

Mechanisms of protein-protein association

Atomistic molecular dynamics study of the association process

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät III
Chemie, Pharmazie, Bio- und Werkstoffwissenschaften
der Universität des Saarlandes

von
Mazen Ahmad

Saarbrücken

2012

Tag des Kolloquiums: 27.11.2012

Dekan: Prof. Dr. V. Helms

Berichterstatter: Prof. Dr. V. Helms
Prof. Dr. R. Zimmermann
Prof. Dr. M. Zacharias

Vorsitz: Prof. Dr. U. Müller

Akad. Mitarbeiter: Dr. S. Hansen

Abstract

The associations between proteins are very important in the signaling pathways inside the cell. Our knowledge about the detailed mechanism of the association process is very limited due to the lack of suitable experimental methods that would allow studying the association at high time resolution. In this PhD thesis I present results from molecular dynamics simulations to study the mechanism for two types of protein-protein associations. (1) Protein complexes with a hydrophobic interface. (2) Protein complexes with a charged interface. The extensive MD simulations that were started from the unbound proteins reproduced the experimentally known structures of the complexes. This real time dynamics study gave me the possibility to study the mechanism of the binding process at picosecond time resolution to distinguish the existence of two main mechanisms for the association process. For the first type of complexes we observed that the interplay of reducing the dimensionality of the search process and the hydrophobic dewetting help to turn a seemingly complicated binding process into a well-organized bimodal binding process. For the second type of complexes I found out that the water in the interfacial gap forms an adhesive hydrogen-bond network between the interfaces. Furthermore, the interfacial gap solvent generates an anisotropic reduced dielectric shielding with a strongly preferred directionality for the electrostatic interactions along the association direction.

Kurzfassung

Assoziationen zwischen Proteinen sind äußerst entscheidend für die Signalwege einer Zelle. Unser Wissen über den detaillierten Mechanismus des Assoziationsprozesses ist aufgrund des Mangels an geeigneten experimentellen Methoden, die das Analysieren der Assoziation mit hoher zeitlicher Auflösung erlauben würden, sehr begrenzt. In dieser Dissertation präsentiere ich die Ergebnisse von Molekulardynamik-Simulationen, die durchgeführt wurden um den Mechanismus für zwei Arten von Protein-Protein Assoziationen zu untersuchen. (1) Proteinkomplexe mit hydrophober Kontaktfläche. (2) Proteinkomplexe mit geladener Kontaktfläche. Mit den ungebundenen Proteinen gestartet, reproduzierten die umfangreichen MD-Simulationen die experimentell bestimmten Komplexstrukturen. Diese Echtzeit-Dynamikstudie ermöglichte es mir den Mechanismus des Bindungsprozesses mit einer zeitlichen Auflösung im Pikosekunden-Bereich zu untersuchen und damit die Existenz von zwei Hauptmechanismen des Assoziationsprozesses zu unterscheiden. Für die erste Art von Komplexen beobachteten wir, dass das Zusammenspiel der Reduzierung der Dimensionalität des Suchprozesses und der hydrophobe Entnetzung hilft einen scheinbar komplizierten Bindungsprozess in einen gut organisierten bimodalen Bindungsprozess zu verwandeln. Für die zweite Art von Komplexen fand ich heraus, dass das Wasser in der Kontaktflächenspalte ein klebriges Wasserstoffbrücken-Netzwerk zwischen den Kontaktflächen bildet. Außerdem erzeugt das Lösungsmittel in dieser Spalte eine anisotrop reduzierte nichtleitende Abschirmung mit einer stark bevorzugten Richtungsabhängigkeit elektrostatischer Interaktionen entlang der Assoziationsrichtung.

Acknowledgements

I would like to thank my supervisor Prof. Dr. Volkhard Helms for offering me the opportunity to work in his group. I appreciate his patience, constructive discussions, inestimable suggestions for correcting my writings.

Likewise I would like to thank Prof. Dr. Richard Zimmermann for his valuable support.

I want to thank my colleagues in Saarbrücken especially the members of “Tea Club” for several enjoyable discussions. Furthermore, I thank Wei Gu for his collaboration and especially for introducing the field of Molecular Dynamics simulation to me.

I would like to thank my parents, my sisters and my brother for their love, care and dedication.

Finally, I am grateful to my wife Hala for her unlimited love and support, and to my son Ali, who has given me the energy to continue my research.

Contents

List of Figures.....	vii
Important Symbols and Abbreviations	ix
1 Introduction.....	1
2 Kinetics of the association process.....	4
2.1.1 The role of the diffusion in the association process: the collision theory	5
2.1.2 The intermediate states	9
3 Driving forces for protein associations	13
3.1 Structural view of Protein–Protein Interaction	13
3.1.1 Diversity of Protein–Protein complexes.....	13
3.1.2 Residue contribution to the binding energy	14
3.2 Statistical thermodynamics of the association process	15
3.2.1 Entropy loss upon association.....	16
3.3 The intermolecular forces between proteins	19
3.3.1 The electrostatic interactions	19
3.3.2 The Hydrophobic effect.....	21
3.3.3 The hydrogen bonding	22
3.3.4 The van der Waals interactions	24
3.4 Modeling of the bimolecular interactions.....	25
3.4.1 Molecular mechanics	25
3.4.2 Molecular dynamics.....	28
4 Dielectric permittivity in biological systems.....	29
4.1 Background: electric polarization	29
4.1.1 Localized charge distribution: the electric dipole moment	29
4.1.2 The electric field in in polarizable matter: the macroscopic electric field	32
4.1.3 Langevin equation for the orientation polarization	36
4.2 The electric permittivity	38
4.2.1 The constitutive electrostatic relation.....	38
4.2.2 The static dielectric static permittivity	39
4.2.3 Calculation of the static dielectric static permittivity from MD simulations.	40
4.3 Derivation of a generalized fluctuation formula for calculating the dielectric constant in the interfacial gap.....	42

4.3.1	The relation between the total dipole moment and the directing field	43
4.3.2	The relation between E and E_d	43
4.4	Evaluation of the dielectric permittivity by a polarization formula	44
5	Mechanism of Fast Peptide Recognition by SH3 Domains	46
5.1	SH3 domain complexes as an example for complex with hydrophobic- charged interface	46
5.2	MD simulations of the binding process	47
5.3	MD simulations recovered the experimentally stereospecific structures	48
5.3.1	Crystal structure binding mode	49
5.3.2	Binding mode with opposite orientation	49
5.3.3	Binding in the new pocket	49
5.4	The mechanism of the binding process	52
5.4.1	The importance of electrostatic interactions during the diffusive phase	52
5.4.2	The role of hydrophobic dewetting	56
5.5	Conclusion.....	60
6	Simulating the association between charged proteins.....	62
6.1	Introduction.....	63
6.2	MD simulations of the binding process	64
6.2.1	Setting up the binding simulations.....	64
6.2.2	MD simulations recovered the experimental stereospecific structures.....	65
6.3	Hydrogen bonded network between the proteins	69
6.3.1	Water-mediated interfacial connectivity.....	70
6.3.2	Pairwise water mediation.....	73
6.4	Calculation of the dielectric constant of water molecules in the interfacial gaps .	76
6.4.1	Dielectric permittivity of the water in the interfacial gap using the fluctuation formula	77
6.4.2	Dielectric permittivity of the water in the interfacial gap using the polarization formula	80
6.4.3	Studying the regime of the response of the polarization to the externally applied electric fields	83
6.5	Discussion.....	85
6.5.1	The spatial variation in the dielectric constant.....	86
7	Discussion, Conclusion and Outlook.....	88
7.1	Accelerating the association by reducing the activation enthalpy	88

7.2	Accelerating the association by reducing the activation entropy: Reduction of the dimensionality of the search process.....	89
7.3	How can the dielectric permittivity of the water molecules between the proteins affect the association process?	90
	Bibliography.....	94

List of Figures

Figure 1. The phases of the association process.	5
Figure 2. The transition state along the reaction pathway.	8
Figure 3. Φ -Value analysis.....	10
Figure 4. Φ -Value analysis for the association process.	10
Figure 5. Visualization of transient encounter complexes in protein-protein association.	11
Figure 6. Hydrophobicity on small and large scale.....	22
Figure 7. The nature of the hydrogen bonding.	24
Figure 8. Definition of hydrogen-bond geometry used in Grid and XPLOR force field.....	27
Figure 9. Definition of the hydrogen-bond geometry used in the YETI force field.	27
Figure 10. Localized charge distribution.....	31
Figure 11. The types of polarizations.....	33
Figure 12. The component of the macroscopic electric field in the volume that is occupied by a permanent dipole.	36
Figure 13. The Langevin function.	37
Figure 14. Structure of the binding pocket and binding modes of SH3 domains.	47
Figure 15. Starting structures for MD simulations.	48
Figure 16. The recognized binding modes for the peptide on the surface of the CRK-SH3 domain.	50
Figure 17. RMSD during the simulation time for the binding pathway leading to the crystal structure binding mode.	51
Figure 18. Contact maps between the domain and the peptide for different snapshots in comparison to the native crystal contacts.	51
Figure 19. Snapshots along the binding pathway that leads to the crystal orientation.....	53
Figure 20. Minimum distance between peptide and domain in the 13 binding simulations. .	54
Figure 21. Control simulation for the association of an uncharged peptide.....	55
Figure 22. Control simulations starting from intermediates after mutating the charged residues in the peptide into alanine.....	55
Figure 23. Water density in the hydrophobic interfacial gap with the peptide at fixed distance.	58
Figure 24. Control simulation with charged interfaces.....	59
Figure 25. Schematic representation of the mechanism of the binding process.....	61

Figure 26. Charged residues in Barnase- Barstar complex.	64
Figure 27. Generation of the starting structures for the simulations.	65
Figure 28. Recovered crystal structure by MD simulations.	66
Figure 29. Conformational snapshots along the binding pathway.	67
Figure 30. Multiple pathways for the association process.....	68
Figure 31. Water density in the interfacial gap	69
Figure 32. Development of the water bridges between the interfaces.....	70
Figure33. Water-mediated interfacial connectivity.	72
Figure 34. Comparison of the interfacial water-mediated between hydrophilic and hydrophobic interfaces	73
Figure 35. Water-mediated connectivity between the interfacial residues with known long-range cooperative interaction	74
Figure 36. Comparison of the water mediation between the simulations and the crystal structure.....	75
Figure 37. Dielectric permittivity of the water in the interfacial gap.....	79
Figure 38. The local dielectric behavior relative to the value for bulk water for the complex using the polarization formula.	82
Figure 39. Induced polarization of the water molecules in the interfacial gap as a function of the applied electric field.	84

Important Symbols and Abbreviations

\AA	Angstrom
\tilde{A}	Avogadro's number
BD	Brownian Dynamics
D	electric displacement field
E	macroscopic electric field (Maxwell's field)
E₀	externally applied field
E_{act}	the activation energy
E_d	directing field
E_i	internal field
E_{self}	self-field
F	Helmholtz free energy
F	force
G^0	molar free energy of the standard state
G	free energy
h	Planck's constant
H	enthalpy
H^*	enthalpy of the transition state
K_d	dissociation constant
k	Boltzmann constant
k_{ass}	apparent association rate constant
k_{dis}	apparent dissociation rates constant
k_{on}	association rate constant
k_{off}	dissociation rate constant
k_{dif}	the rate of the collisions
$L()$	Langevin function
m	mass
m	Molecular dipole moment
M	Total dipole moment
MD	Molecular Dynamics
MM	Molecular mechanics
n	refractive index

PRE-NMR	paramagnetic relaxation enhancement NMR spectroscopy
PRM	proline rich motifs
Q	partition function
QM	Quantum mechanics
q	charge
q_{ct}	translational partition function
q_r	rotational partition function
q_v	vibrational partition function
P	polarization
P_μ	orientational polarization
P_α	induced polarization
R	Onsager reaction field
R	gas constant
S	entropy
S^*	entropy of the transition state
SH3	SRC Homology 3 Domain
T	Temperature
U	potential energy
α	polarizability
ρ	Charge density
μ	permanent dipole moment
ϵ	Dielectric constant
χ_e	electric susceptibility
Φ	electric potential
ξ	local microscopic electric field
Λ^3	volume of the quantum mechanical cell

1 Introduction

Biological organisms possess the unique property of being able to self-organize themselves in a dynamic way that maintains the vitality of the organisms. The ability of biological systems to reach stability and vitality at the same time is achieved by the communication between the components of the systems. These communications and signals in biological systems are mediated through the association between molecules. The two main features characteristic for the communications inside the cell are specificity and efficiency. In fact, one of the largest achievements in the fields of biology and medicine of the last century was unravelling the origin of biological specificity by introducing the concept of the receptor. This concept of the receptor explained how signals can be submitted specifically in the biological world via the interaction between the receptors and other molecular “messengers”. There are many molecules that play the role of messengers inside the biological systems. These molecules can be used to transfer the signal inside the cell or between the cells of multi-cellular organisms. Moreover, those highly developed organisms that developed a circulation system have the endocrine system which secretes messenger “hormones” to achieve a global function.

The association between biomolecules to form stereospecific complexes is apparently a problem with high complexity because it includes diffusion of the molecules and adopting the correct orientation before binding. The inherent complexity of this binding process leads back to the question how efficiency of signaling mechanisms is achieved. This efficiency assures that the signaling will be efficient inside the cell so that the ratio between the signal and noise is as low as possible. In other words one may ask the questions: does nature play dices and wait for the chance to find the solution for the complexity of the binding process? Or are there mechanisms to guide the binding process downhill to reach the specific complex?

The associations between proteins are very important in the signaling pathways inside the cell. Our knowledge about the detailed mechanism of the association process is very limited due to the lack of suitable experimental methods that would allow studying the association at high time resolution. The association process can be presented as a movie where we see only the first and the last snapshots. The experimental methods provide us with the structures of the free molecules and the complexes only. In particular, the results from structural biology methods represent to us mostly the static picture of biological processes. Typically, these are only able to detect the highly populated structures and not the scarcely populated intermediates. Fortunately, some new methodological advances e.g. new NMR techniques [3] were able to reveal the nature of some snapshots from lowly populated structures which are interpreted as intermediates of the association process.

Molecular simulations are becoming a very important scientific tool to fill the gap between theory and experiment. Simulating the real world problem is an old dream. Already in 1820 Laplace stated [4, 5]:

“Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective positions of the beings which compose it, if moreover, this intelligence were vast enough to submit these data to analysis, it would embrace in the same formula both the movements of the largest bodies in the universe and those the lightest atom; to it nothing would be uncertain, and the future as the past would present to its eyes.”

The continuous developments in the power of computational methods and hardware are pushing the dream of Laplace into reach. The developments in the last years enable us, for

example, to simulate microseconds of real time dynamics for biological systems of thousands of atoms. However, understanding the association process via computer simulations is still very expensive given the current computational resources. The reason is the atomistic scale of the association process and the need to explicitly represent water, rendering the description of the bimolecular system to be simulated relatively expensive. Moreover, the probabilistic nature of the binding process enforces us to run many simulations over relatively long simulation times to be able to capture the occurrence of the binding process. Luckily, when I started my PhD it was the time when computers became fast enough to simulate a relatively large system over time scales of hundreds of nanoseconds which turned out to be sufficient to monitor such binding processes.

In this PhD thesis I present results from molecular dynamics simulations on atomistic scale to study the mechanism and the driving forces for two types of protein-protein associations. (1) Protein complexes with a hydrophobic interface (SH3 domain). (2) Protein complexes with a charged interface (Barnase-Barstar complex). The extensive MD simulations that were started from the unbound proteins reproduced the experimentally known structures of the stereospecific complexes. This real time dynamics study gave me the possibility to study the mechanism of the binding process at picosecond time resolution and at a spatial resolution on the atomic scale and to distinguish the existence of two main mechanisms for the association process. For the first type of complexes we observed the role of the hydrophobic dewetting during the last steps of the binding process [1]. For the second type of complexes I found out that an adhesive water network between the two protein surfaces mediates the binding of the hydrophilic interfaces [2]. Thus, nature does not seem to play dices and wait for the chance to find the solution for the complexity of the binding process. There exist particular mechanisms to guide the binding process downhill to reach the specific complex.

This thesis is organized in several chapters. The second chapter represents a brief introduction to the kinetics of the association process. Understanding the kinetics of the association is very important for the following discussion about the mechanism of the association process. Chapter 3 provides a general introduction to the driving forces of the protein-protein association process. The presentation is split into a structural view of protein complexes, a general discussion of physical factors that affect the association process with a focus on the case of protein-protein association, and a brief summary of the forces acting between proteins. A brief introduction to the field of molecular dynamics simulation is presented in the last section of this chapter.

Chapter 4 provides a general introduction into the theory of electric polarization in solution and in biological system. The derivation of formulas to compute the dielectric permittivity is presented in this chapter. These formulas were used to study the permittivity of the water molecules in the interfacial gaps between two interacting proteins (chapter 6).

Chapter 5 presents the result from the simulations of association between proteins with a large hydrophobic interface. The association of a proline rich motif to an SH3 domain is probed by molecular dynamics simulations. The complete pathway of the association is studied from the diffusion of the unbound proteins to the final complex through transient encounters. The results reveal that the interplay of reducing the dimensionality of the search process due to the electrostatic interactions and the phenomenon of hydrophobic dewetting help to turn a seemingly complicated binding process into a well-organized bimodal binding process. This chapter was published in *Angewandte Chemie International Edition* in 2008 [1].

Chapter 6 presents the results from the simulations of association between charged proteins (Barnase-Barstar complex). The simulations showed that water in the interfacial gap forms an adhesive hydrogen bond network between the hydrophilic protein interfaces. This

network plays an important role in stabilizing the early intermediate states before native contacts are formed. The transformation from these intermediates to the stereo-specific complex is then accompanied by maximization of the interfacial water-mediation. Furthermore, water structure already plays an important long-range role during the diffusive phase in reducing the dielectric shielding of the water. The dielectric constant in the gap between the proteins is strongly reduced up to distances of a few nanometers. Interestingly, the dielectric properties of the water in the interfacial gap are strongly anisotropic with a preferred directionality for the electrostatic interactions along the direction perpendicular to the interfaces. This chapter was published in *Nature Communications* in 2011 [2].

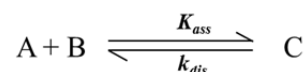
Finally, a general discussion about the mechanisms of the association in the light of my results (chapter 5 and 6) is given in chapter 7.

2 Kinetics of the association process

Understanding the kinetics of the association is very important for the following discussion about the mechanism of the association process. The collision theory is important to understand the role of the diffusion in the association process. The transition state theory is useful to understand the nature of the intermediates and it will be used later to interpret my results and conclusions about the mechanism of associations.

Although the thermodynamic balance between the products and reactants of a biological reaction should be favorable or exothermic for the reaction to occur (or at least only slightly endothermic), the overall exothermicity is not the only important factor determining whether biological interactions take place or not. Some reactions may be thermodynamically very favorable but are kinetically forbidden due to the high energy of the reaction intermediate that is separating the reactant and product states. The rates of the biochemical reactions are in fact very important since they define the biological effect of the reaction. The equations for the kinetics of the association process are the same as the general equations of chemical kinetics.

The association process between two molecules A and B is usually treated as a reversible, one-step process:



Where k_{ass} and k_{dis} are the apparent “experimentally measured” association and dissociation rates. In reality, the association pathway consists of at least three phases: the diffusive phase, the intermediate phases, and the stereospecific phase (see Figure 1). This means that the association process is (at least) a two-step process, in the first step the diffusion brings the molecules together to collide and form encounter intermediates. This step is characterized by rate constant of collisions k_{dif} . The second step is the real binding step where the transition from the intermediates encounter complexes to the stereospecific complex takes place. The binding step is characterized by the intrinsic association rate constant k_{on} and the intrinsic dissociation rate constant k_{off} (see Figure 1). The relation between the experimentally measured association rate k_{ass} and the rates of the two steps model (k_{dif} and k_{on}) is [6, 7]:

$$k_{ass} = \frac{k_{dif} \times k_{on}}{k_{dif} + k_{on}} = \left(\frac{1}{k_{dif}} + \frac{1}{k_{on}} \right)^{-1} \quad (1)$$

As is evident from this equation, the apparent association rate k_{ass} is defined by the slower step. We can distinguish two cases:

$k_{on} \gg k_{dif} \Rightarrow k_{ass} \sim k_{dif}$. In this case the association is “**diffusion-limited**” and the rate depends on the frequency of collisions.

$k_{dif} \gg k_{on} \Rightarrow k_{ass} \sim k_{on}$. In this case the association is “**reaction-limited**” and the rate depends on the activation energy barrier of the reaction.

The fundamental equation for defining the rate of chemical reactions and interactions is the **Arrhenius equation**:

$$k_{ass} = A \times \exp\left(\frac{-E_{act}}{RT}\right) \quad (2)$$

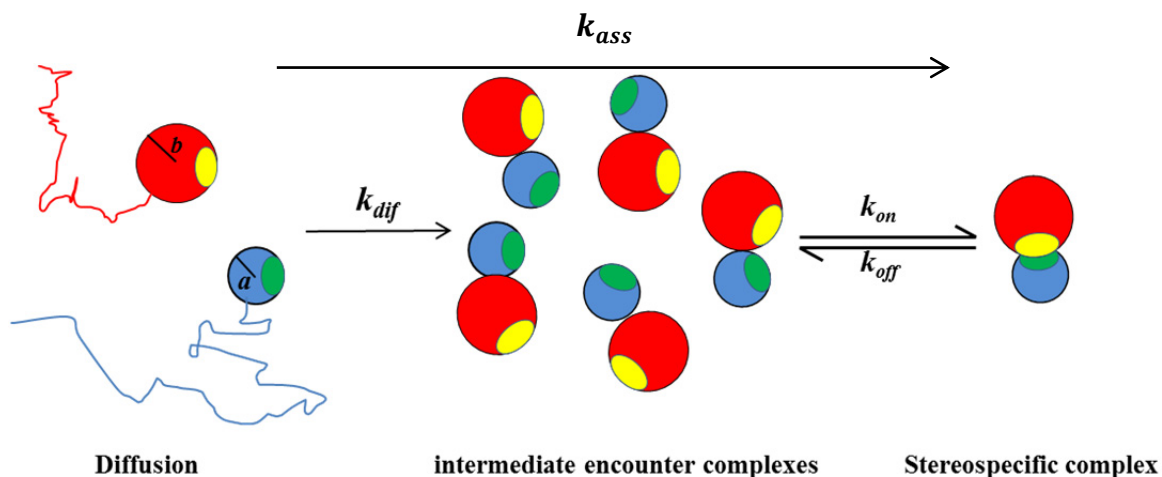


Figure 1. The phases of the association process.

The association pathway consists of three phases: (1) the diffusive phase where both of the reacting molecules are diffusing randomly, (2) the intermediate encounter complexes where the collisions produce loosely bound complexes, (3) and the stereospecific complex. The association process is (at least) a two-step process. The first step is the collision step defined by the rate constant k_{dif} . The second step that is defined by the rate k_{on} is the real binding step where the transition from the intermediates encounter complexes to the stereospecific complex takes place.

The variable A is called the pre-exponential factor. The basic idea is here that the molecules must acquire a certain critical energy E_{act} , "the activation energy", before they can react. The Boltzmann factor $e^{\left(\frac{-E_{act}}{RT}\right)}$ is the fraction of the molecules that have this energy. The Arrhenius equation is very useful to define the effect of the temperature on the rate of the reaction and to calculate the activation energy.

There exist two main microscopic theories in the field of chemical kinetics [8]. The basic difference between them is the way in defining the pre-exponential factor A [9]. The **collision theory** defines the pre-exponential factor by the frequency of collisions between the molecules in the solution and some additional steric factors. In contrast, the **transition state theory** is based on the idea that the energy of the activated complex, "the transition state", defines the rate of the reaction and the pre-exponential factor is related to the entropy of the transition state (see below). In the following I will give a brief summary about the collision theory as it is very useful to understand the role of diffusion in the association process. On the other hand, the transition state theory is very useful for characterizing the phase of intermediate states during the association process.

2.1.1 The role of the diffusion in the association process: the collision theory

In the collision theory, the rate constant for a reaction is assumed to depend on three factors [8, 9]:

$$k_{ass} = k_{dif} \times P \times \exp\left(\frac{-E_{act}}{RT}\right) \quad (3)$$

k_{dif} is the rate of the collisions between the interacting molecules in the solution due to random motion, P is a steric factor accounting for the fact that the interacting molecules should be oriented in a correct orientation to let the interaction take place, and E_{act} is the activation energy term.

2.1.1.1 The frequency of collisions between biomolecules

The **upper limit to the rate of association** cannot be larger than the limit of the collision frequency. Such a limit can be achieved if there exists no steric factor or energetic barrier ($P = 1$ and $E_{act} = 0$) so that the molecules associate into a bound complex whenever they collide.

The diffusion theory provides a good estimation of the collision frequency between molecules in solutions using the hard sphere model. If we consider two spherical molecules A and B of radius a and b and diffusion coefficients D_a and D_b (see Figure 1), the collision frequency factor is given by the **Smoluchowski relation** [10-12].

$$k_{dif} = 4\pi(a + b)(D_a + D_b) \quad (4)$$

This relation gives the misleading impression that the collision frequency is dependent on the size of the molecules. Actually, this factor has only a minimal dependency on the size of the molecules because larger molecules diffuse slower. This can be seen by using the Stokes-Einstein relation $D = RT/6\pi r\eta$ and setting $r = a = b$. This yields estimation for the collision frequency factor dependent only on the viscosity of the solution η :

$$k_{dif} = \frac{8RT}{3\eta} \quad (5)$$

This relation gives a value of $10^{10} M^{-1} S^{-1}$, which is in very good agreement with the experimental finding that the fastest experimental association rates are within the range of $10^9 - 10^{10} M^{-1} S^{-1}$.

The calculation of the rate of collisions through eq. 4 assumes that the molecules A and B do not interact when they are separated and the interaction takes place when they collide (intermolecular distance = $a+b$). The intermolecular interaction can affect the rate of collision by increasing the effective radius of interaction [13]. Debye suggested that the rate of collision can be calculated in this case by substituting $(a+b)$ in eq. 4 by an effective target radius $R^* = 1/\int_{a+b}^{\infty} \frac{1}{r^2} \exp\left(\frac{U(r)}{kT}\right) dr$ [7, 13]. Here $U(r)$ is the intermolecular interaction energy between A and B when they are separated by distance r :

$$k_{dif} = \frac{4\pi(D_a + D_b)}{\int_{a+b}^{\infty} \frac{1}{r^2} \exp\left(\frac{U(r)}{kT}\right) dr} \quad (6)$$

The rate enhancement by an attractive interaction potential in the Debye model is quite modest [14]. For example, if we use the Coulomb interaction potential $U(r) = -Q/r$ eq. 6 becomes:

$$k_{dif} = \frac{4\pi(D_a + D_b) \left(\frac{Q}{kT(a+b)}\right)}{1 - \exp\left(-\frac{Q}{kT(a+b)}\right)}$$

The rate enhancement due to an electrostatic interaction of $9k_B T$ at contact is only 9-fold [14]. Surprisingly, the diffusion limit for the association of a small molecule to a binding pocket on the receptor does not change too much from the frequency of collisions between two proteins. If we assume that the molecule A is a ligand that binds in a round pocket of the same radius on the receptor (molecule B; see Figure 1) then the collision rate is given by the relation [6, 11]:

$$k_{dif} = \frac{4RT}{3\pi\eta} \quad (7)$$

This equation yields a general diffusion limit of $1.6 \times 10^9 \text{ s}^{-1} \text{ M}^{-1}$ for the association rate of a small ligand to a receptor (without interaction). This value is smaller than the rate in eq. 5 by a factor of 2π only.

2.1.1.2 The effect of the steric factors

The previous relations are based on the assumption that a collision between any two parts of the reacting molecules will produce a complex. This can be true for certain interactions between small molecules (e.g. iodine with CCl_4 [11, 15]). However, the pre-exponential factors calculated from collision theory are always higher than the experimental values [9]. This failure is due to the steric restriction on the association process that can be accounted for by introducing a steric factor P . The steric factor was based on the idea that some collisions would be more effective than others, depending on certain directional factors.

The rate constants for the association between macromolecules such as two proteins are heavily influenced by the geometry of the interaction (the steric factor) [8]. The complexity of the binding between macromolecules is higher than that of small molecules because only small parts of each protein (interface) are involved and they should be correctly oriented to form the bound complex because the contacts are specific in the bound complex. This will increase the influence of the steric factor on the association rate constants. Trying to account for the effect of this steric factor on the association through mathematical models is not simple. Alternatively, experiments were done to estimate the association rate of protein-protein association which is produced by the diffusion only (excluding the effect of the electrostatic interactions). Schreiber and Fersht [16] defined the basal association rate constant for the complex of Barnase-Barstar without the contribution of electrostatic interactions which can be masked by high ionic strength of the solution. The basal association rate was found to be about $5 \times 10^5 \text{ s}^{-1} \text{ M}^{-1}$ which is five orders of magnitude lower than the diffusion limit without steric factor (or for the small molecules). Similarly, the basal association rate constant of Hirudin and Thrombin was found to be $5 \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$ [8]. However, the strong electrostatic interactions can again raise the association constants by four orders of magnitude such as observed for the association rate of Barnase and Barstar, $5 \times 10^9 \text{ s}^{-1} \text{ M}^{-1}$. For comparison, the rate of association of H^+ and OH^- is $1.4 \times 10^{11} \text{ s}^{-1} \text{ M}^{-1}$. This point will be discussed later.

2.1.1.3 The transition state theory

The transition state theory expresses the rate of a reaction or of an interaction based on the energy of the postulated transition state relative to that of the reactant state. This transition state is the first-order saddle point along the one-dimensional reaction pathway. This means that it is placed at the energy maximum along this reaction coordinate, but it is an energy minimum in all other coordinates. The rate of the reaction is related to the probability to go over this energetic barrier which is calculated from the energy of the transition state. The importance of the transition state theory relates to the possibility of guessing the structure of the transition state based on the Hammond postulate [17]. This postulates that if there exists an unstable intermediate on the reaction pathway, the transition state for the reaction will resemble the structure of this intermediate.

The transition state theory postulates that the transition state is in thermodynamic equilibrium with the reactant state and populated according to the Boltzmann statistics.

Thus, the rate constants for the association can be calculated according to the free energy difference between the transition state X^* and the ground state X [8]:

$$k_{ass} = \left(\frac{kT}{h}\right) \exp\left(\frac{-\Delta G^*}{RT}\right) \quad (8)$$

ΔG^* is the Gibbs free energy difference between the transition state X^* and the ground (reactant) state X and is called the activation energy (see Figure 2). The activation energies for protein conformational changes and enzymatic processes are often in the range of 10–20 $kcal.M^{-1}$.

The frequency of the transition state decomposition is the same as the vibrational frequency ν for bond breaking: $\nu = \frac{kT}{h}$ where k is Boltzmann's constant and h is the Planck constant. At room temperature (25° C) $\nu = 6.212 \times 10^{12}s^{-1}$ [8, 18, 19] which is the upper limit for the rate constant of any reaction.

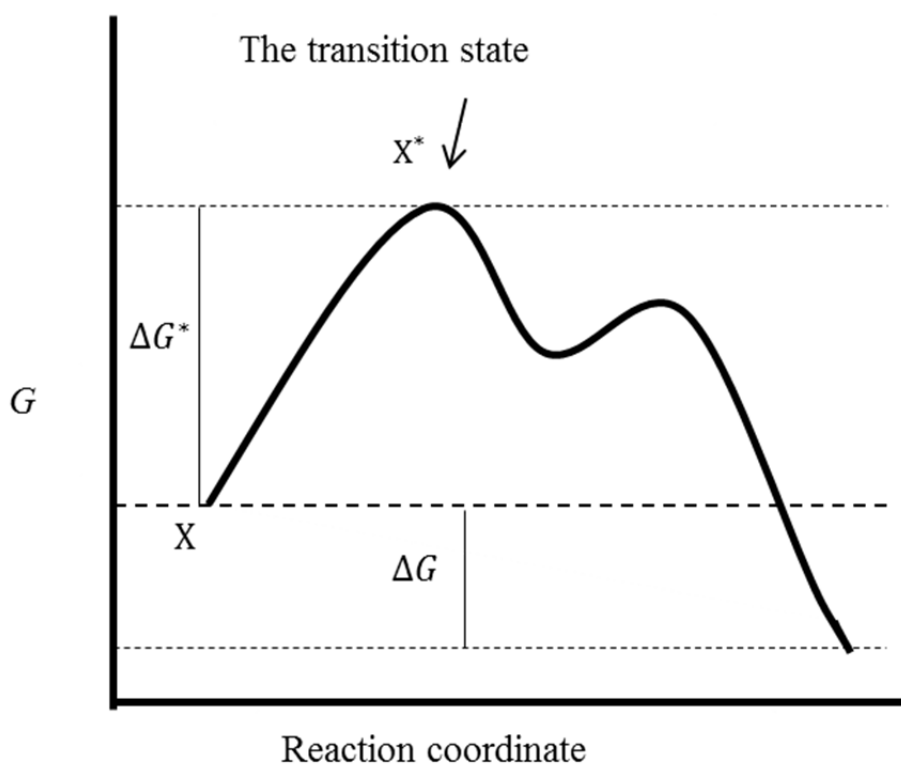


Figure 2. The transition state along the reaction pathway.

The transition state X^* is the position along the reaction pathway with the highest energy (peak). The activation energy ΔG^* is the Gibbs free energy difference between the transition state X^* and the ground state X .

To relate this formula to the geometry of the transition state, we can split up the Gibbs free energy of activation into the activation enthalpy and activation entropy: $\Delta G^* = \Delta H^* - T\Delta S^*$

$$k_{ass} = \left(\frac{k_B T}{h}\right) \exp\left(\frac{\Delta S^*}{R}\right) \exp\left(\frac{-\Delta H^*}{RT}\right) \quad (9)$$

This equation shows that the contribution of the activation entropy to the rate constant does not depend on the temperature and is related to the geometry on the transition state. The main improvement of the transition state theory over the collision theory is the concept of

activation entropy which is an improvement over the less precise concept of the steric factors in the collision theory [9]. In bimolecular reactions, the association of two individual molecules is accompanied by a loss of the entropy thus the activation entropy is usually negative.

2.1.2 The intermediate states

There is little experimental evidence about the intermediate states during the process of protein-protein association. The intermediates are often formed only transiently and are populated at levels that are not amenable to traditional structural approaches. Therefore, mostly theoretical approaches have been used to characterize the nature of the intermediates along the association pathway. Gabdoulhine and Wade introduced the term “diffusional encounter complex” to refer to the end-point of diffusional association in their Brownian dynamics simulations [20]. Spaar and Helms used Brownian Dynamics (BD) simulations to map the encounter complex along the association pathway between Barnase and Barstar and found two minima along the pathway, one leading to association, while the other does not [21].

2.1.2.1 Mapping the interaction in the intermediates

Φ -value analysis is an important concept for defining whether a certain residue contributes to the interaction during the association process [22-24]. Φ -value analysis has been introduced originally in the protein folding field for characterizing the transition state for protein folding [25]. The Φ -value of a particular residue (see Figure 3) is defined as the ratio by how much mutation of the residue affects the activation free energy of the folding process/association process relative to its effect on the folding/binding equilibrium:

$$\Phi = \frac{\Delta\Delta G_{mut}^*}{\Delta\Delta G_{mut}} \quad (10)$$

Here, $\Delta\Delta G_{mut}$ is the effect of the mutation on the free binding energy:

$$\Delta\Delta G_{mut} = -kT \ln \left(\frac{k_d^{mut}}{k_d^{wt}} \right) \quad (11)$$

$\Delta\Delta G_{mut}^*$ is the effect of the mutation on the free energy of the transition state:

$$\Delta\Delta G_{mut}^* = -kT \ln \left(\frac{k_{ass}^{mut}}{k_{ass}^{wt}} \right) \quad (12)$$

k_{ass}^{mut} and k_d^{mut} are the association rate and the association constant for the mutant protein, k_{ass}^{wt} and k_d^{wt} are the association rate and the association constant for the wild-type protein.

A Φ -value of 0 means that the mutant residue has no effect on the free energy of the transition state. This is commonly interpreted in the way that this residue behaves similarly in the transition state as in the unfolded/dissociated state in terms of formed interactions with other residues. On the other hand, a Φ -value of 1.0 means that the effect on the transition state is the same as the effect on the folding/binding equilibrium. This suggests that the residue makes similar interactions in the transition state as in the folded/bound state. Schreiber and coworkers analyzed the Φ values of many hotspot mutations (that have a significant contribution to the binding free energy) for three well studied protein-protein complexes (TEM1-BLIP, barnase-barstar, and Ras-Ral) [14]. They demonstrated that non-

charged mutations always have Φ values close to zero suggesting that the nonpolar residues do not play roles during the intermediate states, whereas charged residues located outside the interfaces have Φ values close to one suggesting that their main role is played during the intermediate states. On the other hand, charged residues located within the interfaces have mixed values [14]. Φ -value analysis can be used with the analysis of double-mutant cycles for analyzing whether contacts and interactions made by the residue during the intermediates states are the same as in the final stereospecific complex [8].

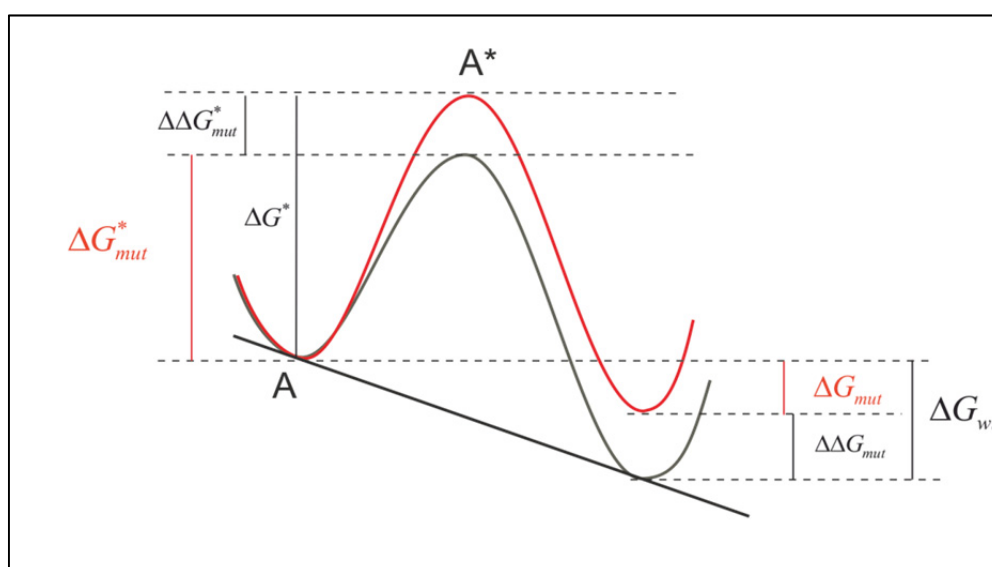


Figure 3. Φ -Value analysis

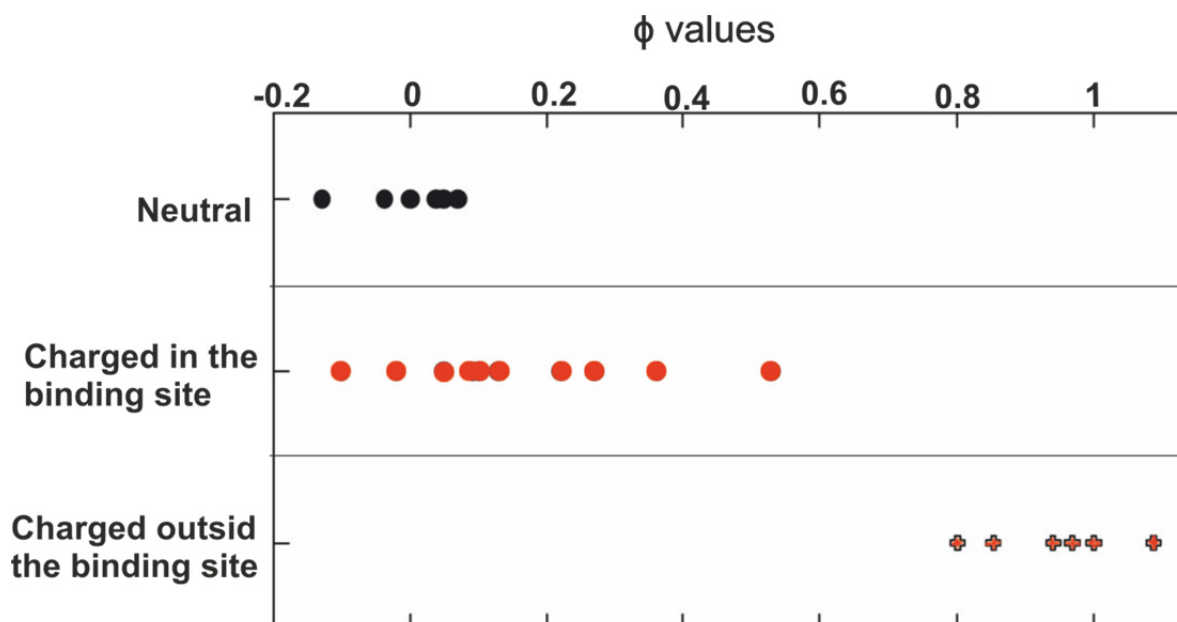


Figure 4. Φ -Value analysis for the association process.

The mutations belong to three different complexes (TEM1-BLIP, Barnase-Barstar, and Ras-Ral). Φ values close to 1 suggest that the involved residues have similar interactions in the transient complex/transition state and in the stereospecific complex, while values close to 0 indicate the residues do not form any contacts in the transient complex/transition state. The values were taken from ref [14].

2.1.2.2 Experimental observation of the intermediates

In the last years, new technical advances in NMR spectroscopy have allowed to obtain structural information about lowly populated, higher-energy conformations that are commonly escape detection by other techniques [3]. The picture that emerged from NMR spectroscopy proofed the existence of intermediates as loosely bonded complexes involving non-specific binding modes. These loosely bound complexes allow the search process for the stereospecific complex to proceed in a space of a reduced dimensionality. The transition from the encounter complexes may proceed as a sliding process of the two proteins until they lock-in into the specific complex. Clore and coworkers used the intermolecular paramagnetic relaxation enhancement NMR (PRE-NMR) to detect the presence of low population transient intermediates under equilibrium conditions [26]. This method was tested for a protein-DNA complex for the association between a sequence-specific transcription factor (the homeodomain of HOXD9) to the non-cognate DNA sites. The PRE data revealed the presence of transient intermediates formed in a stochastic manner at non-cognate sites whose structure is similar to that of the specific complex. The data showed that the homeodomain of HOXD9 slides along the DNA during the search process. In another study, Tang used the PRE-NMR to directly demonstrate the existence and visualize the distribution of an ensemble of transient, non-specific encounter complexes under equilibrium conditions for the protein-protein complex between the amino-terminal domain of enzyme I and the phosphocarrier protein HPr (see Figure 5) [27].

