sum_k \delta_k^m
 \end{aligned} \tag{4.2}$$

where H_P is the protein's Hamiltonian, $\epsilon < 0$ is the energy of interaction between two neighboring monomers, $\delta_m = 1$ if the site m is occupied by a monomer, 0 otherwise, H_ν is the shell Hamiltonian, $\omega < 0$ is the energy of interaction between two vapor molecules, $\delta_i^m = 1$ if the i^{th} solvent molecule is in the special state and is a neighbor of the monomer m and 0 otherwise, and where μ is the chemical potential. We first describe the qualitative features of the system in terms of the collapse temperature in vacuum T_c , and the critical temperature of the solvent T_c^S . Choose the interaction coefficients of the models so that $T_c < T_c^S$. At very low temperature $T < T_c < T_c^S$, all bulk solvent molecules are in their special state and they can all accommodate a solute molecule in their interstices. Therefore there is no thermodynamical work involved in introducing the hydrophobic chain into the solvent lattice. The solvent and chain decouple and the chain collapses at a transition that is the same as that in vacuum. For $T_c < T < T_c^S$ the protein is in a coil phase as almost all the solvent molecules remain in their special state, and hence the chain and solvent do not really couple. The relevant range for our study is $T > T_c^S$ in which, in principle, the chain would be in an extended coil state, and the solvent in its liquid (disordered) phase. In this range, however, we observe a collapse transition which is mediated by solvent fluctuations. We believe that this collapse transition is the analog in the model of a hydrophobic induced collapse. The mechanism of the collapse relies on thermal fluctuations that cause changes in the density of shell molecules surrounding the chain. At low temperature (but still above T_c^S) solvent molecules in the special state nucleate a sufficiently large domain that can drive

the collapse of the chain inside. If the temperature is further increased, the probability of finding a pair of neighboring bulk solvent molecule in their special state becomes uniform throughout the system, and the chain undergoes a transition to its coil phase. This solvent mediated collapse transition is the subject of the remainder of this thesis.

4.2 Algorithm for chain conformation with solvent effect

Numerical simulations of the combined solvent and chain were performed by using Monte Carlo methods. The two systems are simultaneously simulated, and we keep track of both equilibrium averages of solvent and chain. In order to embed the chain into the lattice of solvent molecules, we have implemented a Monte Carlo process of both systems through the BKL algorithm. The lattice of solvent molecule consists initially of a lattice for which all the solvent molecules are in disordered state (all class 10). The system is then thermalized for a certain simulation time, providing an equilibrium configuration for the solvent lattice in the same way as in chapter 1, but with the modified Hamiltonian defined in equation 4.1. Consider a transition for an initial state of the lattice of solvent molecules S to a new configuration S' . The change in the energy $\Delta E_{solvent}$ from the initial state S to the new state S' is given by,

$$\Delta E_{solvent} = \omega \Delta N_{11} - (k_B T \ln [Q - 1] + \mu) \Delta N_Q \quad , \quad (4.3)$$

where ΔN_{11} is the change in the number of bonds and ΔN_Q is the change in the number of molecules that are not in their special state. The Metropolis transition probabilities $P(S, S')$ are given by

$$P(S, S') = \begin{cases} 1 & \text{if } \Delta E_{solvent} \leq 0 \\ e^{-\beta \Delta E_{solvent}} & \text{if } \Delta E_{solvent} > 0 \end{cases} . \quad (4.4)$$

This choice of transition rates satisfies detailed balance as shown in chapter 1. Embedding of the chain is then performed on this equilibrated configuration except that the protein is initially placed linearly between solvent molecules. In order to accommodate the latter, all the neighboring solvent molecules around the chain are forced to be in their special state. The chain is thermalized for a certain simulation time while the lattice of solvent molecules is kept fixed, providing an equilibrium configuration for the chain. Equilibrium average quantities are then measured for many independent configurations of the combined solvent and chain at a given temperature. This procedure is repeated for many temperatures to study the temperature dependence of the quantities of interest. Attempted Monte Carlo moves of the chain consist of a chain conformational change together with its shell of vapor molecules and have been implemented through the BKL algorithm. In the standard Monte Carlo method, a monomer is chosen at random (the pivot) and a symmetry operation is also chosen at random. Rotation is then applied to the left or right segment of the protein around the pivot. Transitions violating the excluded volume condition are rejected. The energy change $\Delta E_{protein}$ between the new conformation and the old conformation is calculated and the transition is accepted with probability $P = 1$ if $\Delta E_{protein} \leq 0$ and with probability $P = \exp(-\beta \Delta E_{protein})$ if $\Delta E_{protein} > 0$. However, because the solvent is disordered in the temperature range of interest, the probability of finding a pair of neighboring solvent molecules that can accommodate a monomer is low. A transition of the protein together with its shell is a rare event, leading to large simulation times. By implementing the transition dynamics through the BKL algorithm, we take advantage of the fact that the transition probabilities of the chain are already known, and we perform at each iteration a transition of the protein together with its shell of vapor molecules. In fact, in the BKL algorithm, we consider all possible transitions from a given initial state of the protein and its shell of solvent molecules and calculate the energy change for all possible transitions with respect to the initial configuration to attribute a probability to each possible transition with the

Metropolis algorithm, to satisfy detailed balance. Consider a transition from an initial state S to a new conformation state S' . The change in the energy ΔE_{HS} from the initial state S to the new state S' is given by,

$$\Delta E_{HS} = \epsilon \Delta N^{(P)} + \omega \Delta N_{11}^{(\nu)} + \mu \Delta N_1^{(\nu)} \quad , \quad (4.5)$$

where $\epsilon, \omega < 0$ and $\mu > 0$, $N^{(P)}$ is the number of contacts between neighboring monomers, $N_{11}^{(\nu)}$ is the number of contacts between neighboring vapor molecules in the shell or between bulk solvent molecules in their special state, and $N_1^{(\nu)}$ is the number of vapor molecules contributing to the shell. The Metropolis transition probabilities $P(S, S')$ are given by

$$P(S, S') = \begin{cases} 1 & \text{if } \Delta E_{protein} \leq 0 \\ e^{-\beta \Delta E_{protein}} & \text{if } \Delta E_{protein} > 0 \end{cases} . \quad (4.6)$$

This choice of transition rates satisfies detailed balance. A transition probability 4.6 is associated with each move of the hydrophobic chain. Transitions violating the excluded volume constraint are assigned zero probability. Only rotations of $\pm\pi/2$ to the right or left of the pivot are allowed so there are $4(L-1)$ possible transitions for the protein and its shell given some initial conformation, where L is the number of monomers. We then construct the quantities

$$R_i = \sum_{j=0}^i P_j \quad i = 1, 2, \dots, 4(L-1) \quad , \quad (4.7)$$

in a similar way to equation 2.20, where P_j is the probability of each transition, and with $R_0 = 0$. We then choose an uniformly distributed random number x in the interval $I = [0, R_{4(L-1)}]$ and identify the i^{th} transition for the protein to be performed so that $R_{i-1} \leq x < R_i$. Motion of the chain and its shell of solvent molecules is done by updating the coordinates and the

energy of all monomers and vapor molecules. Time is then updated by an increment

$$\Delta t = -\frac{\ln(y)}{R_{10}} \quad , \quad (4.8)$$

where y is an uniformly distributed random number in the interval $I = [0, 1]$. The (x, y) coordinates and energy of all the monomers and vapor molecules of the applied transition constitute the new initial condition for the next iteration. The flow chart of the algorithm is shown in table 4.2 and an example of the solvent lattice together with the hydrophobic system for six different temperatures is shown in figure 4.2.

4.3 Collapse transition with hydrophobic interaction

We have performed numerical simulations with the algorithm presented in the last section in which about 20 000 independent configurations were averaged for a protein model of length $L = 4$ to study the dependence over the parameters of the combined model. The dependence of the energy of interaction between neighboring molecules in their special state ω and the number of accessible state Q of the solvent model is explained in chapter 2. Decreasing ω or increasing Q decrease the critical temperature of the solvent T_c^S and change the shape of the transition of the lattice of solvent molecule. The energy of interaction between neighboring monomers ϵ , determines the temperature of the collapse transition of the protein in vacuum, as shown in chapter 3. A new parameter was introduced in the combined model, the chemical potential μ . We described next the temperature dependence of thermodynamical quantities for a small length to capture the effect of the chemical potential on the chain dynamics. The parameters were chosen to be $Q = 50$, $\omega = -100$ K and $\epsilon = -100$ K, for a SARW of length $L = 4$, while the chemical potential is varied.

Flow chart for solvent and hydrophobic system updates (1/3)

1. Initialize the 2-dimensional lattice of solvent molecule (all in class 10) with lattice constant $a = 1$.

 2. Thermalize the solvent
 - a. Construct the numbers $Q_i = \sum_{j=0}^i N_j P_j$, with $i = \{1, \dots, 10\}$, where P_j is the probability to perform a transition to the j^{th} class as determined by equation 4.4, where N_j is the number of molecules in the j^{th} class and where $Q_0 = 0$.
 - b. Choose a uniformly distributed random number x in the interval $I = [0, 1]$. Identify the class of the molecule which will undergo the transition such that $Q_{i-1} \leq x < Q_i$. Choose a molecule in the i^{th} class at random.
 - c. Perform the transition of the molecule by updating its class, +5 if the molecule is initially in its special state and -5 if the molecule is in another state.
 - d. Update the class of the 4 nearest neighbors of the molecule, +1 if the molecule was originally in its special state (all the 4 nearest neighbors of the molecule lose a neighbor in its special state) and -1 if the molecule was not originally in its special state (all the 4 nearest neighbors of molecule gain a neighbor in its special state).
 - e. Update the number of molecules in each class.
 - f. Update the time by an increment of $\Delta t = -\ln(y)/Q_{10}$, where y is a uniformly distributed random number in the interval $I = [0, 1]$.

 3. Repeat Step a - f up to a maximal simulation time providing an equilibrium lattice.

 4. Initialize the protein of length L on the two dimensional square lattice of solvent molecules with lattice constant $a = 1/\sqrt{2}$ by placing the linearly L monomers between solvent molecules, starting at a site (k, l) chosen at random.

 5. Force the neighboring solvent molecules along the protein to be in the special state and calculate the change in the free energy involved ΔF_S .
-

Flow chart for solvent and hydrophobic system updates (2/3)

6. Thermalize the hydrophobic system.

- g. Calculate the conformational energy of the chain, the interaction energy of the vapor molecule of the shell, and the number vapor molecule that contributes to the shell of the protein.

The latter together with the (x, y) coordinate of the all monomers and vapor molecules contributing to the protein's shell corresponds to the initial configuration for one iteration step.

- h. Perform each possible 4 transitions ($\pm\pi/2$ applied to the left or right of the monomer m , with $m = \{1, 2, \dots, L\}$) for all monomers and keep track of the (x, y) coordinates of all the monomers and vapor molecules. If a solvent molecule in the new location is a neighbor of one monomer but is not in its special state, perform a transition for this molecule to its special state. If a solvent molecule was a neighbor of a monomer in the original conformation but doesn't contribute to the shell in the new location and was not in its special state before the immersion of the protein, perform a transition for this solvent molecule to one of the other $(Q - 1)$ orientation state.

- i. Calculate the new conformational energy of the protein as well as the interaction energy between pairs of vapor molecule and the number of vapor molecule contributing to the shell around the protein.

