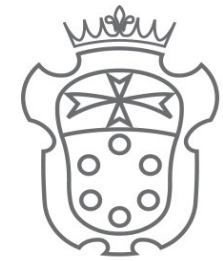

SCUOLA
NORMALE
SUPERIORE

UNIVERSITÀ
DEGLI STUDI
DE L'AQUILA

A computational study on the hydration-shell properties of antifreeze and non-antifreeze proteins

Thesis for the Degree of ***Doctor of Philosophy*** *in*
Methods and Models for Molecular Sciences

Candidate

Akash Deep Biswas

Advisors

Prof. Vincenzo Barone (SNS)

Prof. Isabella Daidone (UNIVAQ)

Academic Year 2020/2021

Classes of Science
Scuola Normale Superiore
University of L'Aquila
Pisa, Italy - 56126

Candidate's Declaration

I hereby declare that the work presented in the thesis entitled **"A computational study on the hydration-shell properties of antifreeze and non-antifreeze proteins"** in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy** and submitted in the classes of science of the Scuola Normale Superiore di Pisa is an authentic record of my own work carried out during a period from 2016 to 2020 under the supervision of **Prof. Vincenzo Barone**, Classes of Science of Scuola Normale Superiore di Pisa and co-supervision of **Prof. Isabella Daidone**, Department of Physical and Chemical Sciences, University of L'Aquila.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute/University.

(Akash Deep Biswas)

This is to certify that the above statement made by the candidate is true to the best of our knowledge and belief.

(Vincenzo Barone)
Place: Pisa **Professor, Classes of Science**

(Isabella Daidone)
Place: L'Aquila **Professor, PCS Department**

Dedicated to my family.

ACKNOWLEDGMENTS

I am delighted to express my respect for and profound thanks to my supervisor Professor Vincenzo Barone, Classes of Science, Scuola Normale Superiore, co-supervisor Professor Isabella Daidone, Department of Physical and Chemical Sciences, University of L'Aquila, Professor Andrea Amadei Department of Chemical Science and Technology, University of Tor Vergata, Rome, Dr. Giordano Mancini, Scuola Normale Superiore. I also thank them for their hard work on the manuscripts without which this work could not have been completed and improve them thoroughly.

I am grateful to Professor Vincenzo Barone, Head of SMART LAB, Scuola Normale Superiore di Pisa, Professor Isabella Daidone, University of L'Aquila, and Professor Andrea Amadei, University of Tor Vergata, for providing all of the facilities, assistance, and encouragement needed to complete the research.

I would like to express my heartfelt appreciation to some of the members of SMART LAB, Scuola Normale Superiore, Scientific Advisor Monica Sanna, Dr. Giordano Mancini, Dr. Sara Del Galdo, Alberto Coduti, member from University of L'Aquila Dr. Laura Zanetti-Polzi, and Avogadro staff members for their prompt assistance and cooperation throughout the investigation.

I am grateful to my parents, Jhunu Rani Biswas and Siba Pada Biswas, my siblings Debashree Biswas Barman and Gaurav Deep Biswas, my relatives Daivik Biswas Barman and Manjit Barman, for their moral support, love, encouragement, and blessings in completing this task.

I'd like to thank my friends Anna Maria Frontino, Gaurav Dimri, Giulio Parrotta, Omer Arif, Vinoshene Pillai Rajan, as well as the department's research fellows, for their assistance and motivation throughout my research. I would also like to express my heartfelt gratitude to my friends and all other people whose names do not appear here for assisting me, either directly or indirectly, in all good and bad times. Special thanks to those friends too for sharing their lunch boxes with me during my schooling.

I am also grateful to my Research Publications' anonymous reviewers. They made very useful observations and suggestions in shaping my research.

Finally, I am thankful and indebted to the Almighty for helping me.

(Akash Deep Biswas)

ABSTRACT

Here we present a computational approach based on molecular dynamics (MD) simulation to study the hydration-shell density of several proteins which include a special group of proteins, namely antifreeze proteins, AFPs. AFPs have the ability to inhibit ice growth by binding to ice nuclei. Their ice-binding mechanism is still unclear, yet the hydration layer is thought to play a fundamental role. In particular, the hydration-shell density of eighteen different proteins comprising eight AFPs is calculated. The results obtained show that an increase in the hydration-shell density, relative to that of the bulk, is observed (in the range of 4–14%) for all studied proteins and that this increment strongly correlates with the protein size, while it does not depend on whether the protein is an AFP or not. In particular, a decrease in the density increment is observed for decreasing protein size. A simple model is proposed according to which almost all of the hydration-density increase is located in pockets within, or at the surface of, the protein molecule. We then further investigated the local properties of the hydration shell around the ice-binding surface (IBS) of the AFPs. We found that the hydration shell density of the ice-binding surfaces is always higher than the bulk density and, thus, no ice-like (i.e. with a density lower than the bulk) layer is detected at the IBS. However, the local water-density around the IBS is found to be lower than that around the non-ice-binding surfaces and this difference correlates to the higher hydrophobic character of the IBS with respect to the non-IBS. We hypothesize that the lower solvent density at the ice-binding site can pave the way to the protein binding to ice nuclei, while the higher solvent density at the non-ice-binding surfaces might provide protection against ice growth. Finally, we tested our hypothesis by studying the dependence of the antifreeze activity of seven AFPs on various structural and chemical properties of the IBS and non-IBS and found that the activity strongly correlates with the difference in the local hydration-shell properties of the non-ice-binding surfaces, rather than of the IBSs.

Contents

Certificate iii

Dedication vi

Acknowledgments vii

Abstract ix

1 Introduction 1

1.1 The hydration shell of proteins 1

1.1.1 Antifreeze Proteins 1

1.1.2 Thermal Hysteresis 2

1.1.3 Types of Antifreeze Proteins 4

1.2 Outline of the thesis 4

2 Computational Tools 7

2.1 Molecular Dynamics Simulation 7

3 Length-Scale Dependence of Protein Hydration-Shell Density 13

3.1 Introduction 13

3.2 Theory 14

3.3 Methods 17

3.3.1 MD simulations 17

3.3.2 Protein volume and hydration-shell density calculations 18

3.4 Results and Discussion 20

3.5 Conclusions 28

4 Hydration Shell of Antifreeze Proteins: Unveiling the Role of Non-Ice-Binding Surfaces 31

4.1 Introduction 31

4.2 Methods 33

4.2.1 MD Simulations 33

4.2.2 Hydration Shell Density Calculations 34

4.2.3 Definition of Binding Surfaces 35
4.3 Results and Discussion 35
4.4 Conclusion 47

5 High water-density at non-ice-binding surfaces contributes to the hyperactivity of antifreeze proteins 49
5.1 Introduction 49
5.2 Materials and Methods 51
5.2.1 Molecular Dynamics Simulation 51
5.2.2 Protein hydration shell density calculation 51
5.2.3 Partial molar volume calculation 52
5.3 Results and Discussion 53
5.4 Conclusions 61

6 Conclusions 63

List of Publications 67

Biblography 69

List of Figures 87

List of Tables 89

Chapter 1

Introduction

1.1 The hydration shell of proteins

The hydration of biomolecular systems is essential in their stabilization, [1] dynamics, [2–4] folding, [5, 6] and function. Similarly, shifting the point of view from the biomolecule to the solvent, proteins strongly modify the structural and dynamical properties of their aqueous surroundings according to both theoretical [7–16] and experimental evidence. [3, 17–30] Despite the extent of the perturbation of the water network around proteins has been investigated by numerous methods in the past half-century, a consensus has still not emerged regarding the spatial range of this perturbation and the structural and dynamical features of the hydrating water molecules. Most measurements show that the thickness of protein hydration shells is in the range of 1–1.2 nm, [11–13, 16, 31] while spectroscopic terahertz experiments suggest the existence of larger, long-ranged, dynamic solvation shells. [23, 32, 33]

Although it is now well accepted that the protein hydration shell has a higher density than the bulk one, there isn't a consensus on the magnitude of this rise. [8, 10, 13, 14, 16, 18] Depending on the specific protein under investigation, on the choice of the experimental or computational technique, and the statistical methods utilized for the analysis, and on the size of the identified hydration shell, the estimated density increment with respect to bulk water can range from 1 % to 50 %. [8,10,14,18]

The hydration waters of a special class of proteins, called antifreeze proteins (AFPs), play a critical role in carrying out their activity, which consists in preventing ice from growing upon cooling below the bulk melting point in the body fluids of the host organism. Given that particular attention is given to this class of proteins in the present work, the AFPs, and their properties, are presented in the next subsections.

1.1.1 Antifreeze Proteins

In the 1950s Scholander, a Norwegian scientist, proposed for the first time a mechanism through which Arctic fishes can live in water colder than the freezing point of their blood based on the presence of an antifreeze agent in their body fluids. [34] After

approximately a decade, Arthur DeVries, an animal biologist, was able to isolate the biomolecule from Antarctic fish that is capable of inhibiting ice growth when partly frozen samples are further cooled down [35]. This kind of biomolecules, nowaday called antifreeze proteins, have been found also in other species that survive in sub-zero temperatures, including plants, [36] bacteria, [37,38] insects, [39] snow alga, [40] snow-mold fungi, [41] and sea-ice diatoms. [42]

Antifreeze proteins have nowadays a wide range of applications, in particular in the medical field of cell-based therapies, which are having a boom as next-generation medicines for intractable and complex disorders. [43] Mammalian cell cryopreservation, which is used to store cells and tissues for long periods of time, is an important part of the production process. The long storage process has been recognized as a possible barrier in the future growth of complex cellular therapy drugs. [44] Ice recrystallization cause the development of large ice crystals at the expense of small crystals, leading to cellular damage and thereby being a significant contributor to cell death. Dimethyl sulfoxide (DMSO) is the most commonly used cryoprotective agent for the cryopreservation of mammalian cells and tissues in cell suspension, as DMSO lowers their freezing point. [45]. Cell survival rates, however, are reduced due to DMSO cytotoxicity and suppression of internal signaling. As a result, the discovery of novel cryoprotectants suited to cryopreservation formats and to substitute or eliminate DMSO material is critical for the future production of cell-based therapies and diagnostics. Thus, naturally occurring proteins have caught the interest of cryopreservation researchers due to their good capacity to prevent ice recrystallization or modify ice nucleation, as well as their ability to interact and stabilize biological membranes, making them interesting molecules to use in cryopreservation protocols. [46,47]

1.1.2 Thermal Hysteresis

Although many details of the mechanisms through which the AFPs carry out their activity are unknown, it is widely accepted that the AFPs, despite the high variability in the primary, secondary and tertiary structures, share the common mechanism of binding to small ice nuclei preventing further growth of ice at the adsorbed positions. [48] The AFPs bind to the ice surface by leveraging many hydrogen bonding and hydrophobic interactions between specific ice planes and the ice-binding surface (IBS), which is commonly flat and highly specialized in ice binding. The rest of the protein surface is known as the non-ice-binding surface (non-IBS). A schematic diagram of the IBS and non-IBS of an AFP is reported in Figure 1.1.

The ability of AFPs to inhibit ice growth is limited, in that upon sufficient cooling there is a sudden and rapid ice growth. The difference between the temperature at

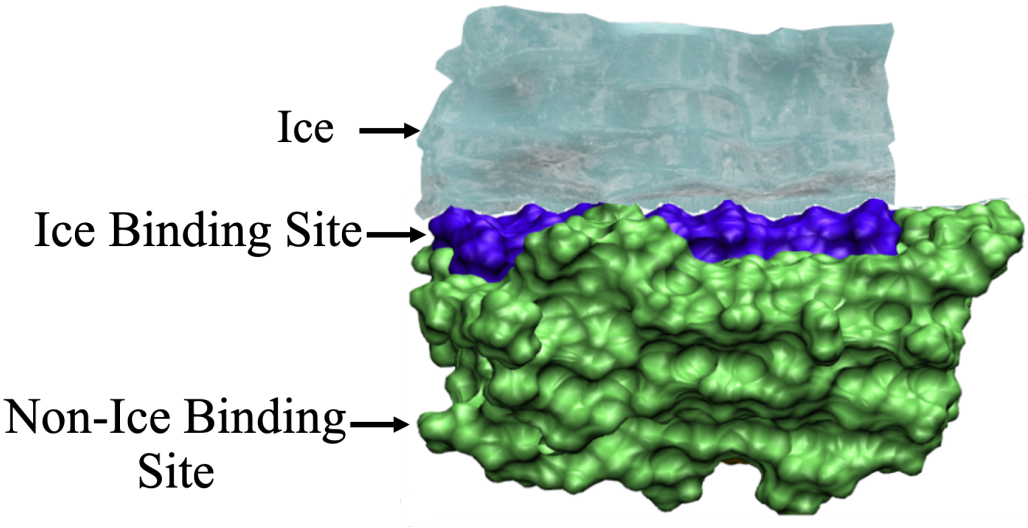

Figure 1.1: Schematic illustration of an antifreeze protein adsorbed to an ice nucleus through the ice-binding-surface (colored in blue). The rest of the protein surface, colored in green, represents the non-ice-binding surface.

which ice starts growing in the presence of AFPs and the bulk melting point has been termed thermal hysteresis (TH), and the temperature at which the ice begins to grow is termed the hysteresis freezing point. The temperature interval of ice growth inhibition is referred to as the hysteresis gap, and the quantitative difference between the melting point and the hysteresis freezing point is referred to as the hysteresis activity.

The dominant hypothesis of how thermal hysteresis occurs is known as the adsorption-inhibition model. [49, 50] According to this model, antifreeze proteins adsorb irreversibly onto the ice surface, limiting ice growth to convex ice fronts between the adsorbed antifreeze proteins. The high surface–to-volume ratio at the interface leads to a shift to lower ice-growth temperature. Based on this model, the hysteresis activity can be related to the adsorbants spacing. [51] Several alternative models, which assume an equilibrium exchange (i.e. a reversible binding) of antifreeze proteins with the ice surface, have been suggested in recent years. [52–54] The underlying distinction between the adsorption–inhibition model and the alternative theories is that the former explains the antifreeze activity through modifications in the surface structure of the crystal resulting in ice–water energy equality within the hysteresis gap, while the latter explains the phenomenon through changes in the ice–water interfacial stress that do not include any ice–water energy equality. Nevertheless, all proposed models assume that the surface water molecules serve as an energy barrier to ice growth.

1.1.3 Types of Antifreeze Proteins

The AFPs are classified on the base of their TH values: moderately active AFPs ($\Delta T < 1$ K), mostly alanine-rich AFPs from fish, fungi and plants, and hyperactive AFPs ($\Delta T > 1$ K), mostly threonine-rich AFPs found in insects. At lower concentrations (<0.5 g/L), hyperactive AFPs are much more active than the moderately-active ones. The classes which the AFPs studied in the present work belong to are described in the following paragraphs; these are types I-IV fish AFPs, hyperactive insect AFPs and fungi AFPs. Representative structures are shown in Figure 1.2.

Type I AFPs are alanine-rich proteins, the structure of which is a single α-helix. These proteins are found in fish Yellowtail flounders (*Limanda ferruginea*), Grubby sculpins (*Myoxocephalus aenaeus*), Winter flounders (*Pseudopleuronectes americanus*), and Shorthorn sculpins (*Myoxocephalus scorpius*). [55] Type II AFPs are globular, cysteine-rich molecules. They are found in Sea raven (*Hemitryptrrus*), smelt (*Osmerus mordax*), and Atlantic herring (*Clupeah harentus harengus*). Type II fish AFPs are further classified as calcium dependent (herring and smelt) or calcium independent (sea raven). [56] Type III AFPs can be present in both Northern and Southern Eelpout (*Macrozoarces americanus*). They are globular proteins, and their primary sequences are not dominated by any specific aminoacidic pattern. Type IV AFPs have been found in the Longhorn sculpin (*Myoxocephalus octodecimspinosis*). They may have 22 amino acid repeats, but no three-dimensional structure has been reported so far. [57]

Hyperactive AFPs have been discovered in insects. They have been initially found in mealworm beetle (*Tenebrio molitor*) and fire beetle (*Dendroides Canadensis*). Both of these proteins have a high content of threonines and cysteines and have a β-helix structure with very similar 12-13 amino acid repeats. Spruce budworm (*Choristoneura fumiferana*) AFP is similar to the two insect AFPs described above, but with some differences: it contains a high content of serines and threonines and has a left-handed-helix arrangement with T-X-T aminoacidic patterns in the ice-binding surface. [58,59]

AFPs can also be found in fungi, as for example in Snow mold fungus from *Typhula ishikariensis*. The corresponding AFP is rich in threonines and has the characteristic T-X-T aminoacidic pattern in the IBS [60].

1.2 Outline of the thesis

Here we present a computational approach based on molecular dynamics (MD) simulation to study the hydration-shell of various proteins, including AFPs and non-

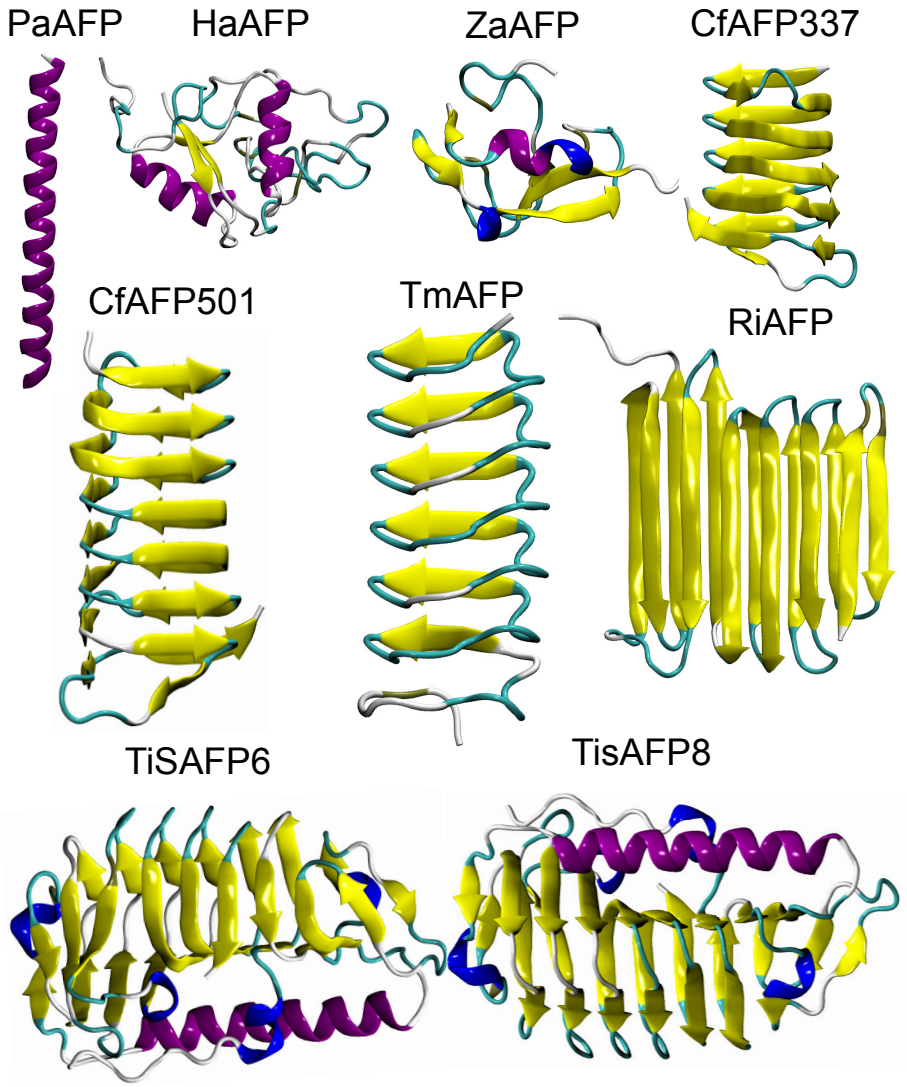

Figure 1.2: Antifreeze proteins structures. Representative configurations are drawn for the different types of AFPs from fish, insects and fungi. α helices are shown in purple, β strands in yellow, 3_{10} helices in blue and coil in cyan. The different AFPs are: fish AFP type I *Pseudopleuronectes americanus* (PaAFP), fish AFP type II *Hemitripterus americanus* (HaAFP), fish AFP type III *Zoarces americanus* (ZaAFP), insect AFP *Choristoneura fumiferana* isoform 337 (CfAFP337), insect AFP *Choristoneura fumiferana* isoform 501 (CfAFP501), insect AFP *Rhagium inquisitor* (RiAFP), snow mold fungus AFP *Typhula ishikariensis* isoform 6 (TisAFP6) and snow mold fungus *Typhula ishikariensis*isoform 8 (TisAFP8).

antifreeze proteins (non-AFPs). In particular, we analyze the dependency of the hydration density on the size of the protein molecule and, for the AFPs, also on the antifreeze activity. In fact, although the mechanism of ice-binding of the AFPS is still not fully understood, the hydration-shell is considered to be of fundamental importance to carry out the antifreeze activity.

The methods utilized are discussed in Chapter 2 of this thesis. The hydration-density of a large set of proteins, containing both AFPs and non-AFPs, are then

characterized and the results reported in Chapter 3. For each protein, the relative density increase with respect to bulk water is calculated and its dependence on the protein size analyzed. In Chapter 4 two of the studied AFPs, one moderately active and one hyperactive, are analyzed in more details in order to characterize the local hydration-density properties around different surface of the proteins. The results are compared with the local hydration properties of two non-AFPs in order to possibly unveil distinct local features around the AFPs surfaces. In Chapter 5 possible correlations between the local hydration-density features around ice-binding and non-ice-binding surfaces and the antifreeze activity are analyzed for a set of seven AFPs, three of which are moderately-active and four hyperactive AFPs. Finally, the conclusions of the thesis are reported in Chapter 6.

Chapter 2

Computational Tools

2.1 Molecular Dynamics Simulation

Systems that are too complex to be defined as quantum states can be adequately described through the use of classical approaches. As soon as one rules out the possibility of a chemical reaction, a low temperature, or an accurate description of hydrogen atoms motion, a system can be safely approximated as a semi-classical system governed by the classical mechanics' laws. The Molecular Dynamics (MD) trajectory of an isolated system is the concatenation of the system's configurations, all of which is integrated with the Newton equations of motion for all of the system's atoms

$$\frac{d^2}{dt^2}\boldsymbol{r}_i = m_i^{-1}\boldsymbol{F}_i \tag{2.1.1}$$

where

$$\boldsymbol{F}_i = -\frac{\partial V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)}{\partial \boldsymbol{r}_i} \tag{2.1.2}$$

is the force acting on the i-th atom of mass m_i, and t is the time.

$V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$ is the force field operating on the system; it is an effective interaction mechanism since it includes the average effect of the electronic degrees of freedom implicitly (i.e. the effect of the electronic distribution). As a result, the definition of the force field (along with the selection of the molecular model) is essential to accurately predicting the system's properties. Many different forms of force fields are accessible for molecular mechanics simulations (such as CHARMM, [61] AMBER, [62–64] or GROMOS [65,66]), all sharing a common practical form of potential energy that involves bonded terms for interactions of atoms connected by covalent bonds and non-bonded terms for long-range electrostatic and van der Waals interactions. A standard molecular force field for a system of N atoms has the following form:

$$V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = \sum_{bonds} \frac{1}{2} K_b (b - b_{eq})^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} K_\phi [1 + cos(n\phi - \delta)] + \sum_{imp.\ dihedrals} \frac{1}{2} K_\xi (\xi - \xi_{eq})^2 + \sum_{coppie} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{\boldsymbol{r}_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{\boldsymbol{r}_{ij}} \right)^{6} \right] + \sum_{coppie} \frac{q_i q_j}{4\pi\varepsilon_0 \boldsymbol{r}_{ij}} \tag{2.1.3}$$

The first term describes the covalent bond stretching interaction, which is represented as a harmonic potential with b_{eq} representing the minimum energy bond length (i.e. the equilibrium length) and K_b representing the force constant.The second term is a three-body interaction corresponding to the bond-angle (θ) deformation expressed as a harmonic potential, where θ_{eq} is the equilibrium bond-angle and K_θ is the force constant. The third and fourth terms are used to represent the four-body proper and improper dihedral angle interactions. The latter is intended to preserve planarity or avoid unintended chiral inversion during a simulation and is represented as a harmonic potential (where, as normal, ξ_{eq} represents the equilibrium value and K_{xi} represents the force constant), while the proper dihedral (ϕ) interaction is represented as a sinusoidal function.The final two terms are sums of the non-bonded atom pairs which describe the efficient non-bonded interactions expressed in terms of van der Waals and Coulombic interactions between the i-th and j-th atoms at the distance $\boldsymbol{r}_{ij}$. The parameters ε_{ij} and σ_{ij} define the Lennard-Jones potential, q_i and q_j are the atom charges, and ε_0 is the dielectric constant in vacuum.