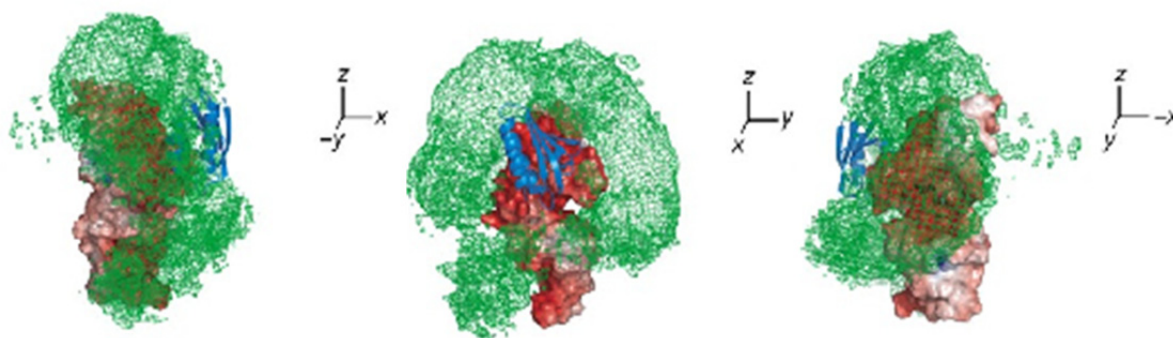


Figure 5. Visualization of transient encounter complexes in protein-protein association.

The distribution of HPr molecules (green) on the molecular surface of E1N (color coded by electrostatic potential) is presented from different sides. The distribution of nonspecific encounter complexes appeared to be qualitatively correlated to the electrostatic surface potentials of the interacting proteins. The location of HPr in the stereospecific complex is shown as a blue ribbon in all panels. The figure is taken from reference [27].

The distribution of nonspecific encounter complexes appeared to be qualitatively correlated to the electrostatic surface potentials of the interacting proteins. An additional study was performed to clarify the screening effect of ions on the electrostatic interaction [28]. The measurements demonstrate that the population of nonspecific encounter complexes is modulated by ionic strength to a larger degree than the stereospecific complex, highlighting the importance of electrostatic interactions in the formation of the nonspecific encounter complex ensemble more than in the stereospecific complex [28]. Volkov determined the conformation of the protein complex between the yeast cytochrome *c* peroxidase and iso-1-cytochrome *c* in solution using PRE-NMR [29]. They reported that the dominant orientation of the protein complex in solution is the same as that in the crystal structure (> 70%) the remaining conformations belong to a dynamic encounter state. At the end of this chapter, it is important to mention that several authors have used the term “encounter complex” describe intermediates along the pathway of protein-protein association, but the precise

meaning of this term may differ depending on the techniques employed by different authors and depending on the systems they studied [14].

3 Driving forces for protein associations

This chapter provides a general introduction into the physical basis of the protein-protein association process. The presentation is split into a structural view of protein complexes, a general discussion of physical factors that affect the association process with a focus on the case of protein-protein association, followed by a general estimation of the entropy loss upon association, and a brief summary of the forces acting between proteins. The hydrogen bonding interaction is explained in more detail because of its high importance for the results to be presented in the following chapters. A brief introduction to the field of molecular dynamics simulation is presented in the last section.

3.1 Structural view of Protein-Protein Interaction

3.1.1 Diversity of Protein-Protein complexes

A classification of protein-protein complexes can be made based on the life-time of the formed complexes [30]. Short-lived complexes can be formed due to the collisions that occur within the crowded matrix of the cell and that create short-lived nonspecific complexes. On the other hand, many oligomeric proteins form long-lived complexes and they appear to be permanently bound where they disassociate only when they enter a degradation pathway [30]. In between these two limits, the protein complexes exist as transient (reversible) complexes in equilibrium between the free and the bound form.

Nowadays, thousands of known structures for protein-protein complexes are available in the PDB data bank [31] and the number is daily increasing. An analysis of these structures showed that the interfaces of the protein complexes are quite different from the binding pockets of small-molecule ligands. Generally, the small molecules bind in deep clefts on protein surfaces [32] whereas the protein-protein interfaces are relatively flat and larger in size [33, 34]. The surface areas of the binding interfaces of the protein complexes are in the range of 700–1500 Å² [33, 34]. The larger interfaces found in protein complexes in comparison to the interfaces between proteins and small molecules imply that the association between proteins leads to a larger disturbance of the water structure around the interfaces upon association. For this reason, water is a very important player in the binding process of two or more macromolecules and needs to be considered by all means. However, the role of water is very complex and cannot be easily explained as in the case of pairwise non-covalent interactions.

The chemical nature of the residues at the interfaces is very important in defining the driving forces for the binding process. The phenomenon termed “hydrophobic dewetting” is considered as the main driving force for protein folding [35-37]. This is in agreement with the hydrophobic nature of the proteins cores. Moreover, the electrostatic interactions through the salt bridges do not show any contribution to the stability of the folding [38] which is expected from the high desolvation energy due to burying charged residues in a medium of low dielectric constant. The driving forces for protein association are not well characterized. However, existence of a large hydrophobic part of the interfaces is observed in many protein complexes such as the case of the pockets of the Proline Recognition Domains which bind a protein with a hydrophobic pocket via proline rich motifs [39, 40]. Thus we can expect hydrophobic dewetting to be the driving force in some cases. The interfaces between the proteins are not as hydrophobic as the core of the proteins. Moreover, the residue composition of most protein-protein interfaces appears to be more similar to that of protein surfaces than to that of protein cores [41]. There are many examples of proteins complexes involving interactions between many charged residues such as the case of the Barnase-Barstar complex. In the last years, several studies have pointed out that dewetting is rather

rare [42-44] because a few polar residues can already prevent the occurrence of dewetting [43, 45]. Thus, hydrophobic dewetting cannot be the only driving force for protein-protein associations.

Statistical analyses of the structures of protein-protein complexes showed that both hydrophobic and hydrophilic interfaces exist. However, the statistics are quite sparse. Typically, between 30-50% of the protein interface area is taken up by hydrophobic amino acids [46, 47]. Ansari and Helms reported that on average about 30% of the interfacial residues are hydrophobic and 70 % of the interfacial residues of protein complexes are hydrophilic including about 37 % charged residues [47]. Xu and Nussinov reported that on average two ion pairs per interface were found in protein complexes [48].

3.1.2 Residue contribution to the binding energy

The contacts between residues in protein complexes make important contributions to the binding free energy. However, the protein part of the total binding energy of the complex cannot simply be added up by summing the pairwise interactions between protein residues [49]. A measure of this non-additivity is the coupling free energy that is due to specific interaction between residues. This coupling free energy is usually measured by the double mutant cycle (see below). However, some nonspecific contributions such as the nonpolar energy due to the hydrophobicity of the residues appear to be additive [50].

The contribution of a certain residue to the binding energy can be measured by site-directed mutagenesis where the residue is usually mutated into alanine. The residues that contribute to the binding energy by more than a threshold value are termed the hot spots. The energetic criterion to define hotspots is usually between $\Delta\Delta G > 1.5 \text{ kcal. M}^{-1}$ [51] and $\Delta\Delta G > 2 \text{ kcal. M}^{-1}$ [52] representing a change in K_d by factors of 12 and 30, respectively.

The interaction energy (coupling energy) between a pair of residues is usually measured by the double mutant cycle [53, 54]. Defining the interaction energy between two residues X and Y requires four measurements of the binding energy: ΔG_{wt} for the wild-type, $\Delta G_{X \rightarrow A, Y \rightarrow A}$ for the double mutant, and $\Delta G_{X \rightarrow A}$ and $\Delta G_{Y \rightarrow A}$ for the single mutants. The change in the binding energy due to the mutations is calculated and the residues are considered to co-operatively interact if the effect of the mutations is more than additive. For this, the interaction energy is defined by:

$$\Delta\Delta G_{X...Y} = \Delta\Delta G_{X \rightarrow A, Y \rightarrow A} - \Delta\Delta G_{X \rightarrow A} - \Delta\Delta G_{Y \rightarrow A}$$

Schreiber and Fersht [55] used double mutant cycles to study the interaction between a subset of five Barnase and seven Barstar residues in the complex of Barnase-Barstar. The coupling energy between two residues was found to decrease with the distance between them. Generally, only residues separated by less than 7 Å have significant coupling energies. The highest coupling energies were found between pairs of charged residues (1.6 to 7 kcal M^{-1}). However, a few pairs showing long-range coupling energies were even separated by distances up to 15 Å. For the first time, this thesis provides a physical explanation for the nature of this long-range interaction (see 6.3.2). The double mutant cycle can also be used to calculate the interaction in the intermediate states if one uses the energy of the transition state instead of the binding free energy.

Important residues for the binding process are mostly located in the core of the interfaces. Chakrabarti and Janin [33] analyzed the architecture of the binding interfaces of 70 complexes and distinguished between small and large interfaces. The small interfaces usually form a single patch on the surface of each component protein. In contrast, larger interfaces are generally composed of multiple patches. Each patch consists of two regions:

the core (55% of the residues), made up by the residues that contain the buried interface atoms, and the rim (45% of the residues), consisting of those residues for which all the interface atoms remain solvent accessible. Generally the rim is very similar to that of the solvent accessible surface whereas the core is depleted by a factor of nearly two in charged residues and enriched by the same factor in aromatic residues. Analyzing the distribution of hot spots showed that more hot spot residues (with $\Delta\Delta G > 2 \text{ kcal.M}^{-1}$) are located in the core area than in the rim area [56]. Guharoy [57] reported a correlation between the contribution of the core residues to the binding energy ($\Delta\Delta G$) and their buried surface area. The correlation factor ($26\text{--}38 \text{ cal.M}^{-1}.\text{A}^{-2}$) is very close to the nonpolar energy that is estimated via the surface area which is derived from the experimental solubility of hydrocarbons ($20\text{--}25 \text{ cal.M}^{-1}.\text{A}^{-2}$) [50].

Despite of the limited contribution of the residues outside the interfacial area to the stability of the complex, the charged residues located outside the interfaces were found to play an important role in the association process prior to the bound complex [14] where the total electrostatic interaction between the protein affects the association rate. This fact has been used in a few cases to engineer protein pairs with faster association rates by increasing the charge complementary of the residues outside the interfaces which can increase the association rate without affecting the stability of the complex [58, 59].

3.2 Statistical thermodynamics of the association process

There are two general features related to the association of biological molecules [11]: (1) Association is achieved by non-covalent interactions. This results in a weak stability of the complex in comparison to covalent bonding. However, the weak interactions enable the association to be reversible. (2) Associations are stereospecific where the two associating molecules fit together as lock and key.

An association usually involves two molecules A and B that form a complex C:



The dissociation constant (equilibrium constant) K_d or K_{dis} is defined at equilibrium by the relative concentrations of these three molecules:

$$K_d = \frac{[A][B]}{[C]} \quad (13)$$

The dissociation constant has the units of concentration (molar unit). The association constant (or the affinity constant K_a) is simply the reciprocal of the dissociation constant. However, the dissociation constant is usually used because of its molar unit. The dissociation constant is an indicator for the strength of the binding thus a small dissociation constant means a tighter binding. The values of the dissociation constant of the biomolecules range from micro-molar μM for weak binding to nano-molar nM and pico-molar pM for strong binding. Although the dissociation constant of the non-covalently bonded complexes is rarely stronger than sub-picomolar, very strong affinity ($K_d = 10^{-15}$) can be found such as in the case of the Biotin-avidin complex [60] and the complex between the Ribonuclease and its inhibitor [61].

The relation between K_d and the binding free energy is easily derived from the molar free energy. The molar free energy for a solute A in an ideal solution is:

$$G = G^0 + RT \ln[A] \quad (14)$$

where G^0 is the molar free energy of the standard state. At equilibrium, the free energy of the product (C) is equal to the free energy of the reactants (A and B):

$$\Delta G^0 = -RT \ln \frac{[C]}{[A][B]} = RT \ln K_d \quad (15)$$

ΔG^0 is the free energy change of the standard state that occurs when 1M of A and 1 M of B associate to form 1 M of C.

To get an idea of the energetic factors that affect the association we have to look at the problem from the viewpoint of statistical thermodynamics. The partition function Q of a molecule A can be broken into various terms [9, 11]:

$$Q = q_{ct}q_tq_rq_vq_{cf}q_s \quad (16)$$

where q_{ct} denotes the contributions of intermolecular contacts with other molecules in the complex, q_t is the translational contribution, q_r is the rotational contribution, q_v the vibrational contribution, q_{cf} the contribution of conformational flexibility, and q_s accounts for solvation effects. The free energy of N molecules is given by [9, 11, 62]:

$$G = -kT \ln \left(\frac{Q^N}{N!} \right) \quad (17)$$

The free energy of the association process is the sum of the free energy of the molecular types A, B and C. Thus we get [9, 11]:

$$G = -kT \ln \left(\frac{Q_A^N Q_B^N Q_C^N}{N_A! N_B! N_C!} \right) \quad (18)$$

At equilibrium, the free energy is at its minimum. Finding the minima of eq. 6 leads to the following equilibrium condition [9, 11]:

$$\frac{Q_A Q_B}{Q_C} = \frac{N_A N_B}{N_C} \quad (19)$$

Comparing this equation with eq. 13 yields:

$$\frac{1}{K_d} = \frac{Q_C V \tilde{A}}{Q_A Q_B} \quad (20)$$

Here, V is the volume and \tilde{A} is Avogadro's number. Now we can break the partition functions into components (eq. 16):

$$\frac{1}{K_d} = \left(\frac{q_{Cct}}{q_{Act}q_{Bct}} \right) \left(\frac{q_{Ct} V \tilde{A}}{q_{At}q_{Bt}} \right) \left(\frac{q_{Cr}}{q_{Ar}q_{Br}} \right) \left(\frac{q_{Cv}}{q_{Av}q_{Bv}} \right) \left(\frac{q_{Ccf}}{q_{Acf}q_{Bcf}} \right) \left(\frac{q_{Cs}}{q_{As}q_{Bs}} \right) \quad (21)$$

The volume V was combined with the translational component because that is the only component to be affected by V (see eq. 24 below). Now it is clear that we can break the free binding energy into the following additive terms:

$$\Delta G^0 = \Delta G_{ct}^0 + \Delta G_t^0 + \Delta G_r^0 + \Delta G_v^0 + \Delta G_{cf}^0 + \Delta G_s^0 \quad (22)$$

3.2.1 Entropy loss upon association

Upon association, the loss of the translational free energy counteracts the binding. The translational part of the partition function is proportional to the volume occupied by the

molecule; the smaller the volume, the more the molecule is restricted and the lower the entropy [9, 62]:

$$q_t = V \left(\frac{2\pi mkT}{h^2} \right)^{\frac{3}{2}} = \frac{V}{\Lambda^3} \quad (23)$$

m is the mass, h is Planck's constant, and Λ^3 is the volume of the quantum mechanical cell. The translational free energy is given by the relation:

$$\Delta G_t^0 = -RT \ln \left(\frac{q_{Ct} V \tilde{A}}{q_{At} q_{Bt}} \right) = -RT \ln \left(\frac{\Lambda_A^3 \Lambda_B^3 \tilde{A}}{\Lambda_C^3} \right) \quad (24)$$

If we consider the binding of a small ligand with a molecular weight of 100 Da to a protein we get as loss in the translational free energy upon association a value of about [11]:

$$\Delta G_t^0 \approx 8.4 \text{ Kcal. } M^{-1}$$

It is important to note that the dependence on mass is low since a ten-fold increase in the mass increases the free translational energy by only $2 \text{ kcal. } M^{-1}$. This energetic contribution is mostly, but not entirely, of entropic origin. The enthalpic part equals $3/2 RT \approx 0.9 \text{ kcal. } M^{-1}$ at room temperature [11].

Similarly to the translational free energy, the loss in the rotational free energy is independent from the mass and counteracts the binding process. It is estimated to be in the range [11]:

$$\Delta G_r^0 \approx 8.2 \text{ Kcal. } M^{-1}$$

This loss is mostly entropic as well with only $3/2 RT \approx 0.9 \text{ Kcal. } M^{-1}$ as the enthalpic contribution.

In contrast, the vibrational free energy contribution favors the bound state. The complex of two molecules has six new vibrational degrees of freedom that were not present in the separated molecules. Depending on the vibrational modes in the complex, the vibrational entropy can be of larger importance and highly variable in the case of protein-protein association. The vibrational entropy can be estimated by normal mode analysis [63]. A rough estimation of the vibrational free energy change upon protein-protein association gave a value of [11]:

$$\Delta G_v^0 \approx -4.6 \text{ to } -6.5 \text{ Kcal. } M^{-1}$$

The total negative contribution of the restriction of the motions upon association (translational, rotational and vibrational) is thus:

$$\Delta G_{t+r+v}^0 \approx 10 - 12.2 \text{ Kcal. } M^{-1}$$

The contribution of forming contacts must exceed this value to allow for binding to occur. Since the enthalpic contribution to the translational and rotational free energy is small, and is conserved during association, we can consider the loss of the free energy due to restricting the motion of the reactants as purely entropic in origin.

3.2.1.1 Conformational changes upon association

Associations can be accompanied by conformational changes that may cause a decrease in the conformational entropy. The conformational changes can either be limited to the mobility of the side chains or also include the backbone. An important example for the latter case is the local folding of unstructured parts of a protein upon binding [64]. However, it came almost as a shock that, nowadays, about 40% of the proteins in the human proteome

are expected to be intrinsically disordered and become fully or partly structured only upon binding to their partners [65, 66]. The functionality of these intrinsically disordered proteins [67] recently opened up a new era in the field of structural biology after fifty years of research have concentrated on defining the structure of the proteins under the assumption that the functionality is encoded in the structure.

The folding of the unstructured parts upon binding will add an additional energetic penalty to the binding free energy by decreasing the conformational entropy. Spolar and Record [64] analyzed experimental data of many proteins and gave an estimation of 5.6 cal.K^{-1} for the loss of the conformational entropy upon folding per residue. In the same study they noticed that the change of conformational entropy is zero for the association processes of many complexes.

The entropic penalty of the conformational changes during the binding process can be very low if its price was paid prior to the binding process. In this context, there exist two main working models about the mechanisms of conformational changes upon binding : the induced fit model and the conformational selection [68]. According to the “induced fit” hypothesis, the conformational changes are induced by the association process [69]. The “conformational selection” hypothesis assumes that the association process takes place once the bound conformation is adopted due to thermal motion. For this, the bound conformation is assumed to be populated, prior to the binding interaction, as one conformation in the ensemble of interconverting conformations at thermal equilibrium [70]. For a long time the induced fit model was the dominating model to explain conformational rearrangements during binding. A new study by Lange et al. [71] revealed that conformational selection, rather than induced-fit, is the mechanism that determines how complexes involving the protein ubiquitin form. In this study, the authors presented a structural ensemble of ubiquitin from a long dynamics simulation of microseconds length that was refined against residual dipolar couplings (RDCs) measured by NMR. The ensemble covered the complete structural heterogeneity observed in 46 crystal structures where ubiquitin was bound to other proteins. A large part of the solution dynamics was concentrated in one concerted mode, which accounts for most of ubiquitin's bound conformations and ensures a low entropic cost upon formation of the complex [71].

The association between biomolecules is not necessarily accompanied by a reduction in the conformational entropy. Several studies showed that the motion can also increase or remain the same within molecular complexes [72]. Zidek reported that the backbone flexibility of mouse major urinary protein increases upon binding of a hydrophobic ligand [73]. The associated increase in backbone conformational entropy of the protein appears to make a substantial contribution toward stabilization of the complex [73]. Chang and McCammon [74] studied the change in configurational entropy (conformational and vibrational) changes upon complex formation between the protein kinase A (PKA) and A-kinase anchoring proteins (AKAP). PKA is a promiscuous protein because it uses a single face to interact with multiple A-kinase anchoring proteins. The results show that the majority of configurational entropy loss was due to decreased fluctuations within rotamer states of the side chains and not due to the usually assumed decrease in the number of rotamer states available to the side chains. Moreover, they found a direct linear relationship between the total configurational entropy and the number of favorable, alternative contacts available within hydrophobic environments. They proposed a general strategy for promiscuous proteins to be proceed through an increase in the configurational entropy by providing alternative contact points at its interface [74].

3.2.1.2 The entropy gain by releasing water from the interfaces

Most proteins have a few hydration sites at their protein surfaces that can bind water rather tightly. Displacing these tightly bound water molecules upon association can be an important driving force for the association. Dunitz [75] used experimental thermodynamic data to estimate the entropic contribution for releasing a water molecule bound to the protein to the bulk water. The entropy of tightly ordered water in crystals is about $10 \text{ cal} \cdot \text{M}^{-1} \text{K}^{-1}$. On the other hand, the entropy of liquid water at room temperature is $16.7 \text{ cal} \cdot \text{mol}^{-1} \text{K}^{-1}$. The difference is $6.7 \text{ cal} \cdot \text{M}^{-1} \text{K}^{-1}$. This means that displacing one tightly bound water molecule upon binding favorably contributes up to $2 \text{ kcal} \cdot \text{M}^{-1}$ to the free binding energy. This factor is not very important when a small molecule binds to a protein as it cannot displace many water molecules. However, the association of proteins can displace a sizeable number of water molecules. The entropic gain due to the release of these water molecules is an important factor in driving the association process. For example, the displacement of 10 water molecules can contribute to the free binding energy up to $20 \text{ kcal} \cdot \text{M}^{-1}$ which is enough to counteract the negative contribution of the translational and rotational free energy ($10 \text{ kcal} \cdot \text{M}^{-1}$).

3.3 The intermolecular forces between proteins

The non-covalent interactions between molecules differ in their range of interaction, specificity, and strength. The strength of non-covalent interactions between proteins ($5\text{-}15 \text{ kcal} \cdot \text{M}^{-1}$) is much weaker than the strength of the weakest covalent interactions (chemical bonding) that range between $47\text{-}190 \text{ kcal} \cdot \text{M}^{-1}$ [76]. However, it is of crucial importance that non-covalent interactions of transient complexes are of this magnitude because this enables the proteins to bind with a sufficient life-time to exert their cellular functions and then to dissociate again.

3.3.1 The electrostatic interactions

In principle, all forces between molecules are electrostatic in origin due to the way how atoms are built up from charged electrons and protons. However, in the molecular modeling field, the term electrostatic is usually reserved for interactions that occur between charged or dipolar particles. The electrostatic interaction between two point charges q_1 and q_2 separated by a distance r in a medium of dielectric constant ϵ is given by Coulomb's law (in CGS units):

$$U_{el} = \frac{q_1 q_2}{\epsilon r} \quad (25)$$

The electrostatic interaction is usually the longest interaction in range. The range of the electrostatic interactions is variable according to the multipole character of interacting molecules. The energy between two charged particles (monopoles) falls off with distance r as $1/r$. The net energy of the electrostatic interactions that involve dipoles or quadrupoles is much shorter in their range. For example, the net electrostatic energy between a charge and a dipole falls off with $1/r^2$; moreover, the energy between two dipoles falls off with $1/r^3$. This explains the important role of charged residues in the association process in spite of the polar nature of many non-charged residues. Relative to the considerable strength of electrostatic interactions in vacuum (at $\epsilon = 1$), the electrostatic interaction is strongly reduced (i.e. screened) in water due to the high dielectric constant of the water ($\epsilon \approx 78.5$). As an example, the electrostatic interaction between an Na^+ and Cl^- ion pair in vacuum at 0.3 nm

distance is about 100 kcal.M^{-1} . If the same interaction occurs in water, the interaction is only about 1.3 kcal.M^{-1} which equals about $2 RT$ at room temperature ($RT \approx 0.6 \text{ kcal. M}^{-1}$).

Calculations based on the methods of continuum electrostatics [77, 78] suggest that Columbic interactions between buried ions pairs are generally not strong enough to compensate for unfavorable desolvation effects. Thus the overall contribution of electrostatic interactions to an assembly process would, in general, be expected to be unfavorable. Moreover, the charge complementarity which is favorable at protein-protein interfaces [79] is not really favorable to optimize the interaction which is controversial; rather it is related to minimization of the loss in the solvation energy due to burial of a single charged residue.

Also electrostatic interactions between buried salt bridges in the protein interior were reported to be destabilizing [38]. Even simple hydrophobic interactions provided more stabilizing energy than the buried salt bridge [38]. Hendsch and Tidor [80] studied the electrostatic contribution to the free energy of folding for 21 salt bridges in 9 protein crystal structures using a continuum electrostatics approach. The majority (17) were found to be electrostatically destabilizing. In another study, Lee and Tidor [63] studied the electrostatic interaction between Barnase and Barstar. They found that the Barstar interface is electrostatically optimized to interact with Barnase by reducing the desolvation energy. Sheinerman and Honig [77] used the continuum electrostatics method to investigate the contribution of electrostatic interactions to the binding of four protein-protein complexes; barnase-barstar, human growth hormone and its receptor, subtype N9 influenza virus neuraminidase and the NC41 antibody, the Ras binding domain (RBD) of kinase cRaf and a Ras homologue Rap1A. In two of the four complexes electrostatics were found to strongly oppose binding (hormone-receptor and neuraminidase-antibody complexes), in one case the net effect was close to zero (Barnase-Barstar) and in one case electrostatics provided a significant driving force favoring binding (RBD-Rap1A) [77]. They concluded that the important factor in defining the contribution of the electrostatic interactions is the extent to which the desolvation of buried charges is compensated by the formation of hydrogen bonds and ion pairs.

Our ability to estimate the importance of electrostatic interactions by continuum electrostatics methods is limited by the knowledge of the relative dielectric constant. Most continuum electrostatic methods use a value of 80 for water and an internal dielectric of 2-6 inside the protein. Moreover, the possible contribution of a structured solvent is not accounted for. The picture that emerged from our simulation of the binding process of two proteins (see 6.3) show that the water mediation is very important in charged interface between Barnase and Barstar. Moreover, the interfaces are not desolvated as is assumed here but rather keep most of their first hydration shells which are responsible for the majority of the solvation free energies. In spite of having evidences from crystal structures of protein complexes about the hydrated nature of the interfaces, the existence of these water molecules is commonly ignored and the crystal water is simply removed before processing the structures. This can lead to closing the water gaps between the interfaces during the energy minimization which is usually performed before the calculations.

The electrostatic interactions are the most important factors in determining the rate of association between proteins. Mutations on non-charged residues led to rather small changes of the binding rates whereas mutations of charged residues can affect association rates by more than 20-fold [58, 59]. Moreover, the relation between the association rate and the electrostatic interactions and the ionic strength was found to follow the Debye-Hückel energy for the interaction between a pair of proteins [58].

3.3.2 The Hydrophobic effect

Hydrophobic interactions are responsible for several biological processes such as the aggregation of amphiphilic lipids into bilayers, the burial of hydrophobic residues in proteins that helps proteins to fold, and the aggregation of protein subunits into multi-subunit quaternary structures [81, 82].

The importance and the nature of the hydrophobic effect is one of the most complicated points in understanding the driving forces in biology. Before going into details about the hydrophobic effect, one should clarify two main points. First, the recent investigations of the hydrophobic effect showed that this effect is complex and cannot be explained by one generic mechanism [83]. Second, the common understanding about this effect is heavily populated with nonscientific explanations based on the concept of “love and hate”. Many terms such as “hydrophobic interaction” and “hydrophobic bond” and “hydrophobic effect” [84, 85] are used to describe the tendency of the hydrocarbons to aggregate in aqueous solution. Indeed, all these terms are not accurate and populated the scientific understanding of these phenomena. Hildebrand [84] objected that the word “phobia” is inappropriate because alkyl chains in micelles of soap are not bonded together by phobia for surrounding water; they stick together just as strongly in the absence of water. Moreover, the recent understanding of the hydrophobic effects shows that this effect originates from the complex behavior of the water and there is no special physical interaction between the hydrophobic materials themselves.

As just pointed out, there is no single generic mechanism for the origin of the hydrophobic effect. The recent understanding of this effect distinguishes between the hydrophobic effect on a small scale (“small hydrophobes”) which is entropic in origin, and the hydrophobic effect on a larger scale which is enthalpic in origin. Both of the mechanisms are related to the effect of the solute on the hydrogen bond network in water [35, 86].

The nature of the hydrophobicity observed for small hydrophobes that drives the clustering of small hydrophobes in water is usually explained by the “iceberg” model suggested by Frank and Evans [87] and formulated later by Kauzmann [88]. According to this model, when a small-solute (like a methane molecule) is inserted in the water, it will exclude the centers of water molecules from a spherical volume. If this volume is small enough to fit inside the network of water hydrogen bonds without breaking them, water molecules can adopt orientations that allow hydrogen-bonding patterns to go around the solute (Figure 6a). The new restriction in the orientation of the water molecule around the solute “the iceberg” will decrease the entropy of the water molecules. The clustering of small hydrophobes in water is induced by the gain in entropy by releasing these water molecules to the bulk water [35, 86].

The hydrophobicity on a larger scale is explained by the hydrophobic dewetting model (Figure 6b). If a large solute (usually larger than 1 nm) is inserted in the water, the low curvature of its surface will make it impossible for adjacent water molecules to maintain a complete hydrogen bonding network with the surrounding liquid. To minimize the loss of hydrogen bonds, water tends to move away from the large solute so that an interface of vapor-like water is formed around the hydrophobe (hydrophobic dewetting) [86]. The significant changes of the total number of hydrogen bonds will lead to a decrease in the enthalpy of the water molecule [35]. When two hydrophobic surfaces come closer than a critical separation distance D_c , a drying transition takes place in the space between the surfaces termed “microscopic capillary evaporation” and the resulting imbalance in pressure causes the two surfaces to attract [35, 82, 86]. Thus, the attraction between the two large hydrophobic surfaces results from the enthalpy gain upon evaporation of the water of the

interfaces. The hydrophobic dewetting is considered as a major component of the forces that fold and stabilize the structure of biomolecules [35, 44, 82, 86, 89-92]. Berne and coworkers [89, 90] used MD simulations to show that the hydrophobic dewetting is taking place during protein folding. We have shown that dewetting is taking place during protein association when two hydrophobic interfaces are close to each other [36].

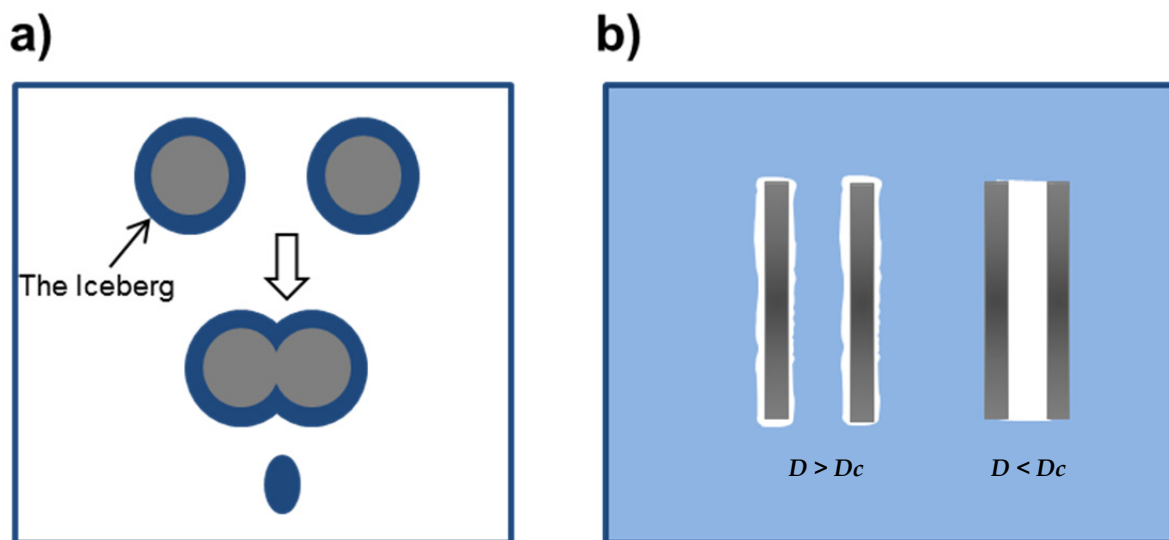


Figure 6. Hydrophobicity on small and large scale

(a) "Iceberg" model of hydrophobicity for small hydrophobes [82]. (b) The hydrophobic dewetting between large hydrophobic surfaces. The figure is taken from reference [82].

3.3.3 The hydrogen bonding

The hydrogen bonding interaction is one of the most important non-covalent interactions in the biological world. In particular, the hydrogen bonding is considered as one of the main players in defining the three-dimensional conformation of biological macromolecules. The secondary structure of proteins and DNA is defined mainly by the hydrogen bonds between the amino acids and nucleotide basis. The unique feature of the hydrogen bonding is its specificity in comparison with other non-covalent interactions. The importance of hydrogen bonding for the stability of protein complexes is well characterized. Janin and co-workers [46] found that protein-protein complex contain about 10 intermolecular hydrogen bonds (approximately one hydrogen bond per 170 \AA^2). About one third of these hydrogen bonds involves at least one charged residue and 13% are formed between two charged groups [46].

A hydrogen bond $D-H \cdots A$ is formed when a hydrogen atom which is covalently bonded to an atom D (donor) interacts with a region of high electron density in another atom A (acceptor) where D is an atom with electronegativity greater than that of hydrogen and A can be any σ or π electron donor. The hydrogen atom is attracted to both atoms, D and A and plays the role of a bridge between them. After the formation of the hydrogen bond, the distance between the bonded atoms is closer than the sum of their van der Waals. On the other hand, the length of the covalent bond D–H is somewhat increased and its IR stretching frequencies are shifted to lower frequencies. However, the definition of the hydrogen bonding is still controversial. The early picture of the hydrogen bond was introduced by Pauling [93] as an electrostatic interaction where only very electronegative atoms (F, O, Cl, N and Br) were considered to be able to form a hydrogen bond. Pimentel and McClellan [94] gave a more comprehensive definition of a hydrogen bond by

considering that donors such as C–H and acceptors such as π electrons can participate in a hydrogen bond which is generally weak.

The directionality of the hydrogen bonding is the tendency to be located on or near the lone-pair orbital axis of the acceptor. The strength of the hydrogen bond increases by increasing the linearity of the angle (D–H \cdots A) [95, 96]. The importance of the directionality is related to the strength (and nature) of the hydrogen bond. The very strong hydrogen bonds are very linear to enable the sharing of the electron pair. On the other hand, the hydrogen bonds with weak to moderate strength are more flexible regarding the directionality.

The enthalpy values reported in the literature for hydrogen bonding range from -0.5 kcal.M^{-1} to -37 kcal.M^{-1} [95, 97, 98]. The hydrogen bonds are usually classified into three types according to the strength of the bond [99]: strong ($< -15 \text{ kcal.M}^{-1}$), moderate (-4 to -15 kcal.M^{-1}) and weak ($> -4 \text{ kcal.M}^{-1}$). Most hydrogen bond energies between small molecules range from -3 to -10 kcal.M^{-1} [100]. Also the hydrogen bonds in the biological world are mostly of weak to moderate strength ($3\text{-}9 \text{ kcal.M}^{-1}$). However, this range of strength is most suitable for biological processes, since these hydrogen bonds are stable enough to provide significant binding energies, but sufficiently weak to allow rapid dissociation [8].

3.3.3.1 The nature of the hydrogen bonding

Giving a brief description of the physical nature of the hydrogen bonding is important for the following discussion about its treatment in force fields. Hydrogen bonding is not a simple interaction. The attractive force of the hydrogen bonding is the result of many different interactions where many components contribute to the hydrogen bonding energy [95, 101, 102]. There are at least four component interaction types in hydrogen bonding [101, 102]: electrostatics ES, polarization PL, van der Waals (dispersion DISP/ electron exchange repulsion EX), and covalency (charge transfer CT). These components provide to the hydrogen bonding a bimodal nature of a partly non-covalent and partly covalent bond. Charge transfer describes the transfer of electrons from an occupied orbital of one molecule to the unoccupied orbitals of the other and this makes the hydrogen bond similar to a covalent bond [102].

Generally, the hydrogen bonding interaction was considered mainly electrostatic and sometimes covalent [102]. The calculations of Coulson and Danielsson showed that the long hydrogen bond is essentially electrostatic with a covalent contribution of only a few percent [95, 103]. Morokuma and Umeyama [101] showed that the contribution of the energy components are strongly distance dependent. At a relatively small separation, electrostatics ES, CT, and PL can all be important attractive components, competing against a large EX repulsion. At longer distances for the same complex the short-range attractions CT and PL are usually unimportant and ES is the only important attraction. On the other hand, the very strong hydrogen bonds (energy $< -20 \text{ kcal mol}^{-1}$) have a quasi-covalent nature with a large charge transfer contribution [102]. In general, the magnitudes of the component contributions to the total energy of the hydrogen bonding is reflected in the strength of the bond (Figure 7) [99]: the very strong H-bond is strongly covalent with strong directionality, the moderate strength is mostly electrostatic with moderate directionality and the weak H-bond is a combination of electrostatics/dispersion with weak directionality.

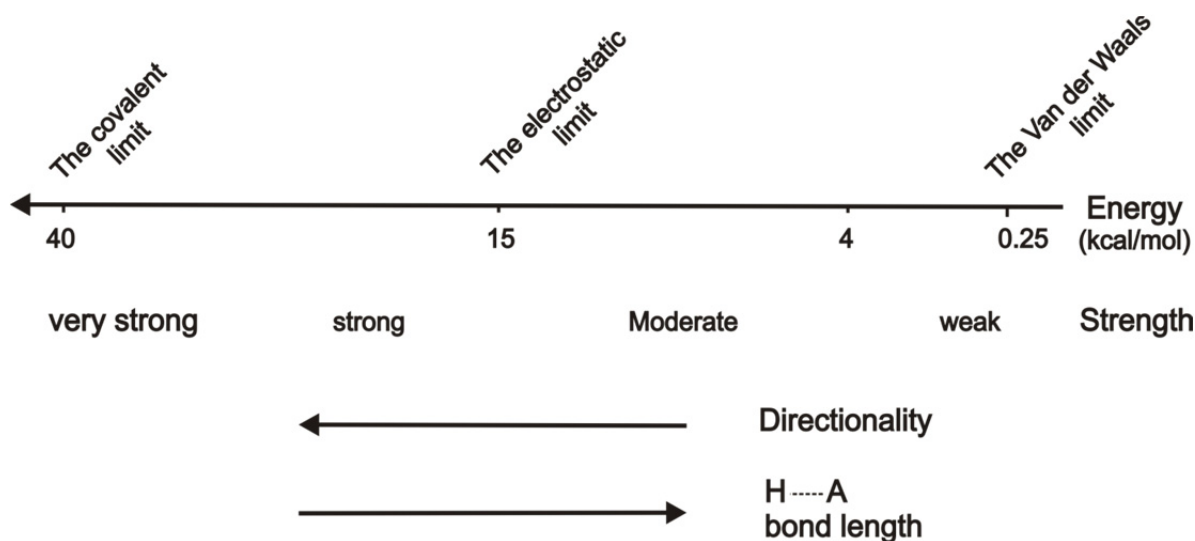


Figure 7. The nature of the hydrogen bonding.

The range of the hydrogen bonding strength is widely spread depending on the nature of the bond and the relative contributions. The very strong hydrogen bonding is mainly due to the covalent bonding. The moderate to strong hydrogen bonding (up to 15 kcal.M^{-1}) is dominated by the electrostatic interaction. The very weak hydrogen bonding is mainly due to the van der Waals interactions.

3.3.4 The van der Waals interactions

The van der Waals interactions are the shortest-ranged interactions. They are active when the atoms are very close. As the separation is reduced, the energy decreases, passing through a minimum. The energy then rapidly increases as the separation decreases. The van der Waals interactions are of complex nature. They consist of the attractive dispersive interactions (London forces) and the repulsive exchange forces. The dispersive interactions are due to instantaneous dipoles which arise from the fluctuations in the electron clouds which in turn induce a dipole in neighboring atoms. The attractive dispersive interaction varies as $1/r^6$. The repulsive exchange forces are of shorter range than the exchange forces. The exchange forces have a quantum mechanical origin to prevent the overlap between the atoms because any two electrons in a system are prohibited to have the same set of quantum numbers.

Various functions were used to model the van der Waals interactions in empirical force fields such as Buckingham potential, Morse potential and Lennard-Jones potential. The 6-12 Lennard-Jones potential is used in many important force fields like OPLSAA [104]:

$$U_{vdw} = 4\varepsilon \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right) \quad (26)$$

The collision diameter σ is the separation at which the energy is zero. The well depth ε is the minimum of the energy. Although the attractive van der Waals energies are small, they can make significant contributions to the binding energy when they are summed over a molecule. For example, a methylene group in a hydrocarbon contributes about 2 kcal.M^{-1} . Thus they can be very important for large hydrocarbons. The van der Waals forces are of central importance between the forces at the macroscopic scale [76].

3.4 Modeling of the bimolecular interactions

Three major categories of force fields are usually used in the field of biomolecular modeling [105, 106]: (i) statistical effective energy functions (EEEE) such the one in Rosetta [107] where the energies are derived from the frequency of contacts between atoms or residues in the protein database and; (ii) empirical effective energy functions (EEEE) such as FOLDX [105] that combine a physical description of the interactions with lessons learned from experiments to derive an energy function that can reproduce the experimental free binding energy; Both the statistical and the empirical energy functions are widely used in the field of protein design because of their cheap computational cost; (iii) physical effective energy functions which are the well-known molecular mechanics (MM) force fields [104, 108-111].