- j. Calculate the energy change $\Delta E_P = \epsilon\Delta N^{(P)} + \omega\Delta N_{11}^{(\nu)} + \mu\Delta N_1^{(\nu)}$ between the old and the new conformation for the hydrophobic system and attribute a probability for each possible $4(L - 1)$ transitions.

Attribute a probability $P = 0$ if the new conformation violates excluded volume condition. Transitions for which $\Delta E_P \leq 0$ are given probability $P = 1$ while transitions for which $\Delta E_P > 0$ are given probability $P = \exp(-\beta\Delta E_P)$.

- k. Construct the number $R_i = \sum_{j=0}^i P_j$ where P_j is the probability of the j^{th} transition, where $i = \{1, 2, \dots, 4(L - 1)\}$ and $R_0 = 0$.

- l. Choose the i^{th} transition to be performed by generating a uniformly distributed random number x in the interval $I = [0, R_{4(L-1)}]$.

Identify the transition to be performed such that $R_{i-1} \leq x < R_i$.

- m. Perform the chosen transition by updating the (x, y) coordinates and energy of interaction of all the monomers and the vapor molecules that belongs to the shell surrounding the protein. The new conformation becomes the new initial conformation for a new iteration step.

- n. Update the time by an amount of $-\ln(y)/R_{4(L-1)}$, where y is a uniformly distributed random number taken the interval $I = [0, 1]$.

 Flow chart for solvent and hydrophobic system updates (3/3)

7. Repeat steps g - n for a maximal simulation time providing an equilibrium configuration.
8. For many independent configuration averages,
 9. Repeat steps a - f to perform transitions of the solvent of lattice molecules for some maximal simulation time, set to provide uncorrelated measurements.
 10. Repeat Step 5 to force the immersion of the protein.
 11. Repeat steps g - n to perform transitions of the hydrophobic system up to some maximal simulation time providing uncorrelated measures.
 12. Measure the thermodynamical quantities P_{11} , c , C_S , and M on the lattice of solvent molecules.
 13. Measure the thermodynamical quantities $\langle R_E^2 \rangle$, $\langle R_g^2 \rangle$, $\langle R'_E \rangle$, $\langle R'_g \rangle$, C_P , C_ν , C_{tot} , $P_L(n)$, $\langle \Delta N_\nu^2 \rangle$, $C(E_P, E_\nu)$, and $C(E_P, N_\nu)$ on the hydrophobic system model.
14. Display the mean of the thermodynamical average of the quantities of interest.

 Table 4.1: Flow chart for hydrophobic system and solvent updates.

Chain conformations together with its shell of vapor molecules

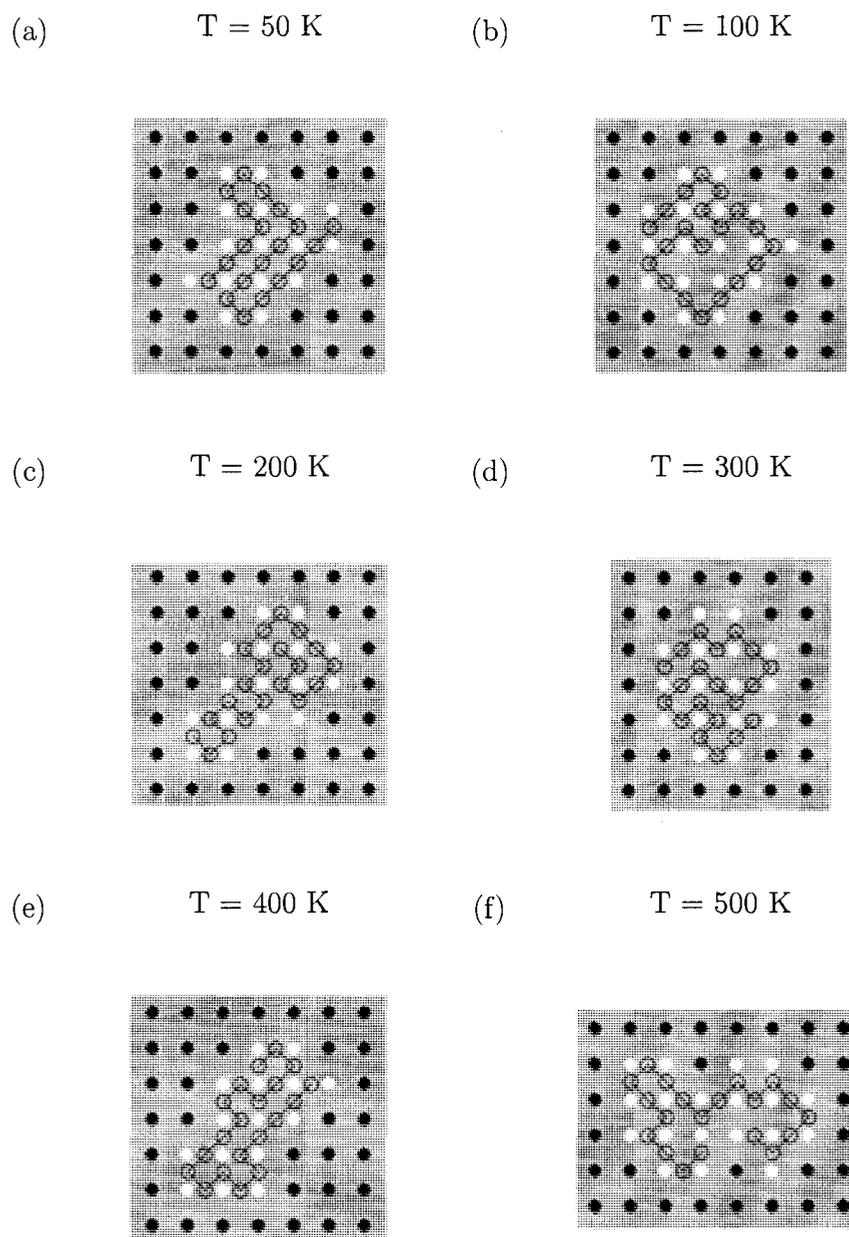


Figure 4.2: Examples of chain conformation (open circles) with its shell of vapor molecule for many temperatures with parameters $\omega = -100$ K, $Q = 50$, $\epsilon = -100$ K and $L = 20$. The white filled circles represent vapor molecules while the black ones are bulk solvent molecules in the liquid phase. At low temperature, the chain is in a collapsed state and entropy maximization of the total system is achieved by removing shell molecules. At high temperature, the protein fluctuates randomly in space and bulk solvent molecules contribute to extend the volume available in space for the protein.

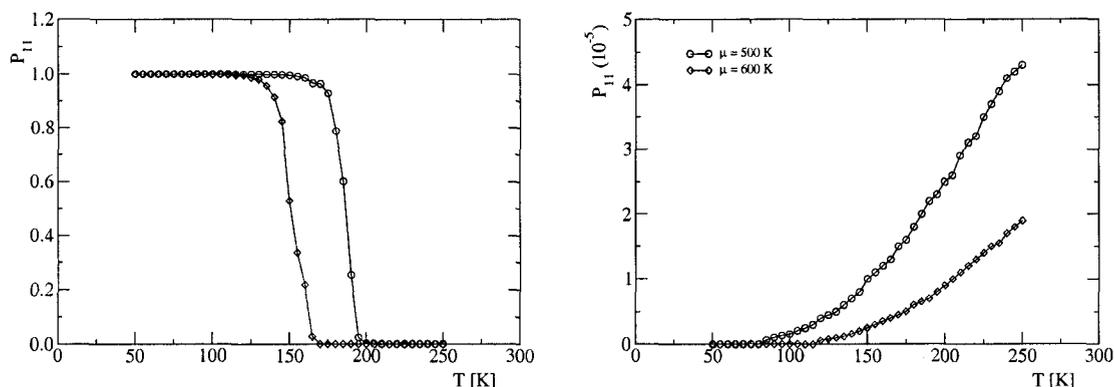


Figure 4.3: Probability that a neighboring pair of solvent molecule be both in the special state (P_{11}) as a function of temperature T [K], calculated with the modified Hamiltonian (equation 4.1), with parameters $\omega = -421$ K and $Q = 50$. The chemical potential is low in the left figure and a phase transition from an ordered state to a disordered state is included in the model. Increasing the chemical potential decreases the critical temperature of phase transition. The same probability function is shown in the right figure as the chemical potential is large. In this regime, the lattice is in a disordered state at low temperature and the probability of hydrogen bonds formation is low due to the entropic and energetic cost of constraining a solvent molecule in its special state.

4.3.1 Solvent model with chemical potential

The probability that a pair of solvent molecules be both in their special state, P_{11} , was calculated to illustrate the dynamics of the solvent model in the presence of the chemical potential μ . The results are shown in figure 4.3. In the low chemical potential regime, the energy of interaction between neighboring vapor molecule dominates the energetic cost to add vapor molecule in the configuration and the system undergoes a phase transition from an ordered phase at low temperature to a disordered phase a high temperature. The function P_{11} is similar to the case of $\mu = 0$. Increasing μ in this regime decreases the critical temperature of phase transition. On the other hand, as μ is large, the energetic cost to add a vapor molecule dominates the energy of interaction between pair of vapor molecule, and there is no phase transition. The solvent is in its liquid phase and increasing temperature increases fluctuations in the number of vapor molecules. The function P_{11} is low at low

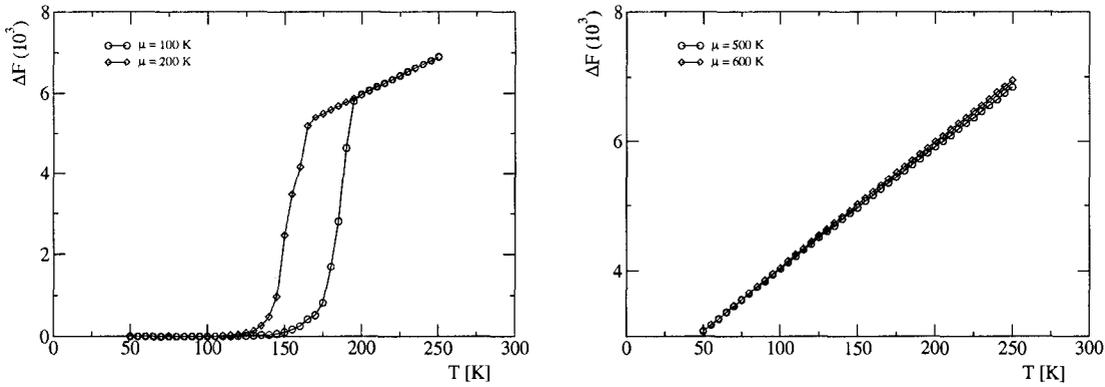


Figure 4.4: Thermodynamical work ΔF_S required to immerse the chain in the solvent model as a function of the temperature T [K]. The results in the low chemical potential regime are shown in the left figure. As $T < T_c^S$, the lattice is in an ordered state and there is no thermodynamical work required to immerse the protein in the solvent. As $T > T_c^S$, the lattice is disordered and the thermodynamical work to immerse the protein increases linearly with temperature. The results for the large chemical potential regime are shown in the right figure. The lattice is in the liquid state for all temperatures and the thermodynamical work increases linearly with temperature.

temperature and slowly increases as temperature is increased. The relevant range for this study is the liquid state and we have set the parameters to satisfy this condition.