The parameters used in the force field framework can be defined in a variety of ways. They can be equipped with *ab initio* measurement findings on small molecular clusters. Alternatively, the parameters can be fitted to experimental data such as crystal structures, energy and lattice mechanics, infrared or X-ray data on small molecules, liquid properties such as density and enthalpy of vaporization, free energy of solvation, nuclear magnetic resonance data, and so on. It is worth noting that all parameters are optimized for specific types of molecules (inorganic molecules, organic molecules, biomolecules, etc.) and thus any force field is typically well suited for specific general conditions (i.e. specific thermodynamic conditions and boundary conditions), and hence the choice of the forcefield depends on the precise system to be simulated.

Any second-order differential equation in 2.1.1 can be rewritten as two first-order differential equations for the particle position $\boldsymbol{r}_i$ and velocity $\boldsymbol{v}_i$, as seen below.

$$\frac{d}{dt}\boldsymbol{v}_i(t) = m_i^{-1}\boldsymbol{F}_i \tag{2.1.4}$$

$$\frac{d}{dt}\boldsymbol{r}_i(t) = \boldsymbol{v}_i(t) \tag{2.1.5}$$

These equations are solved using finite difference method approaches: given the molecular positions, velocities, and forces at time t, the algorithms calculate the positions, velocities, and forces at time $t + \delta t$, and the equations are solved step by step. The time phase δt is determined by the process of solution, but it is normally selected to be significantly shorter than the time it takes a molecule to move its length. Verlet and its computationally effective counterpart leap-frog, [67, 68] Beeman, [69] or the Gear predictor-corrector [70] are the most commonly used algorithms (falling under the general finite difference scheme).

Due to computing costs, a condensed phase system is usually simulated as a box containing $10^4 - 10^5$ atoms, i.e. a much smaller scale than the macroscopic counterpart system. As a result, a significant portion of the system's molecules would be arranged on the surface of the simulation box, subjecting them to forces somewhat different from the bulk molecules. Periodic Boundary Conditions (PBC) [71] are used to prevent any inhomogeneity. The method involves simulating the system in a central box surrounded by an infinite number of translated copies of itself, known as *images*. The molecules in the initial central box and their periodic representations shift precisely the same throughout the simulation. As a result, when a molecule exits the central box, one of its *images* appears from the opposite side. As a consequence, neither spatial borders nor surface molecules exist.

Since intermolecular forces decay rapidly with distance, a maximum distance value (typically 1-2 nm) can be established such that forces can be directly measured within such a distance, but can be ignored for distances greater than this cut-off. Since the electrostatic force has a significant effect on the interaction between particles that are more than 1-2 nm apart, approximated methods for computing long-range interaction contributions must be used. The Ewald summation technique, [71] which employs periodic boundary conditions, is the most commonly used method for simulating condensed phase systems. The cumulative electrostatic energy of N interacting particles and their periodic representations can be calculated as follows:

$$E = \frac{1}{8\pi\varepsilon_0}\sum_{\boldsymbol{n}=0}^{\infty}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{|\boldsymbol{r}_{ij}+\boldsymbol{n}|}\right) \tag{2.1.6}$$

where $\boldsymbol{n} = (n_x L\ n_y L\ n_z L)^T$ with n_x, n_y, n_z integers and L length of the cubic cell edges. To avoid self-interactions, $\boldsymbol{n} = 0$, $i = j$ terms are omitted. It was shown that

the sum over $\boldsymbol{n}$ for such a potential ($\boldsymbol{r}^{-1}$) is only conditionally convergent, and that its limit will differ or even diverge if the order of the terms in the sum is modified. This issue can be solved by subtracting by subtracting to each simulated point charge ρ^q a Gaussian charge distribution ρ^G of equal magnitude and opposite sign with respect to ρ^q, spreading out radially from the charge. These distributions have the effect of screening interactions between neighboring point charges, resulting in a short-ranged interaction energy. The cumulative charge distribution is then calculated *via*:

$$\rho^q = \rho^s - \rho^G \tag{2.1.7}$$

where

$$\rho^s = \rho^q + \rho^G \tag{2.1.8}$$

First, the interaction energy due to ρ^s is calculated in the real space, then, in order to recover the original charge distribution, $-\rho^G$ is computed by means of a Fourier transform. The final interaction energy form

is

$$\begin{aligned} E = & \frac{1}{4\pi\varepsilon_0}\frac{1}{2}\sum_{\boldsymbol{n}}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{|\boldsymbol{r}_{ij}+\boldsymbol{n}|}erfc\left(\frac{|\boldsymbol{r}_{ij}+\boldsymbol{n}|}{\sqrt{2}\sigma}\right) \\ & + \frac{1}{2V\varepsilon_0}\sum_{\boldsymbol{k}\neq 0}\frac{e^{-\sigma^2 k^2/2}}{k^2}|S(\boldsymbol{k})|^2 \\ & - \frac{1}{4\pi\varepsilon_0}\frac{1}{\sqrt{2\pi}\sigma}\sum_{i=1}^{N}q_i^2 \\ & + \frac{|\sum_{i=1}^{N}q_i\boldsymbol{r}_i|^2}{2\varepsilon_0 L^3(2\varepsilon_r+1)} \end{aligned} \tag{2.1.9}$$

where

$$S(\boldsymbol{k}) = \sum_{i=1}^{N} q_i e^{i\boldsymbol{k}\cdot\boldsymbol{r}_i} \tag{2.1.10}$$

and $erfc(|\boldsymbol{r}_{ij}+\boldsymbol{n}|/\sqrt{2\sigma^2})$ is the complementary error function (i.e. $erfc(z) = 2/\sqrt{\pi}$ $\int_z^\infty e^{-t^2}\,dt$) which falls to zero with increasing its argument. As a result, if σ is small enough, the amount over $\boldsymbol{n}$ in the first term is reduced to the single term $\boldsymbol{n} = 0$. The second term is a sum over the reciprocal vectors $\boldsymbol{k}$. Again, if σ is small, the terms in the $\boldsymbol{k}$-space sum are needed to achieve energy convergence. The third term represents a correction function for the self-interactions included in the previous term for $\boldsymbol{n} = 0$. Finally, the last term is the energy contribution of the depolarizing field due to the total electric moment of the point charges, which is compensated by the effect of

the external dielectrics. In fact, the virtual cubic cells are organized as concentric spherical layers beginning with the central box, with the $\boldsymbol{r} \to \infty$ sphere immersed in a continuum dielectric with a dielectric constant of ε_r. If ε_r approaches infinity, this term disappears due to complete compensation of the two impacts. *Particle Mesh Ewald* [72] (PME) is a widely used method for improving the efficiency of the reciprocal amount.

When Newton's equations of motion are combined, total energy is conserved, and if the volume of the simulation box is kept constant, the simulation produces a microcanonical ensemble (NVE). Other statistical ensembles, such as canonical (NVT) and isothermal-isobaric (NPT), however, better describe the experimental conditions than the microcanonical. As a result, methods to constrain temperature and pressure have been developed. The thermal bath coupling method, or *Berendsen coupling* [73], has the great advantage of being simple. This algorithm simulates a coupling of the system with an external thermal bath at the temperature T_0. The interaction between the bath and the system is modulated by a time constant τ, the coupling is obtained by multiplying the velocities for a constant and the temperature T is scaled to the reference temperature T_0 through an exponential law. The *isothermal*, or isogaussian [74], approach allows for precise temperature control. Using this algorithm, a term is applied to the motion equations that acts as a changing friction force in time to maintain the kinetic energy steady. The configurational properties of the canonical ensemble are correctly generated by this approach, but the momenta distribution is not canonical [75].

The *velocity rescaling* method [76] was recently developed. In this approach the velocities of all the particles are rescaled by a properly chosen random factor. This thermostat is similar to Berendsen coupling, with the same scaling using τ, but the stochastic term ensures that a proper canonical ensemble is generated. Typically, the different techniques for doing MD at constant pressure are based on the same rules as the constant temperature system, with the pressure playing the role of the temperature and the atomic velocities playing the role of the atomic positions.

When Newton's equations of motion are integrated, the limiting factor determining the time step to be utilized, is the highest frequency occurring in the system. The vibrations of bonds involving hydrogen atoms are the highest frequency vibrations of solvated biological macromolecules. An O-H bond's bond stretching frequency is usually about 10^4 Hz, so the average time would be of the order of 10 fs. For a fair sampling of a periodic function, samples should be taken at least twenty times per period, so the MD simulation time-step should be approximately ≈ 0.5 fs [77]. Since these bond vibrations are practically uncoupled from all other vibrations in the sys-

tem, constraining them does not notably alter the system's dynamics. SHAKE [78] and LINCS [79] are the most often used techniques for constraining bond vibrations. This is not the case for bond-angle fluctuations, which generate the second-highest frequency vibrations. In fact, constraining bond-angles has a significant impact on many other system fluctuations, including global, collective fluctuations. Furthermore, a method for increasing simulation efficiency can be used: instead of simulating hydrogen atoms in macromolecules bound to the molecule by bonds, angles, and dihedrals, the location of the hydrogens can be produced every MD step based on the position of three neighboring heavy atoms. All forces acting on the hydrogen atom are then redistributed through these heavy atoms, and the mass of each hydrogen is added to the mass of the bound heavy one [80].

Chapter 3

Length-Scale Dependence of Protein Hydration-Shell Density

3.1 Introduction

The hydration of bio-molecular systems plays important roles in their stability [1], dynamics [2–4], folding [81], ligand recognition [5, 6] and function [9, 82]. There have been a variety of theoretical [7–16] and experimental [3, 17–30] studies demonstrating that proteins modify the structural and dynamical organization of their surrounding water molecules. The thickness of protein hydration shells inferred from most studies is in the range of 1-1.2 nm [11–13, 16, 31], although larger, long-ranged dynamical solvation shells have been hypothesized on the basis of terahertz spectroscopical studies [23, 32, 33].

Although it is now rather well accepted that the density of the protein hydration shell is higher than the bulk one [8, 10, 13, 14, 16, 18], there is not a consensus about the magnitude of this increment. Depending on the protein, on the experimental technique, or on the computational approach, used and on the distance from the protein surface at which the density increment is evaluated, different estimates have been given ranging from 1% to 50% [8, 10, 14, 18].

Here, we focus on the dependence of the protein hydration shell density on the protein size, and partly function by analyzing the hydration properties of eighteen different proteins, out of which eight are antifreeze proteins (AFPs), by means of MD simulation and a recently developed analysis to investigate hydration shell densities [13, 16]. Previous work have focused on the length-scale dependence of the hydration density of hydrophobic solutes [83–86] but, to the best of our knowledge, this is the first systematic analysis of the correlation between hydration shell densities and the size and shape of globular proteins, which are systems with amphipathic surfaces.

3.2 Theory

The basic idea for constructing a simple, yet reasonable, model describing the hydration shell-density variations as a function of the protein size and geometry, is to approximate the protein molecule as an ellipsoid and to partition the relevant quantities (i.e. the protein excluded volume V_{ex}, the solvent-accessible volume V_{acc}, and the solvent density ρ_{sh}, within the solvation shell of the protein). The hydration shell is basically divided into two parts: one residing inside the effective protein ellipsoid (V_{ex}^{in}, V_{acc}^{in} and ρ_{in}, respectively), the other residing within the solvation shell but outside the effective protein ellipsoid (V_{ex}^{out}, V_{acc}^{out} and ρ_{out}, respectively) (see Fig. 3.1). For a description of the procedure used to define the effective protein ellipsoid volume, $V_{ell,eff}$ and the solvation shell see the Methods section.

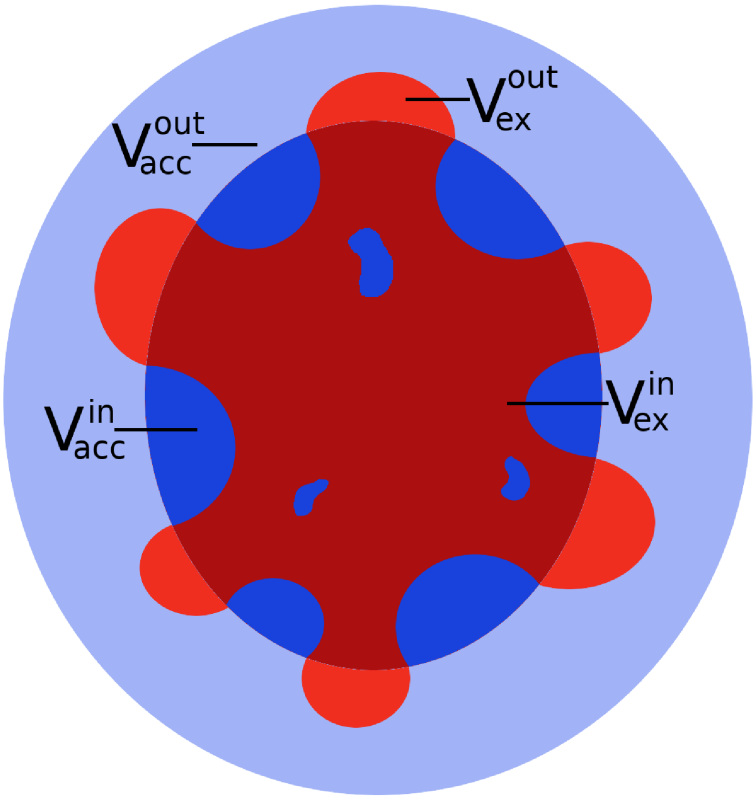

Figure 3.1: Schematic diagram showing the main parameters used in the model. The inner ellipsoid corresponds to the effective protein ellipsoid volume, $V_{ell,eff}$, while the outer ellipsoid corresponds to the volume of the ellipsoid defined by the boundary layer of the hydration shell, V_{sh}. The excluded protein volume, V_{ex}, is represented in red (dark red corresponds to V_{ex}^{in} while light red to V_{ex}^{out}); the solvent-accessible volume, V_{acc}, is represented in blue (dark blue corresponds to V_{acc}^{in} while light blue to V_{acc}^{out}).

Then, we make the following assumptions:

- (i) we assume that the protein excluded volume inside the effective protein

ellipsoid can be expressed as a fixed fraction of the protein ellipsoid volume and that this fraction is constant for the different proteins:

$$\chi = \frac{V_{ex}^{in}}{V_{ell,eff}} \tag{3.2.1}$$

- (ii) similarly, we assume the ratio between the excluded volume due to protein atoms outside the protein ellipsoid and the excluded volume due to protein atoms inside the protein ellipsoid to be a constant independent of the protein type:

$$\gamma = \frac{V_{ex}^{out}}{V_{ex}^{in}} \tag{3.2.2}$$

- (iii) we assume that the inner shell-density (i.e. ρ_{in}) is the same for each protein (this assumption will be confirmed by the MD simulation results, see Results section). For the outer shell-density (ρ_{out}) two cases are considered, one in which it is considered the same for each protein, the other in which it is affected by the shell size which depends, in turn, on the protein size. Hence, ρ_{out} is given either by a constant independent of the protein considered

$$\frac{\rho_{out}}{\rho_b} = r_0 \tag{3.2.3}$$

or by a simple function of the ratio $\alpha = V_{sh}/V_{ell,eff}$

$$\frac{\rho_{out}}{\rho_b} = \frac{a_0 + \alpha}{\alpha + a_1} \tag{3.2.4}$$

where V_{sh} is the volume of the ellipsoid defined by the boundary layer of the hydration shell (see the Methods section), ρ_b is the solvent bulk density, a_0 and a_1 are parameters identical for all the proteins considered and we used the fact that $\lim_{\alpha\to\infty} \rho_{out}/\rho_b = 1$.

From the above assumptions we may obtain the solvent-accessible volume inside the protein ellipsoid, V_{acc}^{in}, and the solvent-accessible volume outside the protein ellipsoid, V_{acc}^{out}, as follows

$$V_{acc}^{in} = V_{ell,eff} - V_{ex}^{in} \approx V_{ell,eff}(1 - \chi) \tag{3.2.5}$$

$$V_{acc}^{out} = V_{sh} - V_{ell,eff} - V_{ex}^{out} \approx V_{ell,eff}[\alpha - (1 + \gamma\chi)] \tag{3.2.6}$$

The total accessible volume within the solvation shell, V_{acc}, can be then expressed as follows

$$V_{acc} = V_{acc}^{in} + V_{acc}^{out} \approx V_{ell,eff}[\alpha - \chi(1 + \gamma)] \tag{3.2.7}$$

Similarly, the density within the solvation shell can be expressed as

$$\rho_{sh} = \frac{\rho_{in} V_{acc}^{in} + \rho_{out} V_{acc}^{out}}{V_{acc}} \approx \frac{1-\chi}{\alpha - \chi(1+\gamma)}\, \rho_{in} + \frac{\alpha - (1+\gamma\chi)}{\alpha - \chi(1+\gamma)}\, \rho_{out} \tag{3.2.8}$$

Finally, we consider the relative increment of ρ_{sh} with respect to the bulk water-density ρ_b, $\eta = \rho_{sh}/\rho_b - 1$

$$\eta(\alpha) = \frac{\rho_{sh}}{\rho_b} - 1 = \frac{\rho_{in}}{\rho_b}\, \frac{1-\chi}{\alpha - m} + \frac{\rho_{out}}{\rho_b}\, \frac{\alpha - m - (1-\chi)}{\alpha - m} - 1 \tag{3.2.9}$$

with $m = (1+\gamma)\chi$.

From Eqs 3.2.1 and 3.2.2 we can express the total excluded volume via the parameter m

$$V_{ex} = V_{ex}^{in} + V_{ex}^{out} \approx \chi V_{ell,eff} + \gamma\chi V_{ell,eff} = (1+\gamma)\chi V_{ell,eff} = m V_{ell,eff} \tag{3.2.10}$$

which corresponds to a linear relation between the excluded volume and the protein ellipsoid volume, hence providing m once fitting the data with a linear regression (see the Results section).

Similarly, by expressing the mean number of solvent molecules inside the effective protein ellipsoid n_{in} via the inner shell density

$$n_{in} = (1-\chi)\rho_{in} V_{ell,eff} \tag{3.2.11}$$

we can obtain $\beta = (1-\chi)\rho_{in}$ by a linear regression of the n_{in} versus $V_{ell,eff}$ plot as provided by MD simulation data (see the Results section).

In order to obtain a reliable estimate of ρ_{in} we compared for each protein the values of n_{in} provided by the protein-solvent simulation with that as obtained removing from a pure solvent simulation (at the same temperature and pressure) the excluded volume of the protein (see Methods section). Given that the accessible volume inside the effective protein ellipsoid is the same for the protein-solvent and pure solvent simulations, the ratio of the mean numbers of solvent molecules is identical to the ratio of the two inner densities, with the inner density of the pure solvent simulation (by definition) identical to the bulk one. It will be shown in the Results section that ρ_{in}/ρ_b over the protein sample used is approximately constant for all proteins, thus confirming that ρ_{in} can be considered invariant for all the proteins and providing its estimate by simply multiplying such ratio (i.e. the mean ratio over the protein sample) by the bulk density.

Therefore, once obtained m, ρ_{in} and β, we can estimate $1-\chi = \beta/\rho_{in}$ and hence

by using either Eq 3.2.3 or Eq 3.2.4 into Eq. 3.2.9 we have either

$$\eta(\alpha) = \frac{\rho_{in}}{\rho_b} \frac{1-\chi}{\alpha - m} + r_0 \frac{\alpha - m - (1-\chi)}{\alpha - m} - 1 \tag{3.2.12}$$

or

$$\eta(\alpha) = \frac{\rho_{in}}{\rho_b} \frac{1-\chi}{\alpha - m} + \frac{a_0 + \alpha}{\alpha + a_1} \frac{\alpha - m - (1-\chi)}{\alpha - m} - 1 \tag{3.2.13}$$

By using Eq. 3.2.12 (the lower-level model) or 3.2.13 (the higher-level model) to fit the values of η versus α as provided by the protein sample, either r_0 or a_1 and a_2 parameters can be evaluated to obtain a fully analytical expression of $\eta(\alpha)$ at the two levels of approximation.

We end this section with a comment on the applicability of our model to larger proteins, i.e. in the range of α values tending to 1. The model has a formal lower limit, α_{min}, because $V_{acc}^{out}/V_{ell,eff}$ needs to be ≥ 0, which implies, according to Eq. 3.2.6, that $V_{acc}^{out}/V_{ell,eff} = \alpha - m - (1-\chi) \geq 0$, and thus $\alpha \geq m + (1-\chi) = \alpha_{min}$. However, it is likely that in the proximity of α_{min} (i.e. close to a vanishing relative outer accessible volume) our model is not fully reliable (see Results section for an estimate of α_{min}).

3.3 Methods

3.3.1 MD simulations

Molecular dynamics simulation of the following fourteen proteins was performed: eight antifreeze proteins (AFPs), namely yeast-AFP (PDB ID: 2LQ0), fish type I AFPI (PDB ID:1WFA), fish type II AFPII (PDB ID:2AFP), fish type III AFPIII (PDB ID: 1HG7), insect *Tenebrio molitor* TmAFP (PDB ID:1EZG), insect *Choristoneura fumiferana* CfAFP (PDB ID: 1M8N), insect *Rhagium inquisitor* RiAFP (PDB ID: 4DT5), arctic yeast *Leucosporidium sp.* ice binding protein LeIBP (PDB ID: 3UYV), and ten non-AFPs, namely Trp-cage (PDB ID: 2J0F), Heliomicin (PDB ID: 1I2U), GB1 (PDB ID: 5JXV), BPTI (PDB ID: 5PTI), Ubiquitin (PDB ID:3M3J), Barnase (PDB ID: 2KF3), Lysozyme (PDB ID: 1LZT) and Myoglobin (PDB ID: 1UFP), HCAII (PDB ID: 3KS3), and COVID-19 main protease (PDB ID: 6LU7). The Gromacs 5.1.4 software package [87] in conjunction with the OPLS-AA force field [88] was used. Each structure was solvated in a periodic dodecahedral box large enough to ensure at least 1.3 nm distance between the protein surface and the box faces. The simulation box was filled with Simple Point Charge (SPC) water molecules [89] and (when needed) a proper number of ions to neutralize the system. The density of

the boxes containing the SPC-protein solutions was calibrated, as done in previous work [13,16,90,91], in order to obtain (within the NVT MD simulations) a pressure of ≈560 bar, corresponding to the pressure provided by an MD simulation of a pure SPC box at 300 K, with a density corresponding to the experimental liquid water density at the same temperature (≈33.3 molecules per nm^3). Each protein was simulated at room temperature (300 K) in the isothermal-isochoric (NVT) ensemble, using an integration step of 2 fs and keeping the temperature constant by the velocity-rescaling algorithm [76]. All bonds were constrained using the LINCS algorithm [79] and for short-range interactions a cut-off radius of 1.1 nm was employed. The particle mesh Ewald method [72] was used to compute long-range interactions with grid search and cut-off radii of 1.1 nm. After solute optimization and subsequent solvent relaxation, each protein was simulated for 100 ns.

In addition, a 10 ns-long NVT simulation of a pure SPC box at 300 K at the reference pressure of 560 bar is performed in order to compare the properties of the protein hydration shells with the properties of fictitious hydration shells filled with bulk water (see below in the Methods section).

Concerning the choice of the water model, we showed in a previous paper [16] for four of the proteins used here that the density results were not dependent on the water model chosen. To do this we performed MD simulations, and subsequent analyses, with both SPC [89] and TIP4P/2005 models [92] at two different temperatures, i.e. room temperature and a temperature just above the melting temperature for each of the two water models. The results with SPC were fully consistent with the ones obtained with the TIP4P/2005 water model.

3.3.2 Protein volume and hydration-shell density calculations

The protein excluded volume, i.e. the volume enclosed by the solvent-accessible surface, was calculated according to the method reported by Eisenhaber et al. [93], as implemented in Gromacs, using a probe radius of 0.14 nm. The protein mean excluded volume, V_{ex}, was obtained for each simulation by averaging over the instantaneous excluded volumes evaluated at each MD time frame.

In order to characterize the hydration-shell density of a protein we approximate the protein molecule as an ellipsoid defined, at each MD time frame, by the eigenvectors and eigenvalues of the 3 x 3 geometrical covariance matrix of the x, y, z atomic coordinates, as described in the recent papers [13, 16]. Briefly, for each time frame of a trajectory the three eigenvectors define the protein ellipsoid axes, with the corresponding lengths provided by the eigenvalues: considering a Gaussian atomic positional distribution along each eigenvector, a semi-axis $a_i = 2\sqrt{\lambda_i}$ was used, with i

= 1, 2, 3 and λ_i the eigenvalue of the i^{th} eigenvector. A set of ellipsoidal layers around the protein defined by the consecutive ellipsoids with semi-axes $a_i^{(n)} = a_i + n\delta$, with fixed increment $\delta = 0.03$ nm, was then considered. By calculating, and averaging, the instantaneous SPC density within each layer (disregarding the possible presence of protein atoms and/or counter-ions) we obtained the solvent-density profile around the protein, within layers of increasing distance from the protein ellipsoid surface (i.e. the layer-solvent-density profile). Given that such solvent density profile does not account for the effect of excluded volume of the non solvent atoms, the density values within layers including a significant number of protein atoms are lower than the actual density. Typically, single protein atoms can be still present in layers at 0.7-0.8 nm from the protein ellipsoid surface while for layers beyond 1 nm basically no protein atoms are detected. Therefore, the differences among the protein SPC density profiles within the first hydration layers (up to $\approx$ 0.4-0.5 nm) largely reflect differences in the spatial arrangement and compactness of protein atoms.