3.4.1 Molecular mechanics

The molecular mechanics force-fields are based on the Born-Oppenheimer Approximation [112] to the Schrödinger equation where the motions of the electrons are separated from the motions of the nuclei. This is generally a good approximation because the nuclei—much heavier than the electrons—will move only very little on the fast timescale of electronic vibration [113]. Then the potential energy is expressed with respect to the nuclei positions. The potential energy of a molecule is constructed as the sum of the different empirical additive terms accounting for the bonding and non-bonding interactions:

$$\begin{aligned} U = & \sum_{bonds(ij)} \frac{k_{ij}}{2} (b_{ij} - b_{0ij})^2 + \\ & \sum_{angles(ijl)} \frac{k_{ijl}}{2} (\phi_{ijl} - \phi_{0ijl})^2 + \\ & \sum_{torsion(ijlk)} \frac{k_{ijlk}}{2} [1 + \cos(n\phi_{ijlk} - \phi_{0ijlk})] + \\ & \sum_{nonbonded} \left(\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \end{aligned} \quad (27)$$

The potential functions are different from a force field to another and are balanced inside the same force field to reproduce experimental data or data from QM calculations. The possible treatment of hydrogen bonding interaction is explained in the following in more detail because the explicit hydrogen bonding energy is used to quantify the water mediation between the proteins in chapter 6 (see 6.3).

3.4.1.1 Treatment of the hydrogen bonding in molecular mechanics force fields

Most of the recent force fields account for the non-covalent interactions by a combination of electrostatic (coulomb) and van der Waals terms. The contribution of the hydrogen bonding is implicitly included in these two terms. Accounting for the polarization effect in an explicit way is possible when a polarizable force field is used. However, the covalent component of the hydrogen bonding interaction due to charge transfer is not included in the nonbonding terms. This covalent part is directional and a directional explicit hydrogen-bond term is needed to account for it [114].

Most of the discussion about the proper treatment of the hydrogen bonding interaction in molecular mechanics force fields took place in the early stages of force field development. In protein force fields, the general tendency was to cancel the explicit hydrogen bonding term and to include it implicitly in the electrostatic (coulomb) and Van der Waals terms. For example, Hagler and Lifson [115] found that the explicit hydrogen bond function was unnecessary. Kollman first included an explicit hydrogen bonding term in his early version of AMBER [110] but this term was turned off later. Similarly, the early version of CHARMM [108] included an explicitly hydrogen bonding term which was turned off later [116]. However, several force fields that are in use today in the fields of organic chemistry and in medicinal chemistry are still using a directional explicit term for the hydrogen bonding such as the MM2 [117], MM3 [118], MM4 [114], Yeti [119] and Grid [120] force fields.

Does the explicit hydrogen bonding term improve the force field? In general, it is not easy to judge whether the explicit treatment of the hydrogen bonding improves force fields. As mentioned above, the explicit term was turned off in many force fields in later years. In a comparative study of the reliability of many different force fields regarding the modeling of hydrogen bonding and π stacking interactions, Paton and Goodman [121] found that the OPLSAA [104] and MMFF94 force fields [122] best reproduced the ab initio energies and geometries for the S22 and JSCH2005 data sets [123]. Both of these force fields, which do not have explicit term for the hydrogen bonding, were superior to the force fields with explicit hydrogen bonding terms such as the MM2 [117] and MM3 [118] force fields. Using the explicit term seems most advantageous for molecules where charge transfer makes a large contribution to the hydrogen bonding energy. This is rarely the case for hydrogen bonds of water molecules or proteins. Indeed, the electrostatic character is dominant for moderate strength bonds such N–H \cdots O, O–H \cdots O and O–H \cdots N bonds [102] which are the usual pattern in proteins and water. Thus we expect that the modern force fields give reasonable results regarding hydrogen bonding in the solvated protein systems studied here.

3.4.1.1.1 Examples of the explicit hydrogen bonding terms in molecular mechanics

The program Xplor-NIH [124] still uses an explicit term for the hydrogen bonding by using a directional Lennard-Jones potential:

$$E_{HB} = \left(\frac{C}{r_{DA}^i} - \frac{D}{r_{DA}^j} \right) \times \cos^m(\theta_{D-H\cdots A}) \times \cos^n(\theta_{AA-A\cdots H}) \times sw(r_{DA}^2, r_{on}^2, r_{off}^2) \times sw[\cos^2(\theta_{D-H\cdots A}), \cos^2(\theta_{h_{on}}), \cos^2(\theta_{h_{off}})] \quad (28)$$

where r_{DA} is the donor-acceptor distance, atoms AA, is the antecedent atom, and I, j, n and m are positive integers. sw is the usual switching function that is used for the Van der Waals in the early version of CHARRM [108] to maintain a continuous and smooth function to the cutoff point. However, this potential is taken from the early versions of CHARMM [108] where the hydrogen bonding was explicitly included in the potential energy and the second angular term was turned off in that early version n=0.

The grid force field [120] used a directional 6-4 CHARMM-like potential to account for the hydrogen bond energy with an angular term for the angle "D – H \cdots A" (see Figure 8).

$$E_{HB} = \left(\frac{C}{r_{DA}^6} - \frac{D}{r_{DA}^4} \right) \times \cos^m(\theta_{D-H\cdots A}) \quad (29)$$

m is usually 4.

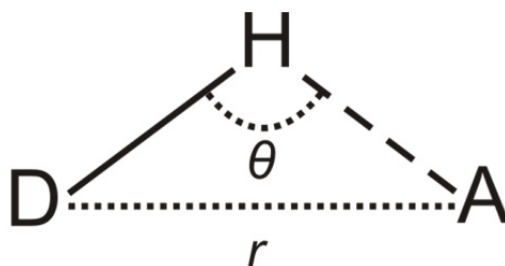


Figure 8. Definition of hydrogen-bond geometry used in Grid and XPLOR force field.

Since hydrogen bonding is due to the interaction of a hydrogen atom with a lone-pair, few force fields included an additional term to account for the relative position (orientation) of the lone-pair that is hydrogen bonded. Examples are the MM4 [114] and the Yeti [119] force fields. The Yeti force field [119] used a potential function for hydrogen bonds (and salt bridges) similar to the previous function with an additional angular dependency of the angular deviation of the H-bond from the closest lone-pair direction at the accepting atom (angle H-A-LP in Figure 9):

$$E_{HB} = \left(\frac{C}{r_{H\dots A}^{12}} - \frac{D}{r_{H\dots A}^6} \right) \times \cos^2(\theta_{D-H\dots A}) \times \cos^4(\omega_{H\dots A-LP}) \quad (30)$$

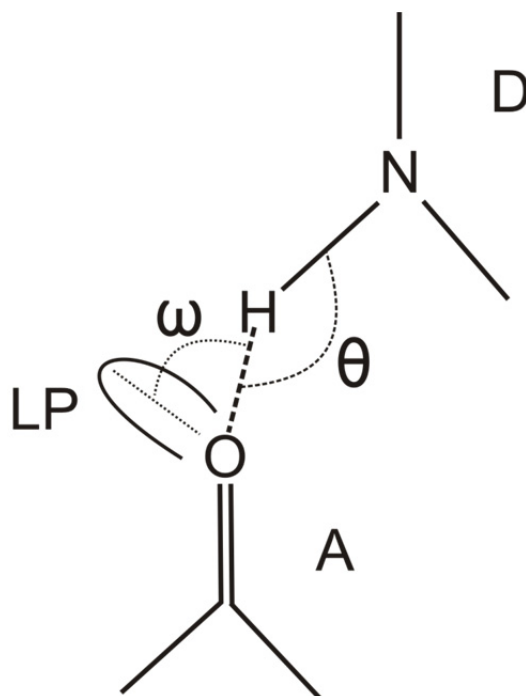


Figure 9. Definition of the hydrogen-bond geometry used in the YETI force field. D is the donating atom, A is the accepting atom, LP is the lone electron pair.

3.4.2 Molecular dynamics

The classical molecular dynamics simulation (MD) is based on Newton's equations of motion

$$\begin{aligned}\frac{d^2\mathbf{r}_i}{dt^2} &= \frac{\mathbf{F}_i}{m_i} \\ \frac{d\mathbf{r}_i}{dt} &= \mathbf{v}_i\end{aligned}\tag{31}$$

Here \mathbf{r}_i and \mathbf{v}_i are the position and the velocity of atom i , $\mathbf{F}_i = \frac{\partial U_i}{\partial \mathbf{r}_i}$ is the force on the atom i .

The Born-Oppenheimer approximation [112] enables us to neglect the electronic motions and only consider the nuclear motions. Thus, quantum effects are generally ignored. The classical MD is suitable for a wide range of problems such as modeling the conformational changes of inorganic and organic molecules and their non-covalent interactions. However, a quantum mechanics (QM) or hybrid (QM/MM) treatment is needed when we want to study important electronic changes of molecular systems such as bond formation.

The equations of motion are solved numerically. This is normally done using an efficient method like the leap-frog algorithm [125]. The forces $\mathbf{F}_i(t)$ at time t are computed from the coordinates at time t using the force field potential. The leap-frog algorithm uses forces $\mathbf{F}_i(t)$ and positions $\mathbf{r}_i(t)$ at time t and velocities $\mathbf{v}_i(t - \frac{\Delta t}{2})$ at time $t - \frac{\Delta t}{2}$; then it updates positions and velocities:

$$\begin{aligned}\mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) &= \mathbf{v}_i\left(t - \frac{\Delta t}{2}\right) + \frac{\Delta t}{m_i}\mathbf{F}_i(t) \\ \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \Delta t \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right)\end{aligned}\tag{32}$$

The initial coordinates are usually obtained from experimental structures, mainly from X-ray crystallography or NMR solution structures. However, the only relevant information available about atomic velocities is the system's temperature T , which determines the velocity distribution. Thus, initial velocities are usually randomly assigned from the standard Maxwellian velocity distribution at a temperature T .

The size of the time step Δt determines the magnitude of the error associated with the integration algorithms. An appropriate time step should be small in comparison to the period of the motions of highest frequency (the fastest motions) in the simulated system. A good rule of thumb for selecting Δt is $\frac{\tau}{\Delta t} \approx 20$ [126] where τ is the period of the fastest motion in the system. The fastest motions in proteins are the stretching vibrations of the bonds connecting hydrogen atoms to heavy atoms (X-H stretching). The frequency of these motions is in the order of 3000 cm^{-1} , implying a time step of $\Delta t = 0.5 \text{ fs}$. On the other hand, the stretching modes between heavy atoms X-X have frequencies around 1500 cm^{-1} implying a time step $\Delta t = 1 \text{ fs}$. Constraint algorithms such as LINCS [127] and SHAKE [128] are usually used to remove the fast (high frequency) motions from the numerical integration to allow for larger time steps to get longer simulations.

4 Dielectric permittivity in biological systems

This chapter provides a general introduction into the theory of electric polarization in solution and in biological system. The derivation of formulas to compute the dielectric permittivity is presented in this chapter. These formulas were used to study the permittivity of the water molecules in the interfacial gaps between two interacting proteins (chapter 6). The derivation of the fluctuation formula (4.3) was published in *Nature Communications* in 2011 [2].

4.1 Background: electric polarization

The main task for the theoretical description of the dielectric permittivity in modern times is finding a macroscopic continuum description of the electrostatic interactions on the basis of the structural information and the dynamic behavior at the atomistic scale. Remarkably, the concepts of the electrostatic theory were developed in the 19th century based on the continuum nature of matter before the atomistic view of matter was understood. However, this macroscopic approach is not applicable to the microscopic level. The microscopic theory of electric polarization is mainly due to Böttcher. In the following, I will follow his approach and notation. As the concept of the electric field in the electric permittivity field is often discussed in a wrong way, I decided to present a detailed background that is needed to understand and judge the derivation presented here.

In the following, the Gaussian (CGS) system of units will be used for consistency with most of the scientific literature. Care should be taken when using the equations with the (SI) units where the equations should be modified to include the permittivity of free space ϵ_0 (see appendix C in Griffiths [129]). In general the ϵ_0 in SI system of units is substituted by $\frac{1}{4\pi}$ in the Gaussian units ($\epsilon_0 \rightarrow \frac{1}{4\pi}$). To simplify the notation in many equations, I will use the vectorial components in the direction of a unit vector \mathbf{e} .

4.1.1 Localized charge distribution: the electric dipole moment

The electric field at a point in vacuum is related to the charge distribution through the Gauss' law. The main electrodynamics relations of the electric field in vacuum are:

$$\begin{aligned} \text{Gauss's law} \quad \nabla \cdot \mathbf{E} &= 4\pi\rho \\ \nabla \times \mathbf{E} &= 0 \\ \mathbf{E} &= -\nabla\Phi \end{aligned} \tag{33}$$

Here \mathbf{E} is the electric field, Φ is the electric potential and ρ is the charge distribution.

In biochemical systems, the electrostatic interactions are much more complex than the simple form of the Coulomb's law that uses charge points. The electrostatic interaction in the real world takes place between charge distributions (such as the electronic cloud around atoms or molecules). The electric field and the electric potential of a localized

charge distribution ρ at point \mathbf{r} (away from the distribution) are defined by the integration over the volume elements $d\tau'$ of volume of the charge density [129]:

$$\text{Coulomb's law} \quad \mathbf{E}(\mathbf{r}) = \int_v \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^2} \hat{\mathbf{r}} d\tau' \quad (34)$$

$$\Phi(\mathbf{r}) = \int_v \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\tau' \quad (35)$$

The unit vector $\hat{\mathbf{r}} = \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}$ points from the volume element $d\tau'$ to the point \mathbf{r} (see Figure 10a). The charge element dq in the volume element $d\tau'$ is $dq = \rho(\mathbf{r}') d\tau'$.

To simplify the electrostatic interaction, a charge distribution (of a net charge Q) can be represented through electric multipoles. The usage of the electric multipoles is an approximation of the charge distribution by using the binomial expansion of the potential in powers of $1/r$ (see Figure 10b):

$$\Phi(\mathbf{r}) = \frac{1}{r} \int_v dq + \frac{1}{r^2} \hat{\mathbf{r}} \cdot \int_v \mathbf{r}' dq + \frac{1}{r^3} \int_v \left[\frac{3(\hat{\mathbf{r}} \cdot \mathbf{r}')^2 - r'^2}{2} \right] dq + \dots \quad (36)$$

The first term of the expansion is the **monopole**. The **monopole** is the common point charge and it contains the total net charge in the distribution $Q = \int dq$. The electric **monopole** is not related to the geometry (size, shape, and density) of the charge distribution. The potential of the monopole ($\Phi_{monopole}(\mathbf{r}) = Q/r$) is the dominant term in the potential as it goes like $1/r$. The second term is the dipole moment:

$$\mathbf{p} = \int_v \mathbf{r}' dq$$

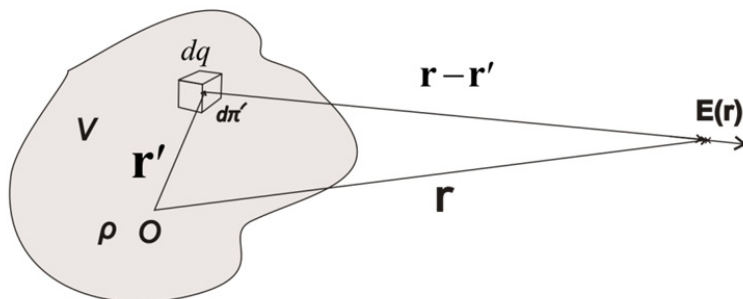
In contrast to the monopole, the dipole moment is determined by the geometry (size, shape, and density) of the charge distribution. When the total charge of the system is zero ($Q = 0$), the dipolar term is the dominant term in the expansion. For this reason, the uncharged molecules still interact with other charge distributions through the dipole moment and the higher terms.

The potential generated by the dipole $\Phi_{dipole}(\mathbf{r}) = \frac{\mathbf{p} \cdot \hat{\mathbf{r}}}{r^2}$ goes like $1/r^2$. The total dipole moment of a system of n point charges q_i is defined relatively to a fixed origin as:

$$\mathbf{M} = \sum_1^n q_i \mathbf{r}_i \quad (37)$$

where \mathbf{r}_i is the vector from the origin to q_i . The dipole moment is directed from the negative charge to the positive charge centers. \mathbf{M} is independent of the choice of the origin if the net charge of the system is zero, but not otherwise.

a) Calculation of the electric field and potential from a localized charge distribution



b) Multipole expansion of the charge charge distribution

ρ	=	+	+	-	+	-	+	-	+
Charge density		Monopole		Dipole		Quadrupole				
		$\Phi \sim \frac{1}{r}$		$\Phi \sim \frac{1}{r^2}$		$\Phi \sim \frac{1}{r^3}$				
$\Phi(\mathbf{r})$	=	$\frac{1}{r} \int_v dq$	+	$\frac{1}{r^2} \hat{\mathbf{r}} \cdot \int_v \mathbf{r}' dq$	+	$\frac{1}{r^3} \int_v \left[\frac{3(\hat{\mathbf{r}} \cdot \mathbf{r}')^2 - r'^2}{2} \right] dq$	+		

Figure 10. Localized charge distribution.

(a) Calculation of the electric field $\mathbf{E}(\mathbf{r})$ and the electric potential of a Localized charge distribution ρ at point \mathbf{r} . The Coulomb's law is used to compute the electric field that is generated due to charges in the volume element $d\tau'$ which contains a charge $dq = \rho(\mathbf{r}') d\tau'$. The field is a superposition of fields from all the volume elements of the charge density. O is the origin. (b) Multipole expansion of the localized charge density into monopole, dipole and higher terms. The electric monopole is the only term that is not related to the geometry (size, shape, and density) of the charge distribution.

Dipole moments are vectors, and they add accordingly $\mathbf{M} = \sum_{i=1}^n \mathbf{p}_i$. The density of the total dipole moment is called polarization:

$$\mathbf{P} = \frac{\langle \mathbf{M} \rangle}{V} \quad (38)$$

Many molecules in nature have permanent dipole moments such as water molecules (Figure 11b) and they are called **polar molecules**. The polarization due to the

inhomogeneous distribution of electrons is called the orientation polarization \mathbf{P}_μ . External electric fields reorient the dipoles of the polar molecules in the same direction as the field. Thus the field that is generated by the dipoles will counteract the external field and reduce the potential energy of the dipole in the field. The potential energy of the dipole in a field is given by the relation:

$$U = -\boldsymbol{\mu} \cdot \mathbf{E} \quad (39)$$

The molecules that do not have permanent dipoles are called **nonpolar molecules**. In these molecules, induced dipoles may be created by applying an electric field (Figure 11c) due to the change in the charge distribution (polarizability). This type of polarization is called **the induced polarization \mathbf{P}_α** . The induced dipole is related to its source, the external electric field:

$$\mathbf{p} = \alpha \mathbf{E}_0 \quad (40)$$

Here α is called the (scalar) polarizability of the particle. The total polarization of a mixture of particles of k different types is simply:

$$\begin{aligned} \mathbf{P}_\mu &= \sum_k N_k \bar{\boldsymbol{\mu}}_k \\ \mathbf{P}_\alpha &= \sum_k N_k \alpha_k (\mathbf{E}_i)_k \end{aligned} \quad (41)$$

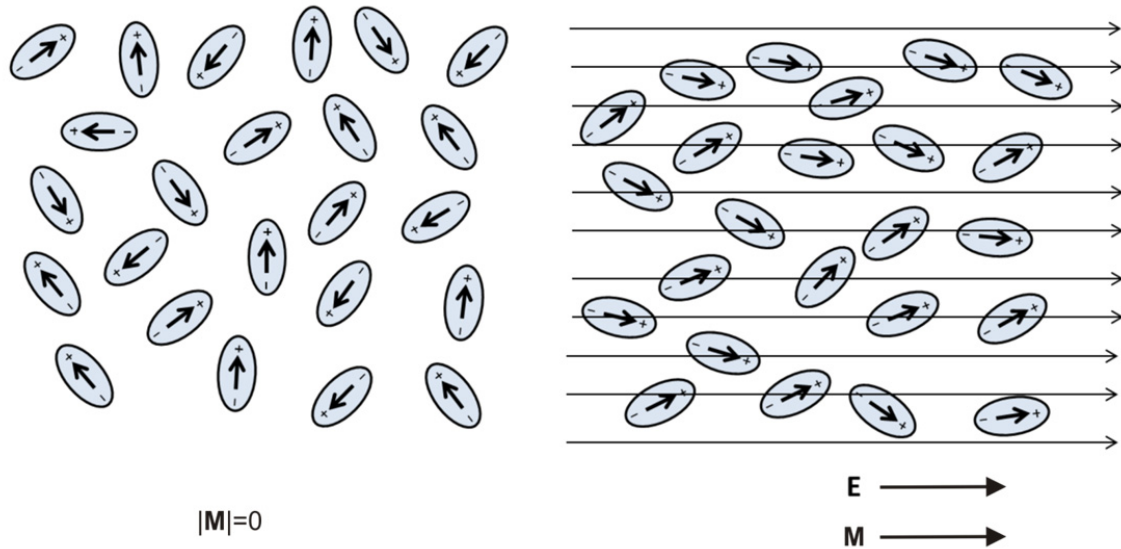
where N_k is the number of particles of type k per volume unit, the index k refers to the k -th sort of particles, $\bar{\boldsymbol{\mu}}$ is the value of the permanent dipole vector averaged over all orientations, α_k is the scalar polarizability of a particle, \mathbf{E}_i (the internal field, see later) is the average field strength acting upon that particle. The index k refers to the k -th kind of particle. The total polarization is the sum of the orientation polarization and the induced polarization:

$$\mathbf{P} = \mathbf{P}_\alpha + \mathbf{P}_\mu \quad (42)$$

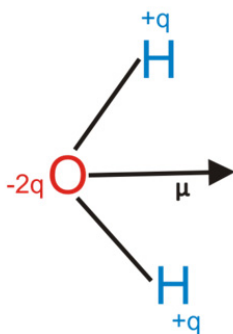
4.1.2 The electric field in in polarizable matter: the macroscopic electric field

The equation that defines the electric field in vacuum (Gauss' law eq. 33) is not applicable inside a dielectric where the source charges are located inside the matter. This problem arises due to the microscopic nature of the matter. The classical approach of Maxwell for the electric field inside the matter was introduced before the details about the atomistic structure of the matter became known. The macroscopic approach of Maxwell pictured matter as a continuum in which virtual cavities were created. Inside these cavities it would be possible to use the original definition of the electric field by introducing a new vector field \mathbf{D} (the displacement field) in such a way that for this field the Gauss' law will be valid [130].

a) Reorientation of the permanent dipoles in the electric field



b) Permanent dipole moment of the water molecule



c) The induced polarization

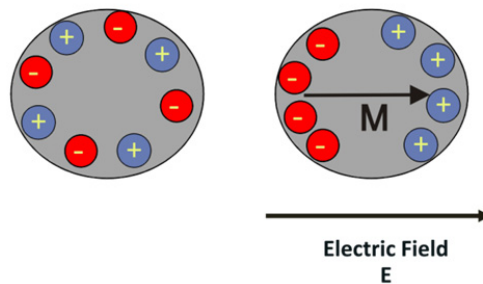


Figure 11. The types of polarizations

(a) Reorientation of the permanent dipoles due to an external electric field. The permanent dipoles are randomly oriented before applying the electric field. Thus the total dipole moment is zero. After applying an external electric field, the dipoles tend to orient in the same direction as the electric field creating a non-zero total dipole moment M . The new dipole M generates an electric field in the opposite direction to the dipole and the external electric field thus counteracts the external electric field. (b) The permanent dipole of the water molecule. (c) Induced dipole by an electric field in a nonpolar molecule. The nonpolar molecule has no dipole when no external electric field is applied. The electric field induces a dipole because the charges (electrons) redistribute in order to minimize the total energy of the system.

$$\nabla \cdot \mathbf{D} = 4\pi\rho \quad (43)$$

The displacement field \mathbf{D} is the field in the volume (cavity) in the absence of the polarized matter. The relation between the displacement field \mathbf{D} and Maxwell's \mathbf{E} is related to the polarization \mathbf{P} :

$$\mathbf{D} - \mathbf{E} = 4\pi\mathbf{P} \quad (44)$$

The Polarization \mathbf{P} is the density of the total dipole moment in the matter ($\mathbf{P} = \frac{\langle \mathbf{M} \rangle}{V}$) and depends on the electric strength \mathbf{E} which polarizes the dielectric.

$$\mathbf{P} = \chi_e \mathbf{E} \quad (45)$$

χ_e is the electric susceptibility of the dielectric. The polarization has the same direction as \mathbf{E} , so the field generated by the polarized matter has opposite direction to \mathbf{E} and works to counteract the macroscopic field \mathbf{E} . The dielectric constant is the factor by which the polarized matter reduces the field inside it.

$$\mathbf{D} = \varepsilon \mathbf{E} \quad (46)$$

The microscopic approach of the electric field in the matter was introduced after the knowledge of the atomistic structure of the matter was discovered by Lorentz [131], Rosenfeld [132], Mazur [133] and de Groot [134]. The microscopic approach pictures the matter as point charges in vacuum and application of the Coulomb' law (eq. 34) assigns a so-called **microscopic field** ξ at every point inside the matter. The **macroscopic electric field** \mathbf{E} (**Maxwell's field**) is defined as the average of the local microscopic electric field ξ taken over the volume elements of the sample [132-137].

$$\mathbf{E} \cdot \mathbf{e} = \frac{1}{V} \int_V (\xi \cdot \mathbf{e}) dv = \langle \xi \cdot \mathbf{e} \rangle_V \quad (47)$$

The main difficulty for computing of the macroscopic field \mathbf{E} directly from the microscopic field is the rapid fluctuations in the microscopic field. Lorentz and Rosenfeld averaged the microscopic field over small intervals of space and time to get rid of the rapid fluctuations [130, 131, 138]. Mazur and de Groot used the methods of statistical mechanics to pass from a detailed to a global description [130, 133, 134]. The local microscopic field fluctuates particularly heavily at positions close to the point charges (singularity problem). Theoretically, this singularity is not a problem in the calculation of the macroscopic electric field as it is calculated over the volume occupied by many molecules and the singularity effect is canceled by symmetry. Practically, the rapid fluctuations in the microscopic field are the main difficulty when applying this approach to get the macroscopic electric field \mathbf{E} from atomistic simulations. Getting a stable value requires a sample of matter as large as needed to represent matter as a continuum which is not practical at all (see the discussion in section 4.2.3).

4.1.2.1 The interaction between a molecule and the component of the macroscopic electric field

The macroscopic electric field at the volume occupied by a molecule is the superposition of the field generated by all other molecules ("the internal field \mathbf{E}_i ") and the field generated by the molecule itself ("the self-field \mathbf{E}_{self} ") [130]:

$$\mathbf{E} \cdot \mathbf{e} = \mathbf{E}_i \cdot \mathbf{e} + \mathbf{E}_{self} \cdot \mathbf{e} \quad (48)$$

The **self-field** \mathbf{E}_{self} is the field that is generated by the dipole in the volume that it occupies. The self-field \mathbf{E}_{self} for a spherical shape dipole (and for any charge distribution in a spherical cavity) is given by the relation [136, 139]:

$$\mathbf{E}_{self} \cdot \mathbf{e} = -\frac{4\pi}{3} \mathbf{P} \cdot \mathbf{e} \quad (49)$$

\mathbf{P} is the polarization in the volume that is occupied by the dipole. The self-field does not affect the direction of the dipole itself and does not induce polarization in the dipole itself.

The **internal field** \mathbf{E}_i is the field that is generated by the surrounding medium around the dipole (all other dipoles and charges except the dipole itself). The **internal field** is responsible for the induced polarization of the polarizable molecules (eq. 40). The internal field can be divided into two parts regarding its effect on the orientation of the dipole itself. Onsager showed that only a part of the internal field influences the orientation of the dipole in the electrical field which was called by Böttcher the **directing field** \mathbf{E}_d [130]. The rest of the internal field was termed the **Onsager reaction field** \mathbf{R} [140]. This reaction field was defined as the field at the position of the molecule that is generated by the surrounding as a response to the field of the molecular dipole of the molecule. Interestingly, this reaction field is oriented always parallel to the dipole of the molecule and thus does not affect its orientation.

The Onsager reaction field in the case of a spherical molecule is given by the relation [130, 140]:

$$\mathbf{R} \cdot \mathbf{e} = \frac{2(\varepsilon_s - 1)}{2\varepsilon_s + 1} \frac{4\pi}{3} \mathbf{P} \cdot \mathbf{e} \quad (50)$$

ε_s is the dielectric constant of the surrounding medium.

Before Onsager [140], Debye considered the internal field to be responsible for the orientation polarization. Thus the Debye relation (**Table 1**) is valid in the case of diluted systems when the reaction field is not strong.

The directing field is the field that is generated by the surrounding medium (all other charges and dipoles) in the cavity of the molecule excluding the field that is generated as a reaction of the medium to the molecule itself. The physical meaning of the directing field \mathbf{E}_d was explained by Böttcher by removing the permanent dipole of the molecule in the following way [130]: *“remove the permanent dipole of a molecule without changing its polarizability; let the surrounding dielectric adapt itself to the new situation; then fix the charge distribution of the surroundings and remove the central molecule. The average field in the cavity so obtained is equal to the value of \mathbf{E}_d ”*.

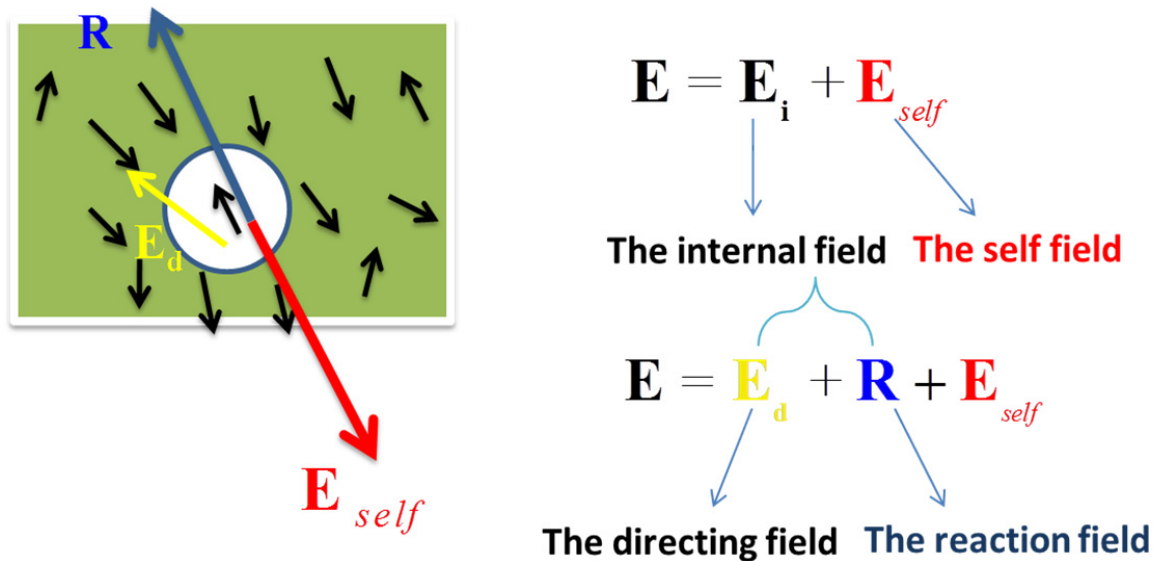


Figure 12. The component of the macroscopic electric field in the volume that is occupied by a permanent dipole.

The self-field \mathbf{E}_{self} (red) is generated by the dipole itself and is points into the opposite direction of the dipole.

The internal field \mathbf{E}_i is the field that is produced by the surrounding medium around the dipole. The internal field has two components: (1) the reaction field \mathbf{R} (the blue arrow) which is the field at the position of the molecule that is generated by the surrounding as a response to the field of the molecular dipole (The self-field) (2) the directing field \mathbf{E}_d (yellow) is the field due to the surrounding medium which is really external to the dipole and not related to any reaction of the dipole with the medium. Both the self-field and the reaction field are oriented always parallel to the dipole. Thus they cannot affect its direction. Only the directing field is responsible for reorienting the dipole.

Using this new nomenclature, we rewrite equation (eq. 48) in the following way:

$$\mathbf{E} \cdot \mathbf{e} = \mathbf{E}_d \cdot \mathbf{e} + \mathbf{R} \cdot \mathbf{e} + \mathbf{E}_{self} \cdot \mathbf{e} \quad (51)$$

4.1.3 Langevin equation for the orientation polarization

When no external electric field is applied, the molecules with permanent dipole moments move and rotate according to their kinetic energy but the averaged value of each permanent dipole vector over all orientations μ is zero as well as the average value of the summed permanent dipole moments. Upon application of an electric field, the electric field will influence the direction of the dipoles and create an orientation polarization (Figure 11a). The energy of the dipole is related to the directing field through the relation:

$$W = -\boldsymbol{\mu} \cdot \mathbf{E}_d = -\mu E_d \cos \theta \quad (52)$$

where θ is the angle between the directions of \mathbf{E}_d and $\boldsymbol{\mu}$. The average orientation of the dipole relatively to the directing field can be calculated according to the Boltzmann

distribution law taking into account that W is the only part of the energy which depends on the orientation of the dipole:

$$\langle \cos \theta \rangle = \frac{\int_0^\pi \cos \theta e^{-\frac{\mu E_d \cos \theta}{kT}} \frac{1}{2} \sin \theta d\theta}{\int_0^\pi e^{-\frac{\mu E_d \cos \theta}{kT}} \frac{1}{2} \sin \theta d\theta} = \frac{1}{a} \frac{\int_{-a}^a e^x dx}{\int_{-a}^a e^x dx} = \frac{e^a + e^{-a}}{e^a - e^{-a}} - \frac{1}{a}$$

$$= L(a) \tag{53}$$

where $a = \frac{\mu E_d}{kT}$

and $x = \frac{\mu E_d \cos \theta}{kT}$

$L(a)$ is termed the “**Langevin function**”. The Langevin function has two limiting values -1 and 1. The behavior of the Langevin function is linear for low values of a . The function shows how the orientation polarization saturates at strong values of the electric field. The function can be written as series in a :

$$L(a) = \frac{1}{3} a - \frac{1}{45} a^3 + \frac{2}{945} a^5 \dots\dots\dots$$

For small values of a , the Langevin function is linear (see Figure 13).

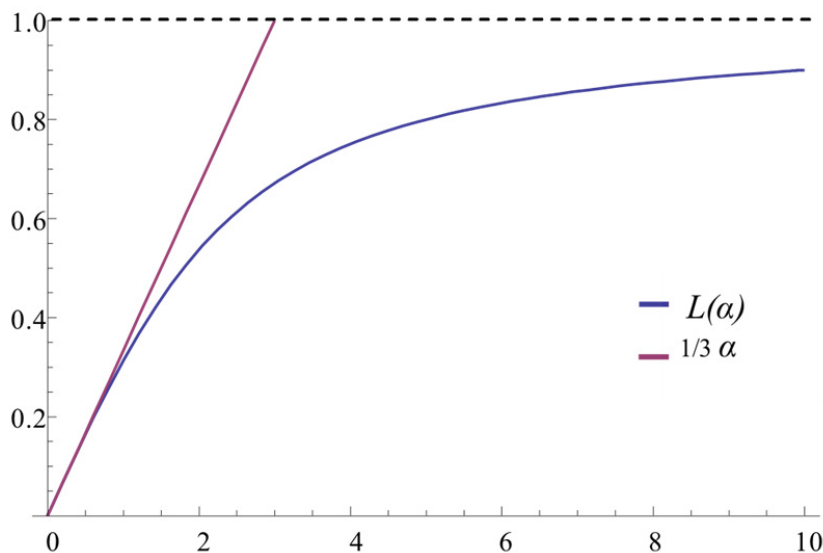


Figure 13. The Langevin function.

The Langevin function has a limiting value of one. For small values of a , the Langevin function can be linearized to $a/3$. Thus for small values of a (i.e. in the case of electric fields), we take the first linear term:

$$\langle \cos \theta \rangle = \frac{\mu E_d}{3kT}$$

and the average dipole value:

$$\boldsymbol{\mu} = \mu \langle \cos \theta \rangle = \frac{\mu^2}{3kT} \mathbf{E}_d$$

Then the orientation polarization is:

$$\mathbf{P}_\mu = N\boldsymbol{\mu} = N \frac{\mu^2}{3kT} \mathbf{E}_d \quad (54)$$

N is the number of dipoles in the volume unit. The Langevin function for the orientation polarization is the basis of the Langevin Dipoles Model for the calculation of the solvation energies [135]. This equation will be used later to check the possibility of having a saturation in the dielectric due to a possible high electric field.

4.2 The electric permittivity

4.2.1 The constitutive electrostatic relation

The macroscopic dielectric permittivity $\boldsymbol{\varepsilon}$ of a sample of matter is generally defined by the **constitutive electrostatic relation** (see equations 44-46):

$$\mathbf{P} = \frac{(\boldsymbol{\varepsilon} - 1)}{4\pi} \mathbf{E} = \chi_e \mathbf{E} \quad (55)$$

Here, the two vectors \mathbf{P} and \mathbf{E} are the average polarization and the average macroscopic electric field in the sample. $\boldsymbol{\varepsilon}$ and χ_e are the dielectric permittivity and the electric susceptibility. The macroscopic electric field \mathbf{E} (Maxwell's field) is defined as the average of the local microscopic electric field $\boldsymbol{\xi}$ taken over the volume elements of the sample [135, 137, 141] (eq. 47).

In general, the polarization \mathbf{P} also has terms in higher powers of \mathbf{E} [130].

$$\mathbf{P} = \chi_e \mathbf{E} + \xi E^2 \mathbf{E} \quad (56)$$

These higher terms are important in the case of very high field intensities where the dielectric saturation and non-linear dielectric effects are observed. We will ignore such

non-linear effects in the work presented below, as the considered fields are of small to moderate strengths.

In isotropic samples, ϵ and χ_e are scalar quantities, whereas in non-isotropic dielectrics, like most solids, liquid crystals, the scalar ϵ and χ_e must be replaced by a tensor:

$$\mathbf{P} = \chi_e \cdot \mathbf{E} \quad (57)$$

Even in samples of arbitrary shape, the constitutive relation was found to be a valid description of the relationship between the local polarization and the local macroscopic electric field [142]. It has been argued that the size of the sample should be large enough to capture the macroscopic feature of the matter. The landmark work of Nienhuis and Deutch [142] showed that the constitutive relation can be applied locally between the local polarization and the local macroscopic electric field in a sample of an arbitrary shape. Unfortunately, it is hard to say what the smallest volume is where the constitutive relation still holds. The work of Hansen [143] showed that the dielectric constant is physically not defined at dimensions that are comparable to the dimension of molecular dipoles.

4.2.2 The static dielectric static permittivity

Before the computer simulation age, the static dielectric theory tried to establish a relation between the dielectric constant of a polar liquid and its molecular dipole moment. This was an attempt to bridge between the microscopic and the macroscopic properties. The starting point for all approaches is rewriting the constitutive relation [130] (eqs: 55, 40, 42 and 54):

$$\frac{\epsilon - 1}{4\pi} \mathbf{E} = \sum_k N_k \left[\alpha_k (\mathbf{E}_i)_k + \frac{\mu_k^2}{3kT} (\mathbf{E}_d)_k \right] \quad (58)$$

The foundation for this relation is actually the development of the understanding the electric field in matter. The main problem in the theory of the static permittivity are the relations between the macroscopic electric field \mathbf{E} , the internal field \mathbf{E}_i and the directing field \mathbf{E}_d (eq. 54). Table 1 contains the main relations used for calculating the dielectric constant. The Clausius-Mossotti relation is valid for non-polar systems that only show induced polarization. It was derived based on the idea of Lorentz by defining the internal field \mathbf{E}_i as the field inside a virtual spherical cavity in the continuum medium $\mathbf{E}_L = \frac{\epsilon+2}{3} \mathbf{E}$. Substituting this field in eq. 58 yields the established Clausius-Mossotti relation (see Table 1). This cavity field (due to Lorentz) is different from the physical cavity field that I will use later. Debye extended the Clausius-Mossotti equation to the case of polar systems by including the orientation polarization (eq. 54). The Debye equation is valid for diluted systems. The reason why his equation does not apply to polar systems of "normal" molecular concentrations is the fact that Debye did not distinguish between the internal field \mathbf{E}_i and the directing field \mathbf{E}_d as he considered the entire internal field to be responsible for the orientation of the dipole. Subsequently, Onsager [140] solved this problem by introducing the idea of the reaction field \mathbf{R} ($\mathbf{R} = \mathbf{E}_i - \mathbf{E}_d$) that does not affect the orientation

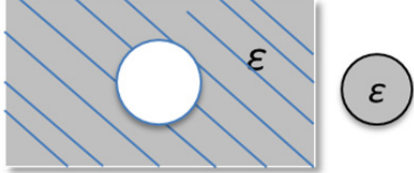
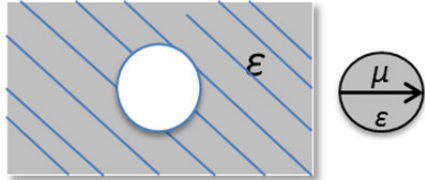
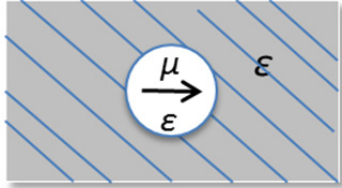

of the dipole (see text in 4.1.2.1). Onsager [140] used the continuum electrostatics relation with a simple model of a spherical molecule inside a continuum medium with the same dielectric constant. Despite the large advancement made by the Onsager relation, the relation gave large errors for associative liquids. This problem was solved later by the modification of Kirkwood [144] and Fröhlich [145] who related the dielectric constant to the fluctuations of the total dipole moment for a spherical sample of liquid.

4.2.3 Calculation of the static dielectric static permittivity from MD simulations

In principle, the constitutive relation (eq. 55) can be used to compute the dielectric constant from a molecular simulation because the polarization and microscopic electric field can be computed explicitly from the simulation trajectory. In practice, the computation of the total electric field is a formidable task due to the large fluctuations of the microscopic field. However, also the volume of the sample has to be large enough to get a smooth value of the macroscopic electric field [129, 132-134]. We have tried to compute the macroscopic electric field explicitly from the simulation trajectory according to its definition in eq. 47. We did this by dividing the sample into slices and computing the local microscopic field in the slices according to eq. 34. The macroscopic field in the gap is the average over all the slices of the sample.

Unfortunately, we found that this method gives very large fluctuations in the calculated electric field especially for small samples. The fluctuations are due to the singularity problem near the point charges. Theoretically, this should be canceled by the symmetry of the electric field around the point charge. Although we tried to improve the accuracy of the numerical integration by using smaller slices this did not give more stable results. At the same time, using more slices increases the computational cost of the numerical integration because one has to loop over all slices and over all atoms in the system. The large fluctuations (errors) that come with this method make the values physically meaningless. We note, though, that computing the macroscopic electric field from atomistic simulations is sometimes possible for certain standard sample shapes (slabs or spherical samples) when tricks are used that avoid the singularity problems [137, 146, 147].