4.3.2 Thermodynamical work

Thermodynamic work is required to immerse the chain in the solvent. This work corresponds to the change in the free energy of the solvent (ΔF_S) as the accommodation of the shell is done,

$$\Delta F_S = \omega \Delta N_{11} - (T \ln(Q - 1) + \mu) \Delta N_Q \quad , \quad (4.9)$$

where ΔN_{11} is the change in the number of pairs of solvent molecules in their special state, and ΔN_Q is the change in the number of water molecules. The results obtained from the numerical simulations are shown in figure 4.4. As μ is low, the lattice is ordered at low temperature and no thermodynamic work is required for the immersion of the solute. In-

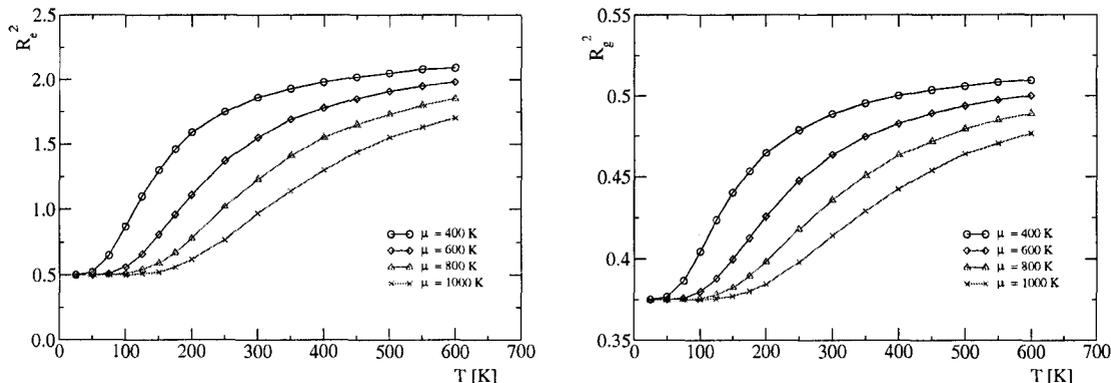


Figure 4.5: Mean squared end to end distance $\langle R_e^2 \rangle$ (Left) and mean squared radius of gyration $\langle R_g^2 \rangle$ (Right) as a function of temperature T [K] with hydrophobic interaction for a small chain length $L = 4$ with $Q = 50$, $\omega = -100$ K and $\epsilon = -100$ K. As the temperature is low, the chain is in a collapsed state due to the entropy cost to add solvent molecules in their special state to the shell surrounding the protein. Increasing the temperature drives a collapse transition from a globular state at low temperature to a coil state at high temperature due to density fluctuation of solvent molecule in their special state contributing to the shell around the chain. Increasing the chemical potential μ increases the temperature of collapse transition with hydrophobic interaction.

creasing temperature allows more vapor molecule in the configuration, and $\Delta F_S > 0$. In the regime of high μ , the lattice is disordered and there is no phase transition. Immersion of the hydrophobic system always involves a thermodynamical work, increasing linearly with temperature.

4.3.3 Distance functions and their derivatives

We generally observe that the collapse transition of the protein with explicit hydrophobic interactions is qualitatively similar to collapse of the SARW in vacuum. Figure 4.5 shows the temperature dependence of the mean squared end to end distance $\langle R_e^2 \rangle$ and the radius of gyration $\langle R_g^2 \rangle$. The distance between pairs of monomers is small at low temperature, and the protein is in a globular phase. Increasing temperature allows wider sampling of conformation space as more vapor molecules are available to allow chain motion, and eventually the protein

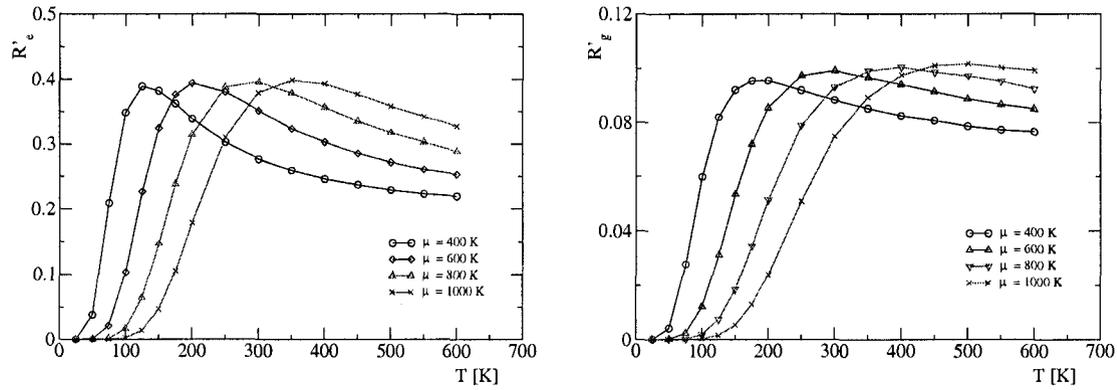


Figure 4.6: Derivative of the mean squared end to end distance $\langle R'_e \rangle$ (Left) and derivative of the mean squared radius of gyration $\langle R'_g \rangle$ (Right) as a function of temperature T [K] with hydrophobic interaction for a small chain length $L = 4$ with $Q = 50$, $\omega = -100$ K and $\epsilon = -100$ K. Increasing temperature increases the derivative of the two distances functions, indicating that the correlation between the distance functions and the conformational energy increases as well. The derivatives reach a maximum close to the critical temperature and decrease slowly afterwards, indicating that the conformational energy is correlated with the distance functions in this region.

undergoes a transition to its coil state at high temperature. Increasing the chemical potential increases the collapse transition temperature. The average quantities of the two measures of length have been calculated in the same way than in chapter 3, by using equation 3.6 and 3.7.

The temperature dependence of the derivatives of the mean squared end to end distance $\langle R'_e \rangle$, and of the radius of gyration $\langle R'_g \rangle$ are also similar to the collapse transition of the SARW in isolation, and are shown in figure 4.6. The derivative of the distance functions is a measure of the correlation of the conformational energy of the chain and the distance functions themselves. The derivative increases as temperature is increased, meaning that the conformational energy and the distance function become correlated up to a maximum, close to the collapse temperature. The two derivatives were calculated in the same way as in chapter 3, by using equation 3.8.

4.3.4 Heat capacity and fluctuations of the shell molecule number

New measures are introduced due to the presence of the layer of vapor molecules surrounding the hydrophobic solute. Define the fluctuation of the number of vapor molecule in the shell $\langle \Delta N_\nu^2 \rangle$ as,

$$\langle \Delta N_\nu^2 \rangle = \langle N_\nu^2 \rangle - \langle N_\nu \rangle^2 \quad . \quad (4.10)$$

Define the heat capacity of the chain C_P , measuring the fluctuations in energy of the protein, the heat capacity of the shell of solvent molecule C_ν , a measure of the fluctuation of the energy of interaction of vapor molecules surrounding the chain, and the heat capacity of the total system composed of the hydrophobic chain and its shell of vapor molecule C_{tot} as

$$C_P = \frac{\langle \Delta E_P^2 \rangle}{k_B T^2} = \frac{(\langle E_P^2 \rangle - \langle E_P \rangle^2)}{k_B T^2} \quad ,$$

$$C_\nu = \frac{\langle \Delta E_\nu^2 \rangle}{k_B T^2} = \frac{(\langle E_\nu^2 \rangle - \langle E_\nu \rangle^2)}{k_B T^2} \quad ,$$

(4.11)

$$\begin{aligned} C_{tot} &= \frac{\langle \Delta E_{tot}^2 \rangle}{k_B T^2} = \frac{(\langle E_{tot}^2 \rangle - \langle E_{tot} \rangle^2)}{k_B T^2} \\ &= \frac{\langle \Delta E_P^2 \rangle}{k_B T^2} + \frac{\langle \Delta E_\nu^2 \rangle}{k_B T^2} + 2 \frac{1}{k_B T^2} (\langle E_P E_\nu \rangle - \langle E_P \rangle \langle E_\nu \rangle) \\ &= C_P + C_\nu + 2 \frac{C(E_P, E_\nu)}{k_B T^2} \quad , \end{aligned}$$

where E_P is the protein energy, E_ν is the energy of the shell molecule, $E_{tot} = E_P + E_\nu$ and $C(E_P, E_\nu)$ is the correlation function between the conformational energy of the chain and the interaction energy between vapor molecule in the shell. The temperature dependence of the the heat capacity of the protein (C_P), the fluctuations of the number of vapor molecules in the shell ($\langle \Delta N_\nu^2 \rangle$), the heat capacity relative to the fluctuation in energy in the shell (C_ν) and the heat capacity of the total system (C_{tot}) are shown in figure 4.7. The the heat capacity of the chain is similar to that of the collapse of SARWs in vacuum, it reaches a

maximum at some temperature and decreases afterwards. However, the temperature for which it reaches a maximum is shifted to higher temperatures when compared to the ones obtained for SARWs in vacuum, indication that the collapse temperature is shifted due to the hydrophobic interaction. The qualitative behavior of the heat capacity of the interaction energy of vapor molecule in the shell are similar to the heat capacity of the chain and the total heat capacity. In particular, they reach a maximum at some temperature close to the temperature of the maximum of the heat capacity of the protein. The fluctuations of the number of vapor molecules in the shell ($\langle \Delta N_\nu^2 \rangle$) also increase up to a maximum and decrease afterwards. This is evidence that a change in vapor density in the shell surrounding the protein drives the collapse transition of the chain.

Define the correlation function of the correlation energy of vapor molecule in the shell and the interaction energy of the hydrophobic molecule $C(E_P, E_\nu)$ and the correlation function of the conformational energy of the chain with the number of vapor molecules in the shell $C(E_P, N_\nu)$ as

$$C(E_P, E_\nu) = \langle E_P E_\nu \rangle - \langle E_P \rangle \langle E_\nu \rangle \quad , \quad (4.12)$$

and

$$C(E_P, N_\nu) = \langle E_P N_\nu \rangle - \langle E_P \rangle \langle N_\nu \rangle \quad . \quad (4.13)$$

The two functions measure the correlation between the two measures. If the correlation function is low, the two measures are independent from each other while they are correlated if the correlation function is high. Results from numerical simulations of the correlation functions are shown in figure 4.9. The two measures behave the same way. At low temperature, the correlation functions are small, indicating that the chain dynamics are independent of the shell. In this range of temperature, the energy of interaction between hydrophobic monomers dominates and the chain is in a collapsed state, independently of the shell confor-