In order to remove the effect on the density profile of the spatial arrangement of the protein atoms that protrude from the protein ellipsoid surface and to compare the solvent density inside each hydration layer with the density of the bulk solvent, we use a previously employed strategy [16] to construct a fictitious hydration shell filled with bulk water. To this aim, the protein coordinates obtained from the protein + solvent MD simulations are inserted into a pure SPC box extracted from a MD simulation of water alone. The water molecules of the pure SPC box overlapping with the protein coordinates are removed, providing a protein + solvent box in which the hydration shell is filled with the bulk solvent. Water molecules of the pure SPC box are removed where the distance between any existing atom and any atom of the inserted protein is less than the sum of the van der Waals radii of both atoms. For each protein, this is done for 10000 different protein coordinates, each one inserted in a different bulk solvent box. The solvent density around the protein as a function of the distance from the protein ellipsoid surface is then calculated for these fictitious protein + solvent configurations and we call it $\rho_{b,fict}$. Finally, the ratio $\rho/\rho_{b,fict}$, i.e. the density variations with respect to the bulk density, is reported as a function of the distance from the protein ellipsoid surface (see Results section).

We define an effective protein-ellipsoid volume, $V_{ell,eff}$, to be used in studying the protein-size dependence of η. This is achieved by adding to each protein ellipsoid semi-axis the solvent radius (we used 0.14 nm as the solvent radius) plus the effective mean thickness due to the excluded volumes of the protein atoms over the ellipsoid surface (0.06 nm) which was obtained via a numerical fit to protein partial molecular volumes, as described in a previous work [13]. Briefly, the effective mean thickness

of the ellipsoid surface was obtained by tuning its value in order to best reproduce the experimental partial molecular volumes over a sample of 14 globular proteins. A more accurate method [94] that could be used to define an effective protein ellipsoid has been proposed based on the calculation of the neutron scattering profile from MD simulation which is used to back-calculate the gyration-tensor of the protein and its closest hydration layers. Nevertheless, such an approach is more time-consuming than the method used here and hence less effective for the analysis of a large number of proteins.

3.4 Results and Discussion

From the MD simulations of the eighteen studied proteins, using the procedure described in the Methods section, we calculated for each protein the layer-solvent-density profile as a function of the distance from the protein ellipsoid surface, see Fig. 3.2 , and the corresponding density variations with respect to the bulk density, $\rho/\rho_{b,fict}$, see Fig. 3.3. For each of the proteins under investigation the density of the solvent is higher than the bulk density in the proximity of the protein ellipsoid surface and decreases at increasing distances, approaching the bulk density (i.e. $\rho/\rho_{b,fict}$ tends to 1). Each layer-density profile approaches the plateau at around 1 nm from the protein ellipsoid surface, thus indicating that beyond such a distance the SPC molecules of the protein-solvent system behave equivalently to the reference pure SPC ones.

By using the mean hydration shell volume, V_{sh}, corresponding to the mean volume of the ellipsoid defined by the boundary layer of the hydration shell (as obtained by adding 1 nm, i.e. the distance from the protein ellipsoid surface at which the density approaches the plateau value of bulk density, to each semi-axis of the protein ellipsoid), the mean number of SPC molecules within such an ellipsoid, n_{sh}, and the mean protein excluded volume, V_{ex}, we can obtain the mean solvent density within the accessible volume of the protein hydration shell, ρ_{sh}, via:

$$\rho_{sh} \cong \frac{n_{sh}}{V_{sh} - V_{ex}} \tag{3.4.1}$$

We can compare ρ_{sh} to the solvent bulk-density, ρ_b, by calculating the relative increase $\eta = \rho_{sh}/\rho_b - 1$. As is shown in Tab. 3.1 the simulations of the studied proteins provided, by means of eq. 3.4.1, a hydration shell mean density in the range of 4-14% higher than the bulk one. This range of η is in agreement with experimental data [18] and previous calculations [8, 10, 13, 16].

By analyzing the data reported in Tab. 3.1, in which also the mean effective-protein ellipsoid, $V_{ell,eff}$, the mean ellipsoidal hydration-shell volume, V_{sh}, are re-

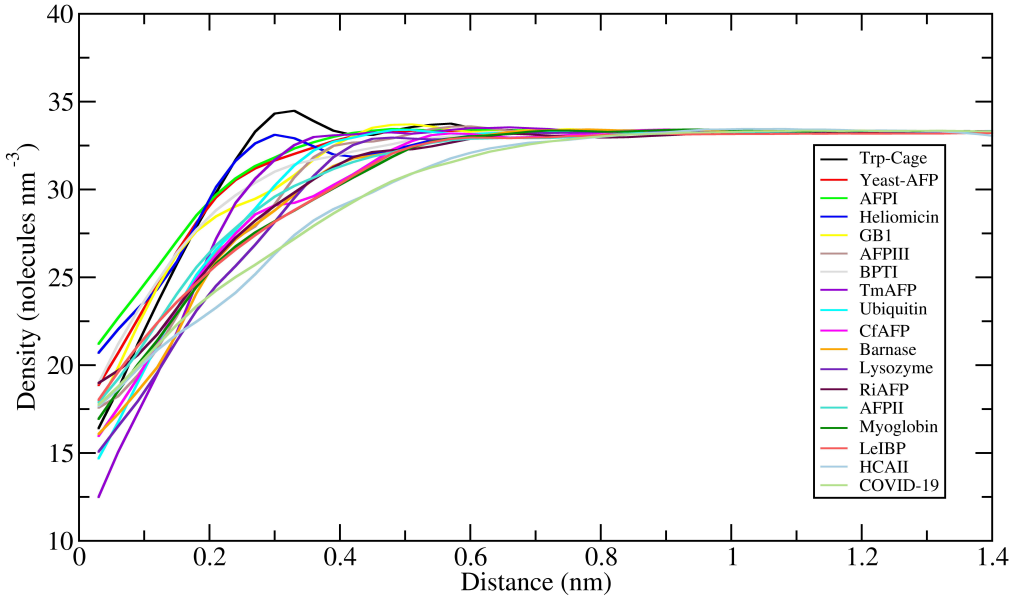

Figure 3.2: Layer density profile of the hydration SPC molecules as a function of the distance from the protein ellipsoid surface calculated for the eighteen proteins: Trp-cage (black), Yeast-AFP (red), AFPI (green), Heliomicin (blue), GB1 (yellow), AFPIII (brown), BPTI (grey), TmAFP (violet), Ubiquitin (cyan), CfAFP (magenta), Barnase (orange), Lysozyme (indigo), RiAFP (maroon), AFPII (turquoise), Myoglobin (dark green), LEIBP (light red), HCAII (light blue), and COVID-19 (light green).

ported, one can notice a positive correlation between η and the size of the protein, as provided by the analysis of the correlation between η and the protein volumes. This behavior is opposite to what was observed for hydrophobic solutes in the same length-scale range [86], for which a decrease in hydration density was observed as a function of the protein size.

On the contrary, no clear dependence of η was found on the protein eccentricity, as provided by the analysis of the correlation between η and the eccentricity of the ellipsoids approximating the protein molecules (see Fig. 3.6A), and on the hydrophobicity of the protein surface, as provided by the analysis of the correlation between η and the hydrophobic fraction at the protein surface (see Fig. 3.6B). In the Supporting Information, Fig. 3.4 and Fig. 3.5, a global view of the chemical character of the surfaces of all proteins is given in terms of hydrophilic/hydrophobic atoms.

In order to interpret the dependence of the hydration shell-density on the protein size through the model proposed in the Theory section, we firstly need to obtain $m, \rho_{in}/\rho_b$ and $1-\chi$, the latter evaluated as $1-\chi = \beta/\rho_{in}$.

To get m we plotted V_{ex} as a function of the effective protein-ellipsoid volume, $V_{ell,eff}$, for all the studied proteins, see Fig. 3.7A, and from a linear fit of the data we

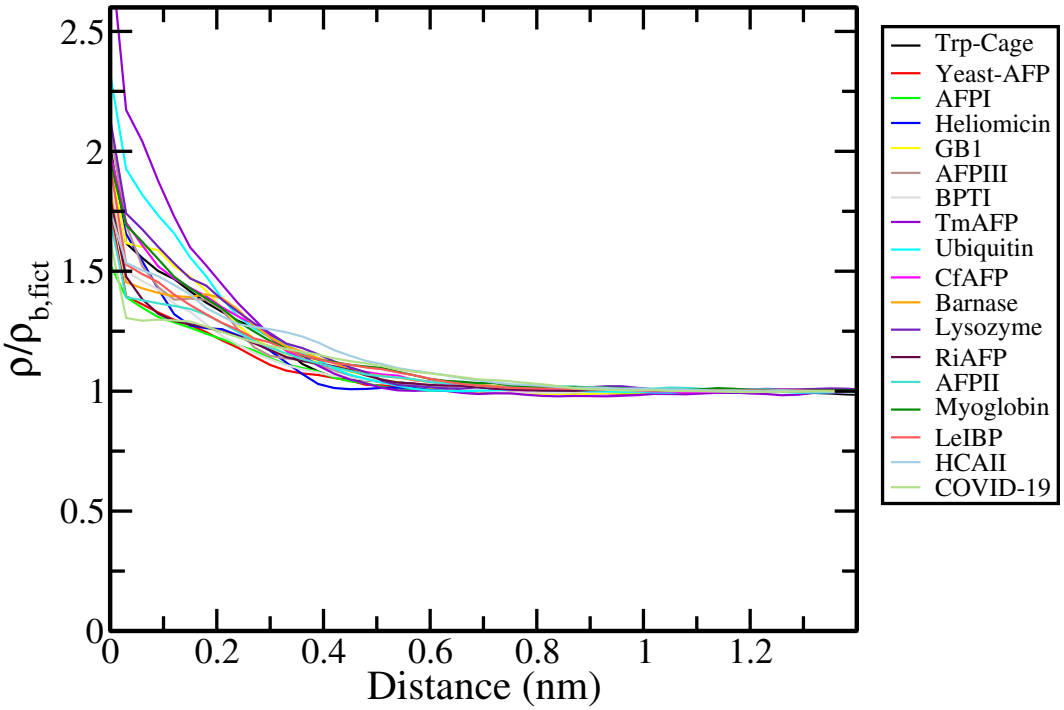

Figure 3.3: Density variations with respect to the bulk density, $\rho/\rho_{b,fict}$, as a function of the distance from the protein ellipsoid surface for the eighteen proteins: Trp-cage (black), Yeast-AFP (red), AFPI (green), Heliomicin (blue), GB1 (yellow), AFPIII (brown), BPTI (grey), TmAFP (violet), Ubiquitin (cyan), CfAFP (magenta), Barnase (orange), Lysozyme (indigo), RiAFP (maroon), AFPII (turquoise), Myoglobin (dark green), LEIBP (light red), HCAII (light blue), and COVID-19 (light green).

obtain m ($m = 0.95$). From the data and the quality of the fit it is evident that the linear relation we assumed between V_{ex} and $V_{ell,eff}$ is a good approximation ensuring that one of our basic assumptions is reasonable.

To get ρ_{in}/ρ_b we used the procedure described in the Methods section, i.e. ρ_{in}/ρ_b is calculated from the ratio of the mean number of solvent molecules inside the protein effective ellipsoid and the corresponding number calculated on the fictitious protein + solvent configurations. The ρ_{in}/ρ_b values, calculated over the protein sample and reported in Fig. 3.8A as a function of $V_{sh}/V_{ell,eff}$, show a rather flat distribution with a mean value of 1.56, thus confirming that ρ_{in} is constant for all the proteins as was assumed in the model. A similar analysis of ρ_{out}/ρ_b, evaluated as the ratio of the mean number of solvent molecules outside the protein effective ellipsoid, but inside the hydration shell, and the corresponding number calculated on the fictitious protein + solvent configurations was performed (see Fig. 3.8B). The comparison between ρ_{in}/ρ_b and ρ_{out}/ρ_b shows that almost all of the hydration shell density increase is concentrated inside the protein effective ellipsoid, being the former $\approx$1.56 and the latter in the range of 1.00 and 1.08. Moreover, ρ_{out}/ρ_b shows a slight dependence on the protein size.

Table 3.1: Residue number, effective protein-ellipsoid volume, $V_{ell,eff}$, ratio between the volume of the shell, V_{sh}, and the effective ellipsoid volume, $V_{ell,eff}$, and relative density increment, η, for the eighteen studied proteins. Volumes are given in nm^3

Protein	Residue N.	$V_{ell,eff}$	$\frac{V_{sh}}{V_{ell,eff}}$	η
Trp-cage	20	4.46	5.95	0.050
Yeast-AFP	25	6.87	4.97	0.052
AFPI	37	8.06	4.86	0.046
Heliomicin	44	9.47	4.38	0.065
GB1	56	11.81	4.00	0.079
AFPIII	66	12.27	3.91	0.071
BPTI	56	12.41	3.94	0.068
TmAFP	82	12.46	4.00	0.083
Ubiquitin	75	14.40	3.70	0.090
CfAFP	121	19.65	3.37	0.088
Barnase	108	20.55	3.30	0.094
Lysozyme	129	22.64	3.19	0.099
RiAFP	143	23.26	3.26	0.089
AFPII	129	25.70	3.06	0.106
Myoglobin	154	28.63	2.99	0.111
LeIBP	241	38.73	2.73	0.123
HCAII	261	46.61	2.56	0.133
COVID-19	306	54.93	2.56	0.125

.

Note: The error on $V_{ell,eff}$ and V_{sh} is ≈0.3%, the one on η is ≈1%. Errors on the different quantities were evaluated through the standard error of their mean calculated over 3 (independent) subtrajectories.

Similarly as for m, by expressing the mean number of solvent molecules inside the effective protein ellipsoid n_{in} via the inner shell density $n_{in} = (1-\chi)\rho_{in}V_{ell,eff}$, we can obtain $\beta = (1-\chi)\rho_{in}$ by a linear regression of the n_{in} versus $V_{ell,eff}$ plot (see Fig. 3.7B) as provided by shell-density calculations (we get $\beta = 11.61$). From β and ρ_{in}, we can evaluate $1-\chi = \beta/\rho_{in} = 0.22$.

Once obtained m, ρ_{in}/ρ_b and $(1-\chi)$ we can use eqs. 3.2.12 and 3.2.13 to fit the values of η as a function of $\alpha = V_{sh}/V_{ell,eff}$ calculated for the different proteins (see Fig. 3.9). From the quality of the fits it emerges that both levels of approximation of the model employed are reasonably accurate, with the higher-level model performing slightly better consistently with the slight size-dependence observed from the explicit calculation of ρ_{out}/ρ_b (see Fig. 3.8). A further validation of our model comes from the fact that if the parameters a_0 and a_1, as obtained from the fitting of η with eq. 3.2.13, are used into the analytical function for ρ_{out}/ρ_b ($\rho_{out}/\rho_b = (a_0+x)/(x+a_1)$) a

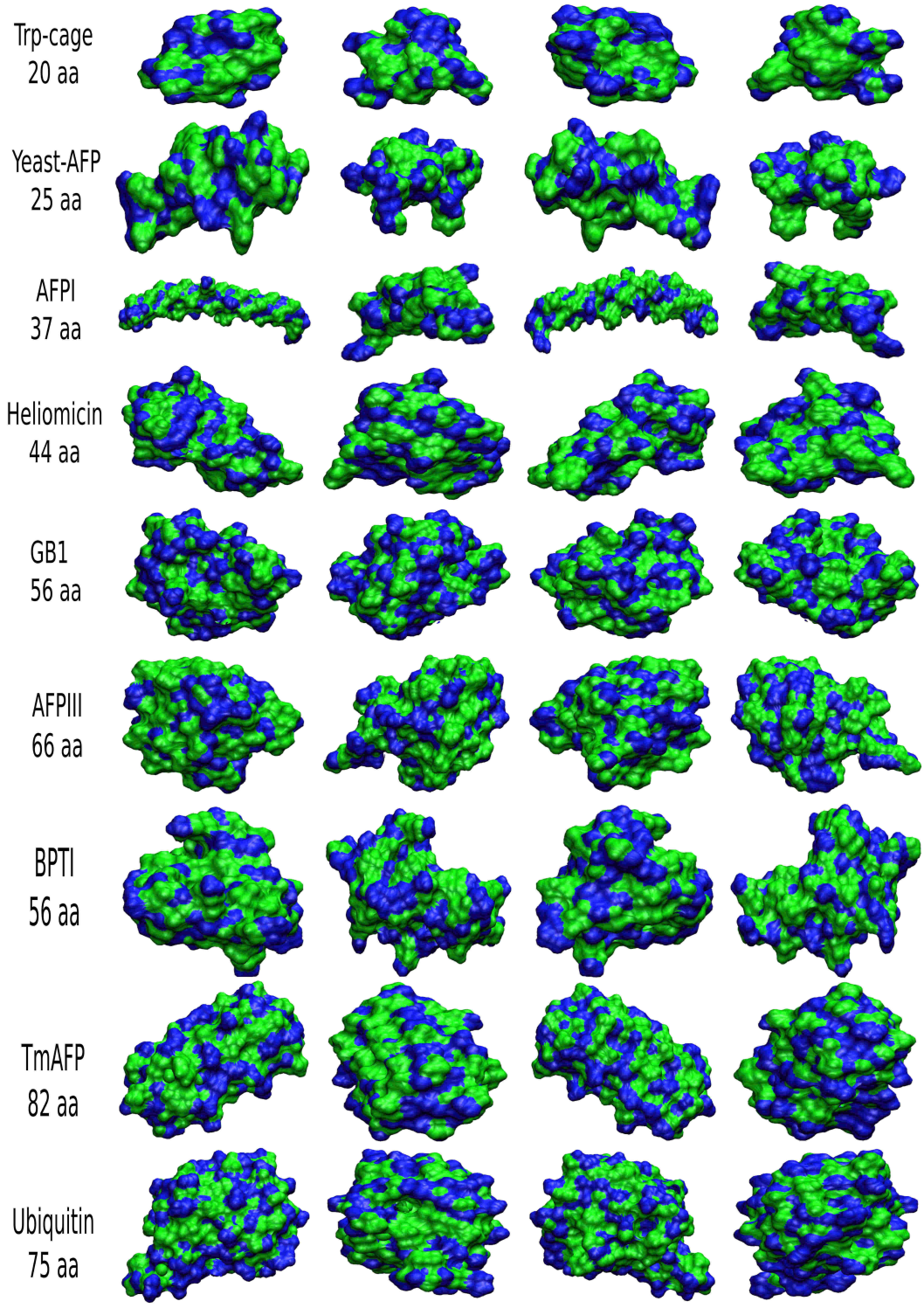

Figure 3.4: Hydrophobic (green) and hydrophilic (blue) atoms at the protein surface.

good reproduction of the ρ_{out}/ρ_b values provided by the MD simulations is obtained (see Fig. 3.8B).

The relevant parameters (i.e. $1-\chi$, ρ_{in}/ρ_b and ρ_{out}/ρ_b which are independent of the level of approximation and are summarized in the caption to Fig. 3.9) can be utilized to understand how the solvent molecules are partitioned inside the hydration

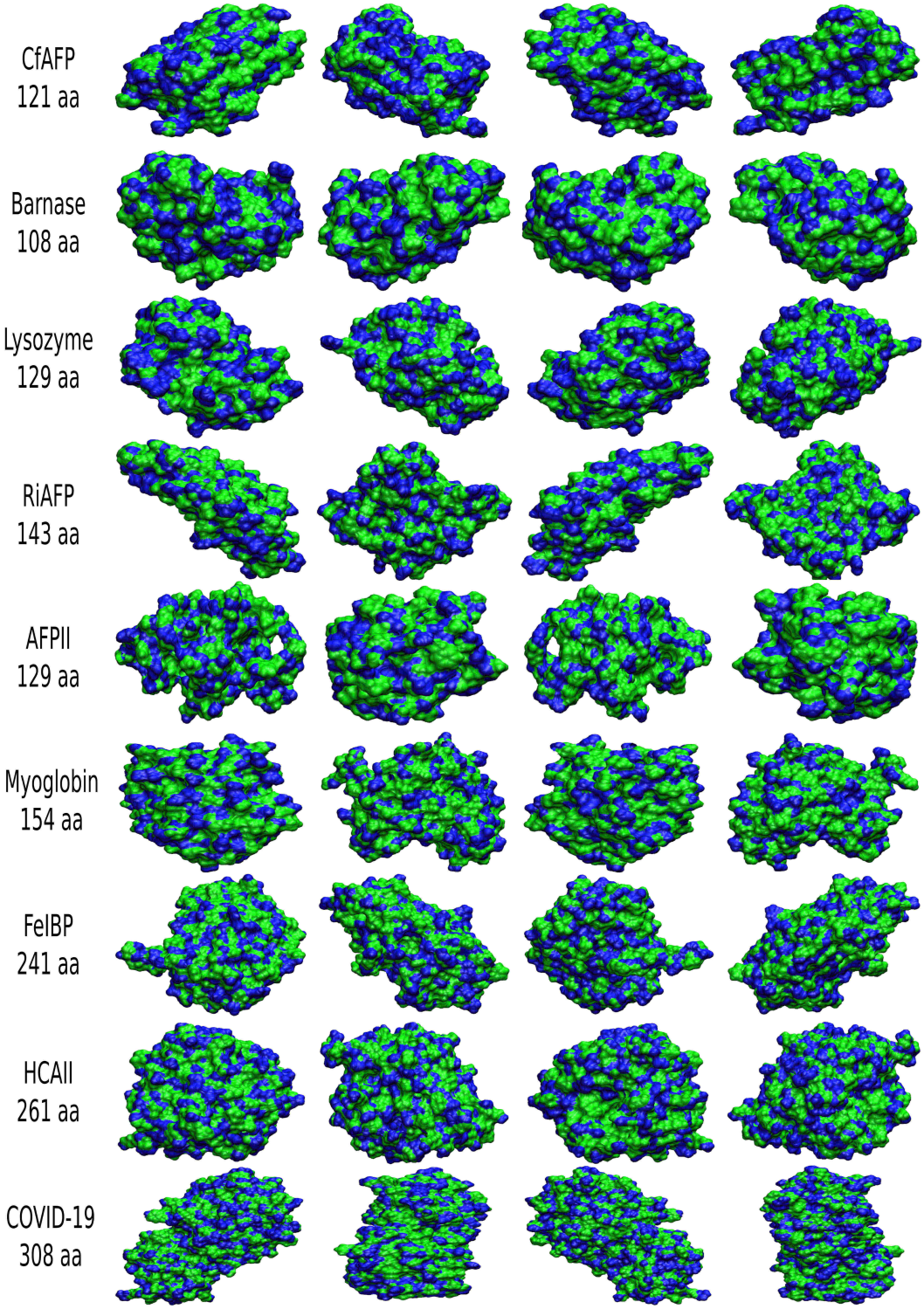

Figure 3.5: Hydrophobic (green) and hydrophilic (blue) atoms at the protein surface.

shell possibly explaining the variation of the solvent-shell density according to the protein size. According to our model ≈22% of the effective protein-ellipsoid volume is accessible to solvent (i.e. $1-\chi = 0.22$) and basically all the solvent density increase is concentrated in this region. In fact, the solvent density in the inner accessible volume is approximately 56% higher than bulk (i.e. $\rho_{in}/\rho_b \approx 1.56$), while it is basically equal

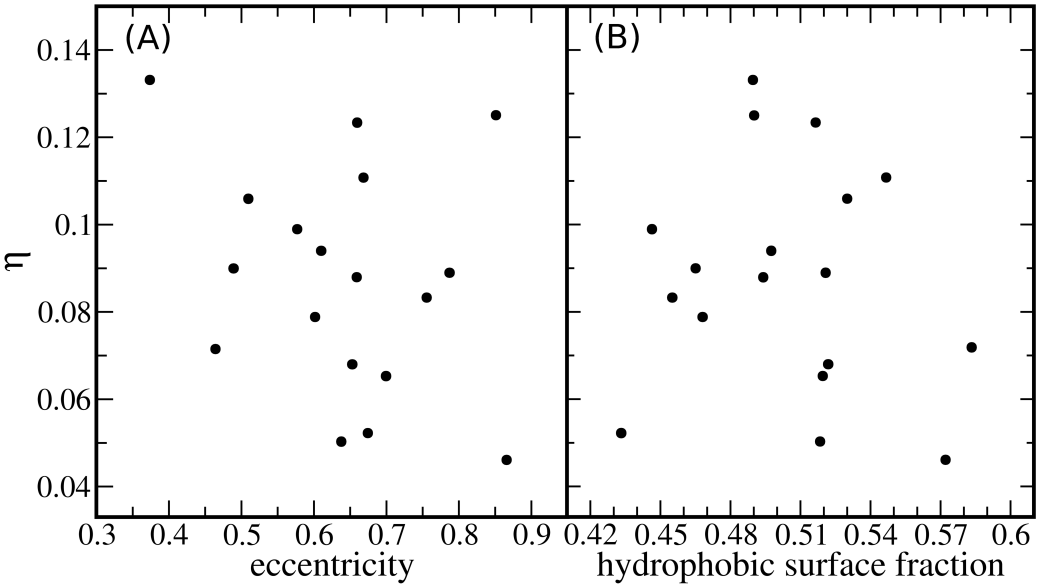

Figure 3.6: η as a function of the ellipsoid eccentricity, defined as $\epsilon^2 = 1 - a^2/c^2$ with a and c the smallest and largest ellipsoid axes, respectively (A), and of the fraction of hydrophobic atoms at the protein surface (B).

to the bulk density in the outer region (i.e. $\rho_{out}/\rho_b = 1.04$ in the case of the lower level of approximation). Thus, almost all of the solvent density increase is confined within the protein and within grooves and pockets at the protein surface, as shown in Fig. 3.10 in which a representative configuration showing the water molecules within the protein ellipsoid is reported.