Table 1. Main relationships in static dielectric theory

System	Relation	Notes
<p>Non-polar systems: Clausius-Mossotti equation</p>	$\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi}{3} N\alpha$	
<p>Polar diluted systems: Debye equation</p>	$\epsilon - 1 = 4\pi N \left(\alpha + \frac{\mu^2}{3kT} \right)$	
<p>Polar systems: Onsager Equation</p>	$\mu^2 = \frac{9kT}{4\pi N} \frac{(\epsilon - \epsilon_\infty)(2\epsilon + \epsilon_\infty)}{\epsilon(\epsilon_\infty + 2)^2}$	<p>Spherical molecules</p> 
<p>Polar systems of interacting molecules: Kirkwood-Fröhlich equation</p>	$g\mu^2 = \frac{9kT}{4\pi} \frac{(\epsilon - \epsilon_\infty)(2\epsilon + \epsilon_\infty)}{\epsilon(\epsilon_\infty + 2)^2}$	<p>Spherical sample</p> 

Notes on the table: g is the Kirkwood correlation factor accounting for the correlation between the dipoles, ϵ_∞ is the dielectric constant of the infinite field frequency (for non-polarizable molecules $\epsilon_\infty=1$). The derivation of these equations depended on the spherical shape of the sample. In most cases, the sample is one spherical molecule except for the case of the Kirkwood-Fröhlich equation where the spherical sample contains many molecules. The shape of the sample is important to define the relations between the components of the electric field. The surrounding medium is considered to be a continuum.

4.3 Derivation of a generalized fluctuation formula for calculating the dielectric constant in the interfacial gap

The main problem in calculating the dielectric properties of the sample volume of interest, the interfacial gap between two proteins, is that the sample is anisotropic and has a non-standard shape so that the derivation of the fluctuation formula has to be adapted to this case. The following derivation follows the approach of Böttcher [2] for the statistical mechanical approach for calculating the dielectric constant. First, a relation is found between the polarization and the part of the macroscopic field that affects the energy of the dipoles in the field and tends to direct them (the directing field; see 4.1.2.1). Then, the relation between the total macroscopic field and this directing field is found for non-standard sample shapes. The following derivation is adapted to our sample. The important points to keep in mind are:

- The shape of the dipoles is spherical. Our simulations were performed using the TIP4P water model where the water molecules have a spherical shape since only the oxygen atom exerts Lennard-Jones interactions. The shape of the molecule is important for the validity of the relations between the electric field components.
- The derivation considers the case of non-polarizable molecules which was the case in our simulation using the OPLSAA force field [104] and the TIP4P[148] water model.
- There is an unavoidable error due to the finite size of the sample.

First we can write the constitutive relation in the following way:

$$\epsilon - 1 = 4\pi \left(\frac{\langle \mathbf{P} \cdot \mathbf{e} \rangle}{\mathbf{E} \cdot \mathbf{e}} \right)_{\mathbf{E}} = \frac{4\pi}{V} \left(\frac{\langle \mathbf{M} \cdot \mathbf{e} \rangle}{\mathbf{E} \cdot \mathbf{e}} \right)_{\mathbf{E}} \quad (59)$$

Here, \mathbf{M} is the total dipole moment of the sample $\mathbf{P} \cdot \mathbf{e} = \langle \mathbf{M} \cdot \mathbf{e} \rangle / V$, and V is the volume of the considered sample.

As mentioned previously, the polarization \mathbf{P} also has terms in higher powers of \mathbf{E} [130]. The higher terms are neglected for weak to moderate field intensities. Thus the constitutive relation considers only the linear term and we can rewrite the constitutive relation as follows [130]:

$$\epsilon - 1 = 4\pi \left(\frac{\partial \langle \mathbf{P} \cdot \mathbf{e} \rangle}{\partial (\mathbf{E} \cdot \mathbf{e})} \right)_{\mathbf{E}} = \frac{4\pi}{V} \left(\frac{\partial \langle \mathbf{M} \cdot \mathbf{e} \rangle}{\partial (\mathbf{E} \cdot \mathbf{e})} \right)_{\mathbf{E}} \quad (60)$$

We now rewrite eq. (60) by expanding the partial derivative $\partial \langle \mathbf{M} \cdot \mathbf{e} \rangle / \partial (\mathbf{E} \cdot \mathbf{e})$ as the product of the two partial derivatives $\partial \langle \mathbf{M} \cdot \mathbf{e} \rangle / \partial (\mathbf{E}_d \cdot \mathbf{e})$ and $\partial (\mathbf{E}_d \cdot \mathbf{e}) / \partial (\mathbf{E} \cdot \mathbf{e})$:

$$\epsilon - 1 = \frac{4\pi}{V} \left(\frac{\partial \langle \mathbf{M} \cdot \mathbf{e} \rangle}{\partial (\mathbf{E}_d \cdot \mathbf{e})} \right)_{\mathbf{E}_d} \left(\frac{\partial (\mathbf{E}_d \cdot \mathbf{e})}{\partial (\mathbf{E} \cdot \mathbf{e})} \right)_{\mathbf{E}} \quad (61)$$

Now, the problem of defining the dielectric constant is broken into two parts: (1) finding a relation between the total dipole moment and the directing field $\partial \langle \mathbf{M} \cdot \mathbf{e} \rangle / \partial (\mathbf{E}_d \cdot \mathbf{e})$, (2) finding the relation between the directing field and the macroscopic field $\partial (\mathbf{E}_d \cdot \mathbf{e}) / \partial (\mathbf{E} \cdot \mathbf{e})$.

4.3.1 The relation between the total dipole moment and the directing field

Kirkwood [144] used Boltzmann averaging to find the relation between a homogenous externally applied field \mathbf{E}_0 and the total dipole moment of an isotropic polar liquid:

$$\frac{\partial \langle \mathbf{M} \cdot \mathbf{e} \rangle_{\mathbf{E}_0, \mathbf{e}=0}}{\partial (\mathbf{E}_0 \cdot \mathbf{e})} = \frac{1}{kT} \langle (\mathbf{M} \cdot \mathbf{e})^2 \rangle_{\mathbf{E}_0, \mathbf{e}=0}$$

where the statistical averages should be taken from a simulation without externally applied field. For an infinite sample of isotropic liquid, the directing field equals the externally applied field as the net direction field is zero. As this is not the case in general, we used in equation 6 the directing field \mathbf{E}_d instead of \mathbf{E}_0 and followed the derivation of King and Warshel [137] (see eq. 14 in reference [137]):

$$\frac{\partial \langle \mathbf{M} \cdot \mathbf{e} \rangle}{\partial (\mathbf{E}_d \cdot \mathbf{e})} = \frac{1}{kT} \left(\langle (\mathbf{M} \cdot \mathbf{e})^2 \rangle_{\mathbf{E}_d} - \langle \mathbf{M} \cdot \mathbf{e} \rangle_{\mathbf{E}_d}^2 \right) \quad (62)$$

The notation $\langle \rangle_{\mathbf{E}_d}$ means that the averages are taken at the given value of \mathbf{E}_d . Since the Hamiltonian in the Boltzmann averaging includes all the possible interactions, the periodic boundary conditions are properly taken into account. This was clearly shown by Smith [149] (see their eq. 5).

4.3.2 The relation between \mathbf{E} and \mathbf{E}_d

For a sample of non-polarizable spherical molecules carrying only a permanent dipole, the directing field \mathbf{E}_d equals the field in the cavity (cavity field \mathbf{E}_c) when the molecules of the sample are removed and the surrounding molecules are relaxed to remove their interactions with the dipoles [130] (the reaction field \mathbf{R}) (see Böttcher [130] p.174). The cavity field in a spherical cavity inside a dielectric medium of dielectric ϵ_s generated as a response to a homogenous externally applied field is known to be uniform and given through the relation [130]:

$$\mathbf{E}_d \cdot \mathbf{e} = \mathbf{E}_c \cdot \mathbf{e} = \frac{3\epsilon_s}{2\epsilon_s + 1} \mathbf{E}_0 \cdot \mathbf{e} \quad (63)$$

On the other hand, the field in a dielectric sphere of permittivity ϵ inside a dielectric medium of permittivity ϵ_s is related to the externally applied field \mathbf{E}_0 through the relation [130]:

$$\mathbf{E} \cdot \mathbf{e} = \frac{3\epsilon_s}{2\epsilon_s + \epsilon} \mathbf{E}_0 \cdot \mathbf{e} \quad (64)$$

Comparing eqs. (63) and (64) gives the relation between \mathbf{E} and \mathbf{E}_d for a spherical sample shape:

$$\mathbf{E}_d \cdot \mathbf{e} = \frac{2\epsilon_s + \epsilon}{2\epsilon_s + 1} \mathbf{E} \cdot \mathbf{e} \quad (65)$$

Substituting eq. (65) and (62) in (61) yields the generalized Kirkwood equation for a spherical sample.

To find the relation between \mathbf{E}_d and the macroscopic field \mathbf{E} in our sample volume (the interfacial gap) we follow the assumption of Onsager [140] that the entire liquid volume is filled up by the solvent molecules. Taking into account that the TIP4P water molecules have spherical shape (where only the oxygen atom exerts Lennard-Jones interactions with the rest of the system), we can consider every water molecule in the sample as a probe of the dielectric constant of the sample volume ε . We will assume that every water molecule is surrounded by the sample volume $\varepsilon_s = \varepsilon$. This assumption is well justified as most of the water molecules are in the sample volume. Moreover, the value of ε_s has a small influence on the calculated ε especially if ε_s has a very close value to ε (see the discussion). Under this assumption, if a homogenous external electric field is applied to the sample, the directing field acting on all molecules in the sample will be homogenous. With $\varepsilon_s = \varepsilon$ eq. (65) becomes:

$$\mathbf{E}_d \cdot \mathbf{e} = \frac{3\varepsilon}{2\varepsilon + 1} \mathbf{E} \cdot \mathbf{e} \quad (66)$$

Substituting eq. (66) and (62) in eq. (61) yields:

$$\boxed{\frac{(\varepsilon - 1)(2\varepsilon + 1)}{3\varepsilon} = \frac{4\pi}{VkT} \left(\langle (\mathbf{M} \cdot \mathbf{e})^2 \rangle_{\mathbf{E}_d} - \langle \mathbf{M} \cdot \mathbf{e} \rangle_{\mathbf{E}_d}^2 \right)} \quad (67)$$

This equation is an analog of the Kirkwood-Fröhlich fluctuation formula for one direction and was derived by King and Warshel [137] for a spherical sample to calculate the dielectric constant inside the protein. It can also be used when external fields are applied. Our derivation shows that the approach of the Kirkwood-Fröhlich formula is well justified in our sample volume. The possible errors from applying this formula are discussed in the discussion of the result (see the discussion 6.4.1.3). Several authors have argued that the Kirkwood-Fröhlich formula is widely applicable independently of the boundary conditions [143] and for an arbitrary sample shape [150, 151].

4.4 Evaluation of the dielectric permittivity by a polarization formula

Alternatively, it is also possible to calculate the dielectric permittivity from the change of the polarization under the application of an external field. For this, we started from the relationship between the polarization \mathbf{P} and the directing field \mathbf{E}_d that follows a Langevin equation [130]. For weak values of \mathbf{E}_d this relationship is linear. Thus we can use the approximation:

$$\frac{\partial \langle \mathbf{M} \cdot \mathbf{e} \rangle}{\partial (\mathbf{E}_d \cdot \mathbf{e})} \approx \frac{\Delta \langle \mathbf{M} \cdot \mathbf{e} \rangle}{\Delta (\mathbf{E}_d \cdot \mathbf{e})} \quad (68)$$

Substituting (63) into (68) and using the assumption $\varepsilon_s = \varepsilon$ as before yields:

$$\frac{\Delta\langle\mathbf{M}\cdot\mathbf{e}\rangle}{\Delta(\mathbf{E}_d\cdot\mathbf{e})} = \frac{\Delta\langle\mathbf{M}\cdot\mathbf{e}\rangle}{\Delta(\mathbf{E}_0\cdot\mathbf{e})} \frac{(2\varepsilon+1)}{3\varepsilon} \quad (69)$$

Comparing 69 , 68, 66 and 61 yields:

$$\varepsilon - 1 = \frac{4\pi}{V} \frac{\Delta\langle\mathbf{M}\cdot\mathbf{e}\rangle}{\Delta(\mathbf{E}_0\cdot\mathbf{e})} \quad (70)$$

This polarization formula is applicable for weak strengths of the externally applied field. The advantage of this formula is that the polarization converges rather fast so that its evaluation does not require long simulation times. In contrast, the fluctuation formula (eq. (67)) needs much longer simulation times to converge [143]. However, the fluctuation formula is more accurate as it takes into account the interactions between the dipoles inside the sample as the Boltzmann averaging is taken over all molecules in the sample volume [130, 144] whereas the Boltzmann averaging in the Langevin equation considers only the orientation of one dipole [130] and neglects the interactions between the dipoles in the sample volume. This is the same difference as that between the Onsager equation [140] and the Kirkwood relation [144] since the Kirkwood correlation factor was introduced in the Onsager equation to correct this point.

5 Mechanism of Fast Peptide Recognition by SH3 Domains

Simulations association of proteins with a large hydrophobic interface

The association of a proline rich motif to an SH3 domain is probed by molecular dynamics simulations. The complete pathway of the association is studied from the diffusion of the unbound proteins to the final complex through transient encounters. The results reveal that the interplay of reducing the dimensionality of the search process and the phenomenon of hydrophobic dewetting help to turn a seemingly complicated binding process into a well-organized bimodal binding process. This chapter was published in *Angewandte Chemie International Edition* in 2008 [1].

Protein-protein and protein-peptide recognition play central roles in the regulation of biological cells and much work has been devoted to unraveling the mechanistic details of these processes during the past decades [21, 58, 152, 153]. However, our understanding of "the binding event" still awaits more detailed information from experiment and advances in the computational performance so that dynamic simulations may be extended to longer simulation times on which these events take place. It is commonly believed that formation of protein-protein complexes follows a pathway from a diffusive phase through one or more intermediate states to the final stereospecific complex. Experimental studies by site-directed mutagenesis [55, 152] and computational studies by Brownian dynamics simulations [21, 153] have shown the important role of long range electrostatic interactions in the diffusive phase. By using NMR paramagnetic relaxation enhancement (NMR-PRE), Clore and coworkers have recently demonstrated the existence and visualized the distribution of an ensemble of transient, non-specific encounter complexes for a relatively weak protein-protein complex [27, 28, 154]. In spite of these advances, our mechanistic understanding of the transformation from nonspecific transient encounter complexes to the final stereospecific stable complex is still limited. Clearly, two of the main challenges for computational modeling of protein binding events are the role of solvent and the time scale of the binding events. In this chapter, I present results on the atomic mechanism of how proline rich motifs (PRMs) are recognized by SH3 domains as a model system for the association between proteins with predominantly hydrophobic interfaces.

5.1 SH3 domain complexes as an example for complex with hydrophobic- charged interface

The SH3 domain is one of the most abundant protein interaction domains in the human genome. It belongs to a large protein family known as the Proline Recognition Domains (Table 2) that are important for the assembly of many intracellular signaling complexes and pathways. This family includes several important domains such as Src homology 3 (SH3) domains, WW domains, EVH1 domains, and GYF domains.

Proline recognition domains are usually parts of larger proteins where they are involved in mediating the recognition of other peptides and proteins. Protein recognition by SH3 domains is known to involve the binding of proline rich motifs, also termed PRMs [39, 40]. The canonical peptide binding pocket of SH3 domains consists of a hydrophobic surface patch including two grooves that accommodate Px and xP residues of the peptide (see Figure 14). The flanking positively charged arginines of the peptide usually form contacts with the negatively charged residues in the RT and n-sCr loops of the SH3 domain. In the crystal structure of the complex the peptide adopts a polyproline type-II helix conformation (PPII), which is the known binding conformation for the PRMs bound to SH3 domains and other proline recognition domains [39, 40].

Organism	SH3	WW	EVH1	GYF
<i>Saccharomyces cerevisiae</i>	25	5	1	3
<i>Caenorhabditis elegans</i>	66	18	2	3
<i>Drosophila melanogaster</i>	90	27	5	2
<i>Mus musculus</i>	163	39	16	2
<i>Homo sapiens</i>	332	80	20	5

Table 2. Abundance of proline recognition domains according to reference [40].

The reasons behind choosing the SH3 domain as a case study for the simulation of protein-protein association is the large hydrophobic interface which is similar in all the SH3 domains and other proline recognition domains. The SH3 interfaces also involve several charged amino acids so that the binding of peptides and proteins is speeded up by long-ranged electrostatic attraction making it accessible to the computational time of molecular dynamics simulations.

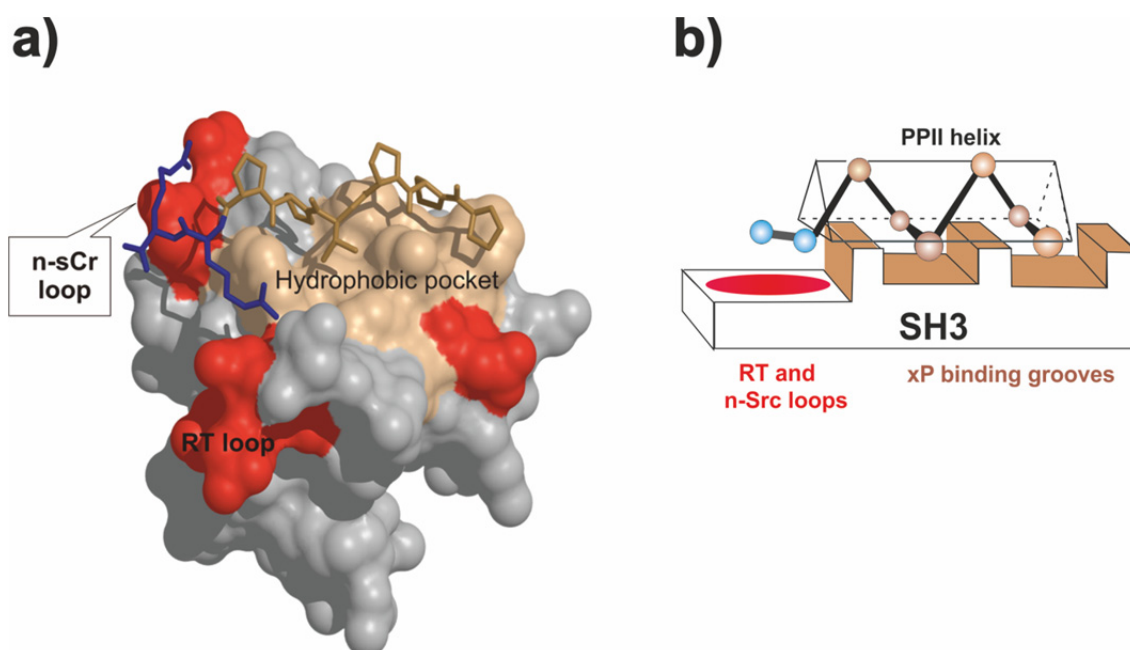


Figure 14. Structure of the binding pocket and binding modes of SH3 domains.

(a) Shows a surface representation of the crystal structure (pdb code 1ckb) of the SH3 domain from N-terminal SH3 domain of C-CRK (Homo sapiens). Surface patches about negatively charged residues are colored red, the hydrophobic part of the binding interface is colored in light brown. (b) Schematic representation of SH3:PPII binding interfaces. The core recognition surface has two hydrophobic xP binding grooves that accommodate the “hydrophobic” proline rich motif that adopts a PPII helix conformation. The negatively charged loops (RT and n-Src) of the SH3 domain are in contact with the positively charged arginine of the peptide.

5.2 MD simulations of the binding process

The starting structure for the MD simulation of the SH3 domain was extracted from a crystal structure of the complex of the C-CRK N-terminal SH3 domain with a PRM (PDB code 1ckb). We used the peptide sequence for the binding motif as in the crystal structure (PPPVPPRR). To minimize the dependency on the starting structure, we chose an extended conformation for the peptide “135° (N-C α -C-N), 180° (C α -C-N-C α), and -135° (C-N-C α -C)” as a starting conformation for the peptide in the simulations. In agreement with a previous finding by Gu and Helms [155], the peptide adopted the PPII helix conformation within a few picoseconds of simulation time before forming the encounter complexes. This conformation was stable during the diffusion phase and in the final bound states, which

were connected by a phase of higher flexibility in the intermediate states where conformational changes in the peptide seem to be helpful during the binding of the peptide to reach its binding pocket. We ran 13 MD simulations with different starting structures, where the peptide and the protein were separated by different minimum distances (13–20 Å), with different orientation for the peptide (0°, 90°, 180°, 270° rotational angles from the native crystal orientation) (see Figure 15). The OPLSAA force field [104] was used for the protein, and the TIP4P water model [148] was used for the explicit solvent. The systems were solvated in cubic water boxes large enough so that the water extended at least 14 Å from the protein surface. Counterions were added to make the system electrically neutral. All simulations were run with version 3.3 of the GROMACS simulation package [156]. Equilibration consisted of 500 steps of steepest-descent energy minimization and an MD simulation of 100 ps length with harmonic position restraints using a force constant of 1000 kJ mol⁻¹ nm⁻² for the heavy atoms in the protein for each system. This was followed by unrestrained MD simulations in the NPT ensemble [1 atm and 310 K using coupling times of 1 ps⁻¹ and 0.1 ps⁻¹]. Long-range electrostatic interactions were computed by the particle-mesh Ewald method [157]. Van der Waals interactions and short range electrostatic interactions were computed within a 14 Å cutoff. A time step of 2.0 fs was used. All simulations were run for 50 ns, and two simulations that came close to the crystal structure were extended to 150 ns. The total length of simulation times is 0.85 μs.

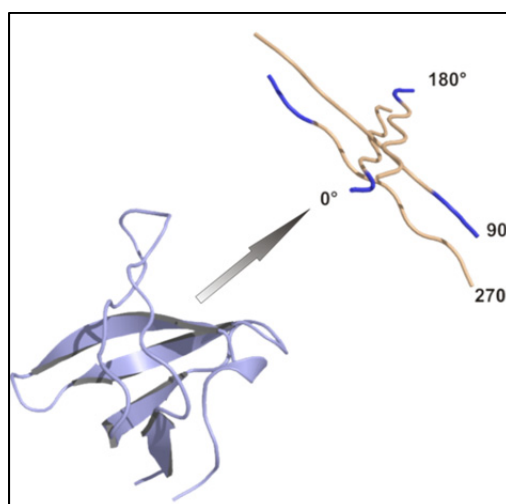


Figure 15. Starting structures for MD simulations.

The peptide was displaced by different distances (13–20 Å) and rotated into different orientations from the native orientation (0°, 90°, 180°, and 270°).

5.3 MD simulations recovered the experimentally stereospecific structures

We ran MD simulations started from different unbound conformations for the C-CRK N-terminal SH3 domain with a PRM for which a crystal structure of their complex is available [158]. In all simulations, we observed a relatively fast diffusive phase leading to the formation of nonspecific encounter complexes (see Figure 16) stabilized by salt bridges between the oppositely charged residues in the domain and the peptide. In six out of thirteen simulations, the encounter complexes led to stable stereospecific complexes involving three different binding modes. We defined these complexes as stereospecific binding modes based on the comparison with experimentally determined structures of complexes for SH3

domains [39, 40, 158]. The determining step for these modes is the formation of the electrostatic transient complex, where the peptide arginine forms salt bridges with the negatively charged residues in the RT or n-sCr loops. This leads either to the same binding mode as found in the crystal structure or to binding in the new pocket (see below), or with the residue D142 leading to the binding with the peptide in opposite orientation.

5.3.1 Crystal structure binding mode

One simulation converged to a conformation close to the crystal structure. Already during the first 50 ns of the simulation it formed a considerable number of native contacts (see Figure 18 and Figure 19). We extended this simulation to 150 ns and it came very close to the crystal structure after 130 ns (Figure 16b). The root mean square displacement (RMSD) for the whole complex backbone from the backbone of the crystal structure is 1.3 ± 0.2 Å averaged over the last 20 ns of the simulation (Figure 17). It contained the known native contacts in the crystal structure (Figure 18) and a polyproline type-II helix conformation for the peptide.

5.3.2 Binding mode with opposite orientation

Three simulations converged within 20-30 ns of simulation time to a conformation with the peptide bound in the same canonical pocket like in the crystal structure but in the opposite orientation (Figure 16C). The peptide adopts a PPII helix conformation as well and is symmetric (opposite orientation) to the one in the crystal structure.

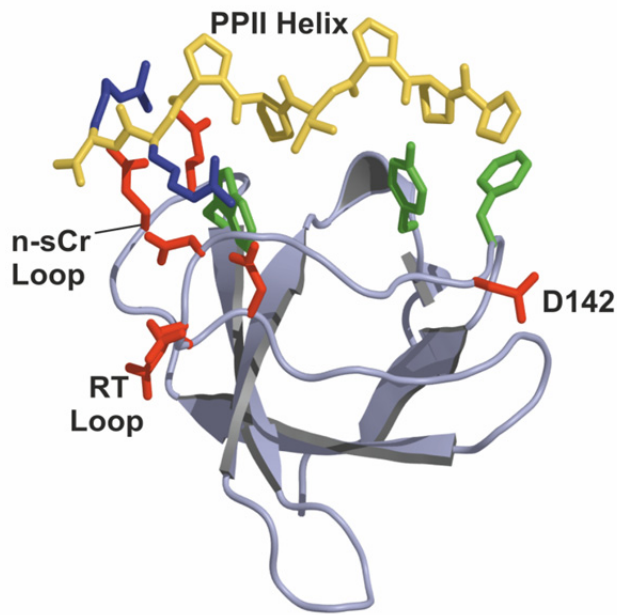
One of the peptide arginine residues formed a salt bridge with residue D142 in the domain during the transient encounter stage, but this contact was not permanent in the final complex. Interestingly, aspartic acid residues are frequently found in this position among SH3 domains. These complexes were stable during the remaining simulation time (until 50 ns). The possibility of SH3 domains to bind the PRMs in two opposite orientations is well characterized [159] and has been found in many other proline recognition domains as well [39, 40]. The structural basis for this is the symmetry of the PPII helix which enables packing into two different orientations in the same binding pocket.

This novel binding mode for the C-CRK SH3 domain described here suggests the possibility of the same peptide to bind in two different orientations to the same SH3 domain. To our knowledge, this has not yet been described experimentally. The static picture from the crystal structure cannot capture the existence of different specific binding modes if they are lowly populated. New technical advances in NMR spectroscopy open the door for getting structural information about rarely populated states [3]. It will be interesting to learn whether those methods can detect the existence of two stereospecific complexes for the same proteins. Judging on the occupancies of the two orientations for the same peptide with the same domain, and thus to characterize the difference in binding free energy, is beyond the aims of this work.

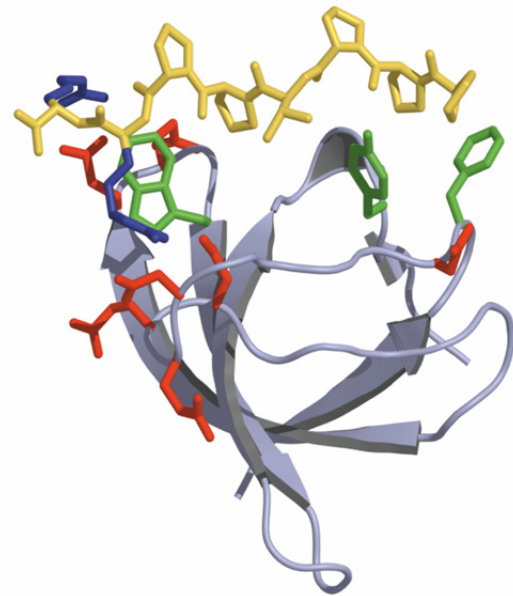
5.3.3 Binding in the new pocket

In two of our simulations the peptide bound within 30 ns of simulation time in a different pocket of the SH3 domain (Figure 16d). These complexes were also stable during the remaining 20 ns of the simulation time. The role of this pocket in peptide binding by SH3 domains has been detected for the first time in a recently determined crystal structure [160] (PDB code 2p4r) where the SH3 domain showed the possibility to bind a peptide with two PRMs. The first motif bound to the canonical pocket and the other motif bound to this face of the domain. Moreover, this pocket has recently been observed in another new structure (PDB code 2drm) as well.

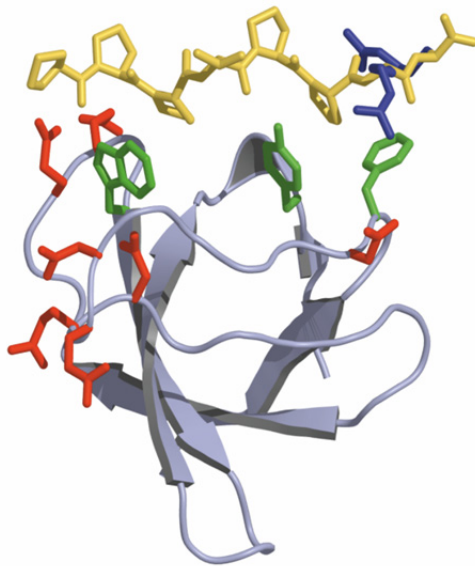
a) The crystal structure



b) recovered crystal structure from an MD snapshot



c) Binding in the opposite orientation



d) Binding mode in the new pocket

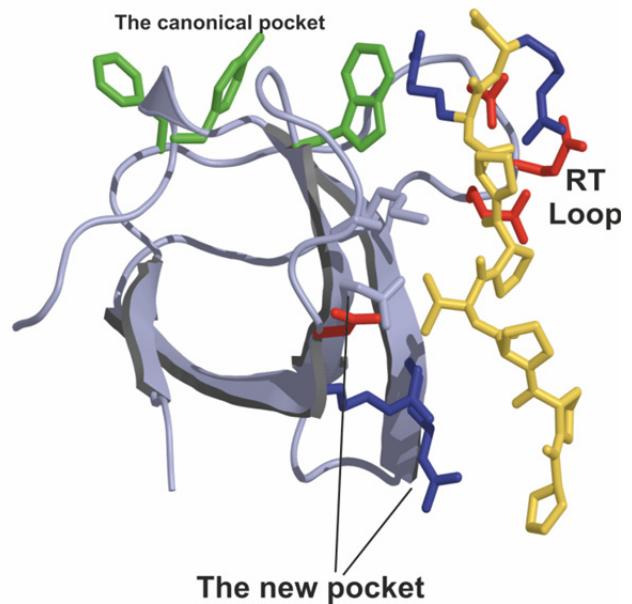


Figure 16. The recognized binding modes for the peptide on the surface of the CRK-SH3 domain.

- (a) Shows the crystal structure of the domain with the binding motif (PDB code 1ckb). The side chains of the negatively charged residues are colored red, the hydrophobic pocket in the domain is shown in green, the PPII helix is yellow, and the arginines are blue.
- (b) Shows the crystal structure binding mode recovered in an MD snapshot at 130 ns.
- (c) Shows the observed peptide binding in the opposite orientation of the crystal structure.
- (d) Shows the observed binding mode in the new pocket. The figure was generated using PYMOL [161].

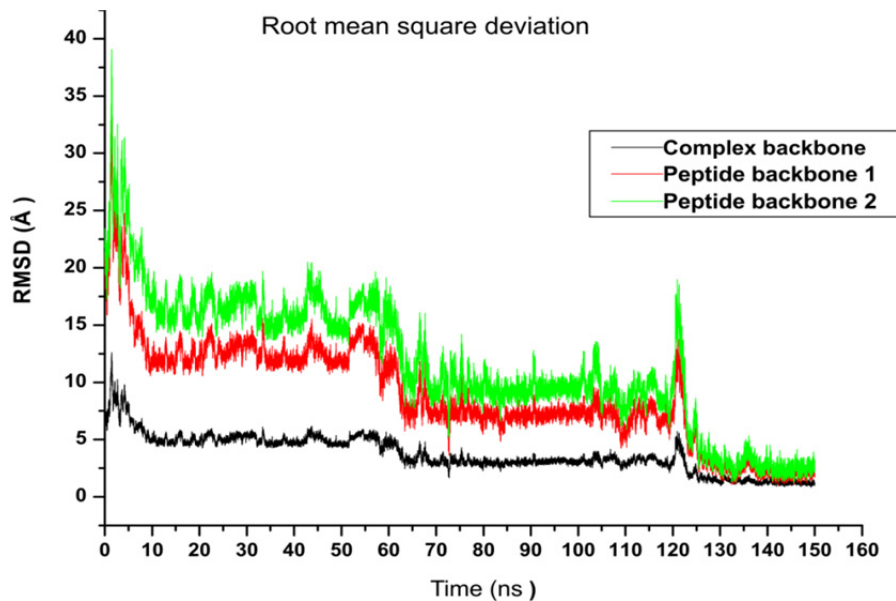


Figure 17. RMSD during the simulation time for the binding pathway leading to the crystal structure binding mode.

The root mean square deviation of the complex backbone (black) and peptide backbone 1 (red) were computed after the backbone of the complex was fitted to the backbone of the crystal structure. The RMSD of the peptide backbone 2 (green) was computed after the backbone of the domain was fitted to the backbone of the domain in the crystal structure.

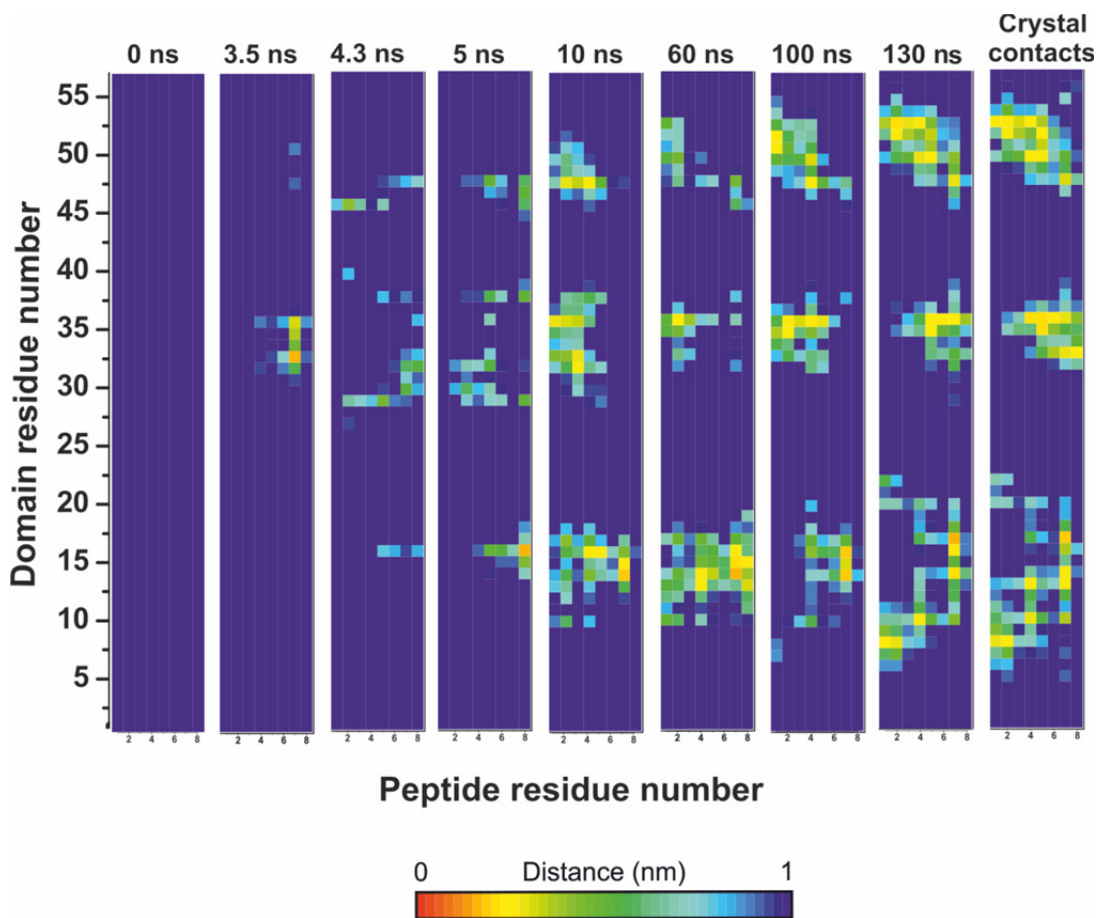


Figure 18. Contact maps between the domain and the peptide for different snapshots in comparison to the native crystal contacts.

5.4 The mechanism of the binding process

5.4.1 The importance of electrostatic interactions during the diffusive phase

Concerning the mechanism of binding, the complex formation pathway was found to consist of three phases. A fast diffusive phase leads to the formation of various electrostatically stabilized intermediate complexes (see Figure 19 and Figure 20), of which some can bind into the stereospecific stable bound complex. Here we define the diffusive phase as the diffusion of both reacting chains before they form short-range contacts. Although the diffusion phase was quite short in our simulations, the simulations showed little dependence on the starting structures of the peptide. In many of the thirteen simulations the peptide completely changed its direction during the diffusive phase (see Figure 19).

As expected, the electrostatic interactions between the oppositely charged residues in both proteins play an essential role. They guide and accelerate the diffusion to terminate by forming a nonspecific complex (Figure 19; Figure 18) stabilized by salt bridges between the side chains of arginine in the peptide (PPPVP \overline{P} RR) and the negatively charged residues in the domain (RT and n-sCr loops or residue D142). In all the binding simulations, the diffusive phase terminated with the formation of a salt bridge between the oppositely charged residues in both proteins and sometimes the N-terminus of the peptide within relatively short simulation time (see Figure 20). Moreover, we did not observe any complete dissociation in any of these simulations after the formation of the encounter complexes (Figure 20).

To investigate the importance of the charged residues in accelerating and guiding the diffusion and stabilizing the encounter complexes, we performed two simulations carrying R \rightarrow A mutations in the peptide and uncharged termini. Despite several collisions took place (as expected from the high peptide concentration), we did not observe formation of stable encounters within the simulation times (see Figure 21).

We have also tested the effect of the electrostatic interaction on the stability of the encounter complexes by running a second group of control simulations. For this, we ran four stability control simulations starting from some encounter complexes (before the hydrophobic collapse) after mutating R \rightarrow A in the peptide. In 3 out of 4 simulations, complete dissociation was observed within relatively short simulation times (< 2 ns) (see Figure 22) in comparison with the previous 13 binding simulations where no dissociation was observed after the formation of the encounter complexes (Figure 20).

These simulations confirm that the electrostatic interactions accelerate the diffusion and stabilize the encounter complexes. The electrostatically accelerated association of proteins has been found to be very rapid in experiments [152]. In particular, electrostatic acceleration increases the affinity by increasing the association rate without affecting the dissociation rate [58]. Such acceleration of diffusion can be critical for protein-protein association if the association is diffusion controlled [7, 20]. Having many negatively charged residues on the surface of the domain therefore gives rise to an ensemble of transient complexes and a binding process with multiple pathways. These observations agree with the picture of encounter complexes emerging from experimental data [27, 28].

The electrostatic nature of the intermediate states in protein-protein association has also been characterized using double mutant cycles [53]. Moreover, NMR-PRE observed a correlation between the spatial distribution of non-specific encounter complexes and the electrostatic potential isosurface [7, 20]. The population of nonspecific encounter complexes

was shown to be significantly more affected by ionic strength than that of the stereospecific complex demonstrating the importance of electrostatic interactions in the formation of the ensemble of nonspecific encounter complexes [28].

Those salt bridges in the transient encounter complexes that led to the stereospecific complex are close to or part of the possible contacts in the last stable complex. This enables these encounter complexes to run a search process with fewer degrees of freedom resulting in a "reduction-in-dimensionality" [153] to find the final stereospecific complex. The main role of electrostatic interactions in diffusion and stabilization of the transient encounter complexes explains the importance of charged residues in the binding motifs for SH3 domains. The positively charged residues in the binding motifs (R, K) are essential for binding, where the consensus (PxxPxR/K) is essential for Class II motifs and (R/KxxPxxP) for class I [39, 40].

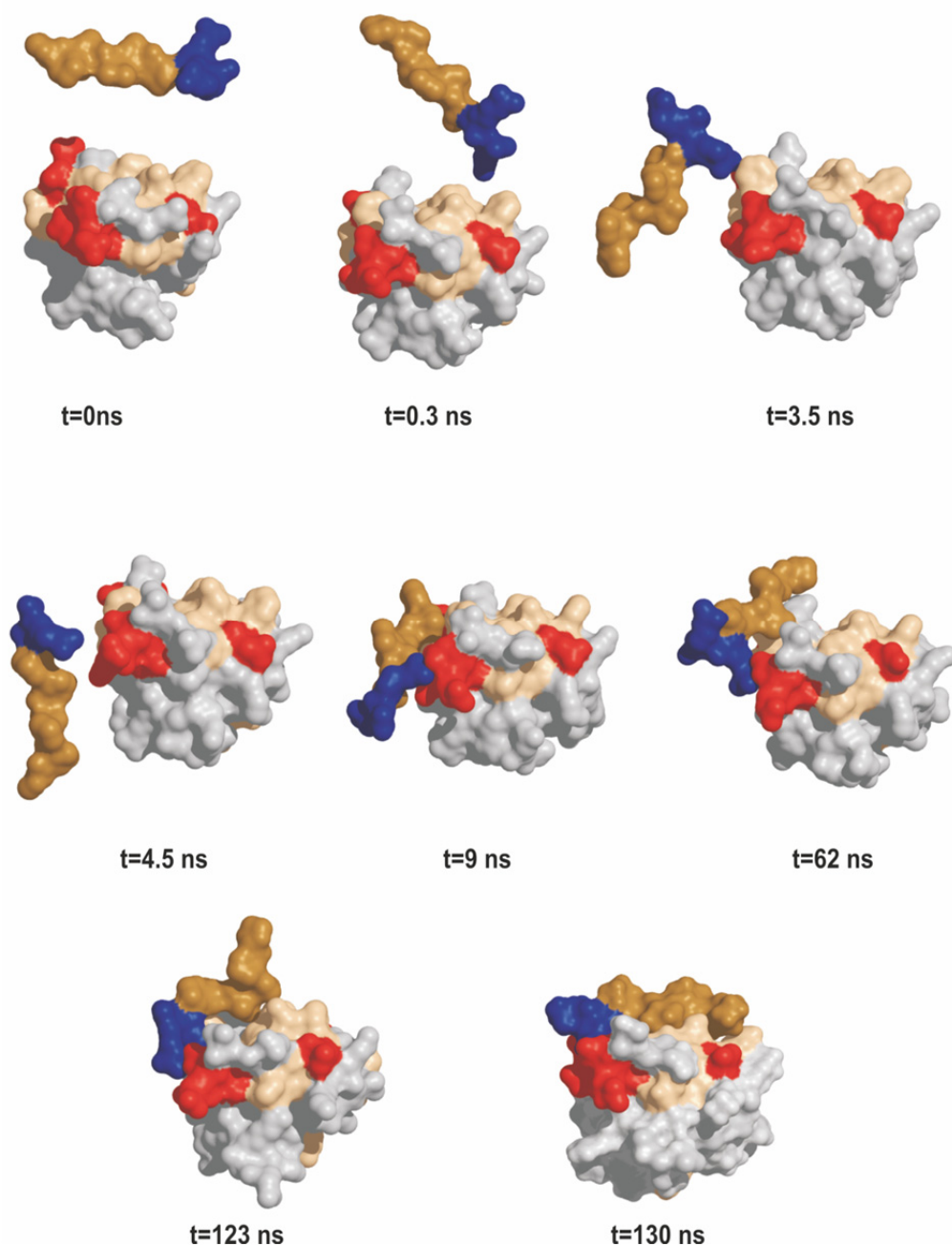


Figure 19. Snapshots along the binding pathway that leads to the crystal orientation.