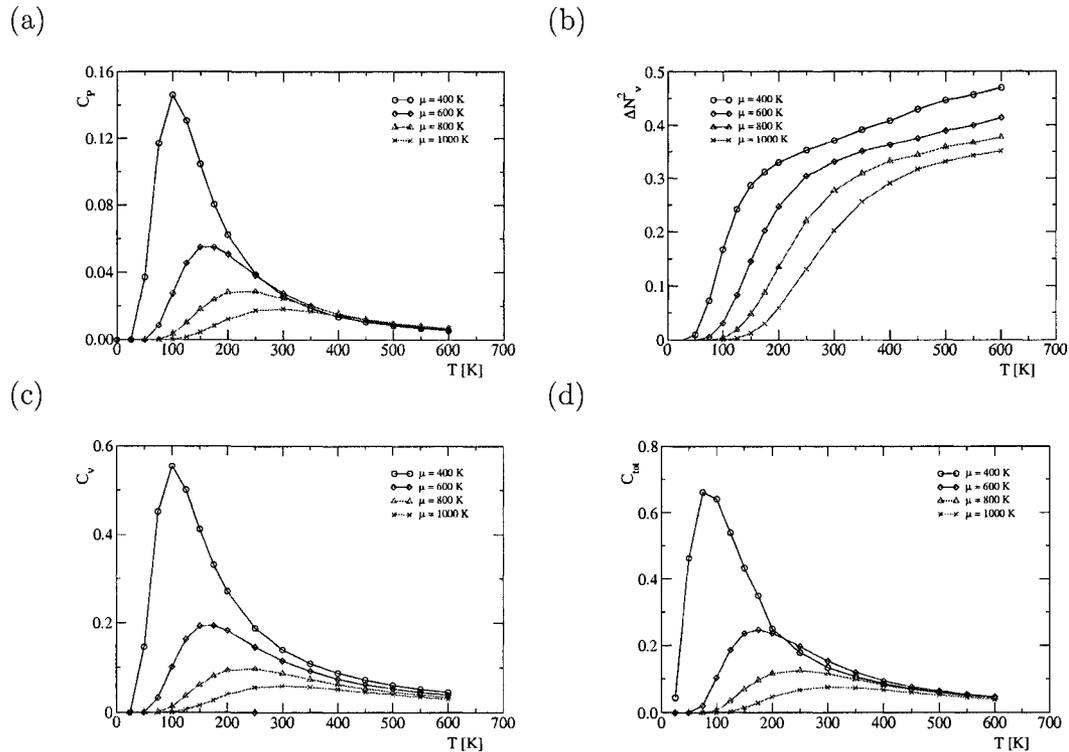


Figure 4.7: (a) Heat capacity of the chain C_P , (b) fluctuation of the number of vapor molecule in the shell $\langle \Delta N_v^2 \rangle$, (c) heat capacity of the interaction energy of vapor molecule in the shell C_v , and (d) heat capacity of the total system C_{tot} as a function of temperature T [K]. The heat capacity of the chain and of the bonds in the shell reach together with the fluctuation vapor molecule in the shell a maximum at the same temperature for a given chemical potential. The heat capacity of the total system follows the same behavior but the temperature of the maximum is higher. Increasing the chemical potential increases the temperature of the maximum and decreases the amplitude for the three measures of the specific heat and for the fluctuation of the number of vapor molecule in the shell.

mation. The correlation functions then increase as temperature is increased. As compared to the end to end distance or the radius of gyration, the correlation function increases as the chain is still in a collapsed state, indicating that the shell of vapor molecule restrict the volume available for the chain. The correlation between the chain and the shell dynamics increases as temperature is increased up to a maximum close to the critical point, and slowly decrease as temperature is further increased.

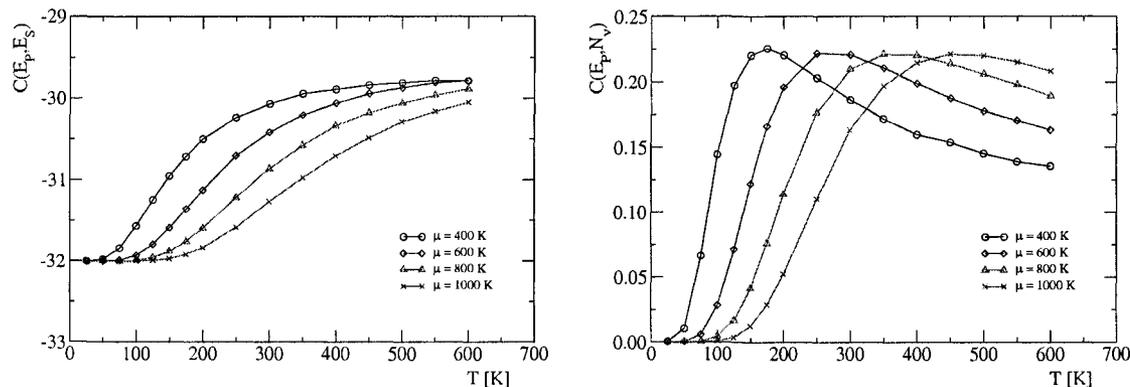


Figure 4.8: Correlation function of the conformational energy of the chain and the interaction energy of the bonds in the shell, $C(E_P, E_S)$, (Left) and between the conformation energy of the chain with the number of vapor molecule in the shell, $C(E_P, N_V)$ (Right). At very low temperature, the correlation function is low as the conformational energy of the chain dominates the dynamics and the chain is in a collapsed state. The correlation function increases at some temperature while the chain is still in a collapsed state, indicating that the shell restricts the volume available for the chain. The correlation energy reaches a maximum close to the critical temperature, and then slowly decreases with temperature. Increasing the chemical potential increases the collapse temperature and thus shifts the temperature at which the correlation functions reach a maximum.

The temperature dependence of the measures of the chain and its shell of vapor molecules for a small chain indicates that the chain undergoes a collapse transition from a globular phase at low temperature to a coil state at high temperature. Fluctuations in the solvent density of the chain drive the collapse; the protein is in a collapsed state at low temperature as the shell restricts the accessible volume for the chain. Increasing temperature, but still below the critical temperature, relaxes this condition and more vapor molecules are allowed around the chain, creating new bonds that can accommodate a hydrophobic monomer and the protein is in a collapsed state. Correlation between the chain and its shell dynamics increases as temperature is increased, and the chain extends. As the temperature is increased, fluctuations of the number of vapor molecules in the shell increase, allowing more available volume for the chain, and the protein undergoes a collapse transition at temperature for which it would have been in a coil state without the solvent. Further numerical simula-

tions with longer protein length were performed to study the critical exponent close to the new critical temperature induced by hydrophobic interaction. The results obtained for the universal exponents of statistical measures on the protein are presented in the next section.

4.4 Critical exponents with hydrophobic interaction

We have carried out numerical simulations for many protein lengths around the collapse region to study the scaling behavior near the collapse temperature. The parameters were set in the same way as in the temperature dependence analysis, namely $\epsilon = -100$ K, $\omega = -100$ K, $Q = 50$ and $\mu = 800$ K. Over 5×10^4 independent configurations were averaged or about 2×10^6 total conformations generated with protein lengths between $[4, 40]$. The critical temperature and the universal exponents are a priori unknown. The same procedure used in chapter 2 to identify the critical temperature and the universal exponents were performed. The results are presented in the next section.

4.4.1 The specific heat and fluctuations of the shell molecule number

The temperature dependence of the specific heat of the chain C_P/L , and fluctuations of the number of vapor molecule in the shell $\langle \Delta N_v^2 \rangle$ are shown in figure 4.9. As was the case for the collapse transition in vacuum, the magnitude of the specific heat increases with chain length. However, the temperature at which it reaches a maximum decreases by a small amount as compared to SARW in vacuum. The temperature for which the specific heat reaches a maximum extrapolates to T_c , the temperature of the collapse transition with hydrophobic interactions in the limit of a long chain. A linear fit applied on the temperature at which the specific heat reaches a maximum versus the square root of the chain is shown in figure 4.10. It extrapolates to a new critical temperature $T_c = (191 \pm 1)$ K, above the tricritical

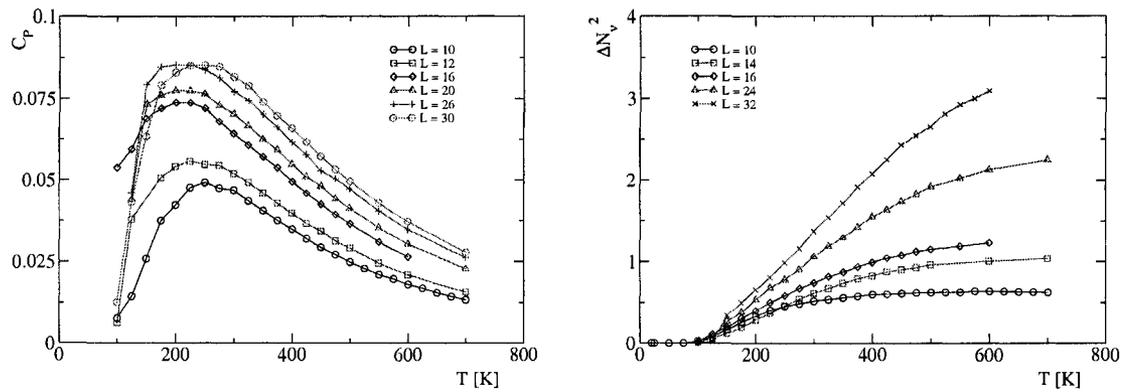


Figure 4.9: Specific heat of the interaction energy of the chain (C_P/L) (Left) and fluctuation of the number of vapor molecule in the shell ($\langle \Delta N_\nu^2 \rangle$) (Right) as a function of temperature T [K]. The magnitude of the specific heat and the fluctuation of the number of vapor molecule in the shell increase as the chain length is increased.

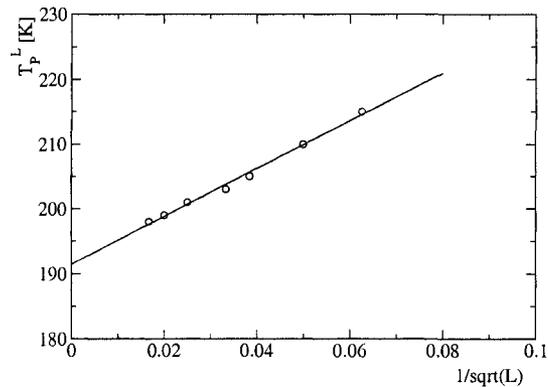


Figure 4.10: Temperature at which the specific heat of the protein reach a maximum T_C^* [K] versus the inversed squared root of the chain length $1/\sqrt{L}$. The solid line extrapolates to $T_c^* = (191 \pm 1)$ K, above the tricritical temperature in vacuum

temperature in vacuum.

Another way to identify the critical temperature is to look at the radius of gyration $\langle R_g^2 \rangle$ divided by $L^{2\nu^*}$, where ν^* is the ν exponent with hydrophobic interaction, versus the temperature. Intersection of the radius of gyration obtained from different length should identify the critical exponent and the collapse temperature. We have performed this analysis

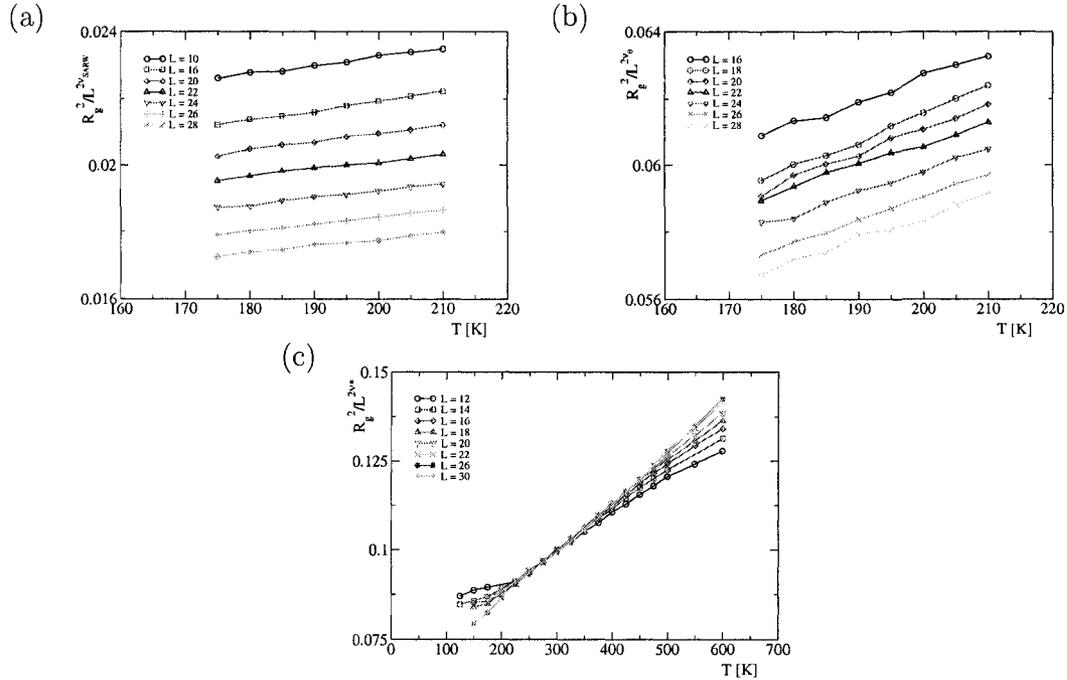


Figure 4.11: Mean squared radius of gyration $\langle R_g^2 \rangle$ divided by $L^{2\nu^*}$ as a function of temperature T [K]. The temperature for which the radius of gyration intersects for all lengths is the collapse temperature. Figure (a) shows the results obtained at $\nu^* = \nu_{SARW} = 3/4$ while figure (b) shows the results for $\nu^* = \nu_\theta = 4/7$. The radius of gyration doesn't intersect at those exponents, denoting that the universal exponents are shifted due to hydrophobic interaction. Our best estimates for the critical temperature is $T_c = 195$ K, where $\nu^* = 0.515 \pm 0.005$, and is shown in figure (c).