The protein-independent density increment of ≈56% found here is in line with previous findings in which the hydration density increase was calculated by means of MD simulation at a distance of 0.2 nm from the protein surface for lysozyme [8] (≈50% increase was found) and from the analysis of the crystal waters in contact with the protein surface for 22 crystal structures [95] (an average ≈20% increase was found).

Our protein data-set contains proteins with a specific function, namely antifreeze proteins. AFPs are able to inhibit ice growth by binding to ice nuclei and, although their ice-binding mechanism is still unclear, yet the hydration layer is thought to play a fundamental role. We have previously shown that the hydration density of two of these AFPs, CfAFP and AFPIII, are very similar to non-AFPs [16]. Here, we extend the study to a total of eight AFPs spanning a rather large length-scale range, and compare the hydration density with non-AFPs. As shown in Fig. 3.9, in which the AFPs proteins are highlighted in red, the behavior of the AFPs does not differ from that of non-AFPs. Indeed, if we separate the data into two sets, one comprising non-AFPs and the other AFPs proteins, and fit our model to the two sets independently we

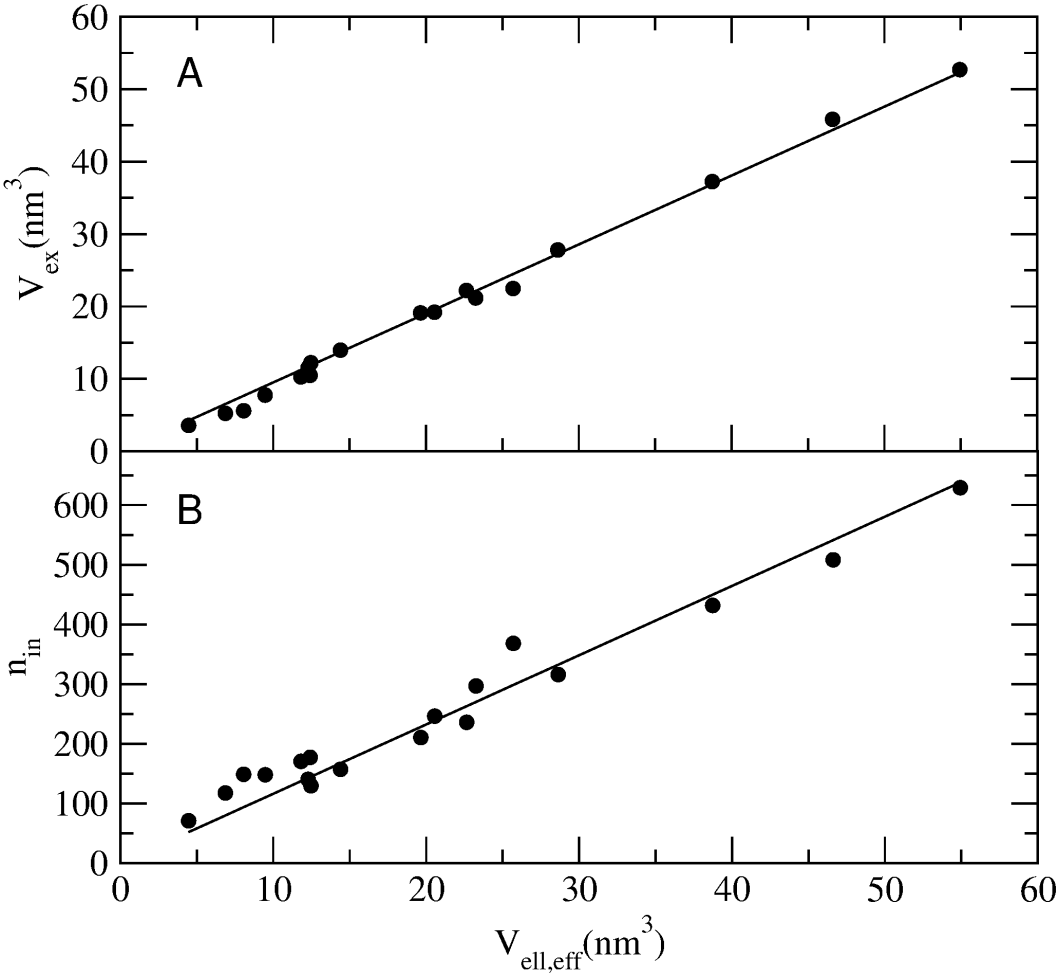

Figure 3.7: A) V_{ex} as a function of the effective protein-ellipsoid volume, $V_{ell,eff}$, for all the studied proteins. From the linear fit $V_{ex} = mV_{ell,eff}$ we obtain $m = 0.95$. B) Mean number of solvent molecules inside the effective protein ellipsoid, n_{in}, as a function of the effective protein-ellipsoid volume, $V_{ell,eff}$, for all the studied proteins. From the linear fit $n_{in} = \beta V_{ell,eff}$ we obtain $\beta = 11.61$. The correlation coefficient is higher than 0.99 for both regressions.

basically obtain the same results. Thus, our results do not show any peculiar behavior of AFPs in terms of the protein-size dependence of the hydration-shell density with respect to non-AFPs.

As explained in the theory section, our model has a formal lower limit $\alpha_{min} = m + (1 - \chi) = 1.17$. However, it is likely that in the proximity of α_{min} (i.e. close to a null relative outer accessible volume) the model is not fully reliable. We might consider a value of α around 1.5 as a lower limit, which corresponds to a maximum number of residues of approximately 500 - 600, hence limiting the applicability of the model to small-to-medium-sized globular proteins.

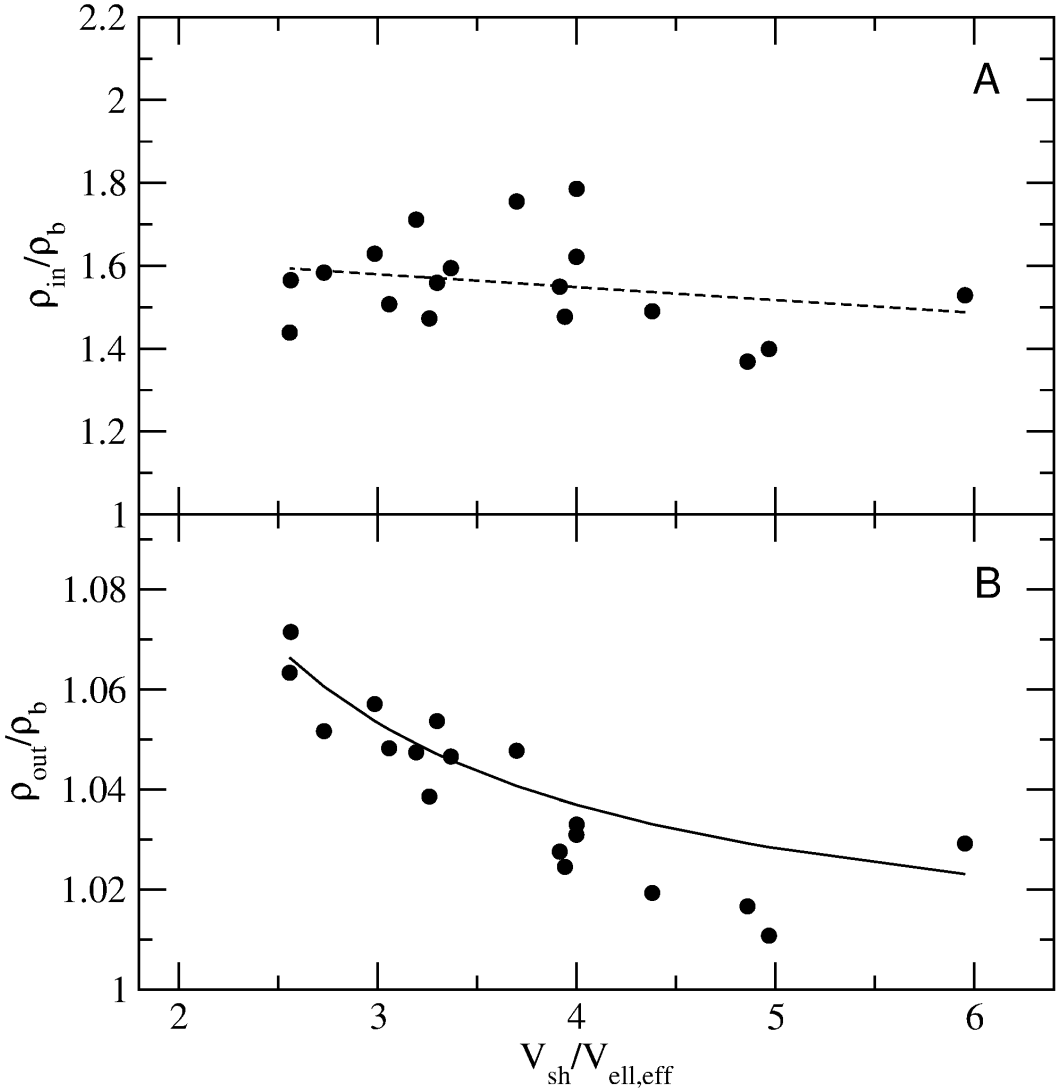

Figure 3.8: (A) ρ_{in}/ρ_b as a function of $\alpha = V_{sh}/V_{ell,eff}$ for all the studied proteins. A linear regression of the ρ_{in}/ρ_b data yields a slope coefficient of -0.031 which is within one standard error of the coefficient (0.032), showing that ρ_{in}/ρ_b can be considered basically constant over the protein sample. (B) ρ_{out}/ρ_b as a function of $\alpha = V_{sh}/V_{ell,eff}$ for all the studied proteins. The black solid line is the analytical function $\rho_{out}/\rho_b = (a_0+x)/(x+a_1)$ with $a_0 = -0.63$ and $a_1 = -0.75$ as obtained from the fit of η with eq. 3.2.12 (see Results section). The analytical expression provides a good reproduction of the ρ_{out}/ρ_b values obtained from the MD simulations.

3.5 Conclusions

By means of molecular dynamics simulation we analyzed the protein hydration-shell density of eighteen proteins, which include eight AFPs, and found an increase of water density with respect to the bulk-density for all the studied proteins. While we did not find any correlation between the magnitude of this increase and the protein eccentricity and protein-surface mean hydophobicity, we found a clear dependence

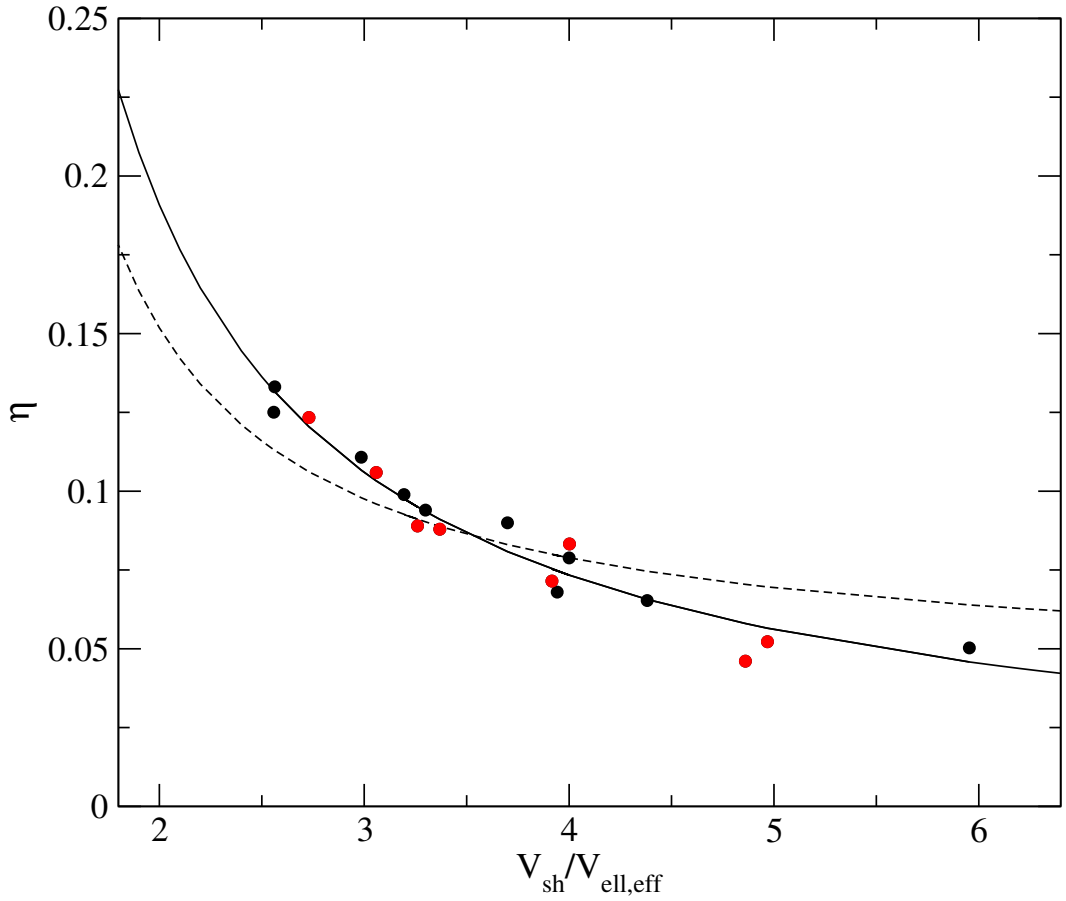

Figure 3.9: (A) η as a function of $\alpha = V_{sh}/V_{ell,eff}$ calculated for the different proteins. The data are fitted with eq. 3.2.12 (dashed line) providing $r_0 = \rho_{out}/\rho_b = 1.04$ and eq. 3.2.12 (full line) providing $a_0 = -0.63$ and $a_1 = -0.75$; the correlation coefficient is 0.97 for both models. The eight AFPs are highlighted in red.

on the protein length-scale. In particular, the increment in hydration-shell density decreases for decreasing protein size (from $\approx$ 14% to $\approx$ 4%) and, within the simple model developed in the present work, this decrease is caused by the protein size only. In fact, the hydration shell-density increase is found to be confined within, or in pockets at the surface of the protein (i.e. inside the effective solvent-accessible volume of the ellipsoid approximating the protein molecule) and to be the same for all proteins. Therefore, since the relative size of this inner solvent-accessible volume, with respect to the total accessible volume of the hydration shell, is lower for smaller proteins, the decrease in water-density is explained by the difference in protein size.

The protein-independent increment of the density of the water molecules confined within the protein effective ellipsoid of $\approx$56% obtained from our model is consistent, yet higher, with previous explicit calculations showing that the density of water molecules in close proximity to the protein surface, and in particular in concave regions, is 20%-50% higher than bulk.

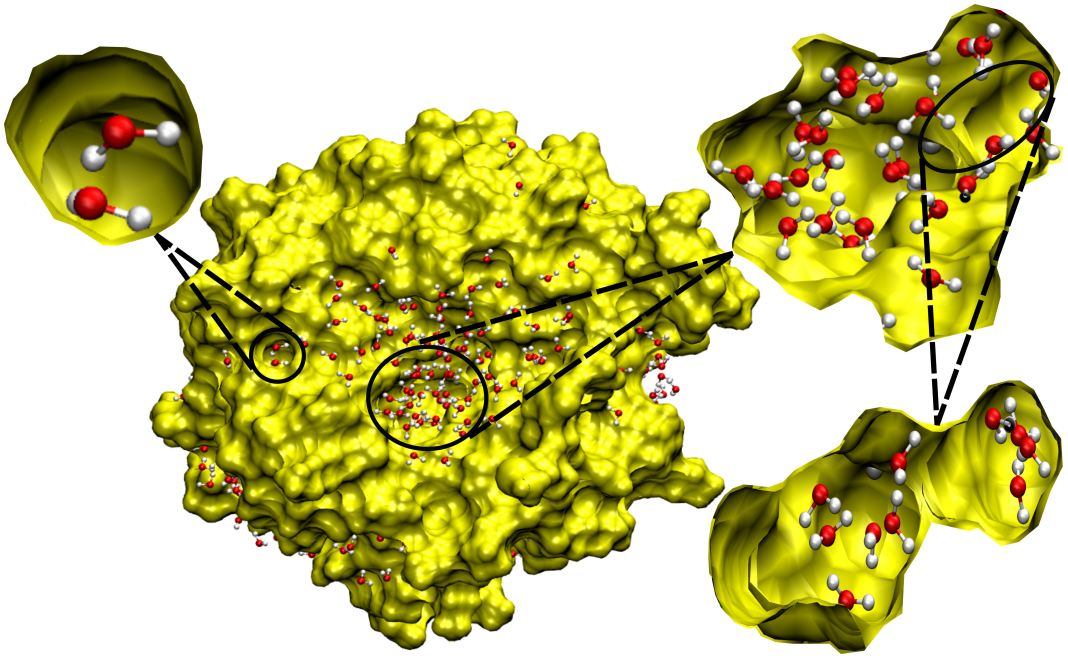

Figure 3.10: Representative configuration of HCAII showing the water molecules within the protein ellipsoid in ball and stick (oxygen and hydrogen atoms are reported in red and white, respectively). Most of the water molecules are localized within grooves and pockets. The behaviour is qualitatively similar for all proteins.

Finally, we show that the antifreeze proteins studied here, which cover a broad range of sizes and types (hyperactive from insects, CfAFP, TmAFP and RiAFP, moderately active from fish, AFPI, AFPII and AFPIII, and moderately active from yeast, yeast-AFP and LeIBP), do not behave differently from non-antifreeze proteins in terms of the size-dependence of their hydration density, in line with our previous work in which we showed that the shell thickness and density of two AFPs did not feature any relevant difference with respect to two non-AFPs [16].

Chapter 4

Hydration Shell of Antifreeze Proteins: Unveiling the Role of Non-Ice-Binding Surfaces

4.1 Introduction

Antifreeze proteins (AFPs) are a vast group of proteins that are able to suppress freezing and are exploited by several organisms to live at subzero temperatures. Despite a remarkable variability in their primary, secondary, and tertiary structures, AFPs share a common mechanism of action: they bind to small ice nuclei, preventing them from growing at the adsorbed positions [48]. AFPs are indeed able to irreversibly bind to the ice surface, exploiting a number of hydrogen bonds and hydrophobic interactions between specific ice planes and a complementary protein surface, the ice-binding site (IBS), a usually flat surface specifically involved in ice-binding.

By binding the surface of a growing ice nucleus, AFPs cause the microcurvature of the ice surface, as water molecules are forced to join the ice surface only between the adsorbed AFPs. As the ice-crystal growth is thermodynamically less favorable on a curved ice surface than a flat one, the ice growth stops until the temperature drops. This results in a local difference between the freezing and the melting temperature, known as thermal hysteresis (TH) [96]. Moderately active AFPs display a TH activity below $\approx 1°$, while the so-called hyperactive AFPs, which are mainly found in insects, can reach TH activities of $5°$ and higher [48, 97].

Although there is a general consensus on the mechanism of action of AFPs, many questions remain open. The nature of the interactions that determine the binding to the ice surface has been widely debated: a spatial match between the polar residues of the IBS and the oxygen–oxygen distance in the ice lattice has been repeatedly noted [37, 98–101], pointing to a critical role of hydrogen bonds in the protein–ice binding. However, IBSs also typically show a relevant hydrophobic content, suggesting that hydrophobic groups might also be involved in ice binding [57, 102–107], with the idea that the release of constrained waters from hydrophobic groups on the IBS

might provide a thermodynamic driving force for ice binding [107]. Another key question is how AFPs are able to recognize ice prior to binding in a vast excess of water. On the basis of molecular dynamics (MD) simulations, it has been proposed that hydration water at the IBS could form an ice-like layer already in solution, facilitating ice recognition [11, 108–114]. Terahertz spectroscopic studies noted the presence of an extended dynamic hydration shell around AFPs and proposed that such a hydration layer plays an important role in the molecular mechanism of the antifreeze activity [32, 33]. Differences between the dynamics of solvation water of IBSs and of the non-ice-binding parts of AFPs have been highlighted [12]. However, other MD simulations [115, 116], and an NMR study [117] found no indication of ice-like structure in liquid water around AFPs. The lack of a clear correlation between the properties of AFP hydration waters and their antifreeze activity has also been highlighted [118]. More in general, the importance of investigating AFP hydration shells to better understand their antifreeze function is increasingly recognized.

In a recent work [119], with the combined use of FTIR spectroscopy in the water O–H stretching region and a postprocessing technique to separate the individual spectral components in vibrational spectroscopy of mixtures (self-modeling curve resolution), the IR spectrum of the solvent perturbed by the solute (i.e., the spectrum of the hydration shell) has been isolated for a type III AFP (ZaAFP). From the analysis of the solute-correlated O–H stretching signal, an estimate of the thickness of the hydration shell and of the number of water molecules inside it was obtained. The resulting 0.95 nm thick hydration shell has been claimed to be very large and to reflect a strong interaction between the AFP and the solvent, possibly correlated with the antifreeze activity. However, the estimated thickness of the hydration shell seems to be in line with the one estimated for non-ice-interacting proteins, that is, ≈ 1 nm [12, 13, 120, 121]. Long-ranged dynamical solvation shells in AFPs have been in fact suggested to reach up to 2 nm from the surface of the solute on the basis of terahertz studies [32, 33]. These last findings are, again, in contrast with computational studies that found a hydration shell thickness of ≈ 1 nm also in a hyperactive AFP (CfAFP) [12].

In this context, we investigate, here, the shell thickness of these two small AFPs, ZaAFP and CfAFP, and compare it to the hydration shell thickness of two nonice-binding proteins: ubiquitin (Ubi) and barnase (Bar). To this aim, we performed MD simulations both at room temperature and at a temperature a few degrees below the melting temperature for the computational water model used (TIP4P/2005 [92]), and we apply a recently developed analysis to investigate hydration shell volumes [13]. Such a methodology also allows obtaining the hydration shell water density as a

function of the distance from the protein surface. The TIP4P/2005 model has been previously used in MD simulations involving AFPs [116, 122]. However, to assess that our results do not depend on the choice of the water model, we also perform MD simulations and subsequent analyses with the simple point charge water model, SPC [89], both at room temperature and at a temperature just above the melting temperature for SPC, Most of the computational works focusing on the hydration shell properties of AFPs investigated hydration waters in terms of structural parameters [11,12,111,116], with the aim of assessing the possible ice-like structure of liquid water at the IBS. Here, we face the same issue from a different point of view, comparing the density of AFPs and non-ice-binding protein hydration shells to the density of bulk water.

4.2 Methods

4.2.1 MD Simulations

MD simulations of CfAFP (PDB ID: 1M8N), ZaAFP (PDB ID: 1HG7), Ubi (PDB ID: 3M3J), and Bar (PDB ID: 2KF3) are performed with the Gromacs software package [87] in conjunction with the AMBER03 force field [123]. Each structure is solvated in a periodic dodecahedral box large enough to ensure at least 1.3 nm distance between the protein surface and the box faces. The simulation box is filled with TIP4P/2005 water molecules [92] and (when needed) a proper number of ions to neutralize the system. The density of the boxes containing the TIP4P/2005-protein solutions is calibrated in order to obtain (within the NVT MD simulations) a pressure of $\approx$ 390 bar, corresponding to the pressure provided by an MD simulation of a pure TIP4P/2005 box at 300 K, with a density corresponding to the experimental liquid water density at the same temperature ($\approx$33.3 molecules per nm^3) [13]. Each protein is simulated at room temperature (300 K) and at low temperature (245 K) that is, 4-7° below the ice-melting temperature for the TIP4P/2005 model (249–252 K) [124]. Simulations at room and low temperature are also performed with a different water model, that is, the SPC water model [89] in conjunction with the OPLS-AA/L force field [88]. With such a water model, the density of the boxes containing the SPC-protein solutions is calibrated in order to obtain (within the NVT MD simulations) a pressure of $\approx$560 bar, corresponding to the pressure provided by an MD simulation of a pure SPC box at 300 K, with a density corresponding to the experimental liquid water density at the same temperature ($\approx$33.3 molecules per nm^3). With SPC, each protein is simulated at room temperature (300 K) and at low temperature (200 K)

that is, 10° above the ice-melting temperature for the SPC model (190 K) [125]. All the systems are simulated in the isothermal–isochoric (NVT) ensemble, using an integration step of 2 fs and keeping the temperature constant by the isokinetic temperature coupling [126]. All bonds are constrained using the LINCS algorithm [79] and for short range interactions, a cut-off radius of 1.1 nm was employed. The particle mesh Ewald method [72] is used to compute long-range interactions with grid search and cut-off radii of 1.1 nm.