The snapshots show the transformation from the starting structure to the final complex through transient encounter complexes. Solvent molecules are not shown for clarity.

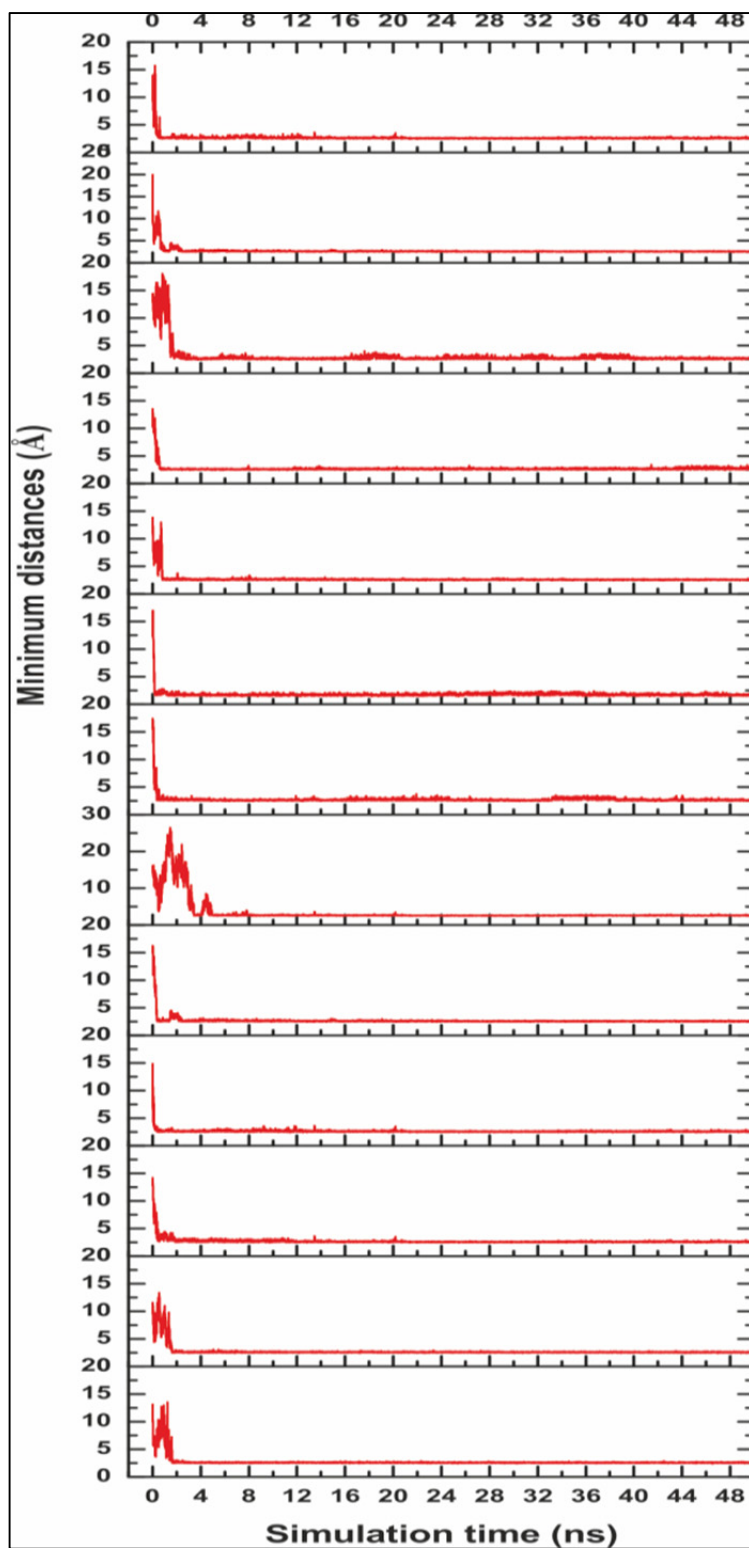


Figure 20. Minimum distance between peptide and domain in the 13 binding simulations.
 We observed relatively fast formation of stable encounter complexes (mostly within 2 ns, maximum 5 ns).

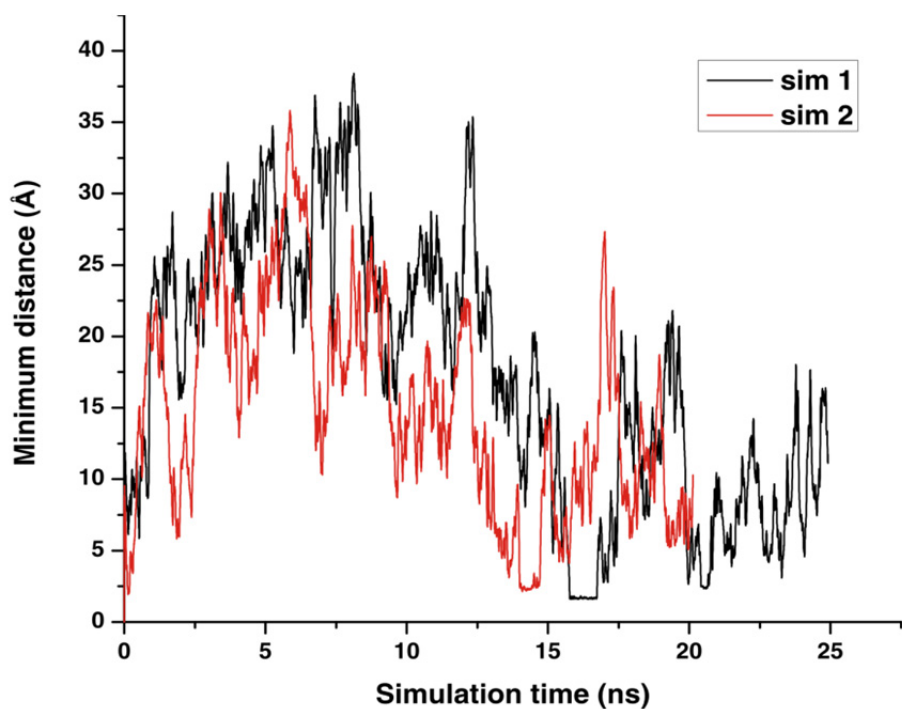


Figure 21. Control simulation for the association of an uncharged peptide.

The minimum peptide-domain distance in the two control simulations is plotted as a function of the simulation time. The simulations were performed as before after mutating two arginines in the peptide to alanines. We did not observe formation of stable encounters during the simulation times (20 and 25 ns).

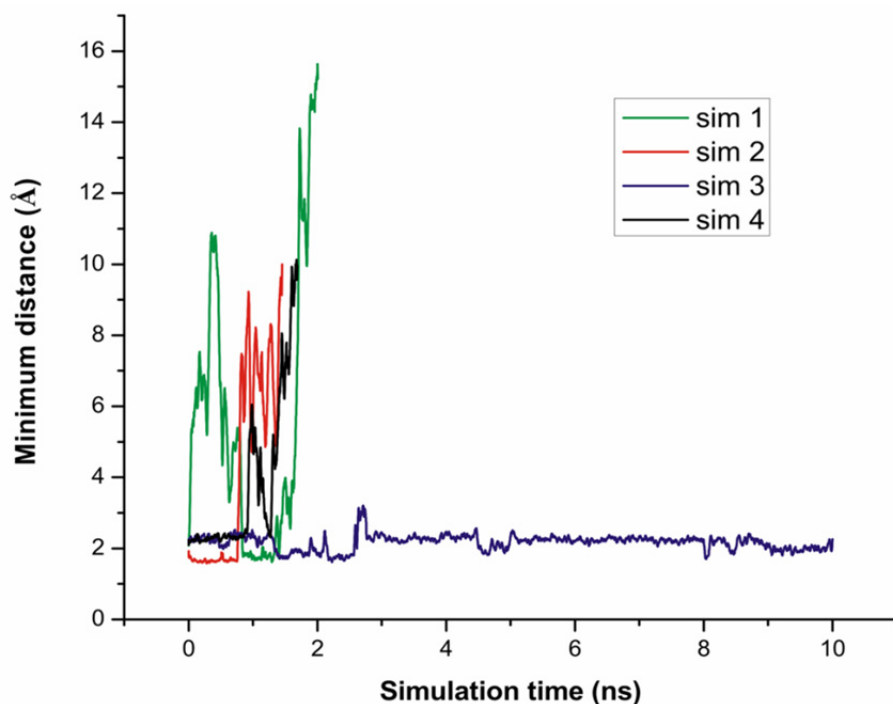


Figure 22. Control simulations starting from intermediates after mutating the charged residues in the peptide into alanine.

The minimum peptide-domain distance in 4 stability simulations is shown. The simulations were performed starting from encounter complexes after mutating arginines in the peptide to alanines. In 3 out of 4 simulations (1, 2 and 4) we observed fast dissociation within < 2 ns. The graph was smoothed by averaging over 10 ps intervals.

5.4.2 The role of hydrophobic dewetting

One of the current unknowns in the area of protein-protein interaction is the transition from the transient complexes to the specific bound complex. This step is usually proposed to be the most difficult step as it is supposed to cross an energy barrier that separates the stereospecific complex from other conformations in the structural ensemble of encounter complexes. The energy barrier is assumed to be mostly due to the desolvation penalty that results from removing the water from the interracial gap. It was already mentioned above that the simulations showed that most of the charged contacts that are known in the crystal structure were formed in the early stages of the binding process. The contacts were mainly electrostatic in the early encounters and the contacts between the hydrophobic parts of the interfaces (hydrophobic collapse) took place in the last stages when the hydrophobic interfaces were close to each other (Figure 19). Having two hydrophobic surfaces pointing toward each other encouraged us to investigate the possible role of dewetting at the hydrophobic interfaces prior to binding.

We calculated the water density in the intermolecular gaps for six snapshots taken from the intermediate stage of that simulation that led close to the crystal structure. To ensure good statistics for the water density in the interfacial gaps, we performed separate MD simulations starting from these snapshots with harmonic position restraints applied to the heavy atoms of the protein to prevent any big change in the conformations of the snapshots. Only the water molecules and the side chains of the proteins could move freely. These MD simulations were performed over 2 ns with harmonic position restraints (using a force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) acting on the heavy atoms in the protein for each system. The water density was averaged over the trajectories between 0.3 and 2.0 ns of simulation time using snapshots stored every 2.5 ps. The gap volumes were defined using the program SURFNET [162] (Figure 23d). Interestingly, the intermolecular gaps showed a significant decrease in water density (Table 3) in the encounter snapshots, which indicates clear partial dewetting of the interfaces before binding.

To characterize the distance dependence of the dewetting, we ran 10 further similar simulations starting from the crystal structure after displacing the peptide by distances of (2.5 – 9.0 Å) along the connection of the two centers of geometry for the peptide and the domain to create "gaps" (Figure 23d). To prevent any big change in the gaps we kept the distance between the peptide and the protein at the same value by applying harmonic restraints as it was done for the simulations started from the snapshots (see above). For comparison we also ran two simulations for the peptide alone (and for the domain alone) and computed the water density inside a virtual gap with the same volume and shape as in the complexes. The virtual gaps were defined as in the case of the complexes after superimposing the protein of the corresponding complex to each snapshot and then computing the gaps. The hydrophobic pockets in the separate domain (Figure 23b) and around the PP II helix in the separate peptide (Figure 23a) showed a reduced water density at the hydrophobic surfaces. The dewetting is more pronounced in the gaps of the complexes, where a significant degree of dewetting is found for all complexes with interfacial distance $< 5 \text{ Å}$ (Figure 23c).

Snapshot (time from where simulation where started)	Average water density (gr/l) \pm standard deviation
102 (ns)	801 \pm 58
104 (ns)	836 \pm 70
106 (ns)	710 \pm 94
121 (ns)	864 \pm 64
122 (ns)	777 \pm 84
125 (ns)	841 \pm 78

Table 3. Density in the interfacial gaps starting from encounter complexes

5.4.2.1 Dewetting is related to the chemical nature of the interfaces

To clarify the importance of partial dewetting as a driving force for protein-protein, we performed in addition six control MD simulations where the nature of the binding interface was altered by introducing charged residues. In these simulations, the peptide was displaced by 3.5 Å from the pocket. In four simulations, charged aspartic acid and/or lysine residues were introduced inside the hydrophobic pocket of the domain as following: mutation set 1 includes two simulations with mutations of the residues 143, 168, 183 and 184 to Aspartic acid. Mutation set 2 includes two simulations with mutations of the residues 143,183 to lysine and mutations of the residues 168, 184 to Aspartic acid. For the simulations with mutations we computed the interfacial water density as described for the wildtype peptide did before. As expected, higher water densities were observed in both mutation sets (948 \pm 49 gr/l with mutation set 1, 946 \pm 51gr/l with mutation set 2 in comparison with 756 \pm 116 gr/l for the wild-type).

Furthermore we analysed whether fast binding process also took place in these proteins with charged mutations. For this, we ran MD simulations of the mutation sets and two further simulations of the wild-type domain served as control. In these simulations, harmonic position restraints were applied only to the domain and the peptide was freely moving. We only observed fast collapse in the simulation with the wild-type pocket (see Figure 24a) where dewetting took place. In all simulations with mutated pocket residues where the water density is only slightly reduced, no collapse was observed within the simulation time (see Figure 24b).

Our simulations suggest that dewetting can guide the further search process to the final specific complex after the initial reduction in dimensionality takes place in the transient state. In the light of these new findings we can now explain why the transient complexes close to the final complex will converge so quickly that they have short life times and low occupancy [27]. Recent work has added further evidence on the relevance of the hydrophobic effect as a major component of the forces that fold and stabilize the structure of biomolecules [35, 44, 82, 86, 89-92]. For example, Lum [86] argued that a vapor like layer forms around large hydrophobic surfaces and showed this effect in simulations. Zhou [90] also showed a distance dependence of dewetting effects during MD simulations for protein folding.

Water density in the intermolecular gaps

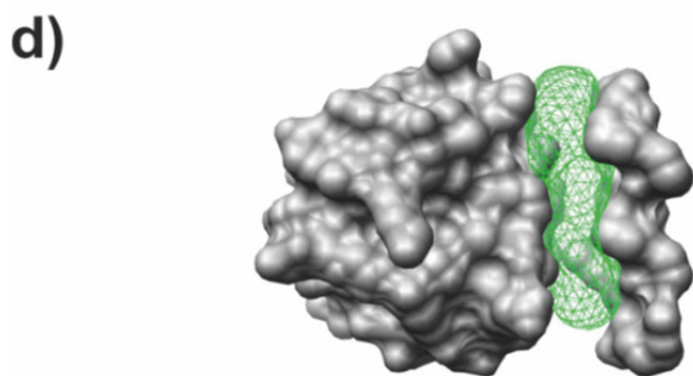
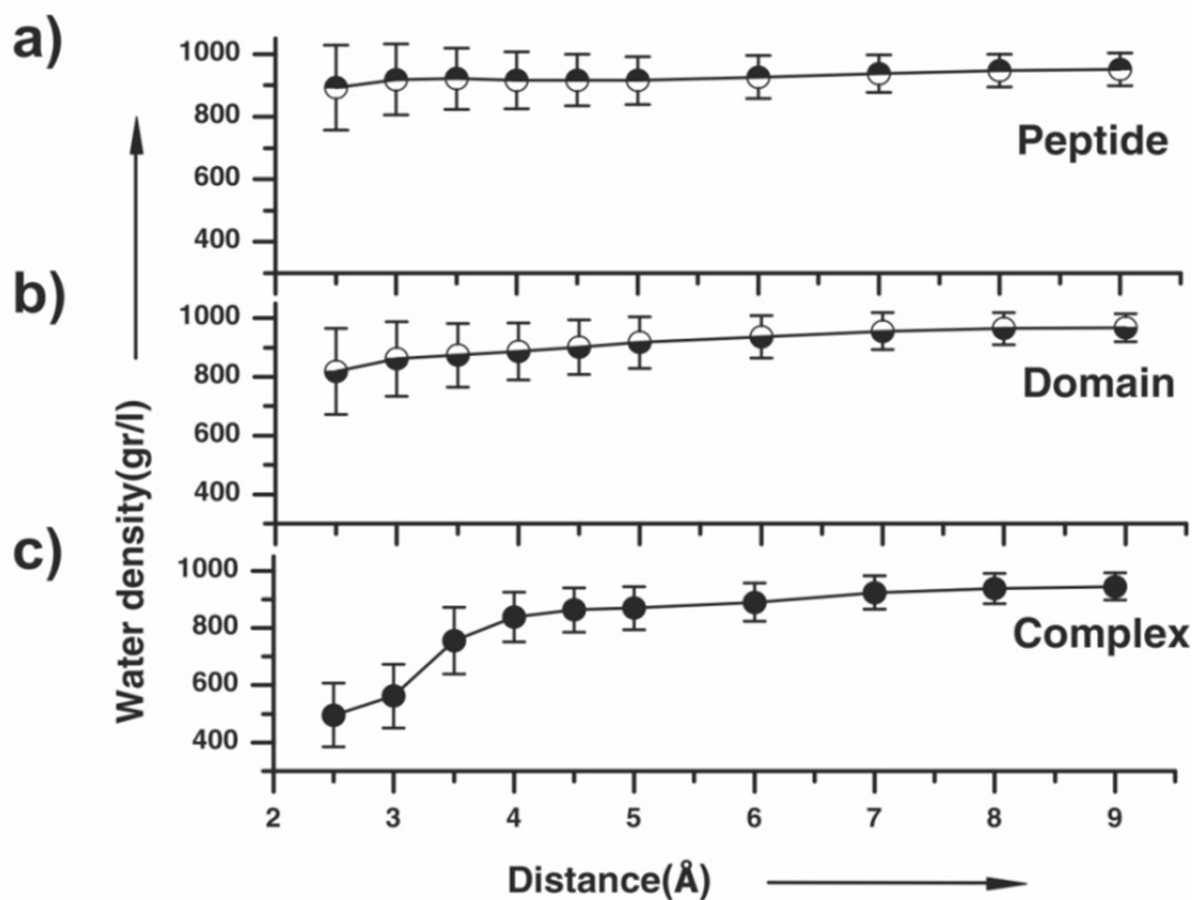


Figure 23. Water density in the hydrophobic interfacial gap with the peptide at fixed distance.

(a) Water density in the interfacial gap for the peptide alone, (b) for the domain alone, (c) for the complex.

The water density was averaged over the last 1.7 ns for each 2 ns position restraint MD simulations. The gap volumes for the free peptide and the free domain were defined by superposition of the corresponding part from the complex simulation to all snapshots of the protein-water system. (d) Representation of the defined interfacial gap between the PPII helix and the hydrophobic pocket in the domain at 4 Å interfacial distance.

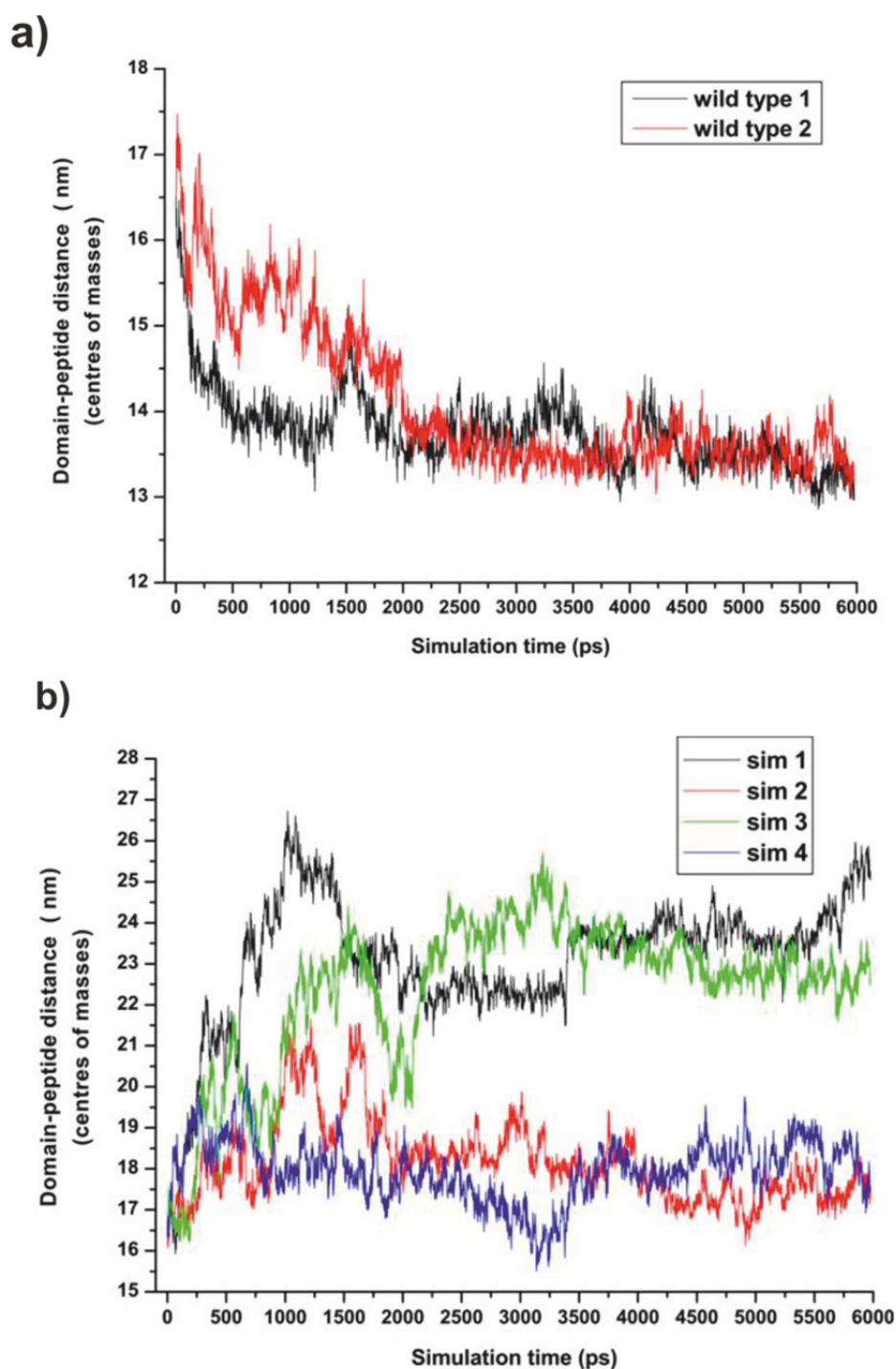


Figure 24. Control simulation with charged interfaces.

The peptide-domain distance between the centers of masses is plotted over the simulation time. All simulations were started after displacing the peptide by 3.5 Å. Harmonic position restraints were applied to the heavy atoms in the domain. **(a)** Two simulations were started from the wild-type complex. Collapse of peptide was observed in both simulations. **(b)** Two types of mutated complexes were simulated (see text). Mutation set 1 (sim1 and sim 2) includes mutations of the residues 143, 168, 183 and 184 to Aspartic acid. Mutation set 2 (sim3 and sim 4) includes mutations of the residues 143,183 to lysine and mutations of the residues 168, 184 to aspartic acid. Higher water densities were observed in both mutations (948 ± 49 gr/l for mutation set 1, 946 ± 51 gr/l for mutation set 2). We did not observe collapse in these simulations within the simulation times.

5.5 Conclusion

At the end of this chapter, it is important to reflect how relevant this model is for understanding protein-protein recognition in general? First of all, the existence of a large hydrophobic aromatic pocket is a common feature in many proline recognition domains. Yet, the hydrophobic nature of protein-protein interfaces is even well known as a general principle. Typically, between 30-50 % of the protein interface area is taken up by hydrophobic amino acids [46, 47]. On the other hand, the presence of salt bridges at protein-protein interfaces is a general feature too because on average two ion pairs per interface were found [48]. Therefore, we suggest that the mechanism found here that guides the association of the C-CRK N-terminal SH3 domain and its peptide binding motif applies to many other protein pairs too.

One lesson we learnt from the simulations is the synergistic nature of the driving forces for binding (Figure 25). The long-ranged electrostatic effects play the main role during diffusion and stabilize the transient complexes formed by the electrostatic parts in the interface. At short distances, this then enables the partial dewetting effect to increase the probability for the collapse of the hydrophobic part of the interface and the convergence to the final specific complex.

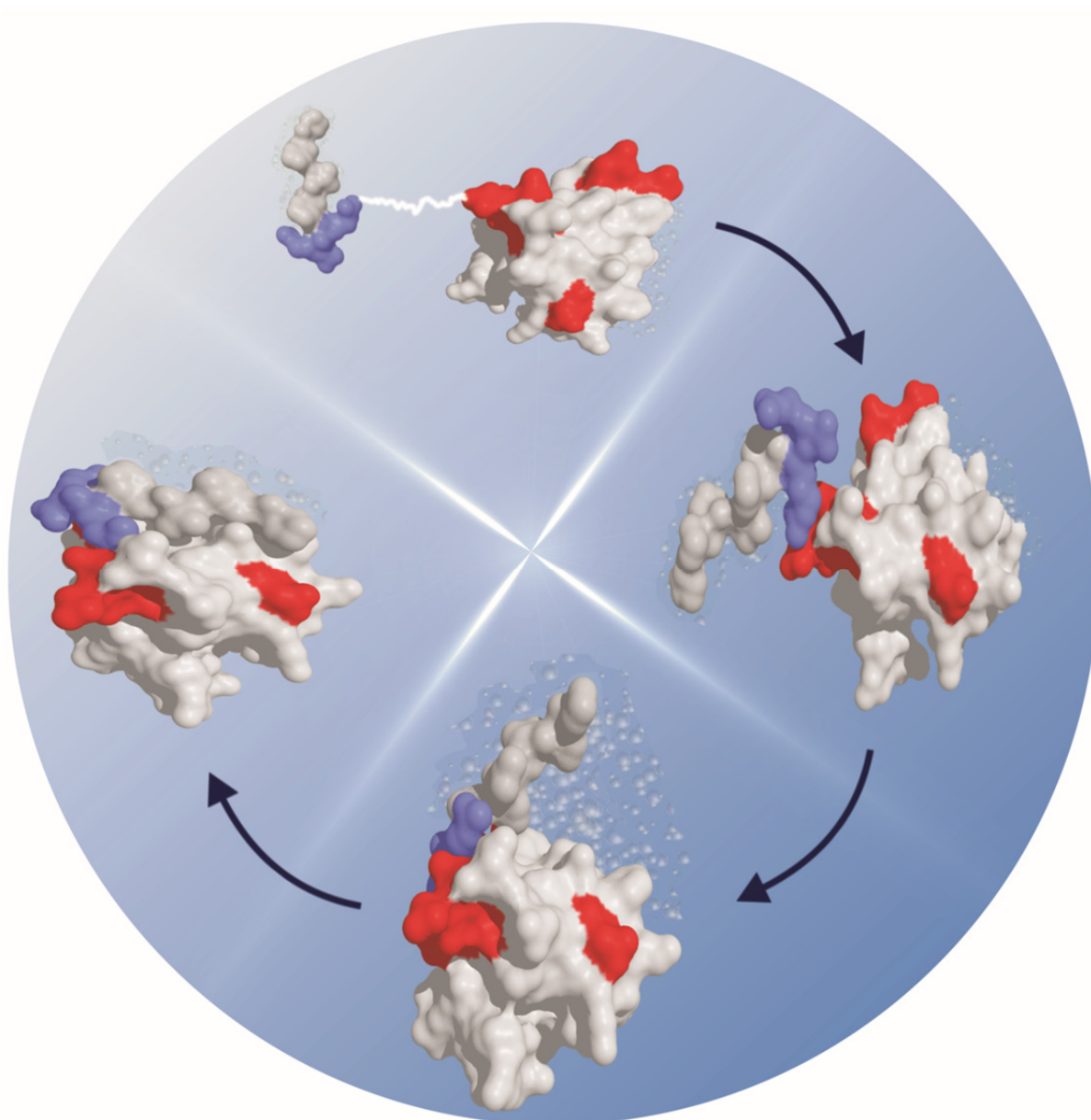
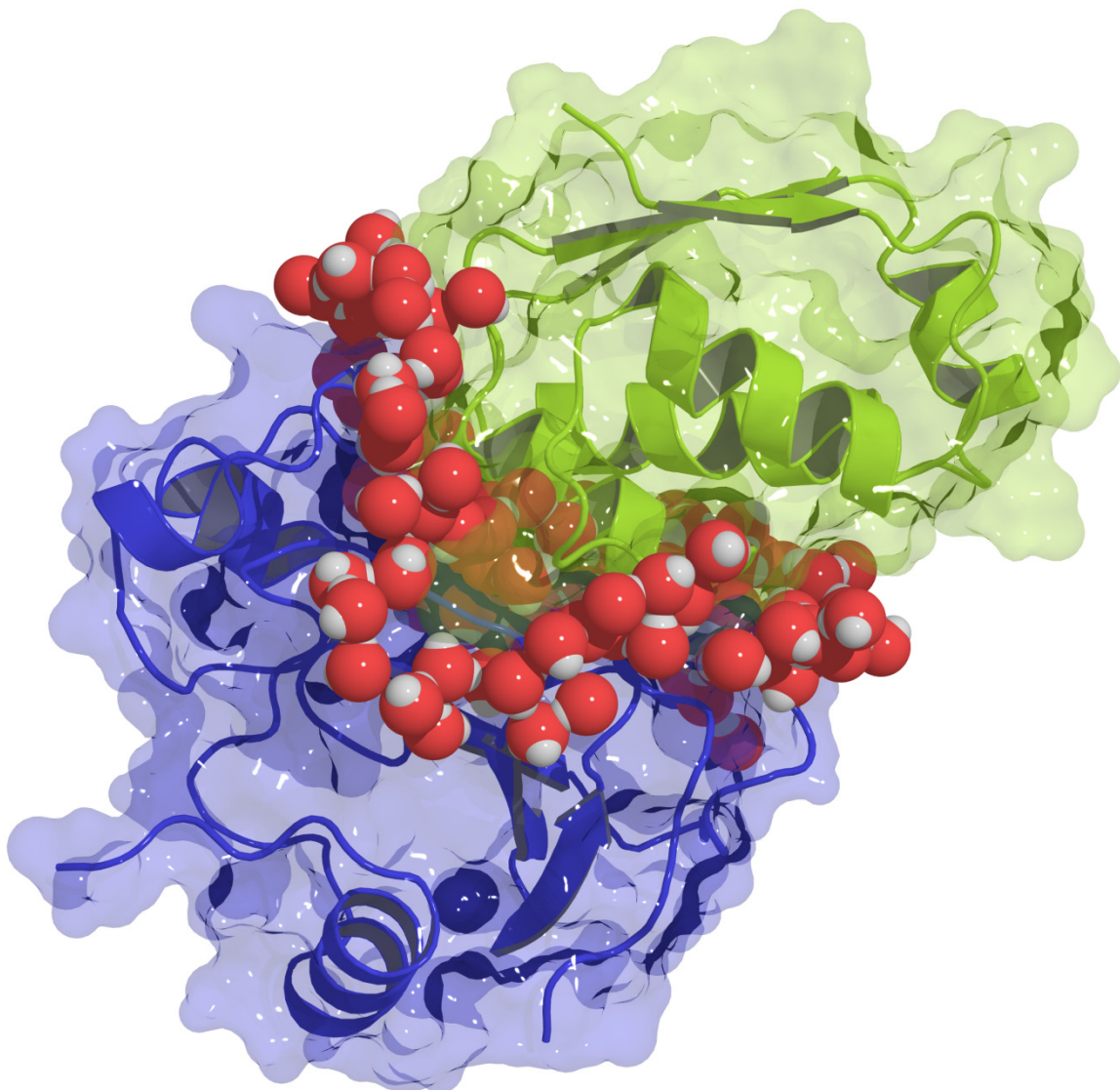


Figure 25. Schematic representation of the mechanism of the binding process.

The positively charged and negatively charged residues are colored blue and red respectively. The hydrophobic interfaces are colored gray. Dewetting of the hydrophobic interfaces of the peptide and the SH3 domain is indicated by white clouds around the interfaces.

6 Simulating the association between charged proteins

The mechanism of the association between charged proteins was studied on an atomistic scale using MD simulations. Our simulations for the binding process of Barnase-Barstar reproducibly recovered the native bound state of the complex and thus provide atomistic insight into the mechanism of binding. The simulations showed that water in the interfacial gap forms an adhesive hydrogen bond network between the hydrophilic protein interfaces. This network plays an important role in stabilizing the early intermediate states before native contacts are formed. The transformation from these intermediates to the stereo-specific complex is then accompanied by maximization of the interfacial water-mediation. Furthermore, water structure already plays an important long-range role during the diffusive phase in reducing the dielectric shielding of the water. The dielectric constant in the gap between the proteins is strongly reduced up to distances of a few nanometers. Interestingly, the dielectric properties of the water in the interfacial gap are strongly anisotropic with a preferred directionality for the electrostatic interactions along the direction perpendicular to the interfaces. Therefore, the water solvent between the two proteins makes an important contribution to the electrostatic funneling of the binding free energy surface for the binding process. This chapter was published in *Nature Communications* in 2011 [2].



6.1 Introduction

Although hydrophilic protein interfaces are often involved in the formation of protein complexes, the mechanism how such surfaces assemble is not well characterized. Most of the previous computational studies on protein folding and association have dealt with hydrophobic interfaces [35, 36, 42, 82, 89, 90]. In that case, hydrophobic dewetting has been shown to be an essential driving force in biology [35-37]. In the last years, however, several studies have pointed out that dewetting is rather rare [42-44] because a few polar residues can already prevent the occurrence of dewetting [43, 45]. Taking into account that on average about 70 % of the interfacial residues of protein complexes are hydrophilic including about 37 % charged residues [47], hydrophilic interfaces are clearly of larger importance for assembly. Moreover, most of the large hydrophobic surfaces tend to be buried during the folding process. The association of hydrophilic surfaces is traditionally thought to result from the direct electrostatic interactions between the binding partners. A recent neutron diffraction study showed an example where association between some small peptides was dominated by charge-charge rather than hydrophobic interactions [163].

Electrostatic interactions are well known to be responsible for the tight and fast binding of certain proteins [14, 55, 152, 164] as well as for stabilizing the intermediate complexes [55]. In the absence of electrostatic interactions, the rate of association between the two proteins Barnase and Barstar is on the order of $5 \times 10^5 \text{ s}^{-1} \text{ M}^{-1}$. The favorable electrostatic interactions between both proteins accelerate the association rate by four orders of magnitude to the experimentally determined value of $5 \times 10^9 \text{ s}^{-1} \text{ M}^{-1}$ [164]. Since this acceleration cannot be assigned to an acceleration in the frequency of collisions as it is expected from the Debye model [14] (see 2.1.1.1), this picture suggests that the electrostatic interactions are playing their role by increasing the ratio of fruitful collisions during the intermediate encounter complexes when the proteins are loosely bound and highly hydrated. The atomistic mechanism of the dynamic transformation from the encounter complexes to the stereo-specific complex is still not understood. Thus we present here results from a series of atomistic molecular dynamics (MD) simulations with explicit solvent representation starting from various unbound conformations. These simulations helped us to understand the development of the binding process during the intermediate state and the nature of the energy funnel that leads to the high ratio of fruitful collisions.

The role of water in the binding process is usually considered to be unfavorable. The energetic penalty due to the desolvation of the interfaces is considered the main energetic barrier for the protein-protein association. Let us consider the complex of Barnase and Barstar, as an example, that is known as one of the tightest binding proteins complexes ($k_d = 10^{-14}$). The continuum electrostatics calculation by Sheinerman and Honig [77] found that the net electrostatic effect is close to zero due to the high desolvation energy. Yet, this model ignores the hydrated nature of the interface which can reduce the desolvation penalty as well as the possible role of water mediation. We will argue below that the simple picture of electrostatic interactions screened by a continuum solvent is not able to explain the complicated nature of hydrophilic interactions. In particular, the continuum description of the water ignores the important role that the structure of the water network plays for the interaction between water and solute molecules. Such a role has been well documented for the case of hydrophobic interactions where it also showed an interesting size-dependence [35, 82]. The recent understanding of the hydrophobic effect as a driving force in biology is due to water structure rather than the interaction between the solutes. In contrast, the mechanistic role of the water structure in hydrophilic interactions is poorly understood especially for large hydrophiles such as proteins.

6.2 MD simulations of the binding process

6.2.1 Setting up the binding simulations

The system we studied is the protein-protein complex of Barnase:Barstar from *Bacillus myloliquefaciens*, which binds via a clearly charged binding interface (see Figure 26). This complex has a very high dissociation constant $k_d = 10^{-14}$, $\Delta G = -19 \text{ Kcal. M}^{-1}$ and $\Delta H = -19.3 \text{ Kcal. M}^{-1}$ [165]. The reason behind choosing this system is the availability of a wealth of experimental and computational data for the association of this model complex.

The unbound conformations were generated by displacing the two proteins from the bound state in the crystal complex (pdb code 1RBS) [164] to an interfacial distance of 1.2–2.3 nm (see Figure 27) and rotating Barstar around the line connecting the centers of masses in some cases (see Table 4). The systems were solvated in a rectangular box large enough so that the water extended at least 1.5 nm from the protein surface resulting in box dimensions of (10.2–12.3 nm) \times (6.9–7.2 nm) \times (6.1–6.3 nm). The OPLSAA force field [104] was used for the proteins, and the TIP4P water model [148] was used for the explicit solvent. Four sodium counter ions were added to make the system electrically neutral. All simulations were run with the GROMACS3.3.1 simulation package [156]. Equilibration consisted of 500 steps of steepest-descent energy minimization and a 100 ps long MD simulation with harmonic position restraints using a force constant of 1000 kJ mol⁻¹ nm⁻² for the heavy atoms in the proteins for each system. This was followed by unrestrained MD simulations in the NPT ensemble [1 atm and 310 K]. For this, the system was coupled to a Berendsen external temperature bath [166] using coupling times of 0.1 ps⁻¹ for proteins and solvent and the pressure was kept constant as well by a weak coupling to a Berendsen pressure bath [166] using a coupling time of 1 ps⁻¹. Long-range electrostatic interactions were computed by the particle-mesh Ewald method [157] using Fourier spacing of 0.12 nm. Van der Waals interactions and short range electrostatic interactions were computed within a 1.2 nm cutoff. A time step of 2.0 fs was used.

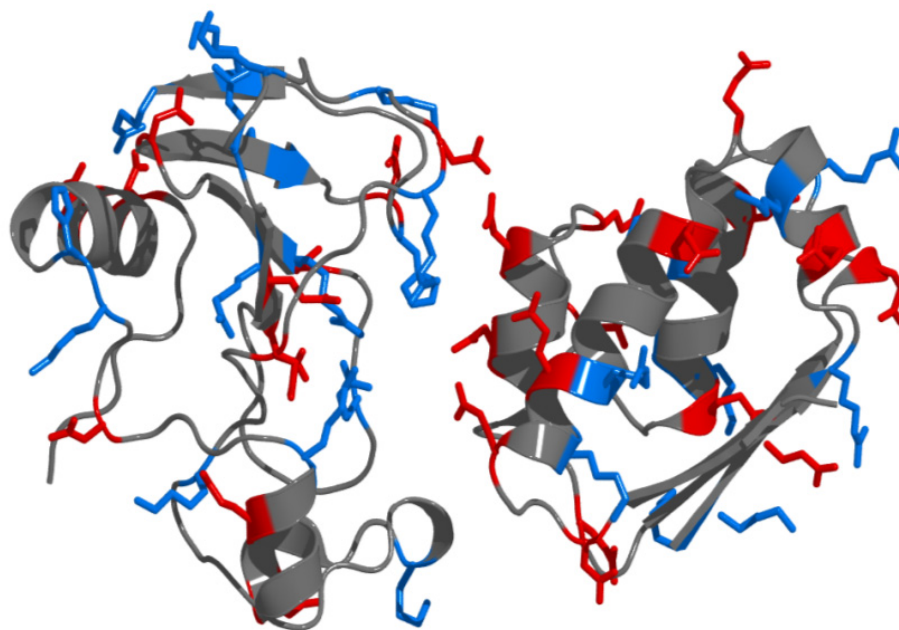


Figure 26. Charged residues in Barnase- Barstar complex.

Positively charged residues are colored blue. Negatively charged residues are colored red. For better visibility, the interfacial distance in the crystal structure of the Barnase-Barstar complex (1brs.pdb) was increased by 5 Å.