by considering various values of ν^* , the results obtained at various ν^* are shown in figure 4.11. There is no cross over between the radius of gyration at $\nu^* = \nu_\theta = 4/7$ and at $\nu^* = \nu_{SARW} = 3/4$ for all lengths. Our best estimate for the exponent is $\nu^* = 0.52 \pm 0.02$, leading to $T_c = (195 \pm 5)$ K, a critical temperature that agrees with the specific heat analysis.

4.4.2 The exponent ν^* with hydrophobic interaction

The exponent ν^* can be obtained by looking at the logarithm of the mean squared end to end distance and the mean squared radius of gyration. In the absence of solvent, the exponent

ν reaches its critical value of $\nu_\theta = 4/7 = 0.5714\dots$ at the tricritical temperature. We have calculated the exponent close to the temperature deduced from the specific heat and the mean squared of gyration analysis. The results obtained for the mean squared end to end distance and the mean squared radius of gyration are shown in table 4.2. The exponent $\nu^*_{\langle R_e^2 \rangle}$ and $\nu^*_{\langle R_g^2 \rangle}$ should be the same at the critical temperature. They differ by 1% at $T = 190$ K, a temperature that agrees with the previous analysis. At this temperature, the average of the two exponents is $\nu^* = 0.52 \pm 0.01$. This exponent is lower than the exponent ν_θ , indicating that the universality class of the collapse transition with hydrophobic interaction is different than the collapse transition in vacuum. The exponent ν^* was calculated the same way than in chapter 3, namely by considering a power law fit of equation 3.16.

4.4.3 The exponent γ^* with hydrophobic interaction

The exponent γ^* is obtained by evaluating the partition function of the protein (equation 3.11) with hydrophobic interaction. Close to the collapse temperature without hydrophobic interaction, the exponent γ approaches its critical value of $\gamma_\theta = 8/7 = 1.142\dots$. The exponent

T [K]	$\nu^*_{\langle R_e^2 \rangle}$	$\nu^*_{\langle R_g^2 \rangle}$	% diff.
175	0.5404	0.5275	2.39
180	0.5426	0.5279	2.72
185	0.5392	0.5281	2.07
190	0.5208	0.5264	1.06
195	0.5176	0.5259	1.61
200	0.4995	0.5248	5.07
205	0.4959	0.5241	5.69
210	0.4882	0.5234	7.21

Table 4.2: Comparison of the exponent ν^* obtained by the mean squared end to end distance $\langle R_e^2 \rangle$ and the mean squared radius of gyration $\langle R_g^2 \rangle$ for many temperatures, obtained from numerical simulations of chain with the solvent for length between $L = [8, 40]$, close to the critical temperature determined by the specific heat and radius of gyration analysis. The last column shows the percentage of difference between the exponent ν^* calculated from the two distance functions. The exponent ν^* differs by about 1% at $T = 190$ K.

T [K]	γ^*
175	1.2498
180	1.2307
185	1.2313
190	1.2253
195	1.2216
200	1.2121
205	1.2058
210	1.2011

Table 4.3: Exponent γ^* for various temperature got from a power fit of the partition function of the protein as a function of the chain length.

γ^* was obtained by a power law fit of the free energy difference as a function of chain length, as in chapter 3. Our best estimates for the exponent γ^* are shown in table 4.3. From this analysis, the exponent γ^* is also changed in the presence of hydrophobic interaction. We calculated $\gamma^* = 1.22 \pm 0.02$ close to the critical point.

4.4.4 Crossover exponent ϕ^* with explicit solvent

The cross over exponent ϕ^* is obtained by a power law fit of the derivative of the mean squared end to end distance and the mean squared radius of gyration (equation 3.17). Close to the collapse temperature and in the absence of solvent, the exponent ϕ reaches a value of $\phi_\theta = 3/7 = 0.4285\dots$ Table 4.4 summarizes our best estimates for the exponent ϕ^* calculated from the derivative of the end to end distance and the radius of gyration. Close to the critical point, the two exponents $\phi^*_{\langle R_e \rangle}$ and $\phi^*_{\langle R_g \rangle}$ should be the same, they differ by less than 1%, at $T = 190$ K. The cross over exponent is also changed with hydrophobic interaction, our best estimate is $\phi^* = 0.513 \pm 0.002$, got at $T = 190$ K.

T [K]	$\phi^*_{\langle R'_e \rangle}$	$\phi^*_{\langle R'_g \rangle}$	% diff.
175	0.6507	0.6152	5.77
185	0.5362	0.4741	13.11
190	0.5114	0.5155	0.78
195	0.4420	0.5086	13.09
200	0.4647	0.5605	17.09
205	0.3757	0.4619	18.67
210	0.3414	0.4576	25.40

Table 4.4: Comparison of the cross over exponent ϕ^* obtained by the derivative of the mean squared end to end distance $\langle R'_e \rangle$ and the mean squared radius of gyration $\langle R'_g \rangle$ for many temperatures, obtained from numerical simulations of chain with chain with the solvent length $L = [8, 40]$.

4.5 Interpretation of the results

A scaling analysis of the critical exponents as well as temperature dependence of the thermodynamic quantities of the chain is presented in this section. It is shown that the collapse temperature with explicit solvent is shifted in the presence of explicit hydrophobic interaction. The critical exponents around the collapse temperature are also modified, indicating that the collapse transition belongs to a universality class different than SARWs in vacuum. At very low temperatures, the interaction energy of the chain dominates, and the dynamics of the protein does not couple with the shell of vapor molecules. The chain is in a collapsed state. As temperature is increased but still kept lower than the critical temperature, the chain is still in a collapsed state while more conformational states would be allowed in vacuum. Energetic costs involved in the addition of a vapor molecule in the shell dominate, reducing the accessible volume for chain conformation, keeping the protein in a collapsed state. Further increases in temperature relax the energetic constraint and more vapor molecules are allowed in the shell. The accessible space increases, allowing more conformational states of the protein, and the chain extends. At very large temperatures, the shell and the bulk solvent can exchange vapor molecules at random, and the chain is in a coil state. The collapse transition from a globular state at low temperature to a coil state at

high temperature is induced by a change in vapor density around the chain. This analysis confirms that solvent effects play a crucial role in the folding process and are essential for a further solution of the protein folding problem.

Chapter 5

Conclusions

Hydration forces in the protein folding problem are not well understood quantitatively, yet they are believed to be essential for a realistic picture of the folding process and thermodynamics. They are responsible for the collapse of the hydrophobic core of a typical protein, and contribute to maintaining the overall structure of native proteins. We have studied in this thesis the collapse transition of a simplified protein model in two dimensions with explicit hydrophobic interactions by immersing the protein model in a two dimensional coarse grained model of solvent. Temperature of the collapse transition of the protein is modified with hydrophobic interaction. Universal scaling of the thermodynamical measures of the protein is also changed, indicating that the collapse transition belongs to a universality class different than the collapse transition of chain in vacuum.

The solvent model consists of a square lattice of solvent molecules with periodic boundary conditions in which each sites are occupied by only one solvent molecule. This model of hydrophobicity was introduced by Widom and coworkers [6, 7]. Each solvent molecule has an orientation variable which can have Q states. Only one orientation state is energetically favored, $Q = 1$ say, called the special state. Solvent molecules interact only with their nearest neighbors with energy ω ($\omega < 0$) if they are both in the special state of

orientation, and with energy $\nu = 0$ otherwise. Neighboring solvent molecules in the special orientation state model the cage like structure of water which can accommodate a solute molecule without breaking any of the hydrogen bonds. A mapping between this model and the Ising model was established and temperature dependence of statistical quantities of the solvent were measured. The solvent model leads to an attractive potential of mean force between solute molecules in which its magnitude increases as the temperature is increased while its range decreases.

The protein is modeled as a two dimensional SARW with weakly attractive interaction ϵ , and was implemented through the pivot algorithm. The protein model undergoes a phase transition from a globular phase at low temperature to a coil phase at high temperature in vacuum. Around the collapse temperature, statistical measures follow universal scaling laws. The critical exponents and the tricritical temperature have been measured and are in excellent agreement with the exact values in two dimensions.

Collapse transition of the protein with hydrophobic interaction was then studied, by embedding the chain model into the solvent model in the liquid phase. Since a solute molecule can only be accommodated between neighboring solvent molecules in their special state, the immersion of the protein was done together with a shell of solvent molecules in their special state surrounding the protein. Those solvent molecules only have one orientation state and model vapor molecules. The shell surrounding the protein can exchange vapor molecules with a solvent reservoir at chemical potential μ . At very low temperature, energy of interaction of the chain dominates, the protein is uncorrelated with the solvent and the chain is in a collapsed state. Increasing temperature would allow more conformational states for the chain in isolation but hydrophobic interaction restricts the accessible volume for the protein. The dynamics of the shell and the protein are now correlated and the protein is

kept in a collapsed state. At some temperature, more vapor molecules are allowed in shell, extending the available space for the chain, and it extends. At very large temperature, the bulk solvent and the shell can exchange vapor molecule at random, and the chain is in a coil state. Fluctuations of the density of the shell vapor molecule drive the collapse transition, and a shifts in the critical temperature is observed as compared to the collapse transition in vacuum. Universal exponents have been calculated around the new critical point. The value of the three exponents considered are not conserved near the critical point, and the universality class of the collapse transition of the protein with hydrophobic interaction is different than the one in vacuum.

This contribution confirms that solvent mediated interactions are essential to have a realistic picture of the folding process. While a three dimensional version of this simple model will not be able to predict structural motifs (e.g. alpha-helices or beta-sheets) observed in proteins, it emphasis on the important role of hydration forces in the folding process, and the needs of a complete understanding of their effects for further advances in the protein folding problem. The results motivate a further study of a more realistic picture of the folding process by including electrostatics interactions between molecules, a three dimensional analysis, or further improvement of the protein in the model.