After solute optimization and subsequent solvent relaxation, each protein is simulated for 100 ns at both temperatures. In addition, 100 ns-long NVT simulations of a pure TIP4P/2005 box at 300 K and at 245 K and of a pure SPC box at 300 and 200 K at the reference pressure of 390 and 560 bar, respectively, are performed in order to compare the properties of the protein hydration shells with the properties of fictitious hydration shells filled with bulk water (see the Results and Discussion section). The 245 K MD simulation of pure TIP4P/2005 and the 200 K MD simulation of pure SPC are also used to evaluate the bulk solvent density at low temperature and at a pressure of 390 and 560 bar, respectively, and provide a water density of 33.1 and 34.3 molecules per nm^3, respectively.

4.2.2 Hydration Shell Density Calculations

In order to characterize the solvent density around the protein, we use the approximation of treating the protein molecule as an ellipsoid defined, at each MD time frame, by the eigenvectors and eigenvalues of the 3 x 3 geometrical covariance matrix of the x, y, z atomic coordinates [13]. We then consider a set of ellipsoidal layers around the protein defined by the consecutive ellipsoids with semi-axes $a_i^n = a_i + n\delta$ with fixed increment $\delta = 0.03$ nm. By calculating at each MD frame the instantaneous TIP4P/2005 density within each layer (disregarding the possible presence of protein atoms and/or counter-ions) and averaging over the MD trajectory, we obtain the solvent density profile around the protein within layers of increasing distance from the protein ellipsoid surface (i.e., the layer solvent density profile, see Figure 4.6). The distance from the protein surface at which the layer solvent density reaches the bulk plateau value (≈33.3 molecules per nm^3 at 300 K and ≈33.1 molecules per nm^3 at 245 K, see MD Simulations subsection) defines the hydration shell boundary layer, that is, the thickness of the protein hydration shell and corresponding shell volume V_{shell} (more detail on the definition of the hydration shell thickness is given in the Supporting Information). The mean value of the number of TIP4P/2005 molecules within each layer, that is, as a function of the distance from the protein ellipsoid surface, is also calculated along the MD simulations. In addition, the protein-excluded volume

V_{ex} , that is, the volume enclosed by the solvent-accessible surface, can be obtained by Gromacs using a probe radius of 0.14 nm, according to the method reported in Eisenhaber et al. [93]. By calculating V_{ex} at each MD time frame of the simulations, we obtain the protein mean volume and the corresponding thermal distribution. From V_{shell}, V_{ex} and the number of TIP4P/2005 molecules inside the hydration shell, N_{shell}, the mean solvent density within the accessible volume of the protein hydration shell can be obtained via

$$\left\langle \frac{N_{shell}}{V_{shell} - V_{ex}} \right\rangle \approx \frac{\langle N_{shell} \rangle}{\langle V_{shell} \rangle - \langle V_{ex} \rangle} = \rho_{shell} \tag{4.2.1}$$

The mean solvent density inside the hydration shell can be compared to the bulk solvent density ρ_{bulk}, providing the relative density increment η inside the hydration shell with respect to the bulk density

$$\eta = \frac{\rho_{shell} - \rho_{bulk}}{\rho_{bulk}} \tag{4.2.2}$$

4.2.3 Definition of Binding Surfaces

For each protein, two surfaces have been defined for the calculation of the hydration surface density ρ_{surf}. For the two AFPs, one of the two surfaces corresponds to the ice binding site (IBS). The IBSs of CfAFP and ZaAFP have been defined according to the literature [127, 128]. The remaining surfaces have been selected as flat as possible and with a solvent accessible surface area (SASA) similar to IBS one. The residues included in each of the eight surfaces are listed below. In Figure 4.1 a representative sketch of the eight surfaces is reported.
CfAFP IBS: residues V5 T7 T21 T23 I37 T39 V52 T54 T67 T69 T82 T84 T99 T101
CfAFP NIBS: residues Q10 T12 T26 D28 Q42 S44 N57 D59 Q72 N74 S87
ZaAFP IBS: residues Q9 L10 P12 I13 N14 T15 A16 T18 V20 M21 Q44
ZaAFP NIBS: residues N1 Q2 V27 T28 V30 G31 I32 P33 E35 D36 R39
Bar Surface 1: residues T24 K25 S26 E27 Q29 A30 G32 W33 V34 A35 S36
Bar Surface 2: residues P62 G63 L64 S65 G66 R67 D91 W92 L93
Ubi Surface 1: residues Q31 D32 L33 E34 G35 I36 P37 P38 D39 Q40
Ubi Surface 2: residues M1 Q2 L15 E16 V17 E18 P19 S20 D21 T22 N25 K29

4.3 Results and Discussion

In order to characterize the solvent density around the four proteins studied (two AFPs, ZaAFP, and CfAFP and two non AFPs, Ubi, and Bar), the proteins and their

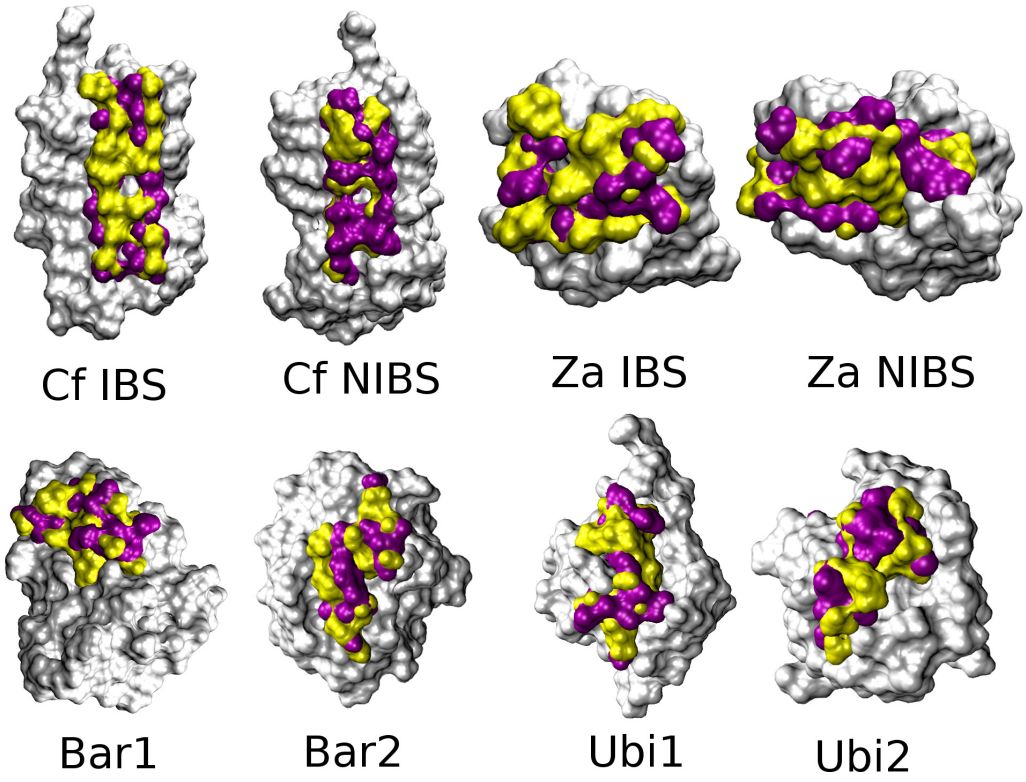

Figure 4.1: Representative sketch of the eight surfaces defined. CfAFP IBS and NIBS; ZaAFP IBS and NIBS; Barnase Surface 1 and Surface 2; Ubiquitin Surface 1 and Surface 2. The hydrophobic and hydrophilic portions of the surfaces are highlighted in yellow and purple, respectively. The rest of the protein is represented as white surface.

hydration shells are modeled as ellipsoids (see Figure 4.2). The solvent density ρ around the protein as a function of the distance from the protein ellipsoid surface is then calculated (see Hydration Shell Density Calculations subsection).

To compare the solvent properties inside the hydration shell with the properties of bulk solvent we use a previously employed strategy [11] to construct a fictitious hydration shell filled with bulk water. To this aim, the protein coordinates obtained from the protein + solvent MD simulations are inserted into a pure TIP4P/2005 box extracted from a MD simulation of water alone. The water molecules of the pure TIP4P/2005 box overlapping with the protein coordinates are removed, providing a protein + solvent box in which the hydration shell is filled with the bulk solvent. Water molecules of the pure TIP4P/2005 box are removed where the distance between any existing atom and any atom of the inserted protein is less than the sum of the van der Waals radii of both atoms. For each protein and at both temperatures, this is done for 10000 different protein coordinates, each one inserted in a different bulk solvent box. The solvent density around the protein ρ_{fict} as a function of the

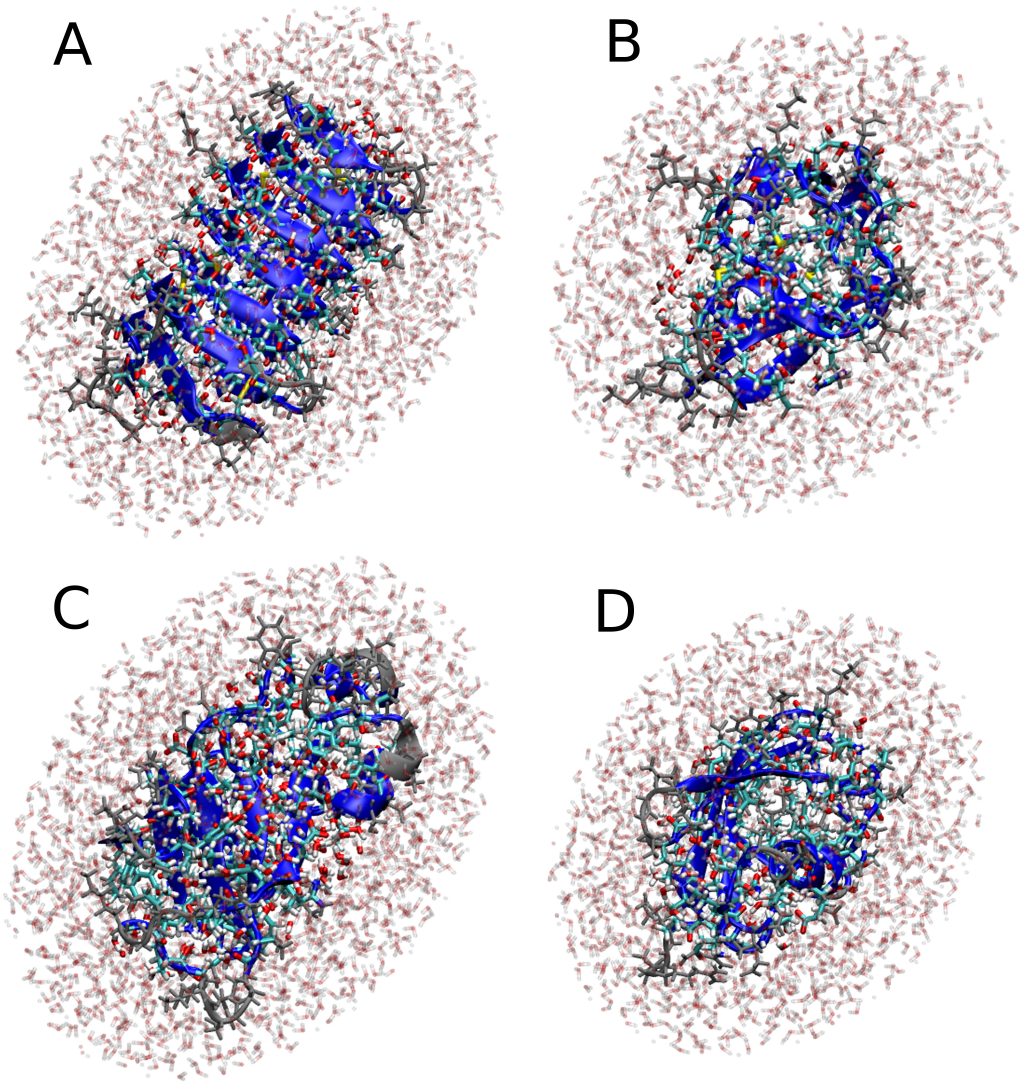

Figure 4.2: Representative structure of the four proteins investigated inside their 1 nm-thick hydration shell at 300 K: A, CfAFP; B, ZaAFP, C, Barnase; D, Ubiquitin. The protein+solvent atoms inside the ellipsoid that approximates the protein are highlighted in licorice with blue cartoon backbone. The hydration shell solvent molecules outside such an ellipsoid are represented in transparent licorice and the protein atoms that protrude from the protein ellipsoid surface are represented in gray licorice.

distance from the protein ellipsoid surface is then calculated for these fictitious protein + solvent configurations (see the Hydration Shell Density Calculations subsection). Such a strategy allows to investigate and compare the hydration shell properties of the different proteins removing possible effects because of the topological properties of the protein surfaces. In addition, it also allows to remove the effect on the density profile of the spatial arrangement of the protein atoms that protrude from the protein ellipsoid surface.

The ratio ρ/ρ_{fict} as a function of the distance from the protein ellipsoid surface

is reported in Figure 4.3 (the corresponding trends for ρ are reported in Figure 4.6). The figure shows that, as previously observed [129], at the protein surface, the density of the solvent is higher than the bulk density and decreases at increasing distances, approaching the bulk density, as expected. At 300 K (Figure. 4.3 a and b) a ≈12% increase is observed in the first 0.5 nm hydration layer in all the four proteins, in line with previous data [129]. The bulk density is reached at 300 K at ≈1 nm distance from the protein surface (see also panels a and b and corresponding insets of Figure 4.6). Beyond such a distance, the TIP4P/2005 molecules of the protein–solvent system behave equivalently to the bulk TIP4P/2005 ones. Such a hydration shell thickness is in good agreement with previous calculations on non-AFPs at 300 K with the same methodology [13] and on CfAFP [12] in the temperature range 240–300 K with a different approach. It is also in good agreement with the previously mentioned IR-based experimental estimate of a 0.95 nm hydration shell thickness for ZaAFP at room temperature of Sun and Petersen [119]. Most notably, the hydration shell thickness of the two AFPs is not different from the one of the two non-AFPs, showing that, in terms of density, AFPs do not feature any long-ranged solvation shell.

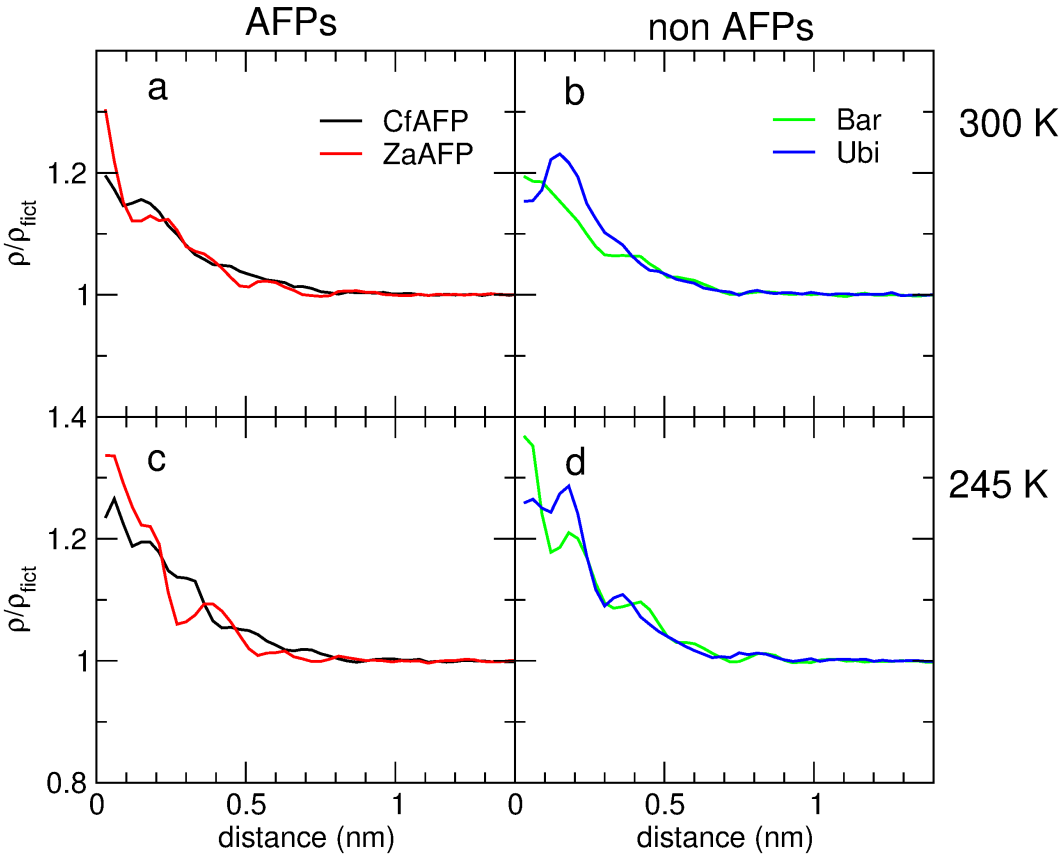

Figure 4.3: Ratio ρ/ρ_{fict} as a function of the distance from the protein ellipsoid surface for the four proteins at 300 K (a,b) and at 245 K (c,d) with TIP4P/2005 water model. AFPs are reported in panels (a,c) (CfAFP, black; ZaAFP, red) and non-AFPs are reported in panels b and d (Bar, green; Ubi, blue).

To assess that the results obtained do not depend on the choice of the water model, we repeated MD simulations and subsequent analyses at 300 K for all four systems with the SPC water model [89]. The ratio ρ/ρ_{fict} as a function of the distance from the protein ellipsoid surface is reported in Figure 4.4, a and b (the corresponding trends for ρ are reported later in this thesis, Figure. 4.5, to be compared with Figure. 4.3. The results are absolutely consistent with the ones obtained with the TIP4P/2005 water model. In fact, both the shell thickness (≈1 nm, see also next following paragraph on shell thickness) and the density increase in the first 0.5 nm hydration layer (≈12%) perfectly match the corresponding data obtained with the TIP4P/2005 water model for all the four proteins.

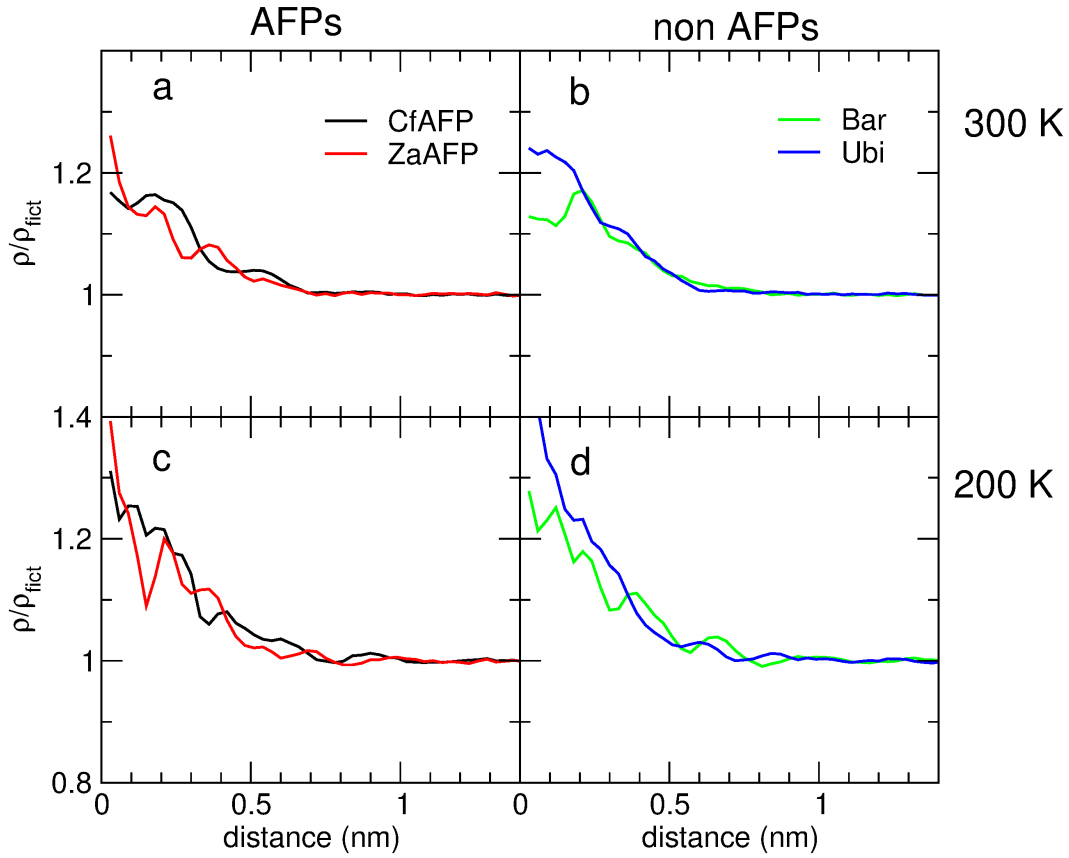

Figure 4.4: Ratio ρ/ρ_{fict} as a function of the distance from the protein ellipsoid surface for the four proteins at 300 K (a,b) and at 200 K (c,d) with SPC water model. AFPs are reported in panels (a,c) (CfAFP, black; ZaAFP, red) and non-AFPs are reported in panels b and d (Bar, green; Ubi, blue).

Given the agreement between the results obtained with the two water models, we per- formed with the SPC water model new simulations also at 200 K, i.e., at a temperature just above the melting temperature for SPC (190 K) [125]. Figure 4.4 c and d (and Figure 4.5, c and d and corresponding insets), show that the results at a temperature just above the melting are similar to the ones obtained at a temper-

ature just below the melting with the TIP4P/2005 water model. In fact, for all the four proteins a similar solvent density increase in the first 0.5 nm hydration layer is observed (≈16%) and the bulk density is reached at a distance of ≈ 1.15 nm.

Also the values of ρ_{shell} and η, together with the corresponding mean values of the protein excluded volume $\langle V_{ex} \rangle$, hydration shell volume $\langle V_{shell} \rangle$ and the number of SPC molecules inside the hydration shell $\langle N_{shell} \rangle$ are in agreement with the ones obtained with the TIP4P/2005 water model (see Table 4.1 and Table 4.2)

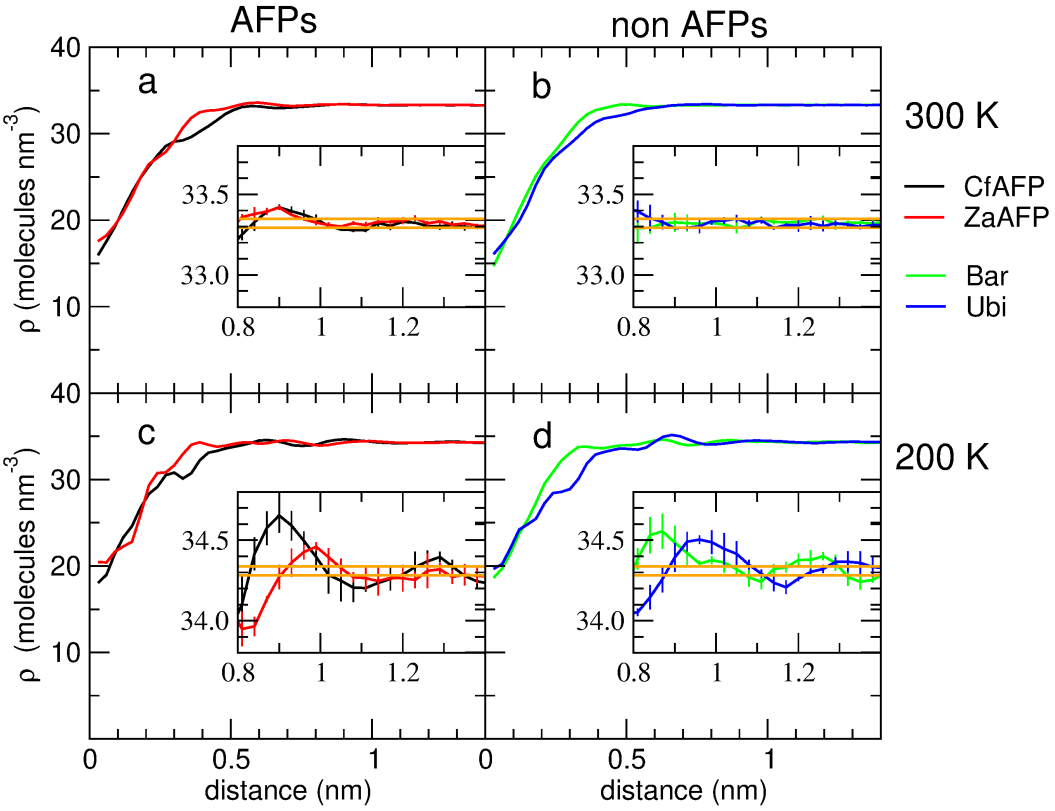

Figure 4.5: Ratio ρ/ρ_{fict} as a function of the distance from the protein ellipsoid surface for the four proteins at 300 K (a,b) and at 200 K (c,d). AFPs are reported in panels (a,c) (CfAFP, black; ZaAFP, red) and non-AFPs are reported in panels b and d (Bar, green; Ubi, blue).