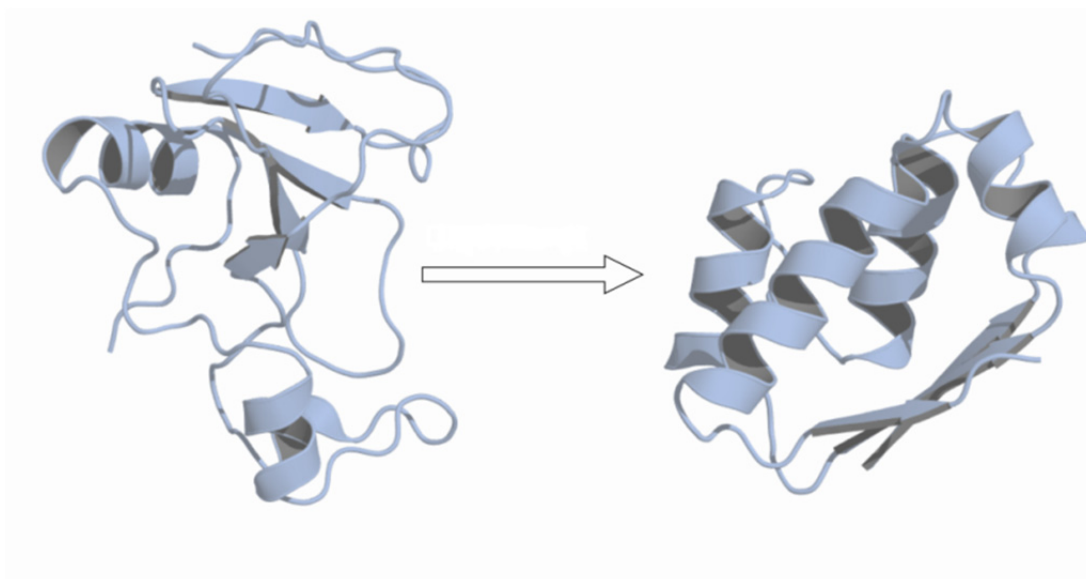


Figure 27. Generation of the starting structures for the simulations.

Barstar (on the right) was displaced from its position in the bound complex along the vector connecting the center of geometry of the two proteins.

6.2.2 MD simulations recovered the experimental stereospecific structures

Nine independent unbiased atomistic molecular dynamics simulations in explicit solvent were started from different unbound conformations. The total length of the simulations is 1.64 μ s. Five simulations out of nine resulted in bound conformations very close to the crystal complex within the simulation time (see Table 4). The five simulations that came close to the crystal structure were extended over hundreds of nanoseconds whereas those simulations that did not come close to the crystal structure were stopped within the first 100 ns. Three of these simulations had a minimum root mean squared deviation (RMSD) less than 0.13 nm from the backbone coordinates of the crystal complex so that they are practically indistinguishable from the native bound conformation (Figure 28). The binding process took place on time scales of hundreds of nanoseconds and consisted of three different phases as we observed in a previous work [36]. Diffusion led to intermediate transient encounter complexes [27], of which some proceeded very close to the stereo-specific complex. The diffusive phase only lasted a very short time with the first contacts forming within 2 ns. The different successful simulations showed that the binding process takes place through many different pathways (Figure 30).

After the short diffusive phase, intermediate complexes were formed with a spatial orientation close to the stereo-specific complex (RMSD < 0.5 nm). The early intermediate complexes only had a few native crystal contacts formed (Figure 29a, b). This is in agreement with double mutant cycle analysis for this complex according to which the transition state for association occurs before most interactions are established [55]. The main conformational changes during the transition to intermediate states are rearrangements of the side chains and of the interfacial water molecules. The conformational changes toward the final stereo-specific complex took place on a time scale of hundreds of nanoseconds; most of the native contacts were formed within 50 to 200 ns (Figure 29a, b).

Simulation	simulation time (ns)	Minimum RMSD (Å)	Average RMSD during the last 10 ns (Å)	Starting position distance(Å)*, rotation**(degree)
SIM1	430	1.2	1.6	15, 0
SIM2	234	1.2	2.0	14, 0
SIM3	275	1.3	2.5	16, 45
SIM4	150	2.6	3.5	20, 45
SIM5	200	3.5	4.1	13, 0

Table 4. List of the successful simulations.

* The displacement distance along the line connecting the centres of geometry.

** The rotational angle of Barstar around the line connecting the centres of geometry.

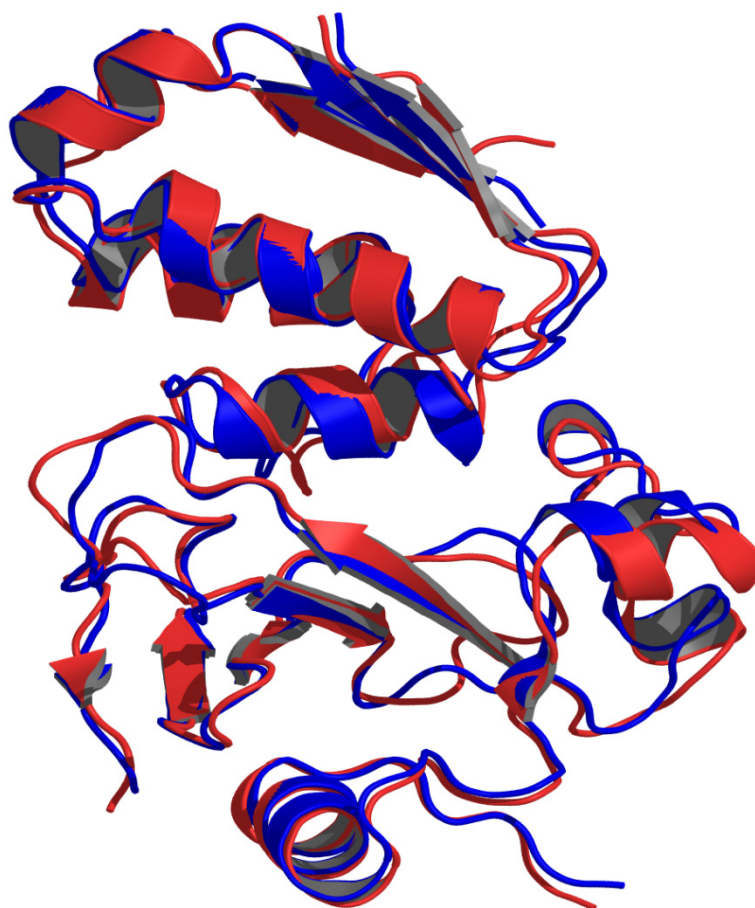


Figure 28. Recovered crystal structure by MD simulations.

A snapshot at 340 ns from simulation SIM1 (red) superimposed on the crystal structure 1BRS (blue) illustrating that the MD simulations almost perfectly recovered the crystal structure.

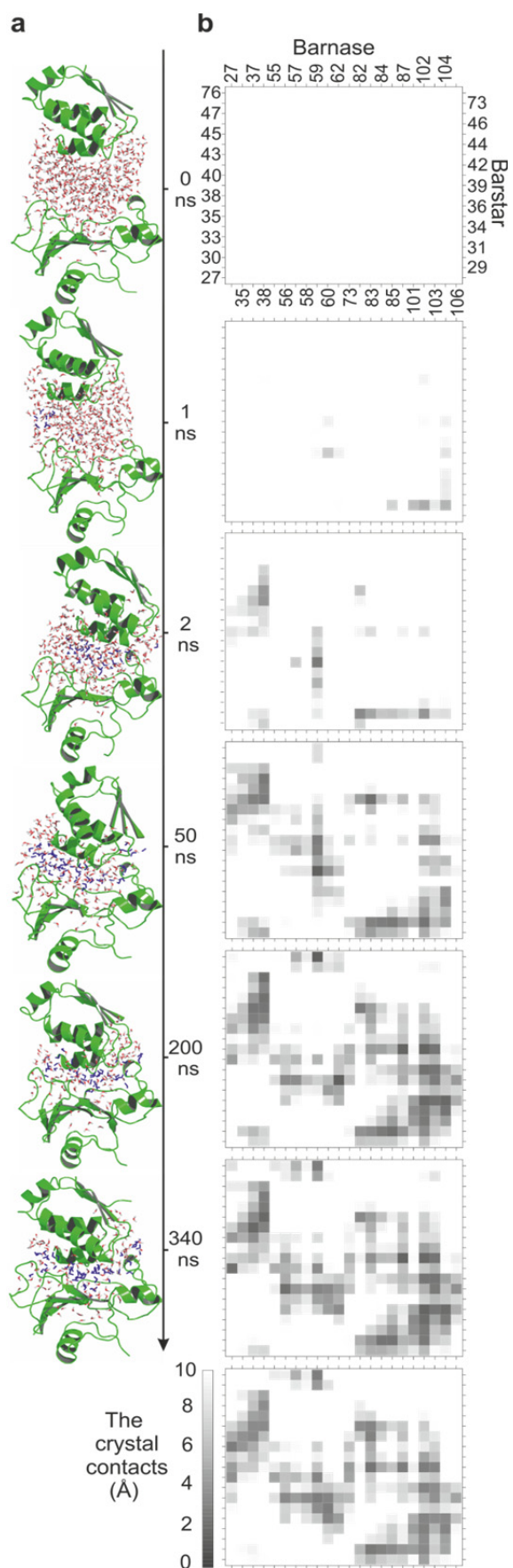


Figure 29. Conformational snapshots along the binding pathway.

(a) Snapshots at different simulation times showing the transformation from the starting structure to the final complex through binding intermediates (from sim1). Only solvent molecules in the interfacial gap are shown. The water molecules that belong to the H-bonded hydration shells of both proteins are colored blue. The graphics were generated using PYMOL [161].

(b) Contact maps between Barnase:Barstar for different snapshots in comparison to the native crystal contacts. The minimum distance between heavy atoms is computed for each pair of the interfaces.

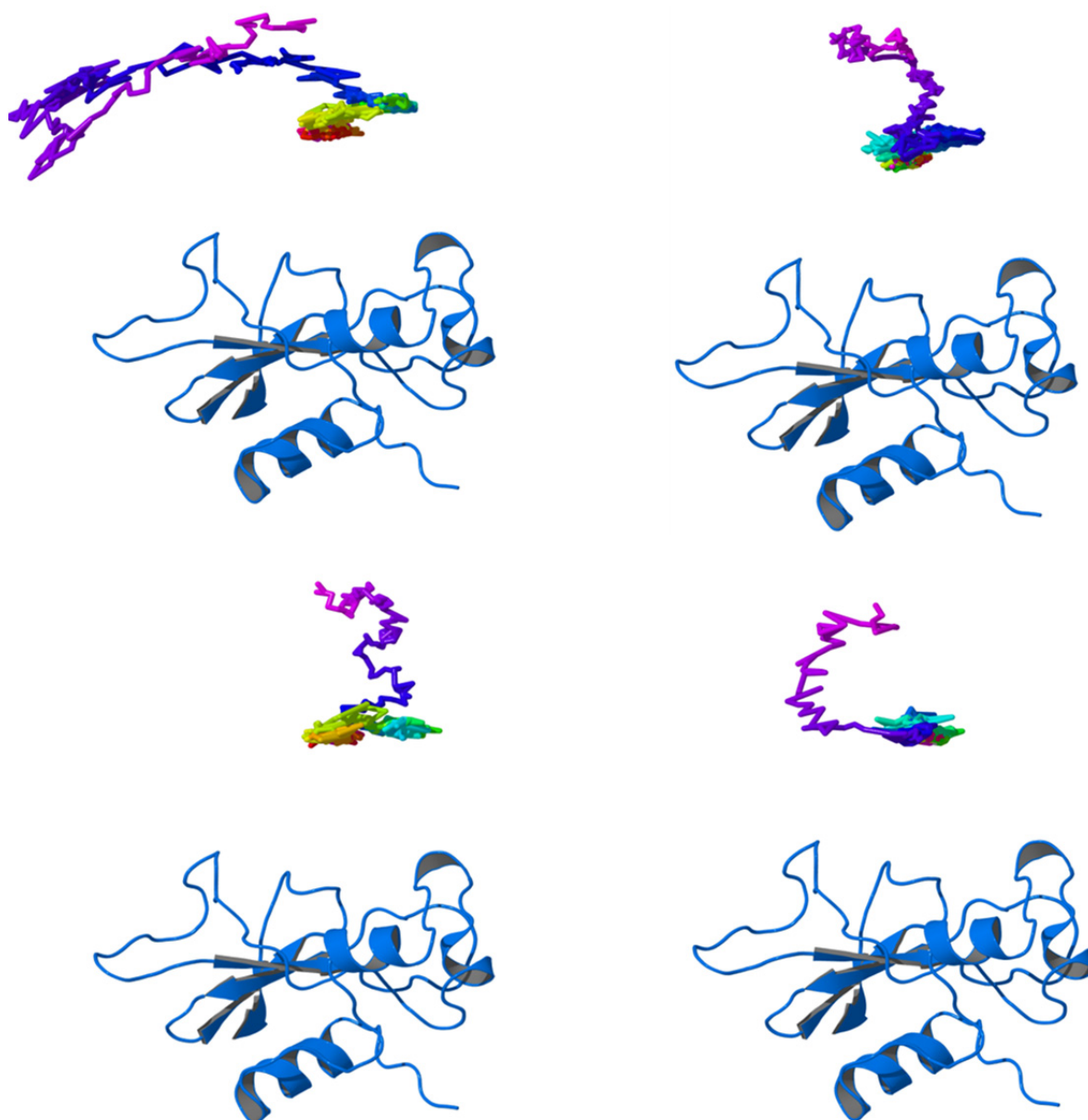


Figure 30. Multiple pathways for the association process

These binding trajectories were generated by superimposing all simulation snapshots on the starting structure of Barnase. Thus, the trajectories visualize the relative motion of the center of geometry of Barstar around Barnase for four different successful simulations which lead to the stereospecific complex. The trajectories are color coded from violet to red to indicate the time line of the binding process. The starting structure for the simulations is colored violet. The stereospecific complex is colored red.

6.3 Hydrogen bonded network between the proteins

During the simulations, the most noticeable property of the intermediate complexes is the highly solvated state of the interfaces. 60–80 water molecules were constantly present in the interfacial gap (within 0.5 nm of both interfaces) showing that the interfaces did not undergo dehydration but rather remained highly hydrated even in the final stereo-specific complex. To further characterize the properties of these interfacial waters, we computed the water density in the interfacial gap and its distance dependence (Figure 31a). In comparison to our previous study on hydrophobic interfaces (see chapter 5.4.2), the interfacial water density did not show any dewetting for the Barnase:Barstar system even at close separation distances. On the contrary, the water density was never lower than that in bulk and was even a bit higher at closer distance. Moreover, the fluctuations of the water density for the complex were comparable to those for the individual proteins and even smaller at close distances (Figure 31b). This means that the hydration shells do not become softer as was found for hydrophobic interfaces [167, 168].

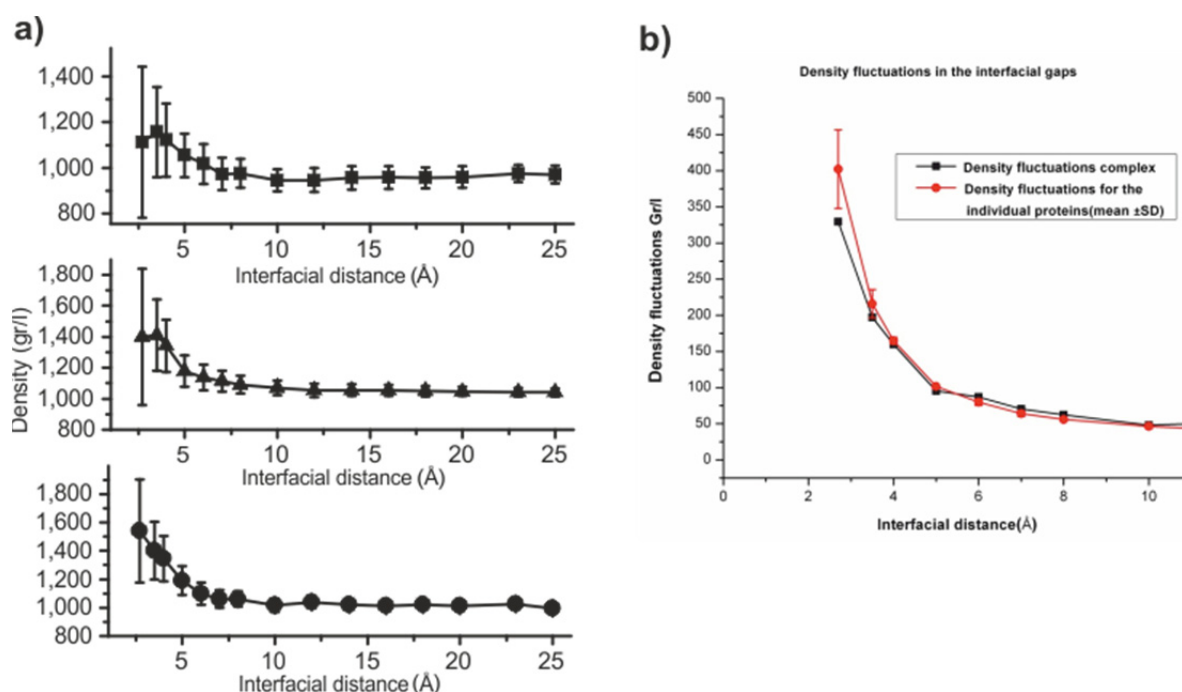


Figure 31. Water density in the interfacial gap

(a) Water density in the interfacial gap for different interfacial distances for the complex (squares) and around the individual proteins (Barnase (triangles) and Barstar (circles)). The water density was averaged over the last 1.7 ns for each 2 ns position restrained MD simulation. The error bars represent the standard deviation when taking the average. The gap was defined as in the calculation of the electric susceptibility (see 6.4.1.2). The gaps for the independent chains were defined by superposition of the corresponding part from the complex simulation to all snapshots of the independent protein-water system. (b) The fluctuation of the density of water in the interfacial gap. The density fluctuation of the individual proteins was calculated as the average of the fluctuation of the individual proteins.

Close analysis of the hydrogen-bonded water hydration shells (up to two shells for each protein) showed that the binding process started by the fusion of the hydrogen-bonded water shells (Figure 32 ; see the blue-colored water in Figure 29a). Early on, tight water bridges form that make indirect contacts between the interfaces before the direct contacts are established (Figure 29b). This shaking hands behavior indicates an important adaptation of the water structure in the interfacial gap during the binding.

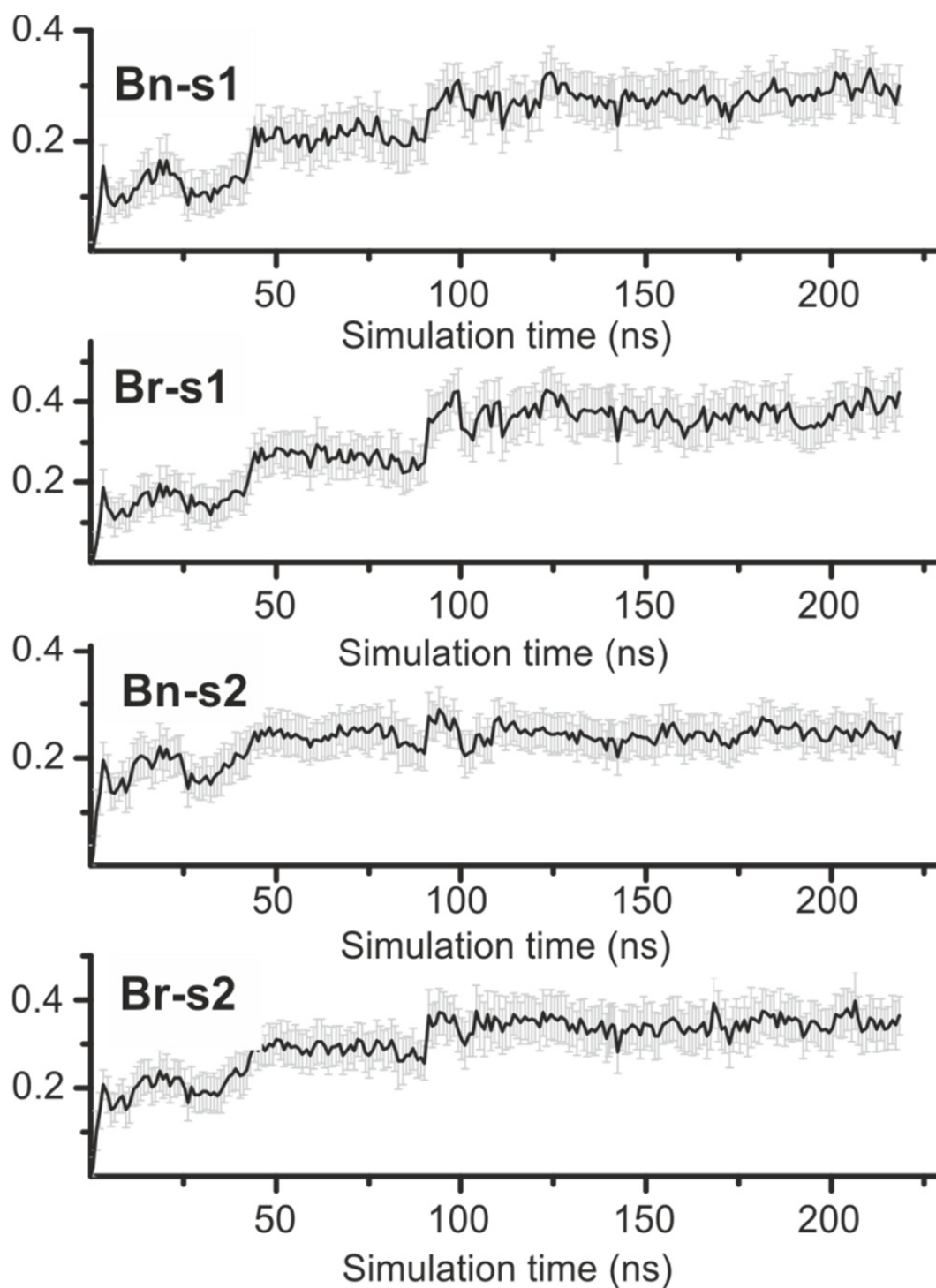


Figure 32. Development of the water bridges between the interfaces

The hydrogen-bonded water shells were defined as the water molecules that are directly bound to the interfaces with hydrogen bonds (the first shell s1) and the water molecules bound to s1 (the second shell s2). A water molecule in a shell was considered shared if it belonged to the s1 or s2 shells of both proteins. The shared fraction with the other chain was computed for each shell for both Barnase (Bn) and Barstar (Br). The values were averaged over time intervals of 1 ns and the standard deviations are presented as the error bars.

6.3.1 Water-mediated interfacial connectivity

Water can mediate the interactions between two proteins by forming water bridges of different lengths and strengths. To account for the importance of water mediation, we introduce here a new way of characterizing the connectivity between the interfaces via the

hydrogen bond network using the concept of maximum flow from mathematical graph theory [169]. This flow is a scalar measure for the connectivity between any two nodes in a network with weighted edges taking into account the possible pathways and the weights of them (see Figure 33b).

To study the water mediation between the interfaces, we defined a hydrogen bonding network using as nodes all hydrogen bond donating and accepting atoms at the interfaces and the water molecules in the interfacial gap (up to 0.5 nm distances from both interfaces). The hydrogen bonds were defined based on the criterion that the distance acceptor-donor should be less or equal than 0.35 nm and the angle $H - D \cdots A$ should be less or equal than 30° [170]. The hydrogen bonds were interpreted as edges connecting the nodes with their explicit energy as weights. We used the explicit value for the hydrogen bond energy calculated according to the explicit hydrogen bond term in the XPLOR force field (see 3.4.1.1.1 for explanation) [124]. (Control calculations using the hydrogen-bond term from the GRID force field showed very similar behaviors, data not shown.) As the interfaces each contain many nodes as acceptor or donors we consider each combined interface as one node when calculating the flow between the interfaces. This can be performed by inserting a virtual super node for each interface and connecting all the nodes of the corresponding interface to this node with edges of infinite capacities [169]. In this way the flow between the interfaces (the super nodes) corresponds to the magnitude of the water-mediated connectivity. The calculations were performed using the R package Igraph [171].

Interestingly, the total interfacial water-mediated connectivity was found to have significant values from the early stages of the transient encounter onwards, even before specific interactions are formed (Figure 33a). From that point it increased about two fold to three fold during the encounter complex states especially during the first 100 ns when the rearrangement of the interfacial water molecules improved their role in mediating the binding of the interfaces. Control simulations showed that the interfacial water-mediated connectivity originates mainly from the hydrophilic and charged residues at the interfaces because mutating these residues to more hydrophobic residues abolished most of the interfacial water connectivity (see 6.3.1.1 below ; Figure 34).

6.3.1.1 Control simulation for interfacial water connectivity with more hydrophobic interfaces

Five snapshots were taken from the trajectory sim1. For each snapshot, we ran MD simulations of 1 ns length with harmonic position restraints applied to the backbone atoms of the proteins to get good statistics on the interfacial connectivity. To compare the computed interfacial water-mediated connectivity of the wild-type proteins to more hydrophobic interfaces of the same shape, we reduced the hydrophilicity of the interfaces by introducing experimentally known mutations [55] that are known to be involved in stabilizing the interaction: In Barnase we engineered mutations in four residues K27A, R83Q, R87A and H102A. In Barstar we engineered mutations in three residues D39A, Y29A and E76A [55]. For the mutant protein complexes containing seven mutations, we performed an MD simulation of 2 ns length with harmonic position restraints applied to the backbone atoms of the protein. Data from the last 1 ns was used for the analysis of the interfacial connectivity. Although the mutated interfaces still contain the majority of the hydrogen bonding acceptor and donor atoms (59 donor \times 114 acceptor) in comparison to the wild-type (68 donor \times 123 acceptor) the interfacial water mediated connectivity determined from the MD simulations of the mutant proteins was very low in comparison with the wild-type (**Figure 34**). This demonstrates that the hydrogen bonding network is mainly oriented and stabilized by the charged and hydrophilic residues.

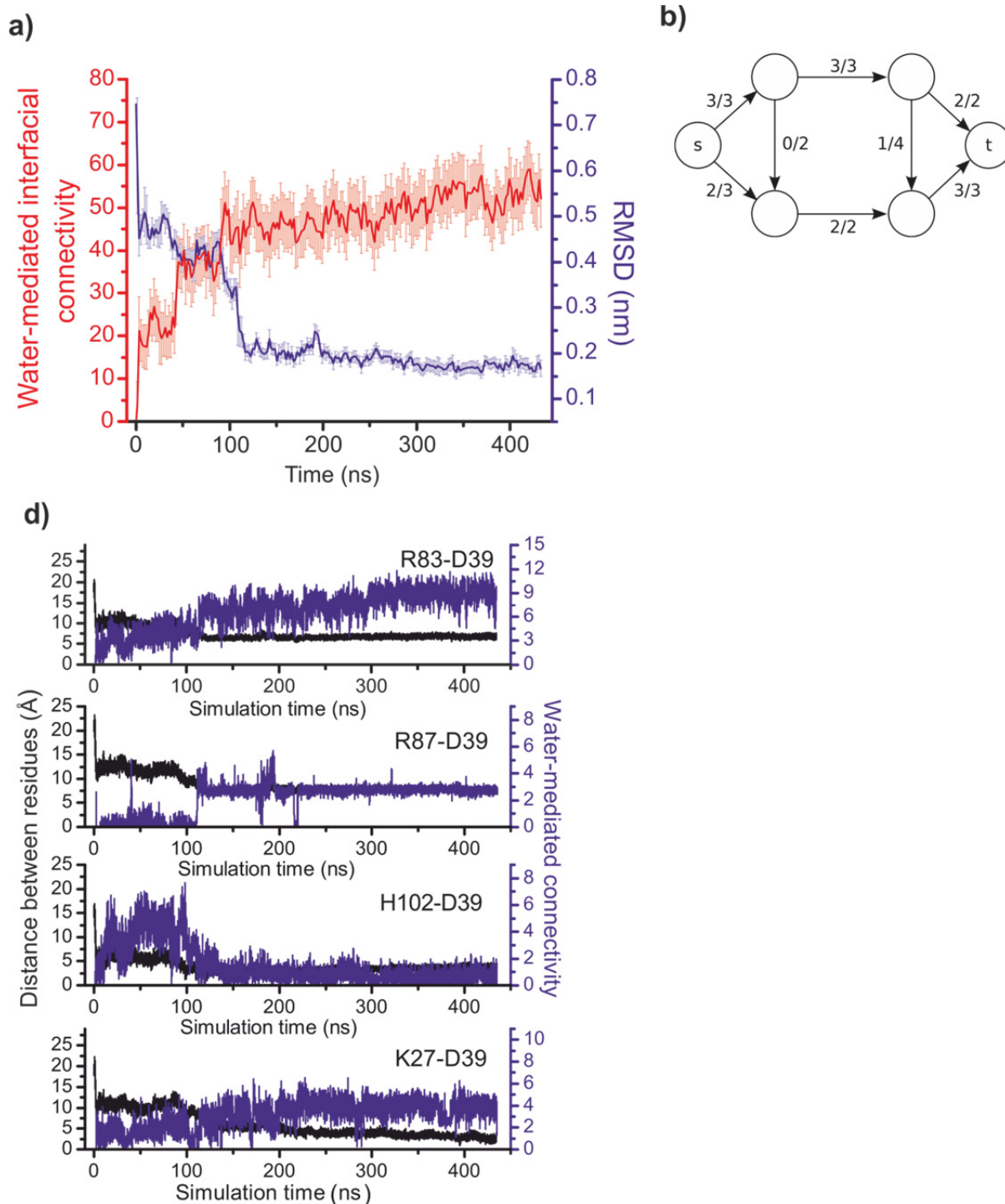


Figure33. Water-mediated interfacial connectivity.

(a) Water-mediated interfacial connectivity (red line from sim1) and the root mean squared deviation RMSD (blue line). The values were averaged over intervals of 2 ns. The error bars represent the standard deviations. The water-mediated interfacial connectivity increased throughout the binding pathway. It shows a clear anti-correlation with the RMSD of -0.75. **(b)** Representation of the idea of the maximum flow in a network. The network consists of many nodes connected by edges of certain weights (capacity). The maximum flow from the source *s* to the sink *t* is defined by the number of edges and their capacities. **(d)** Examples of the pair-wise water-mediated connectivity. Four different pairs are shown with known high cooperative energy. The water-mediated connectivity is formed before specific contacts are established. The connectivity values were averaged over intervals of 1 ns.

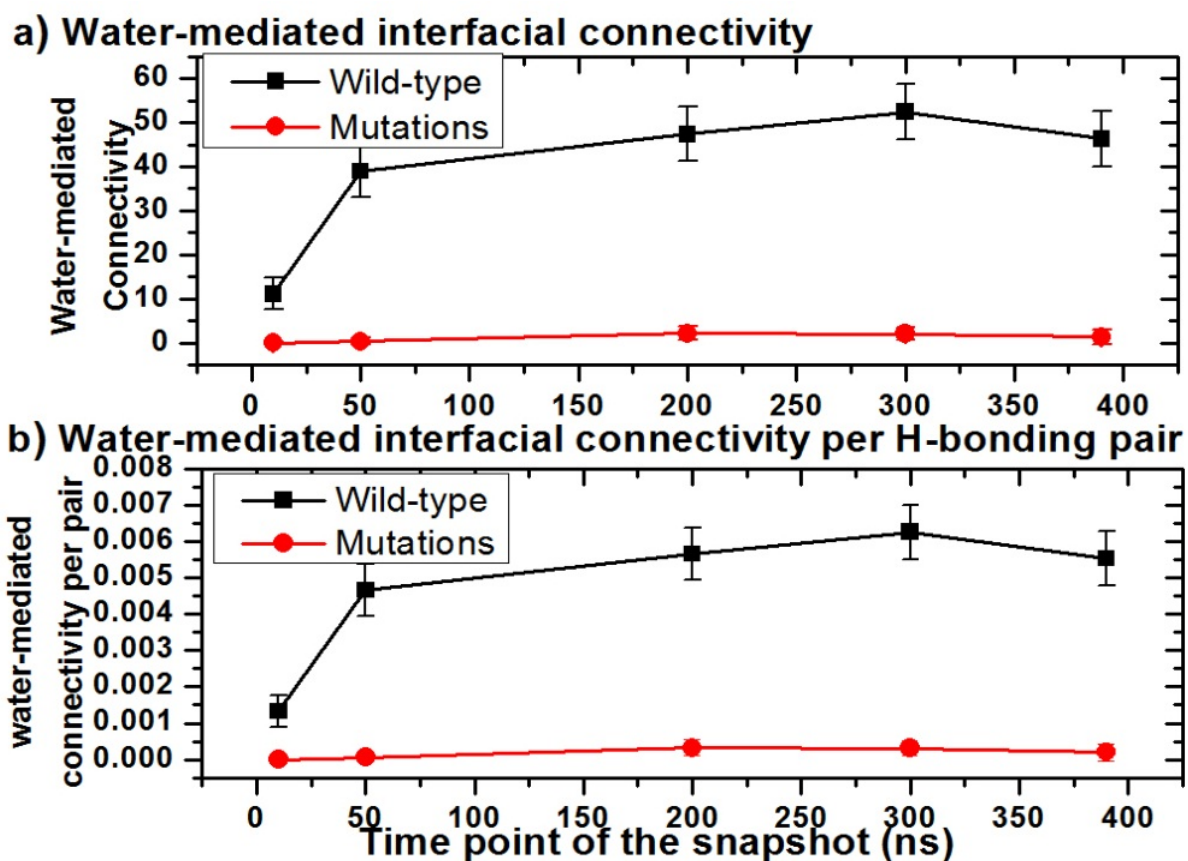


Figure 34. Comparison of the interfacial water-mediated between hydrophilic and hydrophobic interfaces

- (a) Black line: the connectivity for snapshots from the natural binding pathway (from sim1). Every data point corresponds to an MD simulation of 1 ns length starting from the corresponding snapshot with harmonic position restraints applied to the backbone atoms. The connectivity was averaged over snapshots of 0.5 ps intervals. Red line: the connectivity for the snapshots of the complex with seven mutations at the interfaces to reduce their hydrophilicity (see the text in 6.3.1.1). Here every data point corresponds to an MD simulation of 2 ns length with harmonic position restraints applied to the backbone atoms. The connectivity was averaged over snapshots of 0.5 ps length during the last 1 ns of the simulation.
- (b) The values of the connectivity scaled by the number of all possible hydrogen bonding pairs (donor-acceptors). The black line shows the data for the wild-type: number of pairs = 8364 (68 donor \times 123 acceptor). The red line shows the data for the mutations: number of pairs = 6726 (59 donor \times 114 acceptor).

6.3.2 Pairwise water mediation

A detailed analysis of the pairwise water mediation between all possible pairs of interfacial residues showed that most amino acid pairs that form tight, specific interactions in the crystal complex have significant values of the water-mediated connectivity between them already in the early stages of binding before they form direct interactions (examples are presented in Figure 33d). This emphasizes the role of water bridges in guiding the formation of specific interactions and complements the direct interactions of those residues. The water-mediated connectivity stays significant until the stereo-specific complex. Interestingly, some pairs showed experimentally measured cooperative interactions even when they are separated by distances of 0.4-1.5 nm in the bound state [55]. The cooperative nature of such weak long-range interactions is one of the most tantalizing issues in protein-protein interaction. This type of interactions cannot simply be explained by long range electrostatic interactions because many non-charged residues are involved as well. Here, we find

statistically significant water-mediated connectivities for these pairs during the intermediate states and in the stereo-specific complex (Figure 35a). This new finding appears a very plausible model to explain such indirect physical interactions.

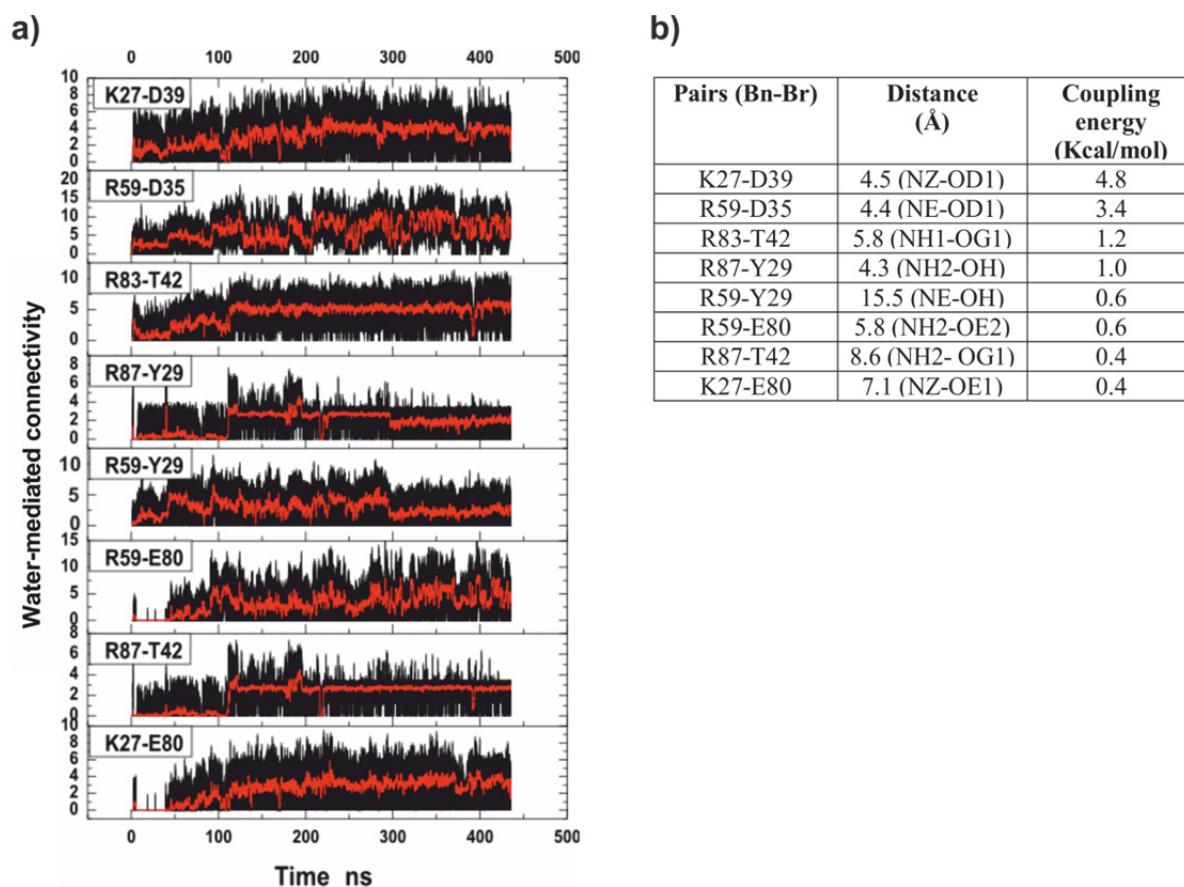


Figure 35. Water-mediated connectivity between the interfacial residues with known long-range cooperative interaction

- (a) Time course of the pairwise water-mediated connectivity between the interfacial residues with known long-range cooperative interaction. The red line connects the averaged values computed over 1 ns intervals.
- (b) Pairs with long-range cooperative interactions. The distances are taken from the crystal structure of the bound complex (1brs.pdb). The coupling energies were measured and reported in reference [55].

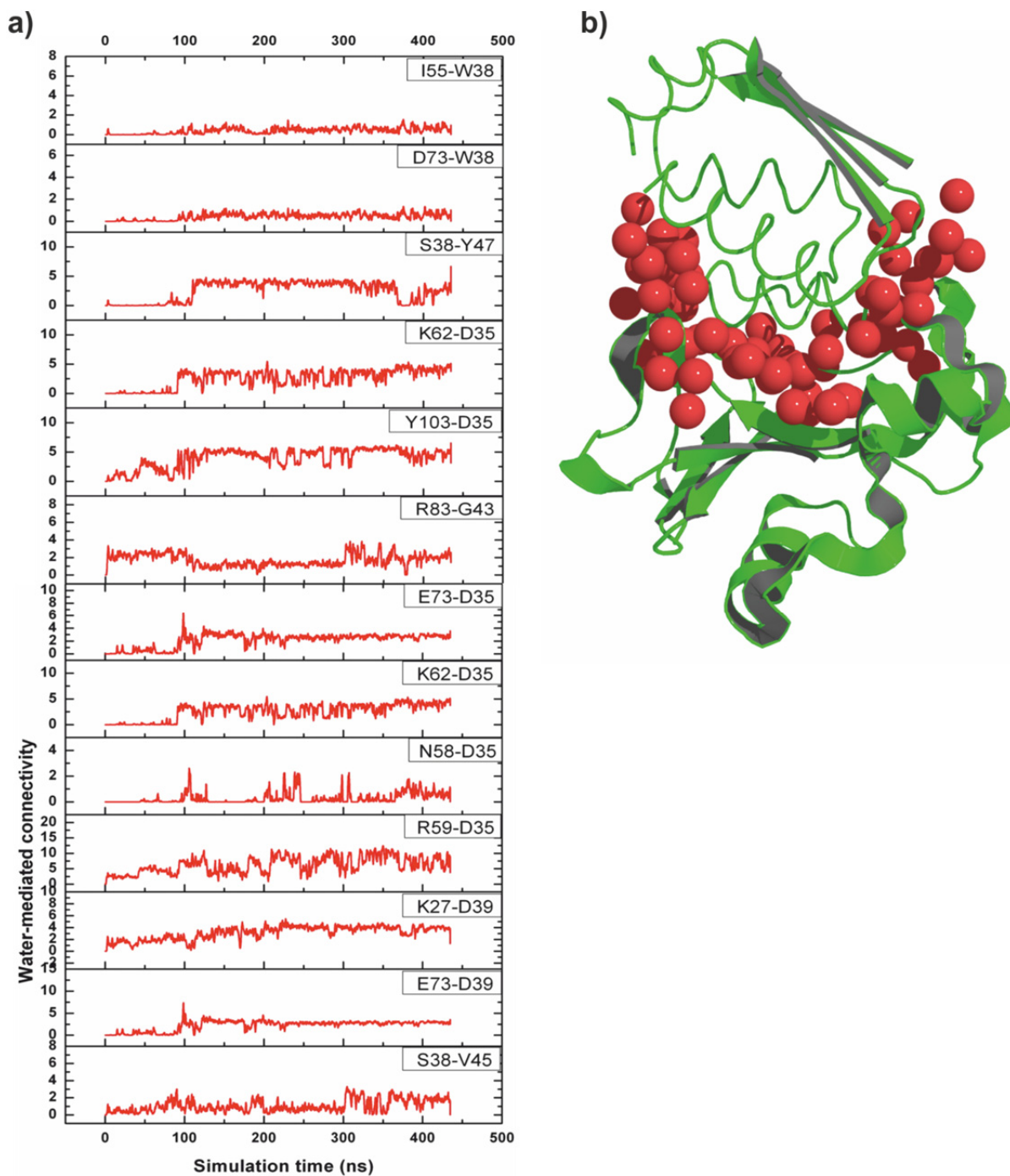


Figure 36. Comparison of the water mediation between the simulations and the crystal structure

- (a) Time course of the pairwise water-mediated connectivity between the interfacial residues that make indirect contacts via water molecules in the crystal structure [164]. The lines connect the averaged values computed over intervals of 1 ns length.
- (b) The interfacial water molecules (shown as red spheres) in the crystal structure of the Barnase-Barstar complex. PDB code 1BRS (chains A and D).