Appendix A

Exact potential of mean force in one dimension

We present in this appendix the exact calculation of the potential of mean force $W(r)$ in one dimension. Consider a one dimensional chain of N solvent molecules, with periodic boundary condition. Each molecule can have Q states of orientation and only one orientation state is energetically favored, $Q = 1$ say. Molecules interact with their nearest neighbors with energy ω ($\omega < 0$) only if they are both in the special state of orientation and with energy $\nu = 0$ otherwise. The Hamiltonian of the system is

$$H = \omega \sum_{\langle i,j \rangle} \delta_{S_i,1} \delta_{S_j,1} \quad , \quad (\text{A.1})$$

where the sum is over all the nearest neighbors of the i^{th} molecule and $\delta_{S_k,1} = 1$ if the orientation state of the k^{th} molecule is $S_k = 1$ and 0 otherwise. The partition function Z is

given by,

$$\begin{aligned}
 Z &= \sum_{S_1=1}^Q \dots \sum_{S_N=1}^Q e^{-\beta\omega \sum_{\langle i,j \rangle} \delta_{S_i,1} \delta_{S_j,1}} \\
 &= \sum_{S_1=1}^Q \dots \sum_{S_N=1}^Q \prod_{i=1}^N e^{-\beta\omega \delta_{S_i,1} \delta_{S_{i+1},1}} \quad .
 \end{aligned} \tag{A.2}$$

Define the transfer matrix V as a $Q \times Q$ square matrix by its matrix elements,

$$\langle S_i | V | S_{i+1} \rangle = e^{-\beta\omega \delta_{S_i,1} \delta_{S_{i+1},1}} \quad . \tag{A.3}$$

Explicitly, the transfer matrix V is given by,

$$V = \begin{pmatrix} b & a & a & \dots & a \\ a & a & a & \dots & a \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a & a & a & \dots & a \end{pmatrix}_{[Q \times Q]} \quad , \tag{A.4}$$

where $b = e^{-\beta\omega} > 0$ and $a = 1$. In those terms, the partition function becomes,

$$Z = \sum_{S_1=1}^Q \dots \sum_{S_N=1}^Q \langle S_1 | V | S_2 \rangle \langle S_2 | V | S_3 \rangle \dots \langle S_N | V | S_1 \rangle \quad , \tag{A.5}$$

where periodic boundary condition was used. Identify the identity operator,

$$1 = \sum_{S_k=1}^Q |S_k \rangle \langle S_k | \quad , \tag{A.6}$$

and insert it in the partition function equation. Then

$$Z = \sum_{S_1=1}^Q \langle S_1 | V^N | S_1 \rangle = \text{Tr} [V^N] \quad . \quad (\text{A.7})$$

As the trace is independent of the matrix representation, we can write V is in diagonal form. In this representation, the diagonal of V will be given by its eigenvalue λ_i with $i = \{1, 2, \dots, Q\}$. The trace of the matrix V^N will be the sum of the eigenvalue λ_i^N . The largest eigenvalue λ_{max} will dominate the sum and we will have as final result for the partition function,

$$Z = \lambda_{max}^N \quad . \quad (\text{A.8})$$

Define P_{11} as the probability for a pair of neighboring molecule be both in their special state irrespectively of the orientation state of the $(N-2)$ other solvent molecules. Without loss of generality, set the molecules located at $(1, 2)$ to be this pair. The result holds if one replace the location of this pair of solvent molecule by a pair located at $(k, k+1)$. The function P_{11} is thus defined as

$$P_{11} = \frac{1}{Z} \sum_{S_3=1}^Q \cdots \sum_{S_N=1}^Q \delta_{S_1,1} \delta_{S_2,1} e^{-\beta \omega \sum_{i=1}^N \delta_{S_i,1} \delta_{S_{i+1},1}} \quad . \quad (\text{A.9})$$

In terms of the matrix element of the transfer matrix V , P_{11} can be rewrite as

$$P_{11} = \frac{1}{\text{Tr} [V^N]} \sum_{S_3=1}^Q \cdots \sum_{S_N=1}^Q \langle S_1 | V | S_1 \rangle \langle S_1 | V | S_3 \rangle \cdots \langle S_N | V | S_1 \rangle \quad . \quad (\text{A.10})$$

Introducing the identity operator leads to the final results for P_{11} ,

$$P_{11} = \frac{\langle S_1 | V | S_1 \rangle \langle S_1 | V^{N-1} | S_1 \rangle}{\text{Tr} [V^N]} = \frac{V_{11}(V^{N-1})_{11}}{\text{Tr} [V^N]} \quad , \quad (\text{A.11})$$

where V_{11} and $(V^{N-1})_{11}$ are the $(1, 1)$ element of the matrix V and V^{N-1} respectively. The probability that two pair of neighboring solvent molecule separated by a distance r be all in their special state, $P(r)$, is calculated in the same way. Without loss of generality, consider the pair of solvent molecules located at $(1, 2)$ and the pair at $(1+r, 2+r)$ to be such of two pairs. The same result is obtained if the first pair of solvent molecules are located at $(n, n+1)$ and the other pair at $(m, m+1)$ with $m-n=r$. The function $P(r)$ is thus defined as

$$P(r) = \frac{1}{Z} \sum_{S_3=1}^Q \cdots \sum_{S_r=1}^Q \sum_{S_{r+3}=1}^Q \cdots \sum_{S_N=1}^Q \delta_{S_1,1} \delta_{S_2,1} \delta_{S_{r+1},1} \delta_{S_{r+2},1} e^{-\beta\omega \sum_{i=1}^N \delta_{S_i,1} \delta_{S_{i+1},1}} \quad . \quad (\text{A.12})$$

In terms of the matrix elements of the transfer matrix V , $P(r)$ can be rewrite as

$$P(r) = \frac{1}{\text{Tr}[V^N]} \sum_{S_3=1}^Q \cdots \sum_{S_r=1}^Q \sum_{S_{r+3}=1}^Q \cdots \sum_{S_N=1}^Q \langle S_1|V|S_1 \rangle \langle S_1|V|S_3 \rangle \cdots \langle S_r|V|S_1 \rangle \langle S_1|V|S_1 \rangle \times \\ \langle S_1|V|S_{r+3} \rangle \cdots \langle S_N|V|S_1 \rangle \quad . \quad (\text{A.13})$$

By inserting the identity operator, we obtain

$$P(r) = \frac{\langle S_1|V|S_1 \rangle^2 \langle S_1|V^r|S_1 \rangle \langle S_1|V^{N-r}|S_1 \rangle}{\text{Tr}[V^N]} = \frac{V_{11}^2 (V^r)_{11} (V^{N-r})_{11}}{\text{Tr}[V^N]} \quad . \quad (\text{A.14})$$

The transfer matrix V as defined by equation A.4 has two positive eigenvalues $\lambda^{(1)} > \lambda^{(2)} > 0$ and the remaining $(Q-2)$ eigenvalues are all 0. Let ψ_i^ν be the i^{th} component of the eigenvector, with $i = \{1, \dots, Q\}$, that belongs to the eigenvalue $\lambda^{(\nu)}$ with $\nu = \{1, \dots, Q\}$. From the structure of V , it follows that since $\lambda^{(1)}$ and $\lambda^{(2)}$ are non zero, then all the ψ_i^1 except ψ_1^1 have

a common value, as do the ψ_i^2 except ψ_1^2 such that

$$\begin{aligned}\psi_2^{(1)} &= \psi_3^{(1)} = \dots = \psi_q^{(1)} \\ \psi_2^{(2)} &= \psi_3^{(2)} = \dots = \psi_q^{(2)}\end{aligned}\quad . \quad (\text{A.15})$$

Then the eigenvalues $\lambda^{(1)}$ and $\lambda^{(2)}$ of V and their associated eigenvectors satisfy

$$\begin{aligned}b\psi_1^{(\nu)} + (Q-1)a\psi_2^{(\nu)} &= \lambda^{(\nu)}\psi_1^{(\nu)} \\ a\psi_1^{(\nu)} + (Q-1)a\psi_2^{(\nu)} &= \lambda^{(\nu)}\psi_2^{(\nu)}\end{aligned}\quad , \quad (\text{A.16})$$

with $\nu = \{1, 2\}$. From this previous equation, the two eigenvalues $\lambda^{(1)}$ and $\lambda^{(2)}$ are the larger and smaller of the roots of

$$\begin{vmatrix} b - \lambda & (Q-1)a \\ a & (Q-1)a - \lambda \end{vmatrix} = 0 \quad . \quad (\text{A.17})$$

We take the eigenvectors to be normalized as $\sum_{i=1}^Q \psi_i^{(\nu)2} = 1$ so that,

$$\psi_1^{(\nu)2} + (Q-1)\psi_2^{(\nu)2} = 1 \quad \nu = 1, 2 \quad . \quad (\text{A.18})$$

The element of any power V^p of V may be express in terms of the eigenvalues $\lambda^{(\nu)}$ and its normalized eigenvectors $\psi^{(\nu)}$ by

$$(V^p)_{ij} = \sum_{\nu=1}^Q \lambda^{(\nu)p} \psi_i^{(\nu)} \psi_j^{(\nu)} \quad . \quad (\text{A.19})$$

But since only $\lambda^{(1)}$ and $\lambda^{(2)}$ are non zero, we have

$$P_{11} = \frac{b}{\lambda^{(1)}} \psi_1^{(1)2} \quad , \quad (\text{A.20})$$

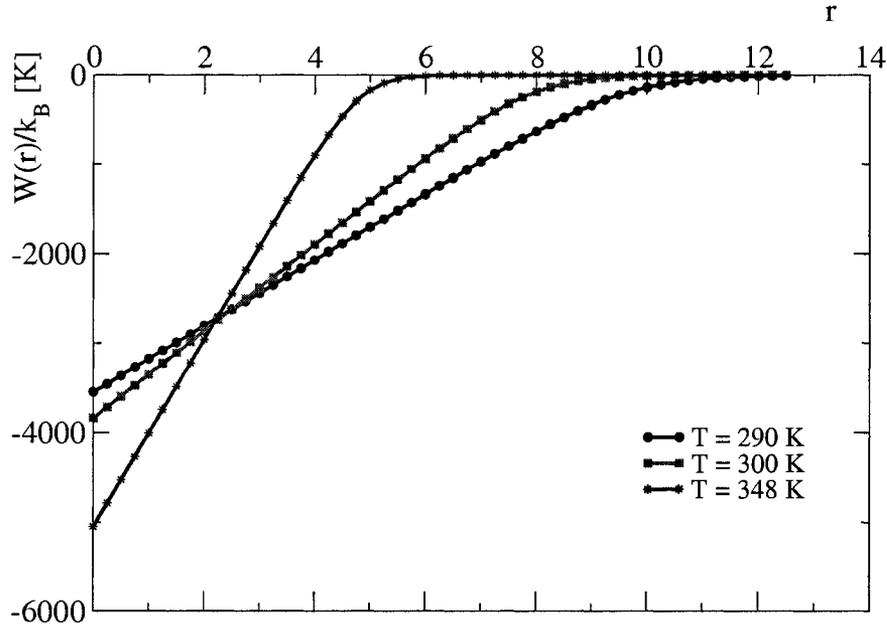


Figure A.1: Potential of mean force $W(r)/k_B$ acting on solute molecule as a function of the distance r between two pair of neighboring solvent molecules both in their special state. The temperature curves were obtained by solving equation A.25 with parameter $Q = 110\,000$ and $\omega = 3000$ K. Increasing temperature increases the magnitude of the potential while its range decreases.

and

$$P(r \geq 1) = P_{11}^2 \left[1 + \left(\frac{\psi_1^{(2)}}{\psi_1^{(1)}} \right)^2 \left(\frac{\lambda^{(2)}}{\lambda^{(1)}} \right)^{r-1} \right], \quad (\text{A.21})$$

in the thermodynamic limit $N \rightarrow \infty$. We thus obtain the solvent mediated potential of mean force,

$$W(r) = -k_B T \ln \left[1 + \left(\frac{\psi_1^{(2)}}{\psi_1^{(1)}} \right)^2 \left(\frac{\lambda^{(2)}}{\lambda^{(1)}} \right)^{r-1} \right]. \quad (\text{A.22})$$

Explicit calculation of $W(r)$ requires evaluation of the ratio $(\psi_1^{(2)}/\psi_1^{(1)})^2$ and $(\lambda^{(2)}/\lambda^{(1)})$.