As AFPs are active at low temperatures, we performed the same analysis at 245 K, that is, $4 - 7°$ below the melting temperature for the TIP4P/2005 water model used (249–252 K [124]) Figure 4.3c,d (and Figure 4.6 c,d and corresponding insets) shows similar results for AFPs and non-AFPs also at low temperature. In fact, for all the four proteins, a similar solvent density increase in the first 0.5 nm hydration layer is observed (≈16%, slightly higher than the 300 K one) and the bulk density is reached at a distance of ≈1.15 nm, slightly higher than the 300 K one (see Supporting Information for details on the shell thickness definition). It has to be noted that at both temperatures, the ratio ρ/ρ_{fict} is always higher than 1, that is, the density inside

Table 4.1: Mean values of the protein excluded volume $\langle V_{ex} \rangle$, hydration shell volume $\langle V_{shell} \rangle$ and number of SPC molecules inside the hydration shell $\langle N_{shell} \rangle$ as obtained from the MD trajectories of the four proteins at 300 K and 200 K[a].

		300 K				
		$\langle V_{ex} \rangle$	$\langle V_{shell} \rangle$	$\langle N_{shell} \rangle$	ρ_{shell}	η
AFPs	CfAFP	19.11	66.18	1706	36.25	0.088
	ZaAFP	11.50	48.07	1306	35.70	0.071
non-AFPs	Bar	19.20	67.67	1767	36.45	0.094
	Ubi	13.94	53.1	1422	36.30	0.089
		200 K				
		$\langle V_{ex} \rangle$	$\langle V_{shell} \rangle$	$\langle N_{shell} \rangle$	ρ_{shell}	η
AFPs	CfAFP	19.2	76.80	2158	37.48	0.093
	ZaAFP	11.46	56.33	1648	36.74	0.071
non-AFPs	Bar	19.30	78.8	2236	37.60	0.096
	Ubi	14.07	62.11	1797	37.42	0.091

Note: [a]For the calculation of V_{shell}, a hydration shell thickness of 1 nm and 1.15 nm is considered at 300 K and 200 K, respectively. All volumes are reported in nm^3. Mean solvent density ρ_{shell} (molecules per nm^3) within the accessible volume of the hydration shell and relative density increment η with respect to the bulk density are calculated according to Equation 4.2.1 and 4.2.2. The error on $\langle V_{ex} \rangle$ and $\langle V_{shell} \rangle$ is $\approx$0.3%, the one on $\langle N_{shell} \rangle$ is $\approx$0.2%, the one on ρ_{shell} is $\approx$0.2% and the one on η is $\approx$1%. Errors on the different quantities are evaluated through the standard error of their mean calculated over 3 (independent) sub-trajectories.

the hydration shell is always higher than the bulk solvent density. Therefore, there is no indication of an ice-like density (i.e., lower than liquid bulk) in the hydration shell of AFPs.

The definition of the hydration shell thickness allows determining the mean solvent density ρ_{shell} within the accessible volume of the protein hydration shell and the relative density increment η inside the hydration shell with respect to the bulk density (Equation 4.2.1 and 4.2.2). The values of ρ_{shell} and η, together with the corresponding mean values of the protein excluded volume $\langle V_{ex} \rangle$, hydration shell volume $\langle V_{shell} \rangle$ and the number of SPC molecules inside the hydration shell $\langle N_{shell} \rangle$ are reported in Table 4.2. Note that the number of molecules inside the hydration shell of ZaAFP (i.e., 1330) is in agreement with the experimental estimate of Sun and Petersen [119], who set a lower limit of the number of water molecules in the hydration shell of the protein to

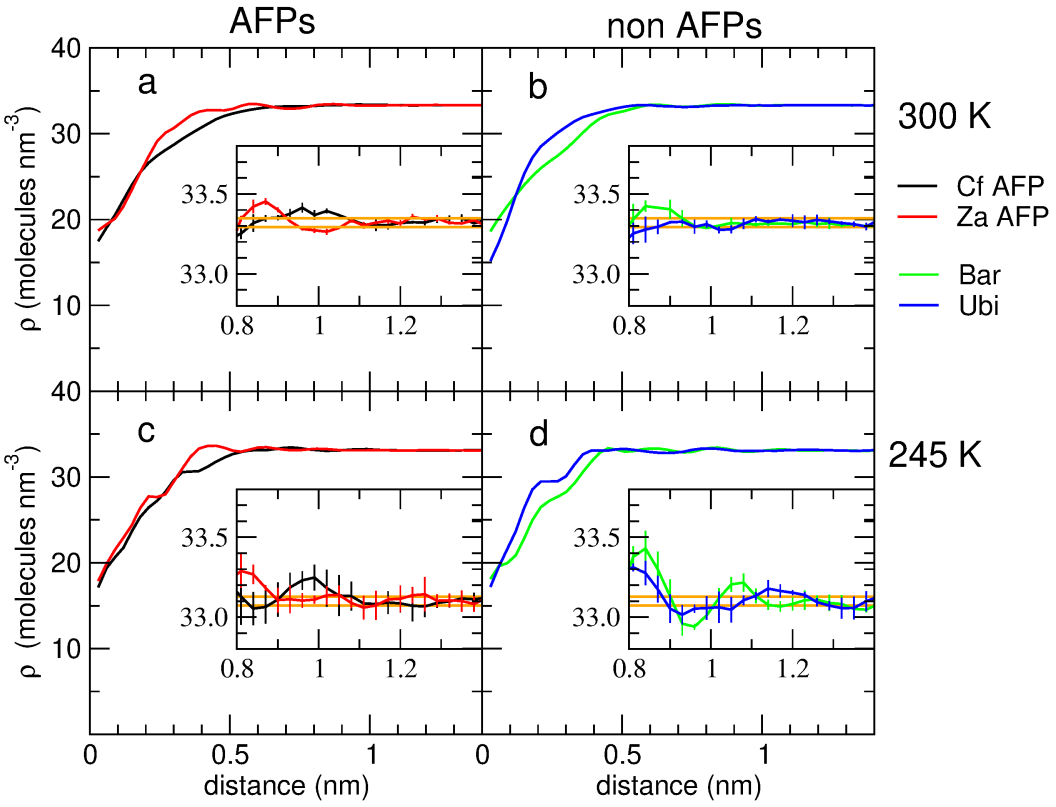

Figure 4.6: Layer density profile of the solvation TIP4P/2005 molecules for the four proteins at 300 K (a and b) and at 245 K (c and d). AFPs are reported in panels a and c (CfAFP, black; ZaAFP, red) and non-AFPs are reported in panels b and d (Bar, green; Ubi, blue). The insets show the convergence of the density profile, with its standard error, within the bulk density noise interval (displayed with two orange lines) corresponding to the 95% confidence.

1100. The data reported in the table show that the hydration shell solvent density and relative density increment with respect to the bulk are essentially indistinguishable for AFPs and non AFPs, both at 300 K and at 245 K. An overall density increment inside the hydration shell of ≈9% can be observed both at low and room temperatures for all the proteins.

In Figure 4.6 and 4.5 the layer density profiles of the four proteins are reported, as obtained by averaging over the MD trajectories the instantaneous TIP4P/2005 (Figure 4.6) and SPC (Figure 4.5) density within ellipsoidal layers around the proteins of increasing dimension (see section Hydration shell density calculations). For both water models, the data obtained at both room (300 K) and low (245 K for TIP4P/2005 and 200 K for SPC) temperature are reported. Such profiles allow determining the thickness of the hydration shell, i.e. the distance from the protein surface at which the layer solvent density reaches the bulk plateau value (≈33.3 molecules per nm^3 at 300 K, ≈33.1 molecules per nm^3 at 245 K with the TIP4P/2005 water model and ≈34.3 molecules per nm^3 at 200 K with the SPC water model, see MD simulations

Table 4.2: Mean values of the Protein-Excluded Volume $\langle V_{ex} \rangle$, Hydration Shell Volume $\langle V_{shell} \rangle$ and Number of TIP4P/2005 Molecules Inside the Hydration Shell $\langle N_{shell} \rangle$ as Obtained from the MD Trajectories of the Four Proteins at 300 K and 245 K[a].

		300 K				
		$\langle V_{ex} \rangle$	$\langle V_{shell} \rangle$	$\langle N_{shell} \rangle$	ρ_{shell}	η
AFPs	CfAFP	19.19	66.84	1735	36.41	0.092
	ZaAFP	11.51	48.64	1330	35.81	0.075
non-AFPs	Bar	19.71	70.97	1881	36.70	0.101
	Ubi	14.00	54.23	1463	36.36	0.091
		245 K				
		$\langle V_{ex} \rangle$	$\langle V_{shell} \rangle$	$\langle N_{shell} \rangle$	ρ_{shell}	η
AFPs	CfAFP	19.19	76.36	2066	36.13	0.092
	ZaAFP	11.61	56.90	1610	35.54	0.074
non-AFPs	Bar	18.91	75.76	2049	36.04	0.089
	Ubi	13.86	62.94	1758	35.82	0.082

Note: [a]For the calculation of V_{shell}, a hydration shell thickness of 1 nm and 1.15 nm is considered at 300 K and 245 K, respectively. All volumes are reported in nm^3. Mean solvent density ρ_{shell} (molecules per nm^3) within the accessible volume of the hydration shell and relative density increment η with respect to the bulk density are calculated according to Equation 4.2.1 and 4.2.2. The error on $\langle V_{ex} \rangle$ and $\langle V_{shell} \rangle$ is $\approx$0.3%, the one on $\langle N_{shell} \rangle$ is $\approx$0.2%, the one on ρ_{shell} is $\approx$0.2% and the one on η is $\approx$1%. Errors on the different quantities are evaluated through the standard error of their mean calculated over 3 (independent) sub-trajectories.

subsection). The layer solvent density approaches indeed the bulk plateau value by increasing the distance from the protein surface. We define the layer solvent density as indistinguishable from the bulk solvent density when the two quantities coincide within their noise. The error on the layer density is evaluated through the standard error of their mean calculated over three sub-trajectories while the bulk density standard error has been previously calculated at 300 K [13] and is 0.014 molecules per nm^3. The same bulk density standard error has been used at 245 K and 200 K. Within such a definition it can be observed from Figure 4.6 and 4.5 that the layer solvent density converges to the bulk solvent density at a distance of $\approx$1 nm from the protein surface at 300 K (insets of panels a and b of both figures) and at a distance of $\approx$ 1.1–1.2 nm from the protein surface at 245 K and 200 K (insets of panels c and d of both figures) for both AFPs and non AFPs. Note that at low temperature the layer density profile

displays a higher fluctuation and standard error, likely as a consequence of the slower dynamics. For the calculation of the hydration shell volume, mean solvent density and relative density increment reported in Table 4.2 for TIP4P/2005 and in Table 4.1 for SPC in the previous section, a hydration shell thickness of 1.15 nm has been considered at 245 K and 200 K, respectively.

According to the previous results, AFPs do not show any difference in terms of solvent density with respect to non-ice-binding proteins, nor considering the hydration shell thickness neither considering the density increment with respect to the bulk. However, such properties have been evaluated on the whole hydration shell and do not take into account that AFPs feature a specific region involved in ice-binding. It has been repeatedly proposed [11, 108–114] that solvation waters at the IBS could display an ice-like layer in solution to facilitate ice recognition, and we thus questioned whether a local density difference at the IBS could be present that remains hidden in the previous global analysis and that could potentially be relevant for AFP activity. We thus analyzed the solvent density at the IBS of the two AFPs and compared it both to the one at non-IBS surface portions of the AFPs and to the one at selected surface portions of the two non-AFPs. We chose non-IBS (NIBS) portions of AFPs and portions of Bar and Ubi having a solvent accessible surface area (SASA) similar to the two AFP IBS and, possibly, a flat surface (more details on the surface definition for the four proteins are provided in the Supporting Information). We then calculated the hydration density around such surfaces, ρ_{surf}, for the four proteins at low and high temperatures as follows. ρ_{surf} is defined as the number of water molecules, N_w, within a 0.55 nm radius cutoff from the heavy atoms on the protein-selected surfaces per unit of SASA ($\rho^{surf} = \langle N_w \rangle / \langle S \rangle$, with N_w and the SASA S averaged along the MD trajectory) [14, 130]. We restricted our analysis to a 0.55 nm thick hydration layer to maximize the possible local differences. The same surface density calculation was performed also for the fictitious protein + solvent configurations in which the hydration shell is filled with bulk water (ρ^{surf}_{fict}). This allowed us to compare the local surface density among different protein portions and also to compare such a density with the bulk density by calculating the relative increment η^{surf}

$$\eta^{surf} = \frac{\rho^{surf} - \rho^{surf}_{fict}}{\rho^{surf}_{fict}} \tag{4.3.1}$$

The results, reported in Table 4.3 and Figure 4.7, show for all the surfaces considered an increment with respect to the bulk density, more pronounced at 245 K, in line with the results on the whole hydration shell reported in Figure 4.3. High variability in such increment can be observed (2.4–9.2% at 300 K, 5.6–11.6% at 245 K).

Interestingly, the two AFP IBSs display a particularly low relative increment at both temperatures. Such a local lower density might be related to the antifreeze activity of the two proteins. However, it has to be remarked that the solvent density at the IBS is still higher than the bulk one, hence not suggesting the presence of any ice-like structure in the vicinity of the ice-binding surface.

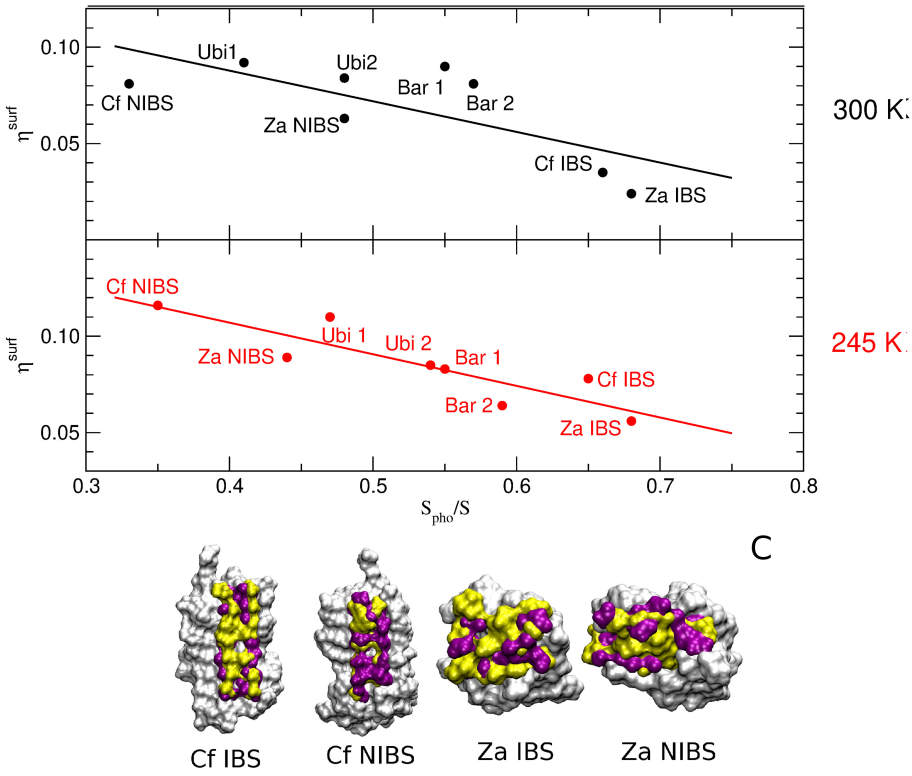

Figure 4.7: Relative increment η^{surf} of the hydration surface density with respect to the bulk density *vs* the fraction of hydrophobic solvent-exposed area (S_{pho}/S) of the 8 surfaces selected for the four proteins. (A) data at 300 K, (B) data at 245 K. (C) Representative sketch of the IBS and NIBS defined for CfAFP and ZaAFP. The hydrophobic and hydrophilic portions of the surfaces are highlighted in yellow and purple, respectively. The rest of the protein is represented as white surface.

Further analyses showed that the surface density increment with respect to the bulk density (i.e., η^{surf}, see Equation 4.3.1) is related to the chemical nature of the exposed surface. As shown in Figure 4.7, η^{surf} is in fact inversely proportional to the fraction of the hydrophobic-exposed surface for all the surfaces investigated. A lower increment corresponds to a higher fraction of the hydrophobic-exposed surface, and the two ice-binding surfaces display the highest hydrophobic content. From the comparison between IBSs and NIBSs in Figure 4.7, it can be also observed that the highly hydrophobic IBS surface is coupled to a high hydrophilic NIBS surface, supplying to the AFPs an amphipathic character [48]. In non-AFPs, the hydrophobic content is on the contrary more similar between the two surfaces (note that for all

Table 4.3: Mean Value of the Solvent Accessible Surface Area $\langle S \rangle$ (in nm^2), Surface Density ρ^{surf} (in Molecules per nm^2) and Relative Increment with Respect to the Bulk Surface Density η^{surf} Calculated for the Four Proteins at 300 and 245 K[a].

		300 K		
		$\langle S \rangle$	ρ^{surf}	η^{surf}
AFPs	CfAFP IBS	8.9	17.7	0.035
	CfAFP NIBS	7.2	18.1	0.081
	ZaAFP IBS	8.3	18.1	0.024
	ZaAFP NIBS	8.0	19.4	0.063
non-AFPs	Bar Surf. 1	8.1	18.6	0.090
	Bar Surf. 2	8.2	18.1	0.081
	Ubi Surf. 1	7.4	17.9	0.092
	Ubi Surf. 2	7.3	17.6	0.084
		245 K		
		$\langle S \rangle$	ρ^{surf}	η^{surf}
AFPs	CfAFP IBS	8.5	18.4	0.078
	CfAFP NIBS	6.9	18.7	0.116
	ZaAFP IBS	8.9	18.5	0.056
	ZaAFP NIBS	8.4	20.4	0.0.88
non-AFPs	Bar Surf. 1	7.7	18.4	0.083
	Bar Surf. 2	7.3	18.7	0.064
	Ubi Surf. 1	7.8	17.4	0.110
	Ubi Surf. 2	6.7	18.5	0.085

Note: [a]The error on $\langle S \rangle$ is ≈0.2%, the one on ρ^{surf} is ≈0.2%, and the one on η^{surf} is ≈1%.

proteins, the value of S_{pho}/S averaged on the two surfaces roughly corresponds to the value of S_{pho}/S for the whole protein).

We also analyze the local arrangement of water molecules within the same 0.55 nm radius cutoff from the heavy atoms on AFP IBS and NIBS. This was achieved by calculating the H–O–O angle distribution, $P(\Theta)$, of nearby water molecules (i.e., with an oxygen–oxygen distance below 0.35 nm) along the MD simulations at low temperature which provides a measurement of the tetrahedral order of the solvent. Similar analyses have been previously used to quantify the degree of order of water at AFP ice-binding sites [109, 111, 131, 132]. In line with previous results on both CfAFP [111] and ZaAFP [132], our results indicate that water is more structured at the IBS with respect to NIBS and bulk water. In fact, as shown in Figure 4.8, the low-

Θ peak, corresponding to the tetrahedral water population [111], is higher at the IBS of both CfAFP and ZaAFP than at both NIBS and in bulk water. It should be noted that this increased local ordering around IBSs with respect to bulk water does not correspond to a decrease in density, as would be the case if the structuring were ice-like. In line with previous findings showing that the height of the low-Θ peak is increased around hydrophobic groups and decreased around hydrophilic groups [109, 132], the intensity of the peak shown in Figure 4.8 also correlates with the hydrophobic and hydrophilic nature of the exposed surface. The highly hydrophilic (65% at 245 K) NIBS of CfAFP displays indeed a less-intense peak than the NIBS of ZaAFP (56% of hydrophilic exposed surface at 245 K), that displays a peak as intense as the bulk one.

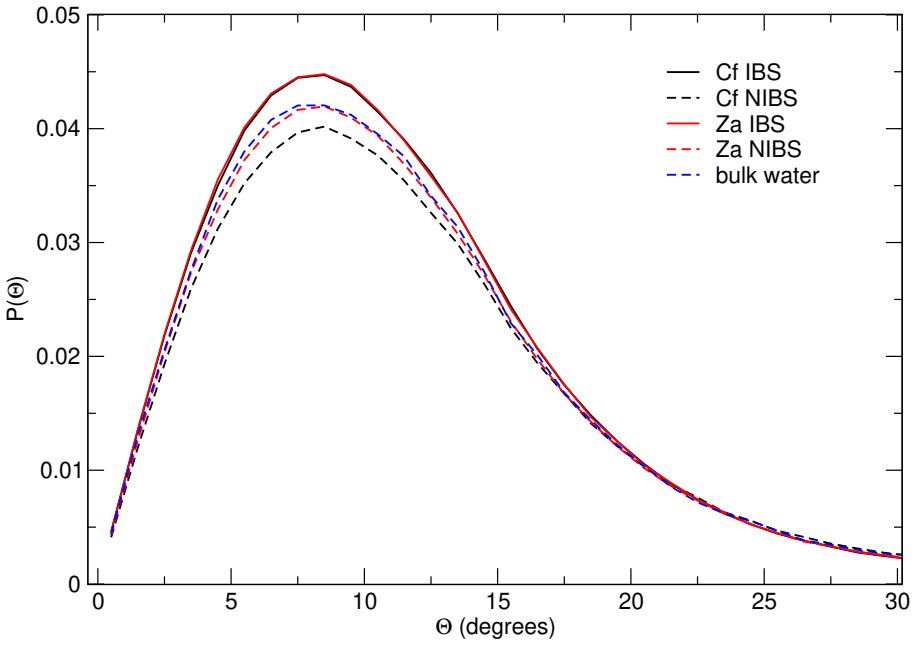

Figure 4.8: H–O–O angle distribution, P(Θ), of nearby water molecules (i.e., with a oxygen–oxygen distance below 0.35 nm) for waters in the first hydration layer (within 0.55 nm from the heavy atoms of the protein) of the IBS (solid lines) and NIBS (dashed lines) surfaces of CfAFP (black) and ZaAFP (red) at 245 K. The low-angle population is increased around the IBSs with respect to bulk water (blue dashed line) and with respect to the NIBSs.

4.4 Conclusion

We analyze here the hydration shell of AFPs in terms of density, comparing it both to the bulk density and to the hydration shell density of proteins with no ice-binding activity. Our results show that the hydration shell of AFPs does not feature any relevant difference with respect to the one of non AFPs. We observe indeed a sim-

ilar shell thickness and relative density increment with respect to the bulk density for all the proteins investigated. The calculated hydration shell thickness for Za-AFP and number of solvent molecules inside such a shell are in very good agreement with the corresponding IR-based experimental estimates, confirming that the technique proposed by Sun and Petersen [119] to separate the spectral component of the solvent perturbed by the solute (i.e., of the hydration shell) is capable to provide reliable information about hydration shells by analyzing the solute-correlated solvent IR spectrum.

We also analyze the solvent density around selected surfaces of both AFPs and non-AFPs to detect possible local differences related to the ice-binding activity of AFPs. Such an analysis shows that the solvent at AFP ice-binding surfaces features a lower density increment with respect to the bulk density and a higher degree of order with respect to the solvent at nonice-binding surfaces. We also show that the lower density increment is related to the high hydrophobic content of AFP ice-binding surfaces. Nevertheless, the solvent density at the IBS is still higher than the bulk one, that is, no ice-like structure is detected at the ice-binding surface. Hence, our data support the recent finding that water at the ice-binding site does not acquire an ice-like or clathrate-like structure in solution [116]. Yet, our data also show that although the overall hydration shell density in AFPs does not differ from the one of non-AFPs, the higher content of hydrophobic (at ice-binding sites) and hydrophilic (at non ice-binding sites) residues in AFPs determines a slightly inhomogeneous density inside the hydration shell of AFPs (lower at ice-binding sites and higher at non ice-binding sites). While the hydrophobic character of the ice-binding surface might facilitate ice-binding, the hydrophilic character of non-ice-binding surfaces might provide protection against ice growth, as was suggested many years ago but on the basis of dynamical analysis [111]. In fact, the higher water density might discourage ice growth around the bound AFPs, contributing to the curvature of the ice surface at the basis of TH. The coupled features of ice-binding and nonice-binding surfaces in AFPs and the resulting density unbalance could thus contribute to the antifreeze activity.