6.4 Calculation of the dielectric constant of water molecules in the interfacial gaps

The role of the electrostatic interaction as a driving force for protein-protein interactions was studied in this work on the association process of Barnase-Barstar as an example for the association of the charged interfaces. Electrostatic interaction may already play an important role during the diffusive phase and when encounter complexes form. During the diffusive phase, the proteins are still separated by water and the magnitude of the electrostatic interactions between the proteins is unclear as we do not know how strongly water screens the electrostatic interaction. Water is generally considered as a very good solvent because of its high dielectric constant. The high value of the dielectric constant of the water means that the electrostatic interactions are weaker by approximately a factor (1/80). The commonly used continuum electrostatic methods for the computation of the electrostatic energy consider all cavities to be filled by water and thus assign the dielectric constant of bulk water. Such a model cannot explain how important the electrostatic interaction is. We expected that the water in between the proteins has a reduced dielectric constant in comparison to the value of bulk water. This would mean that the electrostatic interactions are not strongly screened by the dielectric solvent and thus play an important role in bringing the proteins together during the diffusive phase. The water in between the proteins (in the interfacial gap) showed the formation of a network of hydrogen bonds that mediates between the interfaces. Such an ordering in the structure of the water in the gap should affect its dielectric constant as the water molecules cannot reorient their dipoles easily to counteract the electrostatic interactions. For these reasons we have studied the dielectric constant of the water in the interfacial gap between the proteins when the proteins were kept at different interfacial distances (0.35 – 5.00 nm).

The study tried to address the following points:

- Is the dielectric constant of the water in the interfacial gap lower than the value of the bulk water?
- How does the dielectric constant change with respect to the distance between the proteins?
- Is the dielectric behavior of the water in the interfacial gap isotropic? Or does a preferred directionality exist for the electrostatic interactions in a certain direction?

The calculation of the dielectric constant of the water molecules in the interfacial gap between the proteins was based on the assumption that the sample is clearly anisotropic in one dimension (perpendicular to the interfaces). The sample was assumed to be isotropic in the parallel directions. The result of the calculations shows that the dielectric permittivities in both parallel directions are very similar as we assumed. Under this assumption, the tensor of the dielectric permittivity reduces to the diagonal form with two independent values (perpendicular and parallel) [172].

The calculation of the dielectric permittivity was performed using two different formulas, the fluctuation formula (eq. 67) and with the polarization formula (eq. 70). The detail about our derivation of these formulas is presented in chapter 4. The fluctuation formula gives more accurate results as mentioned before (chapter 4). The advantage of the polarization formula is the fast convergence of the polarization so that its evaluation does not require long simulation times. In contrast, the fluctuation formula (eq. (67)) needs much longer simulation times to converge [143]. An additional advantage of using the polarization formula is the possibility to check if the sample is still responding linearly to the change in the electric field

and is not in the saturation regime. The derivation of the formulas actually assumed that the dielectric of the sample is not in the saturation regime.

6.4.1 Dielectric permittivity of the water in the interfacial gap using the fluctuation formula

6.4.1.1 MD simulations

The starting structures of the simulations were generated from the crystal structure of the complex between Barnase and Barstar [164] using chains A and D. Different starting structures were generated by increasing the interfacial distance along the line connecting the centers of geometry of the interfaces. The interfacial distances ranged from 0.35 nm to 5.0 nm. The x axis is perpendicular to the interfaces while y and z were parallel. Four different MD simulations were performed for each interfacial distance point with two different symmetric external fields (-0.1 and 0.1 V/nm) once in the perpendicular direction to the interfaces (the x axis) and once in the parallel direction (the y axis). These electric fields are much weaker than required for saturating behavior [146]. Moreover, the response of the sample to electric fields of different strength showed that the sample is not dielectrically saturated at such fields (see 6.4.3). The goal of performing four MD simulations for each displacement distance was to get good statistics about the values. As is clear from the derivation, the fluctuation formula is valid even when an external field is applied [137].

The OPLSAA force field [104] was used for the proteins, and the TIP4P water model [148] was used for the explicit solvent. The systems were solvated in a rectangular box large enough so that the water extended at least 1.5 nm from the protein surface. Four sodium counter ions were added to make the system electrically neutral. All simulations were run with the GROMACS simulation package [156]. The energy minimization consisted of 500 steps of steepest-descent energy minimization. Harmonic position restraints were applied to the heavy atoms of both proteins to keep the interfacial distance using a force constant of 1000 kJ mol⁻¹ nm⁻². The MD simulations were performed in the NPT ensemble [1 atm and 310 K]. For this, the system was coupled to a Berendsen external temperature bath [166] using coupling times of 0.1 ps⁻¹ for proteins and solvent and the pressure was kept constant as well by a weak coupling to a Berendsen pressure bath [166] using a coupling time of 1 ps⁻¹. Long-range electrostatic interactions were computed by the particle-mesh Ewald method [157] using Fourier spacing of 0.12 nm. Van der Waals interactions and short range electrostatic interactions were computed within a 1.2 nm cutoff. A time step of 2.0 fs was used. The external electric field was applied as it is implemented in GROMACS. Each simulation was performed for 10 ns. The snapshots for analysis were collected every 0.25 ps from the simulation interval 2 ns to 10 ns thus 32 × 10³ snapshots were used in the analysis of each trajectory.

6.4.1.2 Analysis of the trajectory

The interfacial gap between the proteins was defined as the volume within “the displacement distance” d from both interfaces and not more than 1.5 nm from the line connecting the centers of geometry of the interfaces. A water molecule was considered in the gap if its oxygen atom was located inside the gap. The total dipole of the water molecules inside the gap was calculated using equation (37):

$$\mathbf{M} = \sum_1^n q_i \mathbf{r}_i$$

The dielectric permittivity was calculated for all the simulations corresponding to the different interfacial distances (0.35–5.0 nm) by using the fluctuation formula (eq. 67) that we

have derived for this purpose. The details about the derivation and calculation of the dielectric permittivity are presented in section 4.3. For each distance point, the value of the dielectric permittivity was calculated from four MD simulations of 10 ns length performed with position restrained protein backbone atoms. The four permittivities reported for every distance point are average values from four simulations. The values shown in Figure 37 are taken relative to the value of bulk water for the TIP4P water model [173].

6.4.1.3 Results

Figure 37 (the black line) shows the relative electric permittivity in perpendicular direction to the interfaces for each interfacial distance. The notion “relative” means that it was computed relative to the computed value for bulk water (53 ± 2 with the TIP4P water model). For each distance point, the value of the dielectric permittivity was calculated from four MD simulations. For each distance point, the dielectric permittivity was calculated in the three directions; the perpendicular (the x axis) and the parallel directions (y and z). Interestingly, the water between the two hydrophilic protein surfaces showed a strongly decreased dielectric permittivity that extended beyond 3 nm of interfacial distances. The perpendicular permittivity (the black line) between the two hydrophilic protein surfaces showed strongly decreased values (relatively to the bulk value) that extended beyond 3 nm of interfacial distances. Its value is less than 50% of the bulk value for interfacial distances closer than 1.2 nm. Interestingly, our calculations showed that the dielectric properties of water in the interfacial gap are strongly anisotropic. The comparison with the computed values for the two parallel permittivities (y and z directions) showed that the dielectric properties of water in the interfacial gap are strongly anisotropic. The perpendicular permittivity decreases more strongly than that along the parallel directions up to distances of 3 nm. This means that the water generates a preferred directionality for the electrostatic interactions in the perpendicular direction that drives the interfaces toward each other (association) rather than in the parallel direction.

We note that the computed permittivity did not converge to the value of the bulk water even at 5 nm separation. This can be partially explained by the existence of several shells of water molecules around the individual proteins which contribute with a low permittivity to the apparent average value of the permittivity in the interfacial gap. Thus, we emphasized that even at 3 or 5 nm separation between the two protein surfaces one should not expect to obtain the bulk permittivity. Let us assume an idealized case where the water molecules in a layer of about 1 nm thickness around both proteins are ordered with a permittivity of ca. 25 and the water molecules in the remaining interfacial gap have a bulk permittivity of 54 (which is the value that has been reported in the literature for TIP4P water [148]). Let us also assume that the interfacial gap volume has a shape of a perfect cylinder. Then, at 3 nm distance of the two proteins, averaging over the water molecules in the interfacial gap volume would give a permittivity of $1/3 * (25 + 54 + 25) = 34.7$. At 5 nm distance, one would obtain $1/5 * (25 + 3*54 + 25) = 42.4$. Thus it is not surprising to obtain different values for the perpendicular and the parallel permittivity than from the bulk value even when the proteins are separated by a few nanometers. Moreover, at large distance separation, the interfacial gap has an inhomogeneous dielectric permittivity which may have affected the computed value of the apparent permittivity.

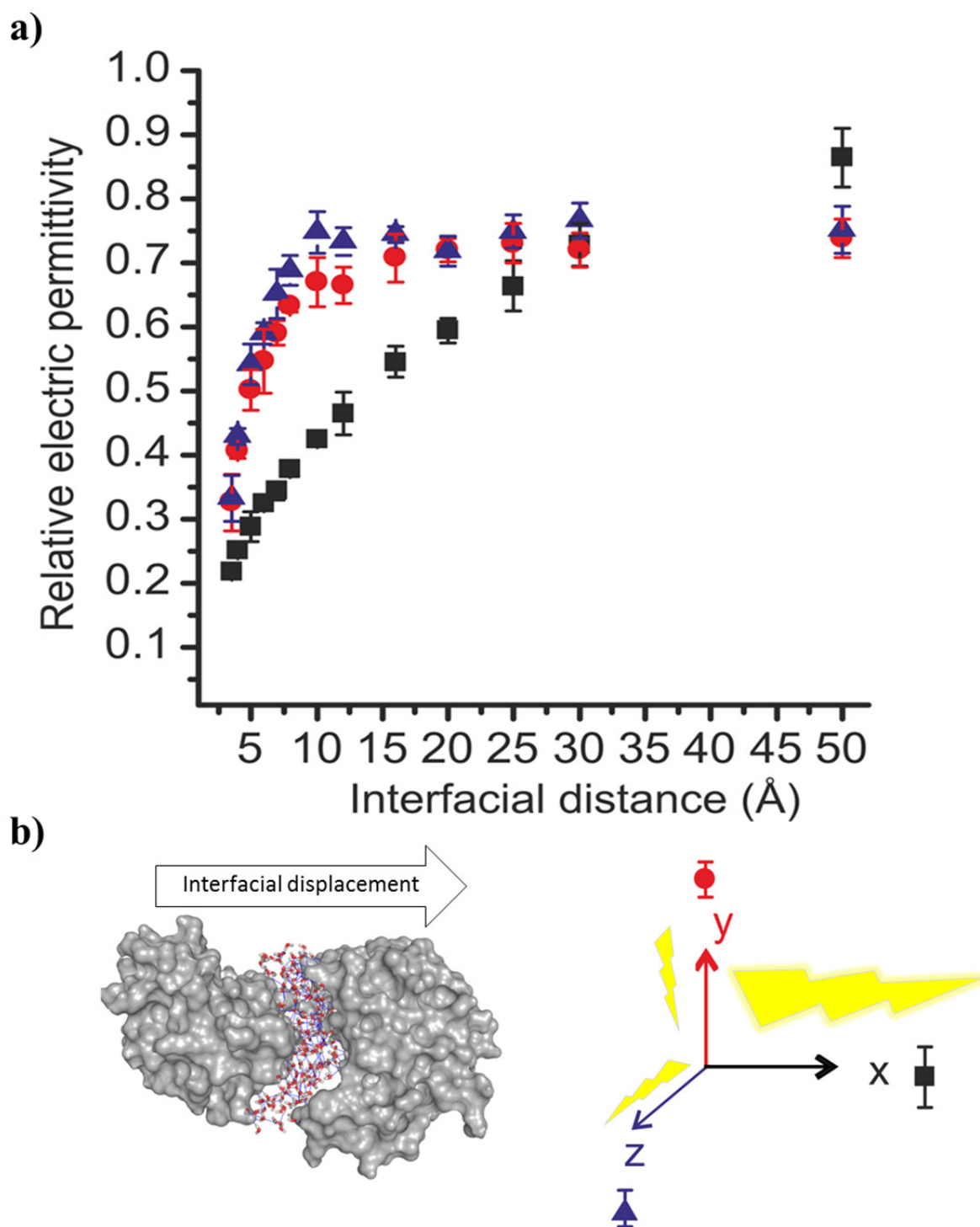


Figure 37. Dielectric permittivity of the water in the interfacial gap

- (a) The local dielectric behavior in the interfacial gaps relative to the value for bulk water. The permittivity was calculated for the perpendicular direction (along the x axis) (black squares) and for the two parallel directions (red circles for the y axis and blue triangles for the z axis). The permittivity of the bulk water modeled by tip4p water model [148] was computed as 53 ± 2 . The values reported are taken as averages from the four simulations for every distance point and the error bars represent the standard deviation.
- (b) Representation of the defined interfacial gap. The interfacial gap was defined as all points within “the displacement distance” d from both interfaces and not more than 1.5 nm from the line connecting the centers of geometry of the interfaces.

To estimate how strongly the inhomogeneous environment around the interfacial gap may affect the results from eq. 67, we first write eq. 67 in the general form. If we do not assume $\varepsilon_s = \varepsilon$ then the derivation leads to (compare eq. 35 in [175]):

$$\frac{(\varepsilon - 1)(2\varepsilon_s + 1)}{2\varepsilon_s + \varepsilon} = \frac{4\pi}{VkT} \left(\langle (\mathbf{M} \cdot \mathbf{e})^2 \rangle_{\mathbf{E}_d} - \langle \mathbf{M} \cdot \mathbf{e} \rangle_{\mathbf{E}_d}^2 \right) \quad (71)$$

Let us consider a sample with computed dielectric constant of $\varepsilon = 25$ according to eq. 67 and let us consider two limiting cases for ε_s . As lower boundary for the dielectric around the interfacial gap we take the value of $\varepsilon_s = 20$ which was computed by Simonson [176] at the surface of various proteins (Simonson gave values of 15 - 25). As upper boundary we take the value of bulk TIP4P water which is $\varepsilon_s \approx 53$. Inserting these two values into the equation above yields $\varepsilon = 27$ and $\varepsilon = 20$ respectively. Thus, varying the dielectric permittivity of the surrounding by a factor of three leads only to small variations in the computed values of epsilon.

Another important point to consider is whether eq. 67 can be applied to a system with ions. This may happen, for example, when an ion migrates into the probe volume during the simulation. Simonson showed that even those situations with ions present in the interfacial gap are handled correctly as the Kirkwood fluctuation formula was shown to be valid for systems with nonzero net charge (see appendix 1 in reference [176]). In such a case the electric momentum of the sample is used. While this momentum depends on the choice of the origin (not like the dipole momentum of polar molecules), its fluctuation is independent of the origin. Moreover, in the system considered here the ions rarely entered into the interfacial gap as this lower dielectric region is energetically much less favourable than the surrounding high dielectric bulk.

For interfacial distances larger than 3 nm, the values of the perpendicular permittivity seem to converge faster to the value of bulk water than that of the parallel permittivity. At this point, we emphasize that these values at larger inter-protein distances are certainly affected by interactions with the periodic copies of the two charged proteins in neighbouring simulation boxes. The effect of the boundary copies on the interaction is unavoidable and cannot be neglected between Barnase and Barstar as was discussed in a recent study by Wang and Helms [177].

6.4.2 Dielectric permittivity of the water in the interfacial gap using the polarization formula

As a check of the stability of our methodology, the dielectric properties of the water in the gap between the interfaces were characterized again using the polarization method (eq. 70):

$$\chi_e = \varepsilon - 1 = \frac{4\pi}{V} \frac{\Delta \langle \mathbf{M} \cdot \mathbf{e} \rangle}{\Delta (\mathbf{E}_0 \cdot \mathbf{e})} = 4\pi \frac{\Delta \langle \mathbf{P} \cdot \mathbf{e} \rangle}{\Delta (\mathbf{E}_0 \cdot \mathbf{e})}$$

Details about the derivation of this equation are presented in section 4.4.

The disadvantage of this equation is that it requires two MD simulations at two different strengths of the externally applied field. Only then, one can calculate the change in the polarization $\Delta \langle \mathbf{P} \cdot \mathbf{e} \rangle$ as a response to a known change in the strength of the external

field $\Delta(\mathbf{E}_0 \cdot \mathbf{e})$. Similar approaches were used by Watts [178] and were applied to an ion channel by Sansom [179]. However, the approach of Watt did not distinguish the directing field from the macroscopic field and was criticized by Neumann [180]. The disadvantage of this equation is that it requires two MD simulations at two different strengths of the externally applied field to be able to calculate the change in the polarization $\Delta\langle\mathbf{P} \cdot \mathbf{e}\rangle$ as a response to a known change in the strength of the external field $\Delta(\mathbf{E}_0 \cdot \mathbf{e})$.

The dependency of the dielectric properties on the interfacial distances was studied for interfacial distances between 0.35 and 3.0 nm. We applied two different symmetric external fields (-0.1 and 0.1 V/nm). These field strengths are still clearly in the linear response regime around the value of the field generated by the system and thus suitable for estimation of the dielectric properties. These two values were used for estimating the changes of the dielectric properties of the interfacial water upon changing the interfacial distance (0.35 – 3.00 nm). For each interfacial gap distance two MD simulations of 4 ns length were conducted with external field strengths of $E_1=-0.1$ and $E_2=0.1$ V/nm. The fields were once applied perpendicular to the interfaces (the x axis). The same procedure and simulations were repeated to estimate the parallel susceptibility along the y axis. For comparison, we also applied the same calculations to the water molecules in comparable virtual gaps of the same volume and shape in a simulation for a water box as well as in simulations for each protein independently.

The same procedures were applied to bulk water simulations and to the individual proteins. The interesting quantity is the ratio:

$$\frac{\chi_{e \text{ complex}}}{\chi_{e \text{ bulk}}} = \frac{\Delta P_{\text{complex}}}{\Delta P_{\text{bulk}}}$$

Here $\Delta P_{\text{complex}}$ and ΔP_{bulk} are the differences of the polarizations between the simulations with E_1, E_2 of the water in the interfacial gap of the complex. The values reported are average values computed from 0.25 ps spaced snapshots taken from the simulation interval 1 ns-4 ns. The interfacial gap was defined as the volume within “the displacement distance” d from both interfaces and not more than 1.5 nm from the line connecting the centres of geometry of the interfaces. A water molecule was considered in the gap if its oxygen atom was inside the gap. For the control simulation of each protein alone (Barnase or Barstar), the gaps were defined by superposition of the corresponding part from the complex to all snapshots of the single protein-water system. For the superposition, the structure of the complex was chosen as the most representative structure obtained by clustering the snapshots from 2 ns of simulation of the complex.

Figure 38 (the black line) shows the relative perpendicular electric susceptibility for each interfacial distance that was computed relatively to the value computed for the corresponding volume and shape in the bulk water. Interestingly, the water between the two hydrophilic protein surfaces showed a strongly decreased dielectric response that extended beyond 3 nm of interfacial distances. Its value is less than 50% of the bulk value for interfacial distances closer than 1.2 nm. We also calculated the susceptibility in a direction parallel to the interfaces in the same way as for the perpendicular direction (red line in Figure 38). Interestingly, this calculation showed that the dielectric properties of water in the interfacial gap are strongly anisotropic. A comparison of the perpendicular and parallel susceptibilities showed that the perpendicular susceptibility decreases more strongly and over longer distances than that along the parallel direction. The perpendicular susceptibility has values of about 50% of the parallel one for interfacial distances less than 1.2 nm. This means that water generates a preferred directionality for the electrostatic interactions in the

perpendicular direction that drives the interfaces toward each other (association) rather than in the parallel direction.

The results for the perpendicular and parallel permittivities are similar to those computed by the polarization method and the previous fluctuation and lead to the same conclusions. This good agreement for the barnase-barstar system is an important validation of the obtained results and for the different behavior of perpendicular and parallel permittivities in particular.

Although the results from the polarization formula are similar to the results from the fluctuation formula, the fluctuation formula is based on a stronger theoretical background. As previously explained, I expect that the polarization formula underestimates the dielectric constant. This underestimation is due to the same reason as that in the Onsager formula that ignores the interactions between the dipoles inside the sample volume. Such interactions can be taken into account by using a fluctuation formula similar to the one derived by Kirkwood. For this reason, we derived a fluctuation formula suitable to our system.

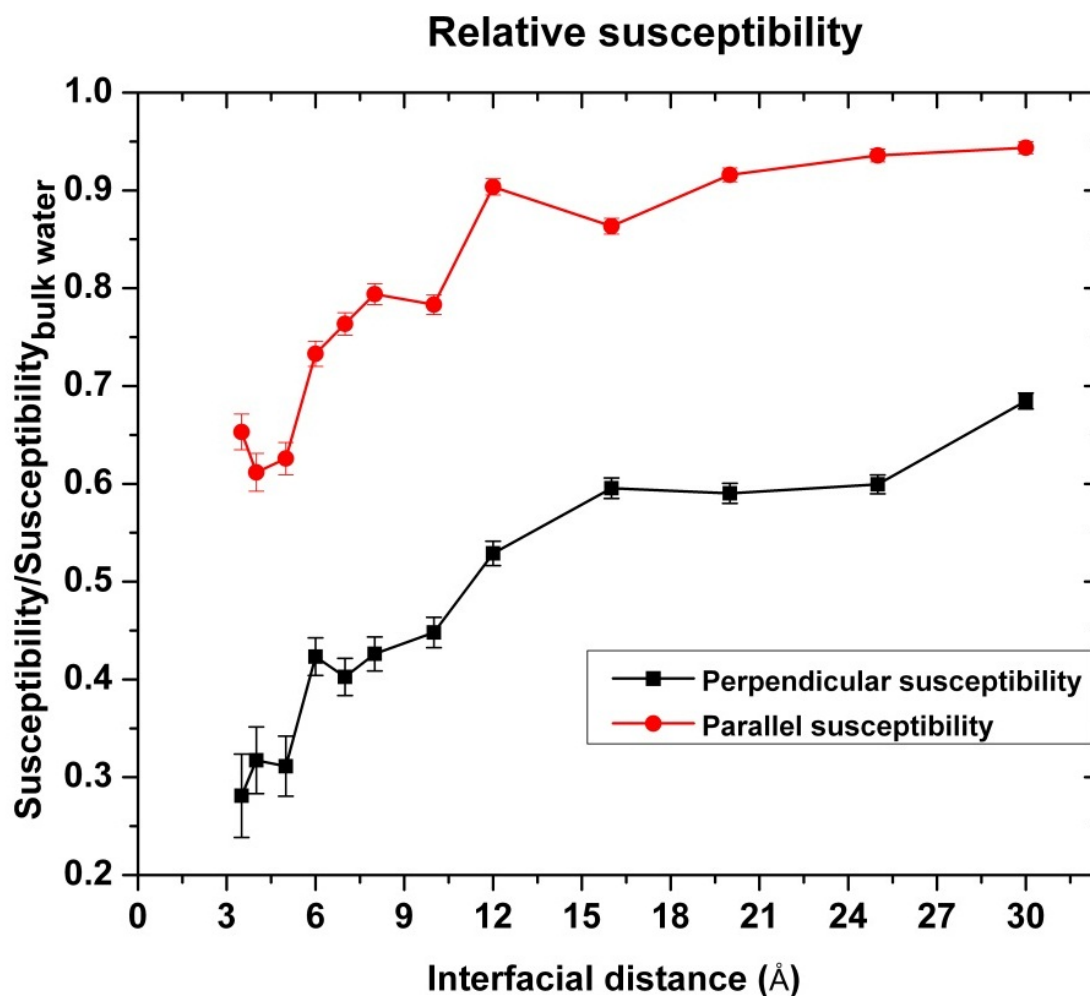


Figure 38. The local dielectric behavior relative to the value for bulk water for the complex using the polarization formula.

The electric susceptibility $\chi_e = \epsilon - 1$ was calculated relative to the value of bulk water. Both perpendicular (the black line) and parallel (the red) susceptibilities were calculated. The error bars represent the error estimated through the propagation of the errors in polarizations. The errors were estimated through 10000 bootstrap sampling.

6.4.3 Studying the regime of the response of the polarization to the externally applied electric fields

The constitutive relation (eq. 55) and the following equations in my derivation are based on the assumption that the change in the polarization still shows a linear response to the changes in the electric field. However, the dielectric materials have a limited ability to be polarized and the saturation behavior appears upon application of a very strong electric field. The saturation behavior is very clear in the Langevin equation (eq. 53) that defines the relation between the orientation polarizations and the strength of the directing electric field, compare eq. 48. In the absence of an external field, the average directing field is zero in for a sample of water molecules. However, in the interfacial gap between the proteins, the directing field is not of zero strength as the electric field generated by the protein will affect the sample. If an additional electric field is applied externally (artificially), this field will be added to the directing field. This opens up the possibility to check whether the sample is in the saturated regime. For this we have tested the response of the polarization inside the gap to different external field strengths from -1.0 to 3.0 V/nm by running 11 MD simulations where the field either pointed in the perpendicular or in the parallel direction. Additionally, we performed simulations without any external field. This gave 23 simulations in total. The simulations were started from the complex with an interfacial displacement of 0.5 nm with harmonic position restraints applied to the backbone atoms with the same parameters and protocols as in the previous simulations that were analyzed by the fluctuation formula (see 6.4.1.1). The simulations were extended to 4 ns length each. The values reported are average values computed from 0.25 ps spaced snapshots taken from the simulation interval 1 ns-4 ns. The interfacial gap was defined as the volume within “the displacement distance” d from both interfaces and not more than 1.5 nm from the line connecting the centres of geometry of the interfaces. A water molecule was considered in the gap if its oxygen atom was located inside the gap. The induced polarization was calculated for every simulation by subtracting the polarization when no external field is applied from the polarization of the sample under the external electric field. The relationship between the induced polarization in the direction of the externally applied field and the strength of the externally applied field was fitted to a Langevin function (see 4.1.3) with two fit parameters A and B.

$$P = A \times L(BE_0) = A \left(\coth(BE_0) - \frac{1}{(BE_0)} \right)$$

The fitting was performed for both the perpendicular (x axis) and the parallel (y axis) directions. We found that the relationship between the induced polarization P and the external field E_0 fits perfectly to the Langevin function (**Figure 39**) for both the perpendicular and the parallel direction. The response is linear for low values of the external field strength. This linear regime around the origin reflects the fact that the sample is not in the saturation regime and still responds to the electric field. Comparing the response of the polarization to the electric fields in the perpendicular and the parallel directions showed that the sample cannot respond to the electric field in the same way. The response is dependent on the direction of the applied electric field. Thus, the anisotropic behaviour is clearly demonstrated. We can expect from **Figure 39** that the dielectric permittivity in the perpendicular direction will be smaller than in the parallel direction because the induced polarization reflects the real dielectric ability of the sample to counteract the external electric field.

Induced polarization by external E-field

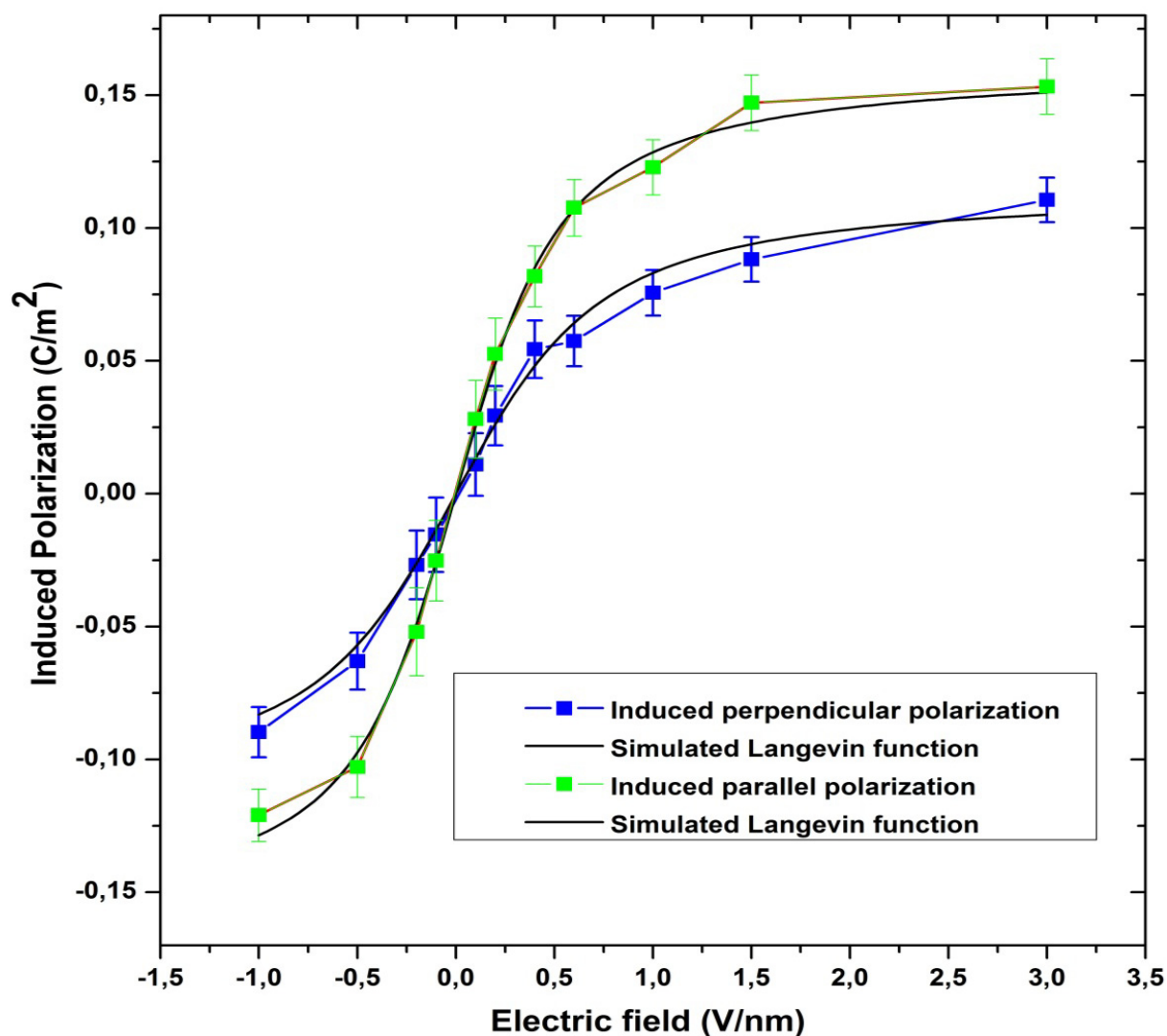


Figure 39. Induced polarization of the water molecules in the interfacial gap as a function of the applied electric field.

Every data point corresponds to an MD simulation of 4 ns length starting from the complex with interfacial displacement of 0.5 nm and applying external E-fields ranging from -1.0–3.0 V/nm with harmonic position restraints applied to the backbone atoms of both proteins. Both the computed parallel (green) and perpendicular (blue) polarization fit perfectly to a Langevin function (the black and red lines respectively). The value of the induced polarization of the water molecules in the interfacial gap is the difference from the observed polarization in the simulation without external field. The response of the polarization is linear at low values of the field strength.

6.5 Discussion

The observations of this study draw the picture of interfacial water acting as molecular glue sticking together the hydrophilic interfaces. In analogy to the well-known hydrophobic effect [88], we will use the term “hydrophilic effect” to describe the multiple roles of water in the binding of hydrophilic interfaces. Our analysis shows that the hydrophilic effect contributes to the binding of hydrophilic interfaces in two favorable ways. The first is a direct energetic contribution through the water mediation between the interfaces that stabilizes the intermediate encounter complexes before the formation of the specific direct contacts as well as the last specific complex. The second is indirect through the reduced dielectric properties and their directionality that is very important during the diffusive phase of binding which is dominated by the electrostatic attraction [152]. It is a striking finding that the dielectric constant of the water in between the two reacting proteins is still significantly lower than the bulk value up to distances of a few nanometres. This strongly enhances the magnitude of the electrostatic interactions during the diffusive phase of binding. Our finding is consistent with atomic force microscopy measurements of the dielectric constant profile perpendicular to a mica surface that revealed such a very long-ranged decrease even up to 10 nm from the surface [174]. The stronger reduction along the perpendicular relative to the parallel directions found here also suggests an important effect of electrostatic focusing on shaping the binding funnel of the free energy landscape for association. The anisotropy of the dielectric properties observed here introduces a new complexity to the continuum treatment of water.

We note that the computed permittivity did not converge to the value of the bulk water even at 5 nm separation. This can be partially explained by the existence of several shells of water molecules around the individual proteins which contribute with a low permittivity to the apparent average value of the permittivity in the interfacial gap. Moreover, at large distance separation, the interfacial gap has a more inhomogeneous dielectric permittivity which may have affected the computed value of the apparent permittivity. The stronger reduction in the perpendicular relative to the parallel directions found here also suggests an important effect of electrostatic focusing on shaping the binding funnel of the free energy landscape for association. The anisotropic electrostatic interactions make the diffusion favorable in the association direction so that it can contribute to the acceleration of the association rate by the reduction of the dimensionality of the search process (see discussion in 7.2). The anisotropy of the dielectric properties observed here introduces a new complexity to the continuum treatment of water.

We also demonstrated a pronounced role of the water mediation network from the early intermediate encounter complexes on. Our analysis showed that the water mediation increases continuously during the transition from the early intermediate complex to the final specific complex and thus provides a thermodynamic gradient toward the stereospecific complex as well. Such a role of bridging water molecules was proposed by Ben-Naim [181, 182] in an early thermodynamic model for the favorable contribution of bridging water between hydrophilic molecules to free energy changes. The finger prints of this role are reflected in the crystal structure of the complex where the positions of 51 water molecules were identified within 0.5 nm of both interfaces [164] (see Figure 36). Nine of these water molecules have relatively low B-factors and are as ordered as protein residues at the interface with the possibility to mediate Barnase-Barstar hydrogen bonds [164]. We propose that the observed mechanism of binding is also relevant for other systems as it is very common for interfaces to be stabilized by bridging water molecules [183-187]. A statistical analysis of a large data set of protein complexes showed that water-mediated polar interactions are as

abundant at the interfaces as direct protein-protein hydrogen bonds [187] with an average of 15 water molecules per 1000 Å² of interface area.

The wet nature of hydrophilic interfaces favorably contributes to protein association in reducing the expensive desolvation penalty that would occur upon binding in the case of dry electrostatic interactions. Dry electrostatic interactions were previously shown to be unfavorable in many studies where salt bridges buried in the protein interior were reported to be destabilizing [38]. Even simple hydrophobic interactions provided more stabilizing energy than the buried salt bridge [38]. This study shows how hydrophilic interfaces may change the physical behavior of water from an isotropic solvent with high dielectric to an adhesive medium with reduced and directional dielectric properties. When nature wants hydrophilic interfaces to bind, bridging water is apparently an essential ingredient.

6.5.1 The spatial variation in the dielectric constant

The reduction of the dielectric constant of the water molecules in the interfacial gap is clearly important for reducing the screening of the electrostatic interactions between the proteins. The importance of the spatial variation in the dielectric constant around ions was characterized from early times on. Born [188] found that the free energy of solvation of an ion of charge q and radius R in a continuum of dielectric constant ϵ is

$$G = \left(1 - \frac{1}{\epsilon}\right) \frac{q^2}{2R} \quad (72)$$

Early calculations of the solvation free energy of ions using the dielectric of the bulk water gave large deviations [189-191]. Laidler [189] showed that using a variable dielectric constant can improve the results from (eq. 72). However, the commonly used solution of this problem in using implicit water model involves adjusting the radius of the ions and atoms to fit the experimental data and using the dielectric of the bulk water [192]. The difficulty of using a spatially variable dielectric constant is due to the difficulty of finding an appropriate method to account for it. Many approaches were presented to calculate the changes in the dielectric constant. The reduction in the dielectric constant was usually attributed to the electric saturation due to the strong electric field [193, 194]. The reason behind the reduction in the dielectric constant due to the electric field is the nonlinear behavior of the polarization in the field strength:

$$\mathbf{P} = \chi\mathbf{E} + \xi E^2\mathbf{E} \quad (73)$$

The general constitutive relation (eq. 55) is:

$$\begin{aligned} \mathbf{D} &= \mathbf{E} + 4\pi\mathbf{P} = \epsilon\mathbf{E} + 4\pi\xi E^2\mathbf{E} \\ \epsilon(E) &= \frac{\partial D}{\partial E} = \epsilon + 12\pi\xi E^2 \end{aligned} \quad (74)$$

$$\Delta\epsilon = \epsilon(E) - \epsilon = 12\pi\xi E^2$$

where $\epsilon(E)$ is the dielectric constant at field strength E , and ξ is the nonlinear electric factor.

Booth suggested a correction to the Kirkwood equation to account for the electric saturation [193, 194]:

$$\varepsilon = n^2 + \frac{28\pi N_0 \mu_0 (n^2 + 2)}{3\sqrt{73}E} L \left(\frac{\sqrt{73}(n^2 + 2)\mu_0 E}{6kT} \right) \quad (75)$$

where n is the refractive index, N_0 number of molecules per volume unit.

This equation was recently used to map the dielectric constant in a membrane protein channel [195]. However, the assumption that the electric saturation is the only reason for the reduction in the dielectric constant ignores the fact that the change in the structure of the solvent can be a very important factor. The ignorance of the structural factors was the reason why the dielectric saturation equations cannot explain the anomalous saturation in dielectric in the case of many solvents [130]. Liszi and Ruff found a relation that includes the structural factors of Kirkwood (g_0, g_1) [196]. Their equation gives predictions for both the normal and anomalous dielectric saturation as well as for the structure-breaking or structure-making effect of the field [196] :

$$\varepsilon(E) = \varepsilon - (\varepsilon - n^2) \left(\frac{\mu_0^2 (n^2 + 2)^2}{20k^2 T^2} g_0^2 - 3g_1 \right) E^2 \quad (76)$$

In our study we found that the reduction in the dielectric constant was due to the changes in the structure of the water molecules in the interfacial gap. The additional work (6.4.3) to check the possible occurrence of electric saturation showed that the electric saturation did not play a role. However, the linear range in the fitted Langevin equation (see Figure 39) is not large. Thus a saturation effect is not far away from occurrence in the first hydration shell where the strength of the electric field is very high. Our findings suggest that the structural factors should be taken into account when developing a method for calculating the local dielectric constant.

7 Discussion, Conclusion and Outlook

In this work I have studied two different mechanisms for protein-protein associations. Now I will try to discuss them from a physical point of view. The formation of protein-protein complexes proceeds in two steps, at least. In the first step, Brownian diffusion makes the two molecules collide and form encounter complexes. In the following steps, some encounter configurations manage to rearrange to the stereospecific complex which is energetically stable enough to be the dominant conformation that can be observed by experiments.

As discussed in chapter 2 (see 2.1.1.1), the diffusional part is usually not a problem and not the source of the complexity of the binding process). The Brownian motions of the proteins always ensure a very high rate of collisions on the order of $10^{10} M^{-1}s^{-1}$ [12, 197, 198]. This collision rate does not depend on the size of the proteins as explained in (2.1.1.1). Moreover, the enhancement of the association rate by electrostatic interactions cannot be explained by the enhancement in the rate of collisions as explained in (2.1.1.1). For example, the Debye model shows that the rate enhancement (for collisions) due to electrostatic interactions of $9kT$ at contact distance is only 9-fold [14]. Electrostatic interactions appear to increase the association rate by increasing the ratio of fruitful collisions not through increasing the rate of collision.

The rate limiting steps for the association process are the transformations from the encounters to the stereospecific complex. The difference between the collision rate and the association rate is related to the steric factor which accounts for the complexity of finding the stereospecific complex and to the energetic barrier to form the encounters if exist.

The free energy landscape governing the association process represents a global transition rather than a single reaction and is expected to have many transition states between the intermediates before arriving at the stereospecific complex [199]. However the association rate is related to how easily intermediate conformations may transition to the funnel of the stereospecific complex. In other words, the question is how the rate of fruitful transitions in the intermediate states can be enhanced. To formulate the problem in physical terms, I recall the equation of the transition state theory (eqs. 8 & 9):

$$k_{ass} = \left(\frac{KT}{h}\right) \exp\left(\frac{-\Delta G^*}{RT}\right)$$
$$k_{ass} = \left(\frac{KT}{h}\right) \exp\left(\frac{\Delta S^*}{R}\right) \exp\left(\frac{-\Delta H^*}{RT}\right)$$

This relation can also be applied to transformations between ground states and different intermediates on the free energy landscape for association, not only to the whole interaction. The beauty of this equation is that it shows that the steric factor of the association process is related to the entropy difference ΔS^* between the intermediate state and the ground state. Thus, there are two ways to accelerate the association process by affecting ΔS^* or ΔH^* (activation enthalpy and activation entropy).

7.1 Accelerating the association by reducing the activation enthalpy

Ground state and the transition state ΔH^* in any reaction along the energy landscape can be reduced by stabilizing the intermediate state by some additional interactions. The advantage

of this fact is that the total electrostatic interactions between two proteins result from all surface residues of the proteins (not only those at the binding interfaces). This can explain how charged residues outside the interfaces can favorably affect the association rate.

The enthalpic barrier in protein association is assumed to originate from the desolvation of the interfaces by breaking the hydrogen bonding between the water and the interfaces before the stereospecific contacts compensate this loss. The stereospecific contacts are mostly hydrogen bonds and they are very short-ranged and directional. Thus they are formed when the proteins have the correct orientations in the stereospecific complex.

In both protein systems that we studied in chapters 5 and 6, the enthalpic barrier ΔH^* is supposed to be low due to many reasons. For the peptide binding to SH3 domain we noticed that long-ranged electrostatic interactions attract the oppositely charged residues in the peptide and the domain. Moreover the hydrophobic dewetting of the interfaces means that the desolvation energy barrier is low because the interfaces are already desolvated in the unbound state. The dewetting transition will reduce ΔH^* because the evaporated water molecules will form more hydrogen bonds in the bulk water. Thus a large enthalpic barrier is unlikely in the case of the association between hydrophobic interfaces.

Also, during the association between Barnase and Barstar ΔH^* is reduced in several ways: (i) the general electrostatic interaction between the proteins is attractive as they have different net charges; (ii) the interfaces did not show any desolvation; (iii) the water mediation is supposed to contribute to reducing ΔH^* .