Define the quantities c , x , S and W in terms of a , b and $(Q - 1)$ by

$$\begin{aligned}
 c &= \frac{b}{a} = e^{\beta(\nu-\omega)} > 1 \\
 x &= \frac{(Q-1)}{c} \\
 S &= \sqrt{1 - \frac{4x}{(1+x)^2} \left(1 - \frac{1}{c}\right)} \\
 K &= \frac{\text{sign}(x-1)}{\sqrt{1 + \frac{4x}{c(x-1)^2}}}
 \end{aligned} \tag{A.23}$$

where $\text{sign}(x-1) = +1$ if $x > 1$ and $\text{sign}(x-1) = -1$ if $x < 1$. The required ratio can thus be expressed as

$$\begin{aligned}
 \left(\frac{\psi_1^{(2)}}{\psi_1^{(1)}}\right)^2 &= \frac{1+K}{1-K} \\
 \frac{\lambda^{(2)}}{\lambda^{(1)}} &= \frac{1-S}{1+S}
 \end{aligned} \tag{A.24}$$

In those terms, the potential of mean force $W(r)$ is given by

$$W(r) = -k_B T \ln \left[1 + \left(\frac{1+K}{1-K}\right) \left(\frac{1-S}{1+S}\right)^{r-1} \right], \tag{A.25}$$

taking as parameters the number of accessible state Q and the energy of interaction $(\nu - \omega)$. Figure A.1 shows the results of equation A.25 for three different temperatures with the parameters set to $Q = 100\,000$ and $\omega = 3000$ K. It is observed that the magnitude of the potential of mean force increases by increasing the temperature while its range decreases.



Appendix B

Mapping between the Ising and the solvent model

A mapping between the solvent model and the ferromagnetic Ising model with an external magnetic field in one dimension is established in this appendix. Following Widom [46], the partition function of the solvent model Z_S is

$$Z_S = \sum_C g_C e^{-\beta E(C)} \quad , \quad (\text{B.1})$$

where the sum is over all molecules in a given configuration C and then over all possible configurations for the lattice and where g_C is the degeneracy of the internal state of the molecule in C . We can rewrite this sum as the sum over all molecules in C that are in their special state of orientation C_1 and over all the molecules that are in another state, $C'(C_1)$. The partition function is then

$$Z_S = \sum_{C_1} \sum_{C'(C_1)} g_{C_1} e^{-\beta E(C_1)} g_{C'(C_1)} e^{-\beta E(C'(C_1))} \quad , \quad (\text{B.2})$$
$$E(C_1) = \omega \sum_{\langle i,j \rangle} \delta_{S_i,1} \delta_{S_j,1} \quad .$$

Only the special state is energetically favored while the other $(Q - 1)$ states of orientation share the same energy, thus $g_{C_1} = 1$ and $g_{C'(C_1)} = (Q - 1)$. Without loss of generality, we set $\nu = 0$ for simplicity. In those terms, the sum over $C'(C_1)$ can be rewrite

$$\begin{aligned} \sum_{C'(C_1)} g_{C'(C_1)} e^{-\beta E(C'(C_1))} &= (Q - 1)^{N_Q} \\ &= e^{\ln[Q-1] \sum_i (1 - \delta_{S_i,1})} \end{aligned} \quad (\text{B.3})$$

where N_Q is the number of molecules that are not in their special state. The energy of molecules that are in their special state can be decomposed as,

$$E(C_1) = \omega \sum_{\langle i,j \rangle} \delta_{S_i,1} \delta_{S_j,1} + \frac{z\omega}{2} \sum_i \delta_{S_i,1} \quad . \quad (\text{B.4})$$

Combining those results,

$$E_C = \omega \sum_{\langle i,j \rangle} \delta_{S_i,1} \delta_{S_j,1} + \left(\frac{z\omega}{2} + k_B T \ln(Q - 1) \right) \sum_i \delta_{S_i,1} + \text{constant} \quad , \quad (\text{B.5})$$

where the constant doesn't affect the thermodynamics of the system. In those terms, the partition function of the solvent model is given by

$$Z_S = \sum_{C_1} \exp \left(-\beta \omega \sum_{\langle i,j \rangle} \delta_{S_i,1} \delta_{S_j,1} - \beta \left(\frac{z\omega}{2} + k_B T \ln[Q - 1] \right) \sum_i \delta_{S_i,1} \right) \quad . \quad (\text{B.6})$$

This is very similar to the partition function of the Ising model,

$$Z_I = \sum_S \exp \left(\beta J \sum_{\langle i,j \rangle} S_i S_j + \beta H \sum_i S_i \right) \quad , \quad (\text{B.7})$$

except that the sum is over all spins up and down. To complete the mapping, we must express Z_I in terms of the sum over the up spins and the down spins,

$$\begin{aligned}
 Z_I &= \sum_{S_I} \sum_{S_I} \exp \left(J\beta \sum_{\langle i,j \rangle} S_i (S_j^\uparrow + S_j^\downarrow)^2 + H\beta \sum_i (S_i^\uparrow + S_i^\downarrow) \right) \\
 &= \sum_{S_I} \exp \left(J\beta \sum_{\langle i,j \rangle} S_i (S_j^\uparrow - (1 - S_j^\uparrow))^2 + H\beta \sum_i (S_i^\uparrow - (1 - S_i^\uparrow)) \right) \quad (\text{B.8}) \\
 &= \sum_{S_I} \exp \left(4J\beta \sum_{\langle i,j \rangle} S_i S_j + 2H\beta \sum_i S_i^\uparrow \right) .
 \end{aligned}$$

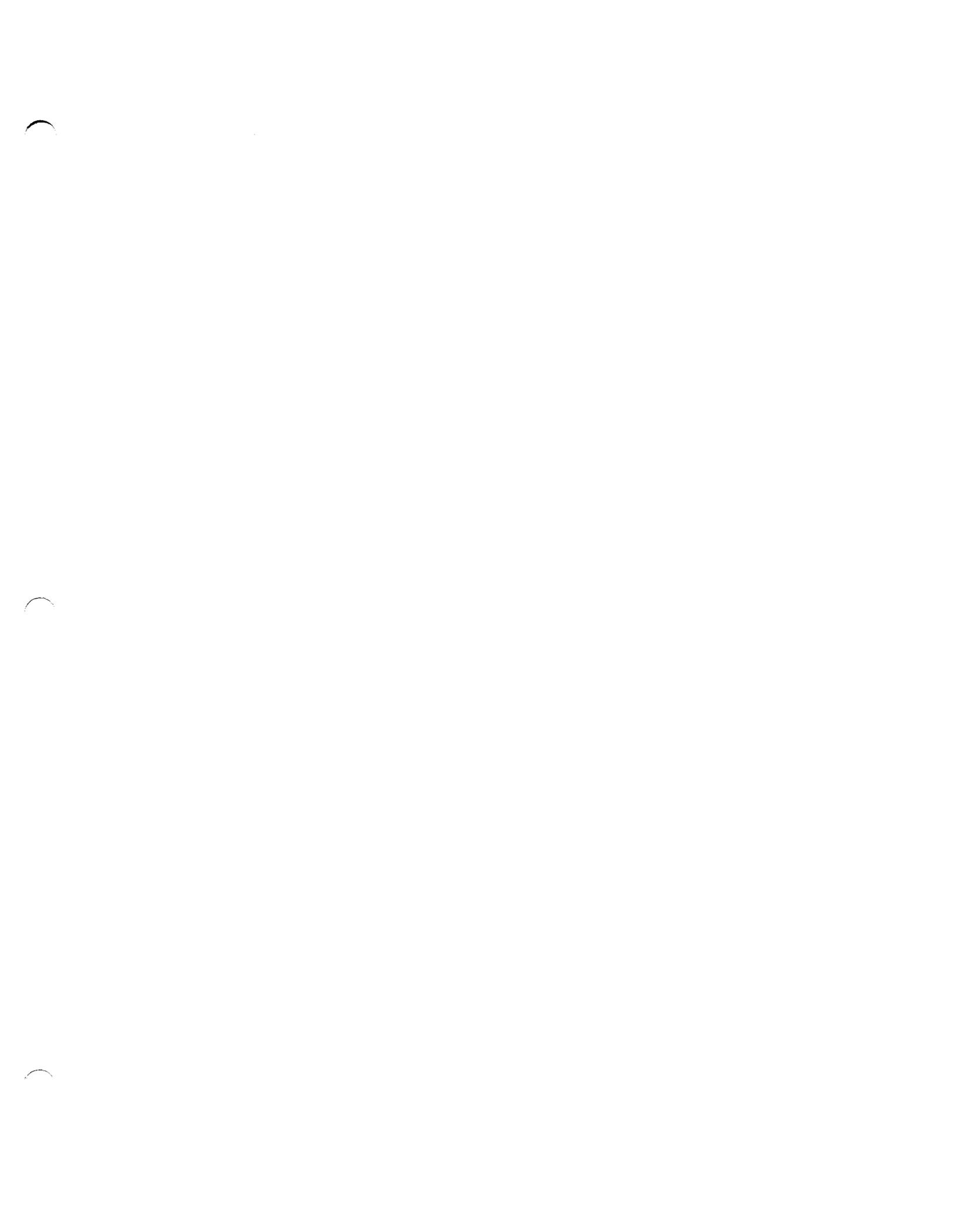
By comparing Z_I and Z , we have

$$\begin{aligned}
 2J &= -\frac{\omega}{2} \\
 2H &= -\frac{z\omega}{2} - k_B T \ln[Q - 1] \quad , \quad (\text{B.9})
 \end{aligned}$$

which establishes the mapping between the solvent model and the ferromagnetic Ising model with an external magnetic field. The same have been obtained by Widom [46]. Phase transition of the ferromagnetic Ising model with an external magnetic field happen as $H = 0$. This leads to a critical temperature for the solvent model,

$$k_B T_c = -\frac{z\omega}{2 \ln[Q - 1]} \quad , \quad (\text{B.10})$$

which agrees with our numerical simulations.