Chapter 5

High water-density at non-ice-binding surfaces contributes to the hyperactivity of antifreeze proteins

5.1 Introduction

Surviving in exceptionally cold environments whose temperatures drop below the melting temperature of ice is an example of an exclusive adaptive mechanism which protects the cells of a variety of organisms including fish, insects, bacteria, plants, and fungi, from getting damaged at temperatures below zero. These organisms exploit a group of proteins, called antifreeze proteins (AFPs), that are able to lower the freezing temperature of the surrounding water. Despite their variability in terms of primary, secondary, and tertiary structures, AFPs share a common mechanism of action which consists in binding to small ice nuclei, preventing their further growing at the adsorbed positions. [133, 134]

AFPs bind to ice exploiting a number of hydrogen bonds and hydrophobic interactions between specific ice planes and a complementary protein surface, the ice-binding surface (IBS). The mechanism proposed for the antifreeze activity is that, by binding to a growing ice nucleus, AFPs cause a micro-curvature of the ice surface between the adsorbed AFPs. [50, 133] Because the ice-crystal growth is thermodynamically less favorable on a curved ice surface than on a flat one, the ice growth stops until the temperature drops. This leads to a local difference between the freezing and melting temperature, ΔT, known as thermal hysteresis (TH). Based on the TH values, the AFPs can be classified into moderately-active ($\Delta T < 1$ K), mostly alanine-rich α-helical AFPs, and hyperactive AFPs ($\Delta T > 1$ K), mostly threonine-rich β-helical proteins found in insects. At lower concentrations (<0.5 g/L), hyperactive AFPs greatly outperform more traditional antifreeze agents, making them of potential interest for use in medicine, agriculture, food processing, and surface protection [135].

A general consensus has been recently reached on the nature of the interactions that determine the binding to the ice surface: a spatial match between the polar

residues of the IBS and the oxygen-oxygen distance in the ice lattice is needed, [37, 98–101] pointing to a critical role of hydrogen bonds. At the same time, a relevant hydrophobic content in the IBS is required, providing an entropically-driven thermodynamic contribution to ice binding. Despite a crucial role played by the AFPs hydration layer in the molecular mechanism of the antifreeze activity is recognized, [32, 33] its structural and dynamical characterization is still incomplete. On the basis of molecular dynamics (MD) simulations, it has been proposed that hydration water at the IBS could form an ice-like layer already in solution, thereby easing ice recognition [11, 108–114]. However, other MD simulations [16, 115, 116, 136] and an NMR study [117] did not find any indication of ice-like structure in liquid water around AFPs. Concerning the dynamics, there is both computational [111, 137] and experimental evidence [33] that water near the IBS displays exceptionally slower hydrogen bond reorientation dynamics compared with other protein surfaces.

At variance with the IBS, detailed studies of the role of the non-ice-binding surfaces (NIBSs) of AFPs are still lacking, despite the undeniable role fulfilled by the solvation water of NIBSs. Previous computational work [11,16,138–140] showed that the solvent at the IBSs features a lower density increment with respect to the bulk density and a higher degree of order with respect to the solvent at the NIBSs. In particular, we recently showed for two AFPs that although the overall hydration shell density does not differ from the one of non-AFPs [16,136], the higher content of hydrophobic (at ice-binding sites) and hydrophilic (at non ice-binding sites) residues in AFPs determines a slightly inhomogeneous density inside the hydration shell of AFPs (lower at ice-binding sites and higher at non ice-binding sites) [16].

While it is widely accepted that the lower density at the IBS favours binding to ice, the higher water density at the NIBS might discourage ice growth around the bound AFPs, as was suggested many years ago but on the basis of dynamical analysis, [111] thus contributing to the curvature of the ice surface at the basis of TH. The coupled features of ice-binding and non ice-binding surfaces in AFPs and the resulting density unbalance could thus contribute to the antifreeze activity by providing protection against ice growth

Here, we provide further evidence of a clear correlation between the hydration density at the IBS and the antifreeze activity for three moderately-active and four hyperactive AFPs. More importantly we show that, rather than differences in the hydration structure at the IBSs, what actually discriminates between moderately-active and hyperactive AFPs is an increased hydration density at the NIBSs of the latter. This increased density around the non-ice-binding parts of the hyperactive AFPs might contribute to enhance their capability of preventing the adsorbed protein

being overgrown by a growing ice surface.

5.2 Materials and Methods

5.2.1 Molecular Dynamics Simulation

The MD simulations of the seven proteins were carried out using Gromacs 5.1.4. [87] The OPLS-AA force field [88] was used for the proteins in conjunction with the SPC model [89] for the water. To neutralize the system an appropriate number of counter ions (when needed) was added to the simulation box. For each AFP, the temperature was held constant at 300 K by the velocity-rescaling algorithm. [76] The density of each box containing the protein and the SPC water molecules was calibrated to get (within the isothermal–isochoric, NVT, ensemble) a pressure of $\approx$ 560 bar, which is the pressure provided by a 100-ns long simulation of a pure SPC box at 300 K with a density corresponding to the experimental liquid water density at the same temperature ($\approx$ 33.3 molecules per nm^3). All bonds were constrained with the LINCS algorithm [79] allowing to use a 2 fs integration step. A cut-off radius of 1.1 nm was used for the short-range interactions and the particle mesh Ewald method [72] was used to compute long-range interactions with grid search and cut-off radii of 1.1 nm. After protein, and subsequent solvent, relaxation, each protein was simulated for 100 ns.

The MD simulation of the pure SPC [89] water, with the density corresponding to the room-temperature liquid water density of $\approx$33.3 molecules per nm^3, was used to create the fictitious protein + bulk-water configurations. For each AFP, the protein coordinates extracted from 10000 frames of the protein + solvent MD simulation are overlaid on 10000 configurations extracted from the bulk-water simulation and the water molecules overlapping with the protein coordinates, i.e. when the distance between any water atom and any atom of the inserted protein is less than the sum of the van der Waals radii of both atoms, are removed. This procedure provides 10000 "fictitious" configurations of protein + solvent in which the hydration shell is filled with bulk-like water molecules. The solvent density, $\rho_{b,fict}$, around the protein as a function of the distance from the protein ellipsoid surface is then calculated for these fictitious protein + solvent configurations.

5.2.2 Protein hydration shell density calculation

In order to characterize the solvent density around the protein, we use the approximation of treating the protein molecule as an ellipsoid defined, at each molecular

dynamics (MD) time frame, by the eigenvectors and eigenvalues of the 3×3 geometrical covariance matrix of the x, y, z atomic coordinates [13]. We then consider a set of ellipsoidal layers around the protein defined by the consecutive ellipsoids with semi-axes $a_i^{(n)} = a_i + n\delta$ with fixed increment $\delta = 0.03$. By calculating at each MD frame the instantaneous water density within each layer (disregarding the possible presence of protein atoms and/or counter-ions) and averaging over the MD trajectory, we obtain the solvent density profile around the protein within layers of increasing distance from the protein ellipsoid surface (i.e., the layer solvent density profile, see Figure 5.1). The distance from the protein surface at which the layer solvent density reaches the bulk plateau value (≈ 33.3 molecules per nm^3 at 300 K) defines the hydration shell boundary layer, that is, the thickness of the protein hydration shell and corresponding shell volume V_{shell}. The mean value of the number of SPC water molecules within each layer, that is, as a function of the distance from the protein ellipsoid surface, is also calculated along the MD simulations. In addition, the protein-excluded volume V_{ex} , that is, the volume enclosed by the solvent-accessible surface, can be obtained by Gromacs using a probe radius of 0.14 nm, according to the method reported in Eisenhaber et al. [93]. By calculating V_{ex} at each MD time frame of the simulations, we obtain the protein mean volume and the corresponding thermal distribution. From V_{shell}, V_{ex} and the number of SPC water molecules inside the hydration shell, N_{shell}, the mean solvent density within the accessible volume of the protein hydration shell can be obtained via:

$$\rho_{shell} = \left\langle \frac{N_{shell}}{V_{shell} - V_{ex}} \right\rangle \approx \frac{\langle N_{shell} \rangle}{\langle V_{shell} \rangle - \langle V_{ex} \rangle} \tag{5.2.1}$$

The mean solvent density inside the hydration shell can be compared to the bulk solvent density ρ_{bulk}, providing the relative density increment η inside the hydration shell with respect to the bulk density:

$$\eta = \frac{\rho_{shell} - \rho_{bulk}}{\rho_{bulk}} \tag{5.2.2}$$

5.2.3 Partial molar volume calculation

The protein partial molar volume, i.e. $v = (\partial V / \partial N)_{p,T,N_{solvent}}$ with V the volume of the system, N the number of solute molecules, at a constant pressure p, constant T, was calculated as [141]:

$$v = \langle V_{ex} \rangle + \frac{\langle N_{shell} \rangle}{\rho_{shell}} + \frac{(N_{SPC} - \langle N_{shell} \rangle)}{\rho_{bulk}} - \frac{N_{SPC}}{\rho_{bulk}} \tag{5.2.3}$$

where N_{SPC} denotes the total number of SPC water molecules contained within each protein simulation box. Equation 5.2.3 can be arranged as follows:

$$
\begin{aligned}
v &= \langle V_{ex} \rangle + \langle N_{shell} \rangle \left(\frac{1}{\rho_{shell}} - \frac{1}{\rho_{bulk}} \right) \\
&= \langle V_{ex} \rangle + \frac{(\rho_{bulk} - \rho_{shell})}{\rho_{bulk}} \frac{\langle N_{shell} \rangle}{\rho_{shell}} \\
&= \langle V_{ex} \rangle - \eta(\langle V_{shell} \rangle - \langle V_{ex} \rangle)
\end{aligned}
\tag{5.2.4}
$$

where η is the relative density increment inside the hydration shell with respect to the bulk density described in the previous section.

5.3 Results and Discussion

Seven AFPs (three moderately-active and four hyperactive) are simulated at a constant pressure and at room temperature (see Methods section). The moderately-active AFPs are: the snow mold fungus AFP from *Typhula ishikariensis* isoform 6 (*Tis*AFP6, PDB ID 3VN3); the fish type III AFP from *Zoarces americanus* (*Za*AFP, PDB ID 1HG7); the snow mold fungus AFP from *Typhula ishikariensis* isoform 8 (*Tis*AFP8, PDB ID 5B5H). The hyperactive AFPs are: the spruce budworm AFP from *Choristoneura fumiferana* isoform 337 (*Cf*AFP337, PDB ID 1L0S); the beetle AFP from *Tenebrio molitor* (*Tm*AFP, PDB ID 1EZG); the longhorn beetle AFP from *Rhagium inquisitor* (*Ri*AFP), PDB ID 4DT5); the spruce budworm AFP from *Choristoneura fumiferana* isoform 501 (*Cf*AFP501, PDB ID 1M8N).

In order to characterize their hydration shell, the proteins are modelled as ellipsoids (details are given in the Supporting Information - SI) and the solvent density, ρ, around them is calculated as a function of the distance from the ellipsoid surface (see Figure S1 and S2 in SI). To compare the properties of water within the hydration shell with those of the bulk water, a previously employed strategy [16, 136] was followed, which is based on a fictitious hydration shell filled with water molecules distributed as they would be in the bulk solvent. This strategy permits to remove possible effects arising from the different topological features of the protein surfaces on the hydration shell features and hence to remove the effect of the spatial arrangement of the protein atoms that protrude from the protein ellipsoid surface on the density profiles. More details on the computational procedure are presented in the Methods section.

The variations of the hydration density with respect to bulk water are followed by computing the ratio $\rho/\rho_{b,fict}$, in which $\rho_{b,fict}$ is the density within the fictitious hydration shell as a function of the distance from the protein ellipsoid surface (see Figure 5.3, panel A). It can be seen that the density of the solvent is higher than the bulk density at the protein surface and decreases at increasing distances, approaching the bulk density of ≈ 33.3 molecules nm^{-3} (corresponding to the experimental liquid

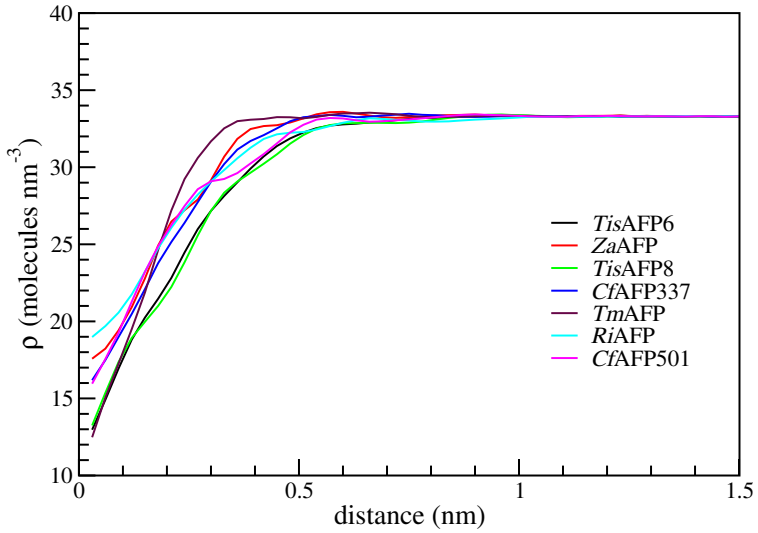

Figure 5.1: SPC solvation molecules layer density profile at 300K for the seven AFPs.

water density at 300 K) at around 1 nm. This analysis shows that the ratio $\rho/\rho_{b,fict}$ is always higher than 1, that is, the density around the proteins is always higher than the bulk solvent density, as previously observed for a subset of these proteins and several non-AFPs. [16, 136]

In order to define the protein hydration shell thickness, and corresponding shell volume V_{shell}, we use the distance from the protein ellipsoid surface of 1 nm at which the layer solvent-density reaches the bulk plateau value. The relative density increment, η, inside the hydration shell with respect to the bulk density, ρ_{bulk}, is calculated as:

$$\eta = \frac{\rho_{shell} - \rho_{bulk}}{\rho_{bulk}} \tag{5.3.1}$$

where ρ_{shell}, the density within the hydration shell, is estimated as:

$$\rho_{shell} = \left\langle \frac{N_{shell}}{V_{shell} - V_{ex}} \right\rangle \approx \frac{\langle N_{shell} \rangle}{\langle V_{shell} \rangle - \langle V_{ex} \rangle} \tag{5.3.2}$$

In the last equation, V_{ex} is the protein-excluded volume (i.e., the volume enclosed by the solvent-accessible surface obtained by using a probe radius of 0.14 nm, according to the method reported in Eisenhaber et al. [93]), N_{shell} is the number of water molecules inside the hydration shell and angle brackets indicate ensemble averages.

From the above data, we can also estimate the partial molar volume, v, through the following expression, that was previously shown to be affected by small statistical errors [13] (its derivation is given in the SI):

$$v = \langle V_{ex} \rangle - \eta(\langle V_{shell} \rangle - \langle V_{ex} \rangle) \tag{5.3.3}$$

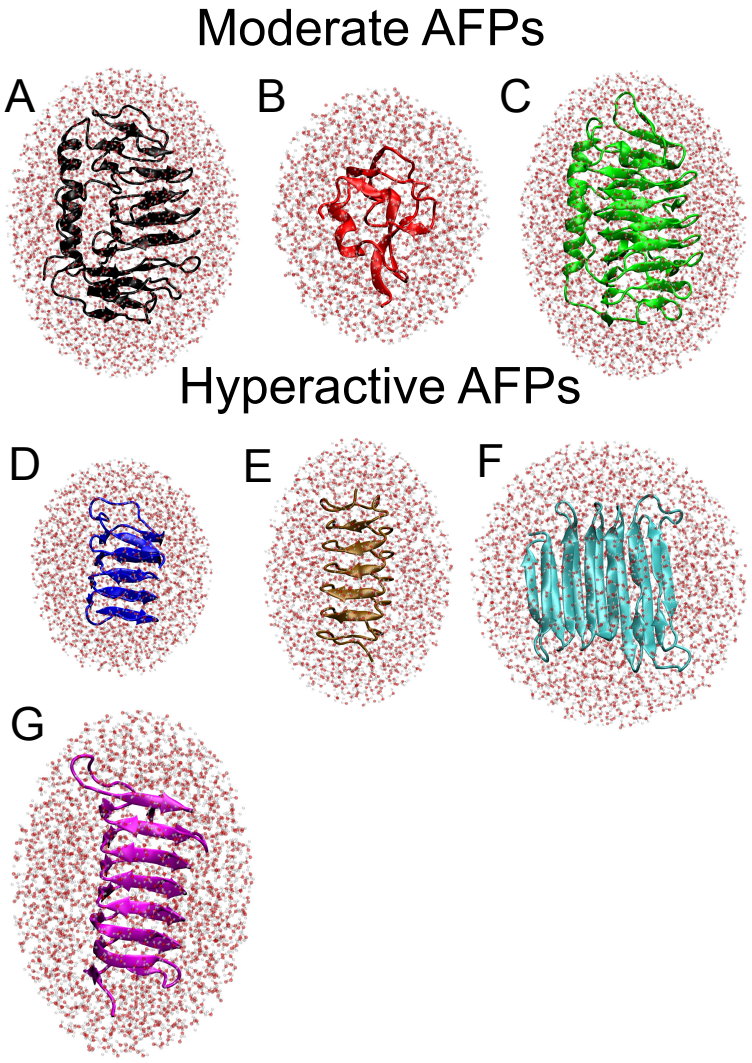

Figure 5.2: The ellipsoidal model generated a representative solvation layer for the seven proteins. The color of the proteins is specified by the density profile shown in Figure S1. A) *Tis*AFP6, B) *Za*AFP, and C) *Tis*AFP8 are moderately-active AFPs with antifreeze activity ¡ 1 K. D) *Cf*AFP337, E) *Tm*AFP, F) *Ri*AFP, and G) *Cf*AFP501 are hyperactive AFPs with antifreeze activity ¿ 1 and ¡ 6 K.

The values of the relative increment in the hydration-shell density, η, and of the partial molar volumes, v, along with the quantities needed for their calculation, for the seven studied AFPs are reported in Table 5.1. η is in the range of 6-12 %, which is in agreement with experimental data [18] and previous calculations. [8, 10, 16, 136]. By reporting η as a function of the partial molar volume v (see eq 5.2.4) a positive correlation can be seen, indicating a length-scale dependence of the density-increase on the protein size (see Figure 5.3 panel B). The dependence of η on the protein size was addressed in details in a previous paper on a larger set of eighteen proteins, out of which eight were AFPs and ten non-AFPs. [136] Conversely, no clear correlation is observed between η and the AFPs activities (see Figure 5.3 in which η is reported as a function of the antifreeze activity, i.e., the experimental thermal hysteresis $\Delta T(K)$

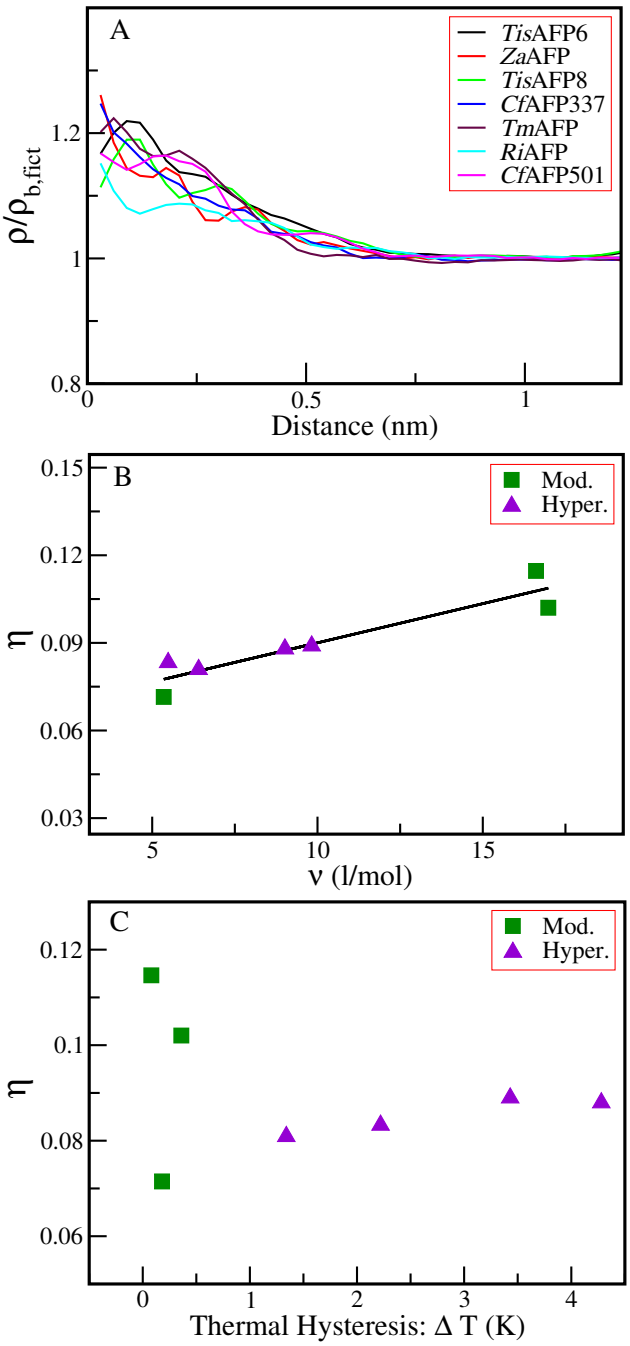

Figure 5.3: Panel A: Ratio $\rho/\rho_{b,fict}$ as a function of the distance from the protein ellipsoid surface calculated for the seven proteins: *Tis*AFP6 (black), *Za*AFP (red), *Tis*AFP8 (green), *Cf*AFP337 (Blue), *Tm*AFP (violet), *Ri*AFP (cyan), and *Cf*AFP501 (magenta). Panel B: Relative surface density increment η as a function of partial molar volume v calculated for each protein molecule using eq 5.2.4. Panel C: η as a function of the antifreeze activity, $\Delta T(K)$, measured at a protein solution concentration of 0.3 g/L. [97] The moderatetly-active (Mod.) AFPs are represented with green squares whereas the hyperactive (Hyper.) AFPs are represented with purple triangles.

at a protein solution concentration of 0.3 g/L).

The lack of any significant correlation between the whole hydration-density of the

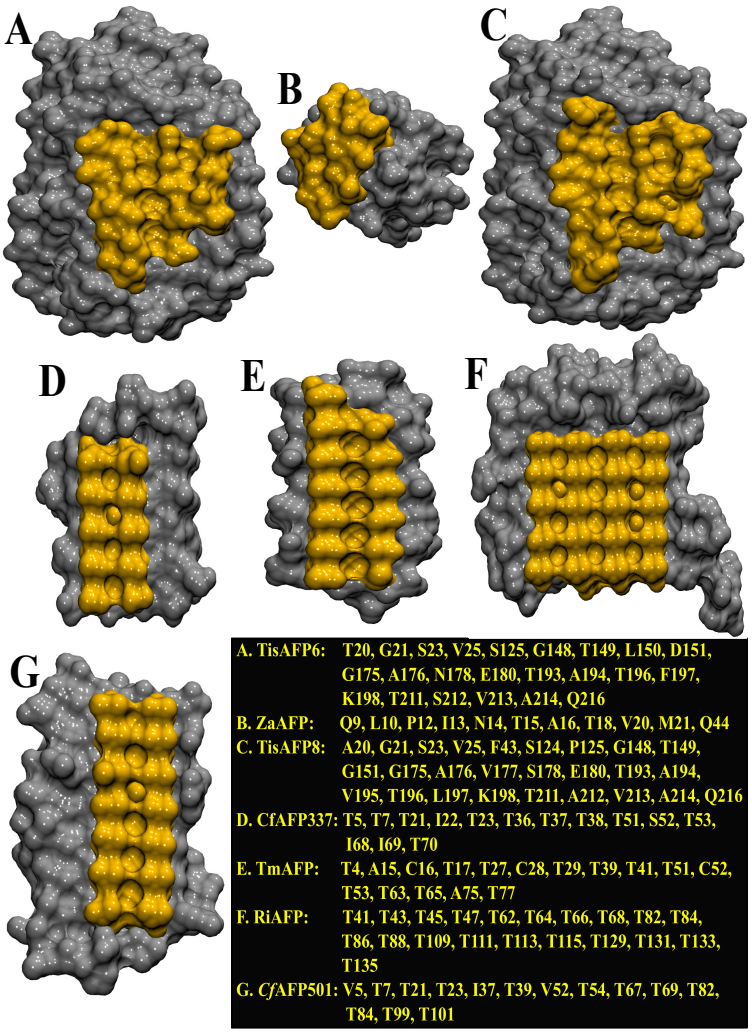

Figure 5.4: The ice-binding surfaces of the proteins are colored in orange and the rest of the protein considered as NIBS are colored in silver. A. *Tis*AFP6, B. *Za*AFP, C. *Tis*AFP8, D. *Cf*AFP337, E. *Tm*AFP, F. *Ri*AFP, and G. *Cf*AFP501. The respective residues of the IBS are also mentioned in the dark column.

AFPs and their activity prompted us to analyze the local density around the IBS, which is the specific surface involved in ice-binding, and compare it to that around the NIBS, i.e., the rest of the protein surface. In order to remove possible effects arising from the topological properties of the IBSs, the seven AFPs studied here were chosen to have rather flat IBSs with a similar surface size, as can be seen from Figure 5.4 and Table 5.2 in which the definition of the IBS and the NIBS, their solvent accessible surface area (SASA) and their hydrophobicity are reported.