From Both cases should be examples where the enthalpic barriers are rather flat during the association of proteins.

7.2 Accelerating the association by reducing the activation entropy: Reduction of the dimensionality of the search process

The difference in entropy between the unbound complexes and the intermediates is negative as the formation of the encounter involves a restriction of the accessible conformations [21]. The important factor for the association rate is the ΔS^* for the transformation reactions between the intermediates along the energy landscape arriving at the stereospecific complex. If a part of the entropy loss takes place before the rate-limiting steps (the transformation between intermediates), the association rate will be significantly enhanced. This is the reason why the search process is more efficient in a space of reduced dimensionality (like a membrane) over 3D space. However, nature uses the reduction of the dimensionality to accelerate many association processes by coupling them with prior fast association to prepay the entropy price.

The advantage of the dimensionality reduction in the search process is very clear when we compare the association between proteins in a two dimensional (2D) space such as a membrane and the association in water (3D). The advantage of reducing the dimensionality of the search process can be seen from the mean diffusion times (τ) to reach a small target of radius b in the middle of a cell of radius L ($L \gg b$) are [7]:

$$\begin{aligned} \tau &= (L^2/3D)(L/b) && \text{in 3 dimensions,} \\ \tau &= (L^2/2D) \ln(L/b) && \text{in 2 dimensions,} \\ \tau &= L^2/3D && \text{in 1 dimension,} \end{aligned}$$

Clearly the reduction in the mean diffusion time is very significant especially when the diffusion takes place in two dimensions instead of three dimensions. In general, the reduction of the dimensionality is not involved through the whole association process. Usually the three-dimensional diffusion (in solution) is coupled to one- or two-dimensional diffusion (e.g. along the surface) [7]. The coupling of the diffusions is very useful because the effective target radius for the three-dimensional diffusion process can be extended by many orders of magnitude if the directed diffusion operates over large distances. The three-dimensional diffusion is the limiting step of the coupled association because the following step (diffusion in lower dimensionality) is much faster. However, many biological structures can act as one- and two-dimensional diffusion guides [7] such as the biological membranes and the one dimensional diffusion in the loops of chromosomal DNA or microfilaments.

Another example of the coupling is the explanation of the fast association rates observed during the association of substrates to the catalytic sites on enzymes. The substrates diffuse in the attractive potential around the enzymes surfaces and form nonspecific complexes. Thus the nonspecific forces held the molecules long enough together to enhance the chance of forming the stereospecific complex [7, 200]. This process can be explained on the basis of ΔS^* : it was mentioned in (3.2.1) that there is an entropic loss due to the association between molecules. Here the price of the loss in the entropy is paid by the enzyme-substrate binding energy which is the first step in the catalysis. The following reaction step will have the advantage of no loss of entropy ($10 - 12.2 \text{ Kcal. M}^{-1}$; see 3.2.1) thus the term $(\Delta S^*/R)$ in eq. 9 will introduce an acceleration by a factor of 6×10^9 [8]. Such a trick in biology is described as a very high effective concentration of the catalytic groups in comparison compared with reactions in solution and is the entropic factor of the catalytic action of the enzymes.

In both mechanisms of proteins associations we have observed a fast diffusion phase. The real search process takes place within the encounter complexes results in a loss of the entropy. However, the encounter will performed the search process in less dimensionality and with less loss in entropy ΔS^* which makes finding the stereospecific easier because a part of the loss of the entropy was prepaid upon forming the encounters.

7.3 How can the dielectric permittivity of the water molecules between the proteins affect the association process?

Now, we come to the surprising anisotropic decrease in the dielectric constant of the water molecules in the case of the binding between Barnase and Barstar. The anisotropic electrostatic interactions funnel the diffusing particles towards each other. This process can be interpreted as a kind of reduction of the dimensionality of the search process which should efficiently accelerate the association process.

The reduction of the dielectric constant of the water molecules in the interfacial gap is clearly important for reducing the screening of the electrostatic interactions between the proteins. Although the pairwise electrostatic interaction is completely enthalpic in origin, the electrostatic interaction in water has an entropic component due to the interaction between the water and the solutes [76]. The entropic nature is clear from the fact that the dielectric constant of water is strongly temperature dependent. If we decompose the entropic contribution to the Coulomb law in a medium of dielectric constant ϵ , $U = \frac{q_1 q_2}{\epsilon r}$ [11]:

$$S = -\frac{\partial U}{\partial T} = -\frac{\partial}{\partial T} \left(\frac{q_1 q_2}{\epsilon r} \right) = -\left(\frac{q_1 q_2}{\epsilon^2 r} \right) \frac{\partial \epsilon}{\partial T} = \frac{U}{\epsilon} \frac{\partial \epsilon}{\partial T}$$

at room temperature $\frac{\partial \epsilon}{\partial T} = -0.46\% K^{-1}$ [11] that means $TS = -1.38U$. This means that the origin of the attraction between charges in water is the entropy of the water and the pairwise interaction between the solutes. This can be understood by the fact that the electric field of the charges induces directionality for the dipoles of the water molecules (orientation polarization; see 4.1.3). We can expect from this simple analysis that the release of the water molecules in the interfacial cavity to the bulk water is an important driving force for the association process. This may explain why the experimental entropy loss of the Barnase: Barstar complex is only $0.3 Kcal.M^{-1}$ while the loss of entropy from restricting is the motion is usually $10 - 12.2 Kcal.M^{-1}$.

A more general picture for the entropy storage in water molecules can be derived from the general thermodynamic equations in a dielectric [130]:

$$\Delta F = \frac{1}{8\pi} \iiint \mathbf{D} \cdot \mathbf{E} dv = \frac{1}{8\pi} \iiint \frac{E^2}{\epsilon} dv \quad (77)$$

$$\Delta S = \frac{1}{8\pi} \iiint E^2 \left(\frac{\partial \epsilon}{\partial T} \right) dv \quad (78)$$

ΔF and ΔS are the changes in the free energy and entropy in the dielectric medium due to the electric field \mathbf{E} . One should remember that $\frac{\partial \epsilon}{\partial T} < 0$. If we use the physical value of bulk water at $300 K^{-1}$ we get again $\Delta S \approx -1.38\Delta F$. Thus a large portion of the self-electrostatic energy of a charge in a dielectric medium is stored in the dielectric medium through the reduction in its entropy. This stored energy can be reused in two different ways in the association process: (i) complete release of the stored entropy of the water molecules that are released due to the binding process; (ii) the electric field of two oppositely charged molecules will be reduced after they associate due to the reduction in the net charge. Such a reduction in the electric field will cause an increase in the entropy of the surrounding water molecules.

The entropy gain upon releasing the water molecules from the gap in between the proteins can be considered as similar to that of released water molecules that were tightly connected to the protein as discussed in (3.2.1.2). However, the entropy gain by releasing the water molecules from the gap is expected to be smaller than that of "frozen" waters bound at interfaces. Dunitz [75] suggested that the upper limit of entropy gain due to the release of a tightly bounded water molecule is $2 Kcal.M^{-1}$. The entropic contribution of the released waters to the free binding energy of ion associations is well studied in the field of electrochemistry because of the availability of experimental data for the hydration entropy of ions [192, 201-203]. Marcus [201, 203] studied the contribution to the entropy changes upon ion pairing. The measurable molar entropy change involved is considered to be made up of four contributions: translational, rotational, electrostatic, and desolvation entropies. The former three can be calculated from the properties of the ions and solvents involved; hence, the fourth is obtained by taking the difference. The analysis showed that the contribution from released water can contribute up to $10 Kcal.M^{-1}$.

An important point here is the fact that this important entropic factor is not taken into account in the continuum electrostatic methods for calculation of the free binding energy such as MM/PBSA [204-208] and CC/PBSA [209]. However, taking this factor into account will be an important improvement to these methods.

The total contribution of the water molecules in the interfacial gap to the binding process is defined by the difference between their free energy in the gap (eq. 77) and the free energy of these in the bulk. Although we know that the entropic changes (eq. 78) are positive, the

enthalpic changes are related to the difference between the interactions of the water molecules within the interfacial gap (water-water and water-protein interactions) and their interactions in bulk. The reduction in the dielectric constant of the solvent in the interfacial gap will reduce the interactions between the water molecules because of the reduction in the reaction field (see eq. 50). We speculate that water molecules will try to compensate this reduction through additional interactions with the protein and increase the water mediation between the proteins. One can expect from the existence of tightly bounded water molecules in the crystal structures and from the observation of the truncation of the water network between the proteins that strong interactions of certain water molecules in the first hydration shell with the protein is larger than the interaction in bulk and enough to pay the price of the loss of entropy. However, further studies are needed to understand the balance between the solvation terms.

I started my work with the question whether nature plays dices. Do binding processes occur by pure chance ? Our simulations and analysis indicate that for electrostatically driven interactions, there always exist mechanisms to simplify the complexity of the binding process downhill to reach the specific complex. However, the most important questions in this field are still far from being answered. How does nature assure the specificity of association? Does evolution optimize the protein interfaces? Is this done by changing the association or dissociation rate? Many questions related to the interaction of proteins are still awaiting a thorough understanding . We are still in the beginning.

Bibliography

1. Ahmad, M., W. Gu, and V. Helms, *Mechanism of Fast Peptide Recognition by SH3 Domains*. *Angewandte Chemie International Edition*, 2008. **47**(40): p. 7626-7630.
2. Ahmad, M., et al., *Adhesive water networks facilitate binding of protein interfaces*. *Nat Commun*, 2011. **2**: p. 261.
3. Mittermaier, A. and L.E. Kay, *New Tools Provide New Insights in NMR Studies of Protein Dynamics*. *Science*, 2006. **312**(5771): p. 224-228.
4. Wallace, P.R., *Physics: Imagination and Reality*. 1991: World Scientific Publishing Company.
5. Laplace, P.S., *Théorie analytique des probabilités*. 1820: Courcier.
6. Shoup, D. and A. Szabo, *Role of diffusion in ligand binding to macromolecules and cell-bound receptors*. *Biophysical Journal*, 1982. **40**(1): p. 33-39.
7. Berg, O.G. and P.H. von Hippel, *Diffusion-Controlled Macromolecular Interactions*. *Annual Review of Biophysics and Biophysical Chemistry*, 1985. **14**(1): p. 131-158.
8. Fersht, A., *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. 1999: W.H. Freeman.
9. Moore, W.J., *Physical chemistry*. 1972: Prentice-Hall.
10. Lauffenburger, D.A. and J.J. Linderman, *Receptors: models for binding, trafficking, and signaling*. 1996: Oxford University Press.
11. Jackson, M.B., *Molecular and cellular biophysics*. 2006: Cambridge University Press.
12. Smoluchowski, V.M., *Mathematical Theory of the Kinetics of the Coagulation of Colloidal Particle*. *Phys. Chem*, 1917. **92**: p. 129-168.
13. Debye, P., *Reaction Rates in Ionic Solutions*. *Transactions of The Electrochemical Society*, 1942. **82**(1): p. 265-272.
14. Schreiber, G., G. Haran, and H.X. Zhou, *Fundamental Aspects of Protein-Protein Association Kinetics*. *Chemical Reviews*, 2009. **109**(3): p. 839-860.
15. Keizer, J., *Diffusion effects on rapid bimolecular chemical reactions*. *Chemical Reviews*, 1987. **87**(1): p. 167-180.
16. Schreiber, G. and A. Fersht, *Rapid, electrostatically assisted association of proteins*. *Nat Struct Biol*, 1996. **3**(5): p. 427-31.
17. Hammond, G.S., *A Correlation of Reaction Rates*. *Journal of the American Chemical Society*, 1955. **77**(2): p. 334-338.
18. Hammes, G.G. and I. Amdur, *Principles of chemical kinetics*. 1978: Academic Press.
19. Hammes, G.G., *Thermodynamics and kinetics for the biological sciences*. 2000: Wiley-Interscience.
20. Gabdoulline, R. and R. Wade, *On the protein-protein diffusional encounter complex*. *J Mol Recognit*. **12**(4): p. 226-34.
21. Spaar, A., et al., *Diffusional Encounter of Barnase and Barstar*. *Biophysical Journal*, 2006. **90**(6): p. 1913-1924.

22. Mateu, M.G., M.M. Sanchez Del Pino, and A.R. Fersht, *Mechanism of folding and assembly of a small tetrameric protein domain from tumor suppressor p53*. *Nat Struct Mol Biol*, 1999. **6**(2): p. 191-198.
23. Wu, L.C., et al., *Two-step binding mechanism for T-cell receptor recognition of peptide-MHC*. *Nature*, 2002. **418**(6897): p. 552-556.
24. Levy, Y., et al., *A Survey of Flexible Protein Binding Mechanisms and their Transition States Using Native Topology Based Energy Landscapes*. *Journal of Molecular Biology*, 2005. **346**(4): p. 1121-1145.
25. Fersht, A.R. and V. Daggett, *Protein Folding and Unfolding at Atomic Resolution*. *Cell*, 2002. **108**(4): p. 573-582.
26. Iwahara, J. and G.M. Clore, *Detecting transient intermediates in macromolecular binding by paramagnetic NMR*. *Nature*, 2006. **440**(7088): p. 1227-1230.
27. Tang, C., J. Iwahara, and G.M. Clore, *Visualization of transient encounter complexes in protein-protein association*. *Nature*, 2006. **444**(7117): p. 383-386.
28. Suh, J.Y., C. Tang, and G.M. Clore, *Role of electrostatic interactions in transient encounter complexes in protein-protein association investigated by paramagnetic relaxation enhancement*. *Journal of the American Chemical Society*, 2007. **129**(43): p. 12954-+.
29. Volkov, A.N., et al., *Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR*. *Proceedings of the National Academy of Sciences*, 2006. **103**(50): p. 18945-18950.
30. Carbonell, P., R. Nussinov, and A. del Sol, *Energetic determinants of protein binding specificity: Insights into protein interaction networks*. *Proteomics*, 2009. **9**(7): p. 1744-1753.
31. Berman, H.M., et al., *The protein data bank*. *Nucleic acids research*, 2000. **28**(1): p. 235.
32. Vajda, S. and F. Guarnieri, *Characterization of protein-ligand interaction sites using experimental and computational methods*. *Current Opinion in Drug Discovery and Development*, 2006. **9**(3): p. 354-362.
33. Prasad Bahadur, R., et al., *A Dissection of Specific and Non-specific Protein-Protein Interfaces*. *Journal of Molecular Biology*, 2004. **336**(4): p. 943-955.
34. Nooren, I.M.A. and J.M. Thornton, *Structural Characterisation and Functional Significance of Transient Protein-Protein Interactions*. *Journal of Molecular Biology*, 2003. **325**(5): p. 991-1018.
35. Chandler, D., *Interfaces and the driving force of hydrophobic assembly*. *Nature*, 2005. **437**(7059): p. 640-7.
36. Ahmad, M., W. Gu, and V. Helms, *Mechanism of fast peptide recognition by SH3 domains*. *Angewandte Chemie-International Edition*, 2008. **47**(40): p. 7626-7630.
37. Dyson, H., P. Wright, and H. Scheraga, *The role of hydrophobic interactions in initiation and propagation of protein folding*. *Proc Natl Acad Sci U S A*, 2006. **103**(35): p. 13057-61.
38. Waldburger, C.D., J.F. Schildbach, and R.T. Sauer, *Are buried salt bridges important for protein stability and conformational specificity?* *Nature Structural Biology*, 1995. **2**(2): p. 122-128.
39. Ball, L.J., et al., *Recognition of proline-rich motifs by protein-protein-interaction domains*. *Angewandte Chemie-International Edition*, 2005. **44**(19): p. 2852-2869.

40. Zarrinpar, A., R.P. Bhattacharyya, and W.A. Lim, *The Structure and Function of Proline Recognition Domains*. Sci. STKE, 2003. **2003**(179): p. re8-.
41. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions*. Proceedings of the National Academy of Sciences, 1996. **93**(1): p. 13-20.
42. Hua, L., et al., *Nanoscale dewetting transition in protein complex folding*. J Phys Chem B, 2007. **111**(30): p. 9069-77.
43. Hua, L., R. Zangi, and B.J. Berne, *Hydrophobic Interactions and Dewetting between Plates with Hydrophobic and Hydrophilic Domains*. Journal of Physical Chemistry C, 2009. **113**(13): p. 5244-5253.
44. Ball, P., *Water as a biomolecule*. Chemphyschem, 2008. **9**(18): p. 2677-85.
45. Giovambattista, N., P.G. Debenedetti, and P.J. Rossky, *Hydration behavior under confinement by nanoscale surfaces with patterned hydrophobicity and hydrophilicity*. Journal of Physical Chemistry C, 2007. **111**(3): p. 1323-1332.
46. Conte, L.L., C. Chothia, and J. Janin, *The atomic structure of protein-protein recognition sites*. Journal of Molecular Biology, 1999. **285**(5): p. 2177-2198.
47. Ansari, S. and V. Helms, *Statistical analysis of predominantly transient protein-protein interfaces*. Proteins, 2005. **61**(2): p. 344-55.
48. Xu, D., C.J. Tsai, and R. Nussinov, *Hydrogen bonds and salt bridges across protein-protein interfaces*. Protein Engineering, 1997. **10**(9): p. 999-1012.
49. Horovitz, A., *Non-additivity in protein-protein interactions*. Journal of molecular biology, 1987. **196**(3): p. 733-735.
50. Chothia, C., *Hydrophobic bonding and accessible surface area in proteins*. Nature, 1974. **248**(5446): p. 338-339.
51. Clackson, T. and J. Wells, *A hot spot of binding energy in a hormone-receptor interface*. Science, 1995. **267**(5196): p. 383-386.
52. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces*. Journal of Molecular Biology, 1998. **280**(1): p. 1-9.
53. Amnon, H., *Double-mutant cycles: a powerful tool for analyzing protein structure and function*. Folding and Design, 1996. **1**(6): p. R121-R126.
54. Carter, P.J., et al., *The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (Bacillus stearothermophilus)*. Cell, 1984. **38**(3): p. 835-840.
55. Schreiber, G. and A.R. Fersht, *Energetics of protein-protein interactions - analysis of the barnase-barstar interface by single mutations and double mutant cycles*. Journal of Molecular Biology, 1995. **248**(2): p. 478-486.
56. Keskin, O., B. Ma, and R. Nussinov, *Hot Regions in Protein-Protein Interactions: The Organization and Contribution of Structurally Conserved Hot Spot Residues*. Journal of Molecular Biology, 2005. **345**(5): p. 1281-1294.
57. Guharoy, M. and P. Chakrabarti, *Conservation and relative importance of residues across protein-protein interfaces*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15447-15452.
58. Selzer, T., S. Albeck, and G. Schreiber, *Rational design of faster associating and tighter binding protein complexes*. Nat Struct Mol Biol, 2000. **7**(7): p. 537-541.

59. Selzer, T. and G. Schreiber, *New insights into the mechanism of protein-protein association*. *Proteins: Structure, Function, and Bioinformatics*, 2001. **45**(3): p. 190-198.
60. Livnah, O., et al., *Three-dimensional structures of avidin and the avidin-biotin complex*. *Proceedings of the National Academy of Sciences*, 1993. **90**(11): p. 5076.
61. Johnson, R.J., et al., *Inhibition of Human Pancreatic Ribonuclease by the Human Ribonuclease Inhibitor Protein*. *Journal of Molecular Biology*, 2007. **368**(2): p. 434-449.
62. Hill, T.L., *An introduction to statistical thermodynamics*. 1960: Dover Publications.
63. Tidor, B. and M. Karplus, *The Contribution of Vibrational Entropy to Molecular Association: The Dimerization of Insulin*. *Journal of Molecular Biology*, 1994. **238**(3): p. 405-414.
64. Spolar, R. and M. Record, *Coupling of local folding to site-specific binding of proteins to DNA*. *Science*, 1994. **263**(5148): p. 777-784.
65. Fersht, A.R., *From the first protein structures to our current knowledge of protein folding: delights and scepticisms*. *Nat Rev Mol Cell Biol*, 2008. **9**(8): p. 650-654.
66. Anthony L, F., *Natively unfolded proteins*. *Current Opinion in Structural Biology*, 2005. **15**(1): p. 35-41.
67. Tompa, P., *Structure and function of intrinsically disordered proteins*. 2009: Chapman & Hall/CRC Press.
68. Boehr, D.D. and P.E. Wright, *How Do Proteins Interact?* *Science*, 2008. **320**(5882): p. 1429-1430.
69. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis*. *Proceedings of the National Academy of Sciences*, 1958. **44**(2): p. 98-104.
70. Ma, B., et al., *Folding funnels and binding mechanisms*. *Protein Engineering*, 1999. **12**(9): p. 713-720.
71. Lange, O.F., et al., *Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution*. *Science*, 2008. **320**(5882): p. 1471-1475.
72. Forman-Kay, J.D., *The 'dynamics' in the thermodynamics of binding*. *nature structural biology*, 1999. **6**: p. 1086-1087.
73. Zidek, L., M.V. Novotny, and M.J. Stone, *Increased protein backbone conformational entropy upon hydrophobic ligand binding*. *Nature Structural & Molecular Biology*, 1999. **6**(12): p. 1118-1121.
74. Chang, C.-e.A., et al., *Entropic contributions and the influence of the hydrophobic environment in promiscuous protein-protein association*. *Proceedings of the National Academy of Sciences*, 2008. **105**(21): p. 7456-7461.
75. Dunitz, J.D., *The Entropic Cost of Bound Water in Crystals and Biomolecules*. *Science*, 1994. **264**(5159): p. 670.
76. Israelachvili, J.N., *Intermolecular And Surface Forces*. 2010: Academic Press.
77. Gilson, M.K., et al., *On the calculation of electrostatic interactions in proteins*. *Journal of Molecular Biology*, 1985. **184**(3): p. 503-516.
78. Honig, B. and A. Nicholls, *CLASSICAL ELECTROSTATICS IN BIOLOGY AND CHEMISTRY*. *Science*, 1995. **268**(5214): p. 1144-1149.
79. Elcock, A.H., D. Sept, and J.A. McCammon, *Computer Simulation of Protein-Protein Interactions*. *The Journal of Physical Chemistry B*, 2001. **105**(8): p. 1504-1518.

80. Hendsch, Z.S. and B. Tidor, *Do salt bridges stabilize proteins? A continuum electrostatic analysis*. Protein science: a publication of the Protein Society, 1994. **3**(2): p. 211.
81. Ball, P., *Water as a Biomolecule*. Chemphyschem, 2008. **9**(18): p. 2677-2685.
82. Ball, P., *Water as an active constituent in cell biology*. Chemical Reviews, 2008. **108**(1): p. 74-108.
83. Hammer, M.U., et al., *The search for the hydrophobic force law*. Faraday Discussions, 2010. **146**: p. 299-308.
84. Hildebrand, J.H., *A criticism of the term "hydrophobic bond"*. The Journal of Physical Chemistry, 1968. **72**(5): p. 1841-1842.
85. Stillinger, F.H., *Water Revisited*. Science, 1980. **209**(4455): p. 451-457.
86. Lum, K., D. Chandler, and J.D. Weeks, *Hydrophobicity at Small and Large Length Scales*. The Journal of Physical Chemistry B, 1999. **103**(22): p. 4570-4577.
87. Frank, H.S. and M.W. Evans, *Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes*. The Journal of Chemical Physics, 1945. **13**(11): p. 507-532.
88. Kauzmann, W.J.A.P.C., *Some factors in the interpretation of protein denaturation*. 1959. **14**: p. 1-63.
89. Liu, P., et al., *Observation of a dewetting transition in the collapse of the melittin tetramer*. Nature, 2005. **437**(7055): p. 159-62.
90. Zhou, R., et al., *Hydrophobic collapse in multidomain protein folding*. Science, 2004. **305**(5690): p. 1605-9.
91. Choudhury, N. and B.M. Pettitt, *The Dewetting Transition and The Hydrophobic Effect*. Journal of the American Chemical Society, 2007. **129**(15): p. 4847-4852.
92. Dzubiella, J., J.M.J. Swanson, and J.A. McCammon, *Coupling nonpolar and polar solvation free energies in implicit solvent models*. Vol. 124. 2006: AIP. 084905.
93. Pauling, L., *The nature of the chemical bond, and the structure of molecules and crystals: an introduction to modern structural chemistry*. 1939: Cornell University Press.
94. Pimentel, G.C. and A.L. McClellan, *The hydrogen bond*. 1960: W.H. Freeman.
95. Lii, J.-H., *Hydrogen Bonding*, in *Encyclopedia of computational chemistry*, P.R. Schleyer, Editor. 1998, John Wiley. p. 1271-1282.
96. Gordy, W. and S.C. Stanford, *Spectroscopic Evidence for Hydrogen Bonds: Comparison of Proton-Attracting Properties of Liquids. II*. Vol. 8. 1940: AIP. 170-177.
97. Mathur, R., et al., *Proton magnetic resonance studies of hydrogen bonding of benzenethiol with several hydrogen acceptors*. The Journal of Physical Chemistry, 1963. **67**(10): p. 2190-2194.
98. Harrell, S.A. and D.H. McDaniel, *Strong Hydrogen Bonds. II. The Hydrogen Difluoride Ion*. Journal of the American Chemical Society, 1964. **86**(20): p. 4497-4497.
99. Grabowski, S., *Hydrogen Bonding - New Insights*. 2010: Springer London, Limited.
100. Joesten, M.D. and L.J. Schaad, *Hydrogen bonding*. 1974: M. Dekker.

101. Umeyama, H. and K. Morokuma, *The origin of hydrogen bonding. An energy decomposition study.* Journal of the American Chemical Society, 1977. **99**(5): p. 1316-1332.
102. Desiraju, G.R., *Hydrogen Bridges in Crystal Engineering: Interactions without Borders.* Accounts of Chemical Research, 2002. **35**(7): p. 565-573.
103. Coulson, C. and U. Danielsson, *Ionic and covalent contributions to the hydrogen bond.* Ark. Fys, 1954. **8**: p. 239-255.
104. Jorgensen, W.L., D.S. Maxwell, and J. TiradoRives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids.* Journal of the American Chemical Society, 1996. **118**(45): p. 11225-11236.
105. Guerois, R., J.E. Nielsen, and L. Serrano, *Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations.* Journal of Molecular Biology, 2002. **320**(2): p. 369-387.
106. Lazaridis, T. and M. Karplus, *Effective energy functions for protein structure prediction.* Current Opinion in Structural Biology, 2000. **10**(2): p. 139-145.
107. Rohl, C.A., et al., *Protein structure prediction using rosetta.* Numerical Computer Methods, Pt D, 2004. **383**: p. 66-+.
108. Brooks, B.R., et al., *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
109. Cornell, W.D., et al., *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.
110. Weiner, S.J., et al., *A new force field for molecular mechanical simulation of nucleic acids and proteins.* Journal of the American Chemical Society, 1984. **106**(3): p. 765-784.
111. Schuler, L.D., X. Daura, and W.F. van Gunsteren, *An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase.* Journal of Computational Chemistry, 2001. **22**(11): p. 1205-1218.
112. Born, M. and R. Oppenheimer, *Zur Quantentheorie der Molekeln.* Annalen der Physik, 1927. **389**(20): p. 457-484.
113. Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide.* 2010: Springer.
114. Lii, J.-H. and N.L. Allinger, *The Important Role of Lone-Pairs in Force Field (MM4) Calculations on Hydrogen Bonding in Alcohols.* The Journal of Physical Chemistry A, 2008. **112**(46): p. 11903-11913.
115. Hagler, A.T., E. Huler, and S. Lifson, *Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals.* Journal of the American Chemical Society, 1974. **96**(17): p. 5319-5327.
116. Smith, J.C. and M. Karplus, *Empirical force field study of geometries and conformational transitions of some organic molecules.* Journal of the American Chemical Society, 1992. **114**(3): p. 801-812.
117. Allinger, N.L., *Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms.* Journal of the American Chemical Society, 1977. **99**(25): p. 8127-8134.

118. Lii, J.-H. and N.L. Allinger, *Directional hydrogen bonding in the MM3 force field. I.* Journal of Physical Organic Chemistry, 1994. **7**(11): p. 591-609.
119. Vedani, A., *YETI: An interactive molecular mechanics program for small-molecule protein complexes.* Journal of Computational Chemistry, 1988. **9**(3): p. 269-280.
120. Goodford, P.J., *A computational procedure for determining energetically favorable binding sites on biologically important macromolecules.* Journal of Medicinal Chemistry, 1985. **28**(7): p. 849-857.
121. Paton, R.S. and J.M. Goodman, *Hydrogen Bonding and π -Stacking: How Reliable are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions.* Journal of Chemical Information and Modeling, 2009. **49**(4): p. 944-955.
122. Halgren, T.A., *MMFF VI. MMFF94s option for energy minimization studies.* Journal of Computational Chemistry, 1999. **20**(7): p. 720-729.
123. Jurecka, P., et al., *Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs.* Physical Chemistry Chemical Physics, 2006. **8**(17): p. 1985-1993.
124. Schwieters, C., et al., *The Xplor-NIH NMR molecular structure determination package.* J Magn Reson, 2003. **160**(1): p. 65-73.
125. Hockney, R.W., S.P. Goel, and J.W. Eastwood, *Quiet high-resolution computer models of a plasma.* Journal of Computational Physics, 1974. **14**(2): p. 148-158.
126. Becker, O.M., *Computational Biochemistry and Biophysics.* 2001: M. Dekker.
127. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations.* Journal of Computational Chemistry, 1997. **18**(12): p. 1463-1472.
128. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.
129. Griffiths, D., *Introduction to electrodynamics.* 1999: Prentice Hall New Jersey,;
130. Böttcher, C., et al., *Theory of electric polarization.* 1973: Elsevier Scientific Pub. Co.
131. Lorentz, H.A., P. Zeeman, and A.D. Fokker, *Collected papers.* 1939: M. Nijhoff.
132. Rosenfeld, L., *Theory of electrons.* 1951: North-Holland Pub. Co.
133. Mazur, P. and B.R.A. Nijboer, *On the statistical mechanics of matter in an electromagnetic field. I: Derivation of the maxwell equations from electron theory.* Physica, 1953. **19**(1-12): p. 971-986.
134. S.R. D.G., *The Maxwell equations: Nonrelativistic multipole expansion to all orders.* Physica, 1965. **31**(6): p. 953-958.
135. Warshel, A. and S.T. Russell, *Calculations of electrostatic interactions in biological systems and in solutions.* Quarterly Reviews of Biophysics, 1984. **17**(03): p. 283-422.
136. Purcell, E., *Electricity and Magnetism McGraw Hill.* New York, 1984.
137. King, G., F.S. Lee, and A. Warshel, *Microscopic simulations of macroscopic dielectric-constants of solvated proteins.* Journal of Chemical Physics, 1991. **95**(6): p. 4366-4377.
138. JAMES, H.M., *Theory of Electrons. L. Rosenfeld.* Amsterdam: North-Holland Pub.; New York: Interscience, 1951. 119 pp. \$2.25. Science, 1951. **114**(2956): p. 217.

139. Böttcher, C., *Theory of Electric Polarization*. 2 ed. Vol. 1. 1973, Amsterdam: Elsevier Scientific Pub. Co.
140. Onsager, L., *Electric Moments of Molecules in Liquids*. Journal of the American Chemical Society, 1936. **58**(8): p. 1486-1493.
141. Purcell, E.M., *Electricity and magnetism*. 1985: McGraw-Hill.
142. Nienhuis, G. and J.M. Deutch, *Structure of Dielectric Fluids. I. The Two-Particle Distribution Function of Polar Fluids*. The Journal of Chemical Physics, 1971. **55**(9): p. 4213-4229.
143. Ballenegger, V., R. Blaak, and J.P. Hansen, *Dipolar Fluctuations in the Bulk and at Interfaces*, in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 2*, M. Ferrario, G. Ciccotti, and K. Binder, Editors. 2006, Springer Berlin / Heidelberg. p. 45-63.
144. Kirkwood, J.G., *The Dielectric Polarization of Polar Liquids*. The Journal of Chemical Physics, 1939. **7**(10): p. 911-919.
145. Fröhlich, H., *Theory of dielectrics: dielectric constant and dielectric loss*. 1949: Clarendon press.
146. Yeh, I.C. and M.L. Berkowitz, *Dielectric constant of water at high electric fields: Molecular dynamics study*. Journal of Chemical Physics, 1999. **110**(16): p. 7935-7942.
147. Glosli, J.N. and M.R. Philpott, *Molecular dynamics study of interfacial electric fields*. Electrochimica Acta, 1996. **41**(14): p. 2145-2158.
148. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
149. Perram, J.W. and E.R. Smith, *Microscopic derivation of fluctuation formulas for calculating dielectric constants by simulation*. Journal of Statistical Physics, 1987. **46**(1): p. 179-190.
150. Ramshaw, J.D., *Existence of the dielectric constant in rigid-dipole fluids: The functional-derivative approach*. The Journal of Chemical Physics, 1977. **66**(7): p. 3134-3138.
151. Alastuey, A. and V. Ballenegger, *Dielectric constant and sample's shape in dipolar fluids*. J. Phys. IV France, 2000. **10**(PR5): p. Pr5-437-Pr5-440.
152. Schreiber, G. and A.R. Fersht, *Rapid, electrostatically assisted association of proteins*. Nat Struct Mol Biol, 1996. **3**(5): p. 427-431.
153. Northrup, S., J. Boles, and J. Reynolds, *Brownian dynamics of cytochrome c and cytochrome c peroxidase association*. Science, 1988. **241**(4861): p. 67-70.
154. Clore, G.M., C. Tang, and J. Iwahara, *Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement*. Current Opinion in Structural Biology, 2007. **17**(5): p. 603-616.
155. Gu, W., et al., *Alternative binding modes of proline-rich peptides binding to the GYF domain*. Biochemistry, 2005. **44**(17): p. 6404-6415.
156. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free*. J Comput Chem, 2005. **26**(16): p. 1701-18.
157. Essmann, U., et al., *A smooth particle mesh ewald method*. Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.

158. Wu, X., et al., *Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk*. *Structure* (London, England : 1993), 1995. **3**(2): p. 215-226.
159. Saraste, M. and A. Musacchio, *Backwards and forwards binding*. *Nat Struct Mol Biol*, 1994. **1**(12): p. 835-837.
160. Janz, J.M., T.P. Sakmar, and K.C. Min, *A Novel Interaction between Atrophin-interacting Protein 4 and β -p21-activated Kinase-interactive Exchange Factor Is Mediated by an SH3 Domain*. *Journal of Biological Chemistry*, 2007. **282**(39): p. 28893-28903.
161. DeLano, W.L., *The PyMOL molecular graphics system*. 2002.
162. Roman A, L., *SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions*. *Journal of Molecular Graphics*, 1995. **13**(5): p. 323-330.
163. McLain, S.E., et al., *Charge-Based Interactions between Peptides Observed as the Dominant Force for Association in Aqueous Solution*. *Angewandte Chemie-International Edition*, 2008. **47**(47): p. 9059-9062.
164. Buckle, A.M., G. Schreiber, and A.R. Fersht, *Protein-protein recognition - crystal structural-analysis of a barnase-barstar complex at 2.0-angstrom resolution*. *Biochemistry*, 1994. **33**(30): p. 8878-8889.
165. Frisch, C., et al., *Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation*. *Journal of Molecular Biology*, 1997. **267**(3): p. 696-706.
166. Berendsen, H.J.C., et al., *Molecular-dynamics with coupling to an external bath*. *Journal of Chemical Physics*, 1984. **81**(8): p. 3684-3690.
167. Godawat, R., S.N. Jamadagni, and S. Garde, *Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations*. *Proceedings of the National Academy of Sciences*, 2009. **106**(36): p. 15119-15124.
168. Sarupria, S. and S. Garde, *Quantifying Water Density Fluctuations and Compressibility of Hydration Shells of Hydrophobic Solutes and Proteins*. *Physical Review Letters*, 2009. **103**(3): p. 037803.
169. Cormen, T., et al., *Introduction to algorithms*, 1990, MIT press Cambridge, MA.
170. Luzar, A., *Resolving the hydrogen bond dynamics conundrum*. *Journal of Chemical Physics*, 2000. **113**(23): p. 10663-10675.
171. Csardi, G. and T. Nepusz, *The igraph software package for complex network research*. *InterJournal Complex Systems*, 2006. **1695**.
172. Stern, H.A. and S.E. Feller, *Calculation of the dielectric permittivity profile for a nonuniform system: Application to a lipid bilayer simulation*. *The Journal of Chemical Physics*, 2003. **118**(7): p. 3401-3412.
173. Neumann, M., *Dielectric relaxation in water. Computer simulations with the TIP4P potential*. *The Journal of Chemical Physics*, 1986. **85**(3): p. 1567-1580.
174. Teschke, O., G. Ceotto, and E.F. de Souza, *Interfacial water dielectric-permittivity-profile measurements using atomic force microscopy*. *Physical Review E*, 2001. **64**(1): p. 011605.
175. Deleeuw, S.W., J.W. Perram, and E.R. Smith, *Computer-simulation of the static dielectric-constant of systems with permanent electric dipoles*. *Annual Review of Physical Chemistry*, 1986. **37**: p. 245-270.

176. Simonson, T., *Dielectric relaxation in proteins: Microscopic and macroscopic models*. International Journal of Quantum Chemistry, 1999. **73**(1): p. 45-57.
177. Wang, L., et al., *Downhill binding energy surface of the barnase-barstar complex*. Biopolymers, 2010. **93**(11): p. 977-985.
178. Watts, R.O., *Electric polarisation of water: Monte Carlo studies*. Chemical Physics, 1981. **57**(1-2): p. 185-195.
179. Sansom, M.S.P., et al., *The dielectric properties of water within model transbilayer pores*. Biophysical Journal, 1997. **73**(5): p. 2404-2415.
180. Neumann, M., *Dipole moment fluctuation formulas in computer simulations of polar systems*. Molecular Physics, 1983. **50**(4): p. 841-858.
181. Ben-Naim, A., *Strong forces between hydrophilic macromolecules: Implications in biological systems*. The Journal of Chemical Physics, 1990. **93**(11): p. 8196-8210.
182. Durell, S.R., B.R. Brooks, and A. Ben-Naim, *Solvent-Induced Forces between Two Hydrophilic Groups*. The Journal of Physical Chemistry, 1994. **98**(8): p. 2198-2202.
183. Cheung, M., A. García, and J. Onuchic, *Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse*. Proc Natl Acad Sci U S A, 2002. **99**(2): p. 685-90.
184. Levy, Y. and J.N. Onuchic, *Water mediation in protein folding and molecular recognition*. Annual Review of Biophysics and Biomolecular Structure, 2006. **35**: p. 389-415.
185. Papoian, G.A., J. Ulander, and P.G. Wolynes, *Role of water mediated interactions in protein-protein recognition landscapes*. Journal of the American Chemical Society, 2003. **125**(30): p. 9170-9178.
186. Tame, J.R.H., et al., *The role of water in sequence-independent ligand binding by an oligopeptide transporter protein*. Nature Structural Biology, 1996. **3**(12): p. 998-1001.
187. Rodier, F., et al., *Hydration of protein-protein interfaces*. Proteins, 2005. **60**(1): p. 36-45.
188. Born, M., *Volumen und hydrationswärme der ionen*. Zeitschrift für Physik A Hadrons and Nuclei, 1920. **1**(1): p. 45-48.
189. Laidler, K.J. and C. Pegis, *The Influence of Dielectric Saturation on the Thermodynamic Properties of Aqueous Ions*. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 1957. **241**(1224): p. 80-92.
190. Buckingham, A.D., *A theory of ion-solvent interaction*. Discussions of the Faraday Society, 1957. **24**: p. 151-157.
191. Buckingham, A.D., *Molecular quadrupole moments*. Quarterly Reviews, Chemical Society, 1959. **13**(3): p. 183-214.
192. Bockris, J.O.M. and A.K.N. Reddy, *Modern Electrochemistry 1: Ionics, 2nd Edition*. 1998: Springer.
193. Booth, F., *The Dielectric Constant of Water and the Saturation Effect*. The Journal of Chemical Physics, 1951. **19**(4): p. 391-394.
194. Booth, F., *Dielectric Constant of Polar Liquids at High Field Strengths*. The Journal of Chemical Physics, 1955. **23**(3): p. 453-457.
195. Aguilera-Arzo, M., et al., *Dielectric saturation of water in a membrane protein channel*. Physical Chemistry Chemical Physics, 2009. **11**(2): p. 358-365.

196. Liszi, J., L. Mészáros, and I. Ruff, *The field dependence of the Kirkwood factor and the nonlinear dielectric behavior of some liquids*. The Journal of Chemical Physics, 1981. **74**(12): p. 6896-6901.
197. Ganti, G., J.A. McCammon, and S.A. Allison, *Brownian dynamics of diffusion-controlled reactions: the lattice method*. The Journal of Physical Chemistry, 1985. **89**(18): p. 3899-3902.
198. Northrup, S.H., et al., *Dynamical theory of activated processes in globular proteins*. Proceedings of the National Academy of Sciences, 1982. **79**(13): p. 4035-4039.
199. Schreiber, G., *Kinetic studies of protein-protein interactions*. Current Opinion in Structural Biology, 2002. **12**(1): p. 41-47.
200. Chou, K.C. and G.P. Zhou, *Role of the protein outside active site on the diffusion-controlled reaction of enzymes*. Journal of the American Chemical Society, 1982. **104**(5): p. 1409-1413.
201. Marcus, Y., *The solvation number of ions obtained from their entropies of solvation*. Journal of Solution Chemistry, 1986. **15**(4): p. 291-306.
202. Marcus, Y., *Solvent release upon ion association from entropy data*. Journal of Solution Chemistry, 1987. **16**(9): p. 735-744.
203. Marcus, Y., *Solvent Release upon Ion Association from Entropy Data. II*. The Journal of Physical Chemistry B, 2006. **111**(3): p. 572-580.
204. Massova, I. and P.A. Kollman, *Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies*. Journal of the American Chemical Society, 1999. **121**(36): p. 8133-8143.
205. Massova, I. and P.A. Kollman, *Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding*. Perspectives in drug discovery and design, 2000. **18**(1): p. 113-135.
206. Srinivasan, J., et al., *Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices*. J. Am. Chem. Soc., 1998. **120**(37): p. 9401-9409.
207. Wang, W. and P.A. Kollman, *Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model*. Journal of Molecular Biology, 2000. **303**(4): p. 567-582.
208. Wang, J., et al., *Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA*. J. Am. Chem. Soc., 2001. **123**(22): p. 5221-5230.
209. Benedix, A., et al., *Predicting free energy changes using structural ensembles*. Nature Methods, 2009. **6**(1): p. 3-4.