Appendix C

Q-Potts state model in the Bragg-William approximation

We derive the critical temperature of the Q Potts state model with $Q \geq 3$ using the Bragg-William approximation. Phase transition for the Q-Potts state model is first order and involves a latent heat of transition[47]. Consider a linear chain of N spins, each of which can have Q states. Spins interact with their nearest-neighbors only if they are in the same state with energy J ($J > 0$). The Hamiltonian of the system is given by

$$H = -J \sum_{\langle i,j \rangle} \delta_{S_i, S_j} \quad , \quad (\text{C.1})$$

where $\delta_{S_\mu, S_\nu} = 1$ if $S_\mu = S_\nu$ and 0 otherwise. Let n_ξ be the number density of the ξ state with $\xi = \{1, \dots, Q\}$, defined as $n_\xi = N_\xi/N$, where N_ξ is the number of spins in the ξ state and N is the total number of spin. The free energy in the Bragg-William approximation [47] is given by

$$F = \frac{-zJN}{2} \sum_{\xi=1}^Q N_\xi^2 + Nk_B T \sum_{\xi=1}^Q N_\xi \ln[N_\xi] \quad , \quad (\text{C.2})$$

where z is the lattice coordination number. At large temperature, equipartition of states lead to $n_1 = \dots = n_Q = 1/Q$. We required in general that $\sum_{\xi=1}^Q N_{\xi} = 1$. A possible parametrization for the number density is

$$\begin{aligned} N_1 &= \frac{1}{Q} (1 + (Q-1)m) \\ N_2 &= \frac{1}{Q} (1 - m) \\ N_3 &= \frac{1}{Q} (1 - m) \\ &\vdots \\ N_Q &= \frac{1}{Q} (1 - m). \end{aligned} \tag{C.3}$$

In those terms, we have

$$\sum_{\xi=1}^Q N_{\xi}^2 = \frac{1}{Q} (1 + (Q-1)m^2) \tag{C.4}$$

and

$$\sum_{\xi=1}^Q N_{\xi} \ln[N_{\xi}] = \frac{(1 + (Q-1)m)}{Q} \ln \left[\frac{1 + (Q-1)m}{Q} \right] + \frac{(Q-1)(1-m)}{Q} \ln \left[\frac{(1-m)}{Q} \right]. \tag{C.5}$$

We can rewrite the last expression in a more convenient form namely,

$$\sum_{\xi=1}^Q N_{\xi} \ln[N_{\xi}] = \frac{(1 + (Q-1)m)}{Q} \ln [1 + (Q-1)m] + \frac{(Q-1)(1-m)}{Q} \ln [(1-m)] - \ln [Q] \tag{C.6}$$

Equation C.6 can be expanded in power series in small m as

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots, \tag{C.7}$$

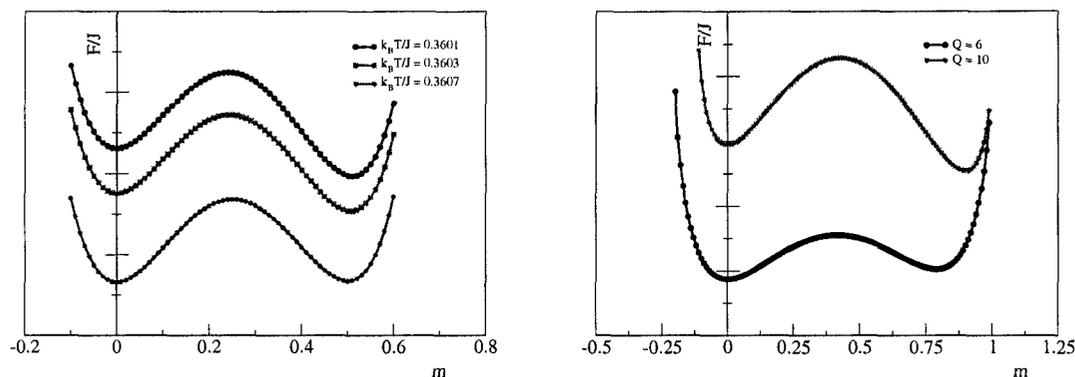


Figure C.1: Free energy F/J as a function of the order of parameter m of the Q-Potts state model in 1 dimension. The left figure show three curves corresponding to $k_B T/J = 0.3601$, $k_B T/J = 0.3603$ and $k_B T/J = 0.3607$ and where Q was fixed to $Q = 3$. The value for which $m = 1/2$ corresponds to the minimum in the free energy. The same calculation is shown in the right figure for $Q = 6$ and $Q = 10$. The minimum of the free energy corresponds to $m = 0.8$ and $m = 0.888$ respectively. The three minima suggest a dependence of the form $m = (Q - 2)/(Q - 1)$ close to the critical point.

leading to the free energy

$$\begin{aligned} \frac{F}{N} = & -\frac{zJ}{2Q} - k_B T \ln[Q] + m^2 \left(\frac{(Q-1)}{2Q} \right) (Qk_B T - zJ) - m^3 \left(\frac{(Q-1)(Q-2)k_B T}{6} \right) \\ & + m^4 \left(\frac{(Q-1)(Q^2 - 3Q + 3)k_B T}{12} \right) + \dots \end{aligned} \quad (\text{C.8})$$

The occurrence of the power 3 of the order of parameter in the free energy expansion indicates that the transition is first order. The critical temperature of transition can be found numerically. Figure C.1 shows the free energy as a function of the order of parameter m , for many numbers of accessible state Q . It is observed that the free energy has a minimum around $m = (Q - 2)/(Q - 1)$. It turns out that this value is exact. The critical temperature can be obtained by looking at the first derivative of the free energy with respect to the order

of parameter,

$$\frac{\partial(F/N)}{\partial m} = -\frac{zJm(Q-1)}{Q} - \frac{k_B T(Q-1)\ln[1-m]}{Q} + \frac{k_B T(Q-1)\ln[1+m(Q-1)]}{Q} = 0 \quad (C.9)$$

and by evaluating the last expression at $m = (Q-2)/(Q-1)$. This leads to,

$$k_B T_c = \frac{zJ(Q-2)}{2(Q-1)\ln[Q-1]} \quad (C.10)$$

First order transition involves at latent heat of transition. As $m = 0$, the entropy of the system is $k_B \ln[Q]$. At the transition temperature, the entropy of the system will be $k_B \ln[Q] - k_B \ln[2](Q-2)/Q$. Thus the latent heat is

$$L = T\Delta S = \frac{J(Q-2)^2}{2Q(Q-1)} \quad (C.11)$$

From our parametrization, only one state was energetically favored, as in the solvent model. To obtain the estimate of the critical temperature for the solvent model, we simply set $J = -\omega$ and get,

$$k_B T_c = -\frac{z\omega(Q-2)}{2(Q-1)\ln[Q-1]} \quad (C.12)$$

where z is the lattice constant.

Bibliography

- [1] H. Lodish, A. Berk and al. *Molecular Cell Biology*, 5th ed. W.H. Freeman and Co., New York, 2004, 973 p.
- [2] S.D. Dubin and G. Feher. *Annu. Rev. Phys. Chem.* 47, 171 (1996).
- [3] Y. Levy and J. Onuchic. *Ann. Rev. Biophys. Struct.* 35, 389 (2006).
- [4] R. M. Levy and E. Gallicchio. *Annu. Rev. Phys. Chem.* 49, 531 (1998).
- [5] L. R. Pratt. *Annu. Rev. Phys. Chem.* 53, 409 (2002).
- [6] A.B. Kolomeisky and B. Widom. *Faraday Discuss.* 112, 81 (1999).
- [7] G.T. Barkema and B. Widom. *J. Chem. Phys.* 113, 2349 (2000).
- [8] G.T. Barkema, G.M. Schutz, B. Widom et al. *Physica A* 291, 24 (2001).
- [9] M. Lal. *Molec. Phys.* 17, 57 (1969).
- [10] N. Madras and A. D. Soka . *J. Stat. Phys.* 50, 109 (1988).
- [11] F. H. Stillinger and A. Rahman. *J. Chem. Phys.* 4, Vol 60.4, 15 (1974).
- [12] H. C. Berendsen, J.R. Grigera and T.P. Straatsma. *J. Phys. Chem.* 4, 91, 6269 (1987).
- [13] W. L. Jorgensen, J. Chandrasekhar, et al. *J. Phys. Chem.* 79, 926 (1983).
- [14] H. W. Horn, J. D. Madura, et al. *J. Phys. Chem.* 120, 9665 (2004).
- [15] W. C. Still, A. Tempczyk et al. *J. Am. Chem. Soc.* 112, 6127 (1990).
- [16] J. Vila, R.L. William et al. *Proteins.* 10, 199 (1991).
- [17] R.H. Zhou, B.J. Berne. *Proc. Natl. Acad. Sci.* 99, 12777 (2002).
- [18] H. Nymeyer, A.E. Garcia. *Proc. Natl. Acad. Sci.* 100, 13934 (2003).

-
- [19] V. S. Pande, E. Lindahl. *Proc. Natl. Acad. Sci.* 101, 6456 (2004).
- [20] B. Widom, K. Koga and P. Bhimalapuram. *Phys. Chem. Chem. Phys.* 5, 3085 (2003).
- [21] B. Widom. *J. Chem. Phys.* 39, 2808 (1963).
- [22] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford Univ. Press, New York, 1987, 274 p.
- [23] A.B. Bortz, M.H. Kalos and J.L. Lebowitz *J. Comp. Phys.* 17, 10 (1975).
- [24] J. Rudnick and G. Gaspari. *Elements of the Random Walk*. Cambridge, London, 2004, 329 p.
- [25] A. Baumgartner. "Simulations of Macromolecules", in *The Monte Carlo Method in Condensed Matter Physics*, edited by K. Binder. Springer-Verlag, Berlin, 1998, 418 p.
- [26] P. J. Flory. *Statistical mechanics of chain molecules*. Oxford Univ. Press, New York, 1989, 459 p.
- [27] P. G. de Gennes. *J. Phys. Lett.* 36, L55 (1975).
- [28] P. G. de Gennes. *J. Phys. Lett.* 39, L299 (1978).
- [29] B. Duplantier and H. Saleur. *Phys. Rev. Lett.* 59, 539, (1987).
- [30] Lax M. and Brender C. *J. Chem. Phys.* 67, 1785 (1977)
- [31] F.T. Wall, F. Mandel. *J. Chem. Phys.* 63, 4592 (1975).
- [32] F. Mandel. *J. Chem. Phys.* 70, 3984 (1979).
- [33] H. Meirovitch. *Macromolecules*. 16, 1628 (1983).
- [34] H. Meirovitch. *J. Chem. Phys.* 89, 2514 (1988).
- [35] N. Madras and A. D. Sokal . *The Self-Avoiding Walk*. Birkhauser, Boston, 1993, 425 p.
- [36] P. Grassberger and R. Hegger. *J. Phys.* 5, 597 (1995).
- [37] K. Binder and D.W. Heermann. *Monte Carlo Simulation in Statistical Physics*. Springer, Berlin, 2002, 180 p.
- [38] K. Kremer, A. Baumgartner and K. Binder. *J. Phys. A*. 15, 2879 (1987) .
- [39] J. des Cloizeaux. *J. Phys.* 37, 431 (1976).
- [40] H. Meirovitch and H.A. Lim. *J. Chem. Phys.* 91, 4, 2544 (1989).
- [41] H. Meirovitch and I. Chang. *Phys. Rev. E*. 48, 5, 3656 (1993).

-
- [42] P. Debye. *J. Phys. Colloid Chem.* 51, 18 (1947).
- [43] B. Farnoux and al. *J. Physique*, 39, 77 (1978).
- [44] P. R. ten Wolde and D. Chandler. *Proc. Natl Acad. Sci. USA*, 99, 6539 (2002).
- [45] G. Gokoglu, H. A. Olgar, E. Aktur, and T. Celik. *International Journal of Modern Physics C*. 16, 1489 (2005).
- [46] G.T. Barkema and B. Widom. *Physica A*. 291, 24 (2001).
- [47] M. Plischke and B. Bergersen. *Equilibrium Statistical Physics*, 3rd edition. World Scientific, Singapore, 2005, 620 p.