For each AFP, the solvent-density around a given surface, ρ_{surf}, is defined as the number of water molecules, N_w, within a 0.55 nm radius cutoff from the heavy atoms on the selected surface per unit of SASA ($\rho_{surf} = \langle N_w \rangle / \langle S \rangle$ with N_w and the SASA, S, averaged along the MD trajectory). We chose a 0.55 nm cutoff, which is lower than the 1 nm thickness of the whole hydration shell, to define the first hydration layer with

Table 5.1: Total number of residues (RN), experimental thermal hysteresis (ΔT), mean excluded volume ($\langle V_{ex} \rangle$), mean hydration shell volume ($\langle V_{shell} \rangle$), hydration-shell density increment relative to bulk density (η) and partial molar volume (v) as obtained from the MD trajectories of the seven proteins.[a]

Protein	RN	ΔT	$\langle V_{ex} \rangle$	$\langle V_{shell} \rangle$	$\langle N_{shell} \rangle$	η	v
*Tis*AFP6	222	0.08	34.57	95.58	2266	0.11	16.61
*Za*AFP[b]	66	0.18	11.50	48.07	1305	0.07	5.35
*Tis*AFP8	223	0.36	34.46	95.85	2254	0.10	16.98
*Cf*AFP337	87	1.34	13.81	53.16	1417	0.08	6.40
*Tm*AFP	82	2.22	12.24	49.87	1358	0.08	5.48
*Ri*AFP	139	3.43	21.16	75.82	1983	0.09	9.82
*Cf*AFP501	120	4.28	19.11	66.18	1707	0.09	9.01

[a] The hydration shell thickness is taken as 1 nm. V_{ex} and V_{shell} are reported in nm^3. The relative density increment η with respect to the bulk density and the partial molar volumes v (reported in l/mol) are calculated according to the eq 3.2.9 and 5.2.4 respectively. The error on $\langle V_{ex} \rangle$ and $\langle V_{shell} \rangle$ is ≈0.3 %, for $\langle N_{shell} \rangle \approx 0.2\%$, and the one on η is ≈1 %. Errors on these quantities are calculated through the standard error of their mean evaluated over 3 subtrajectories.
[b] The number of water molecules inside the hydration shell of ZaAFP (i.e. 1305) is in good agreement with the experimental estimate of a lower limit of 1100 molecules. [119].

the aim of maximizing the local differences in water density at the different surfaces. The surface density calculation is also performed for the "fictitious" configurations in which the first hydration layer is filled with bulk water (ρ_{surf}^{fict}). In this way the local surface density of the IBS and NIBS can be compared among the different proteins by calculating the relative increment with respect to the bulk, η_{surf}, as:

$$\eta_{surf} = \frac{\rho_{surf} - \rho_{surf}^{fict}}{\rho_{surf}^{fict}} \tag{5.3.4}$$

η_{surf} is calculated for both the IBS and the NIBS of each protein and the results are reported in Table 5.2 and Figure 5.5. For all the surfaces considered, a local increment with respect to bulk density, in line with the results on the whole hydration shell reported in Figure 5.3, is observed, suggesting that no ice-like structure (i.e., with a density lower than that of bulk) is present in the vicinity of the ice-binding surface. Moreover, the IBSs feature on average a higher η_{surf} with respect to the NIBSs and, for each protein, the IBS always displays a lower relative increment with respect to the corresponding NIBS (with the only exception of *Tis*AFP6 for which they are comparable).

Further analysis shows that the surface density increment with respect to the bulk density correlates with the chemical nature of the exposed surface. From Figure 5.5,

Table 5.2: Average solvent accessible surface area, $\langle S \rangle$, of the whole protein (WP) and of the ice binding surface (IBS) and average hydrophobic-surface fraction, $\langle S_{pho}/S \rangle$, for the WP, IBS, and the NIBS.

	$\langle S \rangle$		$\langle S_{pho}/S \rangle$		
Protein	WP	IBS	WP	IBS	NIBS
*Tis*AFP6	99.865	10.109	0.562	0.560	0.562
*Za*AFP	40.090	8.171	0.583	0.667	0.562
*Tis*AFP8	98.246	10.697	0.587	0.702	0.573
*Cf*AFP337	48.561	8.019	0.518	0.582	0.505
*Tm*AFP	42.677	7.573	0.455	0.609	0.422
*Ri*AFP	72.970	15.463	0.521	0.540	0.516
*Cf*AFP501	62.845	8.658	0.494	0.667	0.467

Note: The units of S are nm^2. The S and S_{pho}/S values are averaged over 10 000 structures extracted from the 100 ns-long MD trajectory of each protein.

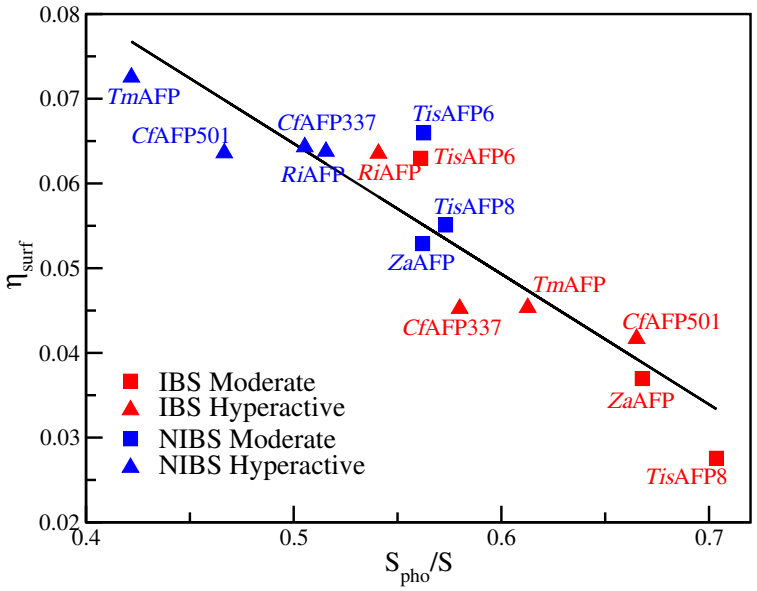

Figure 5.5: Change in the relative water-density increment with respect to the bulk (η_{surf}) as a function of the hydrophobic fraction of the solvent-exposed-area ($\langle S_{pho}/S \rangle$) of the IBSs (represented in red) and NIBSs (represented in blue) of the seven antifreeze proteins.

in which η_{surf} is reported as a function of the fraction of the accessible hydrophobic surface area, S_{pho}/S, for all the IBSs and NIBSs, it can be seen that η_{surf} is inversely proportional to S_{pho}/S. A lower increment corresponds to a higher S_{pho}/S and the ice-binding surfaces display the highest hydrophobic content. Moreover, a highly hydrophobic IBS is coupled to a highly hydrophilic NIBS, conferring to the AFPs a strong amphipathic character. [48] Similar results were previously found on a smaller set, namely two of the seven AFPs studied here [16]

Table 5.3: The relative surface density increment η_{surf} of the IBS and NIBS[a].

	η_{surf}	
Protein	IBS	NIBS
*Tis*AFP6	0.063	0.066
*Za*AFP	0.037	0.053
*Tis*AFP8	0.028	0.055
*Cf*AFP337	0.045	0.064
*Tm*AFP	0.045	0.073
*Ri*AFP	0.064	0.064
*Cf*AFP501	0.042	0.064

[a]Note: The error on the η_{surf} is ≈1 %.

In order to understand whether the properties of the local densities and the chemical character of the IBSs and NIBSs can be related to the antifreeze activity of the AFPs, we report η_{surf} and S_{pho}/S as a function of ΔT (see Figure 5.6). For both moderately-active and hyperactive AFPs the hydrophobicity of the IBS is not only high (with S_{pho}/S in the range of 0.54-0.70) but also similar between the two classes (being on average slightly higher for the moderately-active ones). Moreover, a positive correlation with the activity can be observed. On the contrary, the hydrophobic character of the NIBS, that is lower with respect to the corresponding IBS, is different between the two classes, with S_{pho}/S being lower for the hyperactive than for the moderately-active AFPs.

Together with the high correlation between η_{surf} and S_{pho}/S shown above, similar results are obtained also for the relative increment of solvent density: while around the IBSs the η_{surf} is similar between moderately-active and hyperactive AFPs, with on average slightly lower values for the moderately-active AFP, around the NIBS η_{surf} differs, being higher for the hyperactive AFPs.

These results show that what actually differentiates moderately-active from hyperactive AFPs are the properties of the NIBS, rather than those of the IBS. Indeed, for both classes of proteins the IBS has a high hydrophobic content, and a consequent lower water density, as previously reported by different groups [16, 111, 138, 142, 143]. However, here we show that rather than this, it is the higher hydrophilic content, and consequent higher relative density increase, of the NIBS that differentiates the hyperactive AFPs from the moderately-active ones.

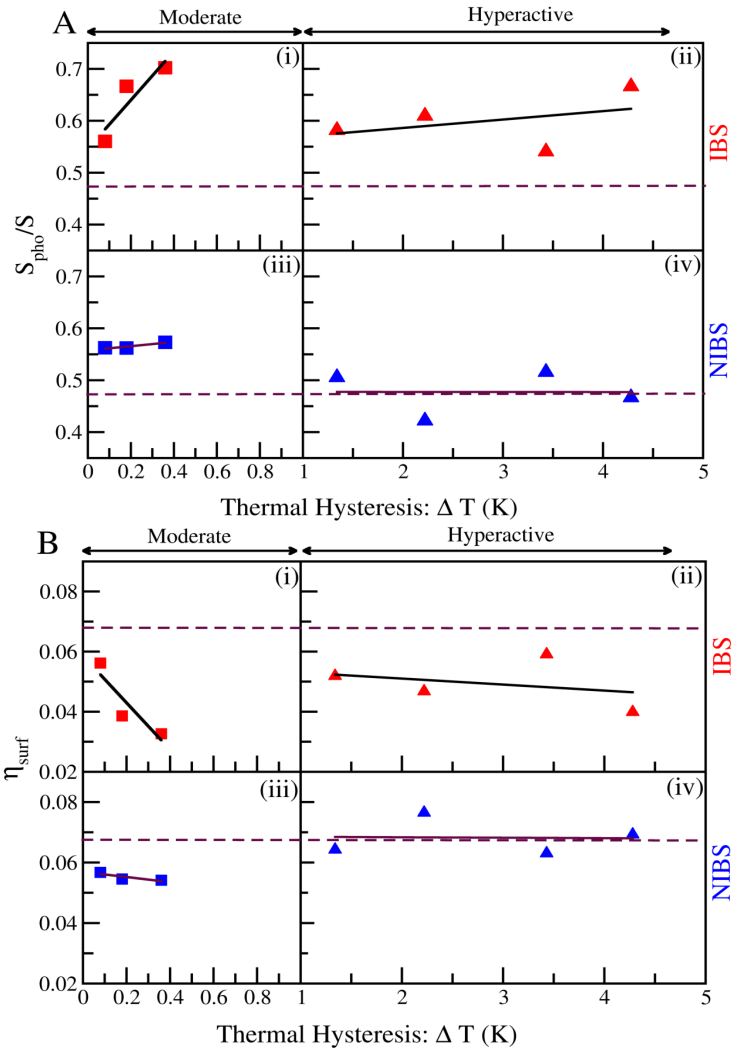

Figure 5.6: (A) S_{pho}/S and (B) η_{surf} of IBS (red) and NIBS (blue) reported as a function of the antifreeze activity of the corresponding protein. The horizontal dashed lines in panels A and B indicate the average S_{pho}/S and average η_{surf}, respectively, calculated over the NIBSs of the four hyperactive AFPs

5.4 Conclusions

The hydration shell density of all the AFPs considered in this work correlates to the experimental liquid water density of $\approx$ 33.321 molecules/nm^3 with a shell thickness of 1 nm (see Fig. Panel A 5.1). We present here the relative density increment η along the entire protein surface, as well as the IBS and NIBS η_{surf} plotted against different parameters such as protein hydrophobicity (S_{pho}/S) and antifreeze behavior ΔT. The relative density increment of whole AFPs is not associated with antifreeze activity, but the η is observed to increase with protein size increase 5.1 Panel B. The hydrophobicity of the IBS for both moderately-active and hyperactive AFPs is not only high but also comparable between the two groups (being on average slightly

higher for the moderately-active ones). The hydrophobic character of the NIBS, which is lower than that of the corresponding IBS, differs between the two groups, with hyperactive AFPs having lower hydrophobicity than moderately-active AFPs. In the case of IBS, $\eta_{surf}$is comparable between moderately-active and hyperactive AFPs; but, η_{surf} is higher for hyperactive AFPs around the NIBS. We hypothesize that the properties surrounding NIBS distinguish between moderately-active and hyperactive AFPs. The IBS has a high hydrophobic content and a lower water level for both types of proteins. We demonstrate that NIBSs are surrounded by higher hydrophilic material with a rise in relative density, resulting in variations between moderately-active and hyperactive AFPs.

Chapter 6

Conclusions

We used molecular dynamics simulation to examine the protein hydration-shell density of eighteen proteins, including eight AFPs, and discovered that all of the proteins tested showed an increase in water density relative to bulk density. While no correlation was found between the density increase and both protein eccentricity and mean hydrophobicity of the protein surface, a strong dependence on the protein length-scale was instead observed. In particular, as protein size decreases, the hydration-shell density increment decreases (from ≈ 14 % to ≈ 4 %), and this decrease is caused solely by the protein size, according to the model developed in this study. In particular, according to our model, most of the hydration shell-density increase is confined inside, or in pockets at the surface of, the protein (i.e. within the ellipsoid used to approximate the protein molecule). Given that the relative size of the solvent-accessible volume inside the protein ellipsoid relative to the overall accessible volume of the hydration shell is smaller for smaller proteins, the reduction in water density can be explained by the variation in protein size only.

According to our model, the protein-independent increase in the density of water molecules confined inside the protein effective ellipsoid is of ≈ 55 % relative to the bulk density. This value is consistent with previous calculations showing that the density of water molecules near the protein surface, especially in concave regions, is 20–50 % higher than bulk. Finally, we found that the antifreeze proteins investigated here, which span a wide variety of sizes and forms (hyperactive from insects, CfAFP, TmAFP, and RiAFP, moderately active from fish, AFPI, AFPII, and AFPIII, and moderately active from yeast, yeast-AFP, and LeIBP), behave similarly to non-antifreeze proteins in terms of size dependence of the the hydration density.

Then, focusing on two of the AFPs studied in the previous part, we analyzed the local density of the water molecules hydrating specific protein surfaces, namely the ice-binding surfaces, IBSs, and the non-ice-binding ones, non-IBSs, to look for possible local variations. We found that the solvent at the ice-binding surfaces has a lower density increment than the bulk density and a higher degree of order than the solvent at non-ice-binding surfaces. We also showed that the lower density increment is linked to the higher hydrophobic content of the IBS relative to the non-IBS. Despite

this, the IBS still has a higher water density than the bulk, indicating that there is no ice-like structure at the ice-binding surface. While this result is in contrast to part of the literature, [11, 108–114] it is in agreement with other recent works. [115, 116] Despite the fact that the average hydration shell density in AFPs is similar to that of non-AFPs, the higher content of hydrophobic (at ice-binding sites) and hydrophilic (at non-ice-binding sites) residues in AFPs results in a slightly inhomogeneous density within the hydration shell of AFPs (lower at ice-binding sites and higher at non ice-binding sites). While the hydrophobic nature of ice-binding surfaces can promote ice binding, the hydrophilic nature of non-ice-binding surfaces may provide defense against ice formation, as suggested several years ago but based on a dynamical analysis [111]. Indeed, the higher water density can prevent ice from forming around the bound AFPs, leading to the curvature of the ice surface at the basis of thermal hysteresis.

The dataset of AFPs was expanded up to seven AFPs (three moderately-active and four hyperactive) to investigate whether the coupled features of ice-binding and non-ice-binding surfaces found in the previously mentioned two AFPs, as well as the resulting density unbalance, is correlated to the antifreeze activity. In particular, we analyzed the local surface relative density increment, and hydrophobicity of the corresponding surface, as a function of antifreeze activity to determine whether the properties of the local densities and the chemical composition of the IBSs and NIBSs are related to the antifreeze activity of the AFPs. The hydrophobicity of the IBS for both moderately active and hyperactive AFPs is not only high (in the range of 0.54 - 0.70), but also comparable between the two groups (being on average slightly higher for the moderately-active ones). On the contrary, the hydrophobic character of the NIBS, which is lower in comparison to the corresponding IBS, differs between the two groups, with the hydrophilicity being higher for hyperactive AFPs than for moderately active AFPs. A similar behaviour is observed for the relative increase in hydration density: while the local surface-density increase at the IBS is similar between moderately-active and hyperactive AFPs, it is on average higher at the non-IBS for the hyperactive AFPs.

Hence, our findings show that the hydration at the non-IBS, rather than at the IBS, is what really differentiate hyperactive AFPs from the moderately-active ones. We hypothesize that, while the lower water density at the ice-binding site can pave the way to protein binding to ice nuclei, irrespective of the antifreeze activity strength, the higher solvent density at the non-IBS of the hyperactive AFPs, with respect to the moderately-active ones, might discourage ice growth around the bound AFPs.

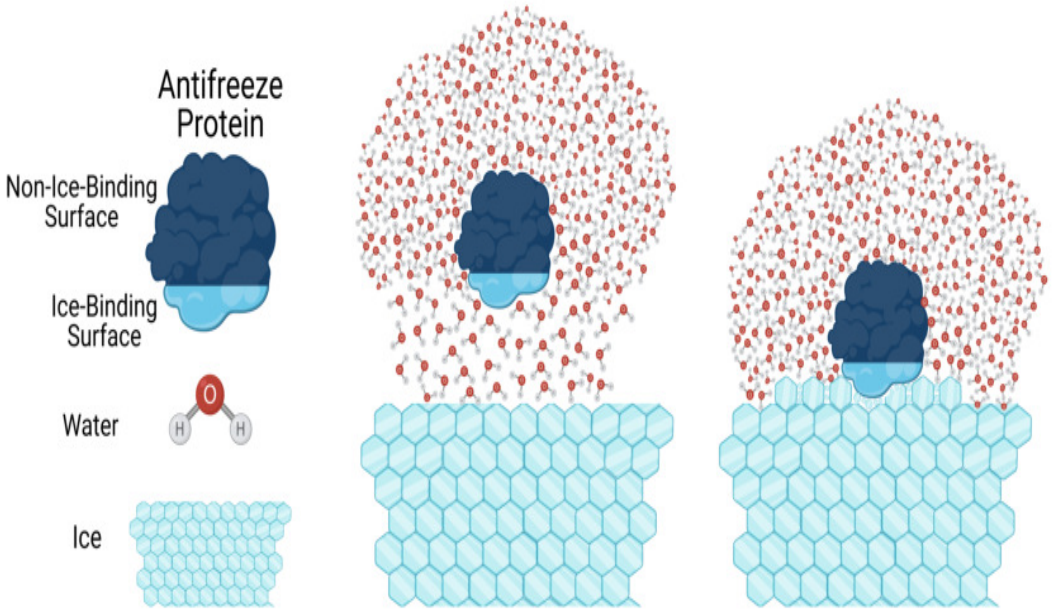

Figure 6.1: Schematic diagram showing the density of water molecules around the ice-binding surfaces and the non-ice-binding surfaces.

List of Publications

Publications:

[1] **Akash Deep Biswas**, Vincenzo Barone, Andrea Amadei and Isabella Daidone, "Length-scale dependence of protein hydration-shell density," ***Physical Chemistry Chemical Physics***, 2020, 22, 7340-7347.

[2] Laura Zanetti-Polzi, **Akash Deep Biswas**, Sara Del Galdo, Vincenzo Barone and Isabella Daidone, "Hydration shell of antifreeze proteins: Unveiling the Role of Non-Ice Binding Surfaces," ***The Journal of Physical Chemistry B***, 2019, 123, 6474-6840

[3] **Akash Deep Biswas**, Vincenzo Barone and Isabella Daidone, "High water-density at non-ice-binding surfaces contributes to the hyperactivity of antifreeze proteins," submitted to ***The Journal of Physical Chemistry Letters***.

Bibliography

[1] Horace R Drew and Richard E Dickerson. Structure of a b-dna dodecamer: Iii. geometry of hydration. *J. Mol. Biol.*, 151(3):535–556, 1981.

[2] James G Kempf and J Patrick Loria. Protein dynamics from solution nmr. *Cell Biochem. Biophys.*, 37(3):187–211, 2002.

[3] S Dellerue and M-C Bellissent-Funel. Relaxational dynamics of water molecules at protein surface. *Chem. Phys.*, 258(2-3):315–325, 2000.

[4] Samuel Toba, Giorgio Colombo, and Kenneth M Merz. Solvent dynamics and mechanism of proton transfer in human carbonic anhydrase ii. *J. Am. Chem. Soc.*, 121(10):2290–2302, 1999.

[5] Sheldon Park and Jeffery G Saven. Statistical and molecular dynamics studies of buried waters in globular proteins. *Proteins: Struct., Funct., Bioinf.*, 60(3):450–463, 2005.

[6] Christian M Stegmann, Daniel Seeliger, George M Sheldrick, Bert L de Groot, and Markus C Wahl. The thermodynamic influence of trapped water molecules on a protein–ligand interaction. *Angew. Chem., Int. Ed.*, 48(28):5207–5210, 2009.

[7] Michael Levitt and Ruth Sharon. Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci.*, 85(20):7557–7561, 1988.

[8] Franci Merzel and Jeremy C Smith. Is the first hydration shell of lysozyme of higher density than bulk water? *Proc. Natl. Acad. Sci.*, 99(8):5378–5383, 2002.

[9] David R. Nutt and Jeremy C. Smith. Dual function of the hydration layer around an antifreeze protein revealed by atomistic molecular dynamics simulations. *J. Am. Chem. Soc.*, 130:13066–13073, 2008.

[10] Anna Kuffel and Jan Zielkiewicz. Why the solvation water around proteins is more dense than bulk water. *J. Phys. Chem. B*, 116(40):12113–12124, 2012.

[11] Anna Kuffel, Dariusz Czapiewski, and Jan Zielkiewicz. Unusual structural properties of water within the hydration shell of hyperactive antifreeze protein. *The Journal of chemical physics*, 141(5):08B605_1, 2014.

[12] Anna Kuffel, Dariusz Czapiewski, and Jan Zielkiewicz. Unusual dynamic properties of water near the ice-binding plane of hyperactive antifreeze protein. *The Journal of chemical physics*, 143(13):10B601_1, 2015.

[13] Sara Del Galdo, Paolo Marracino, Marco D'Abramo, and Andrea Amadei. In silico characterization of protein partial molecular volumes and hydration shells. *Physical Chemistry Chemical Physics*, 17(46):31270–31277, 2015.

[14] Filip Persson, Pär Söderhjelm, and Bertil Halle. The geometry of protein hydration. *The Journal of Chemical Physics*, 148(21):215101, 2018.

[15] Sheh-Yi Sheu, Yu-Cheng Liu, Jia-Kai Zhou, Edward W Schlag, and Dah-Yen Yang. Surface topography effects of globular biomolecules on hydration water. *J. Phys. Chem. B*, 123(32):6917–6932, 2019.

[16] Laura Zanetti-Polzi, Akash Deep Biswas, Sara Del Galdo, Vincenzo Barone, and Isabella Daidone. Hydration shell of antifreeze proteins: Unveiling the role of non-ice-binding surfaces. *J. Phys. Chem. B*, 123(30):6474–6480, 2019.

[17] F Temple Burling, William I Weis, Kevin M Flaherty, and Axel T Brünger. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science*, 271(5245):72–77, 1996.

[18] DI Svergun, S Richard, MHJ Koch, Z Sayers, S Kuprin, and G Zaccai. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci.*, 95(5):2267–2272, 1998.

[19] Neri Niccolai, Roberta Spadaccini, Maria Scarselli, Andrea Bernini, Orlando Crescenzi, Ottavia Spiga, Arianna Ciutti, Daniela Di Maro, Luisa Bracci, Claudio Dalvit, et al. Probing the surface of a sweet protein: Nmr study of mnei with a paramagnetic probe. *Protein Sci.*, 10(8):1498–1507, 2001.

[20] Samir Kumar Pal, Jorge Peon, and Ahmed H Zewail. Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proc. Natl. Acad. Sci.*, 99(4):1763–1768, 2002.

[21] Jorge Peon, Samir Kumar Pal, and Ahmed H Zewail. Hydration at the surface of the protein monellin: dynamics with femtosecond resolution. *Proc. Natl. Acad. Sci.*, 99(17):10964–10969, 2002.

[22] Biman Bagchi. Water dynamics in the hydration layer around proteins and micelles. *Chem. Rev.*, 105(9):3197–3219, 2005.

[23] Simon Ebbinghaus, Seung Joong Kim, Matthias Heyden, Xin Yu, Udo Heugen, Martin Gruebele, David M Leitner, and Martina Havenith. An extended dynamical hydration shell around proteins. *Proc. Natl. Acad. Sci.*, 104(52):20749–20752, 2007.

[24] Matthias Heyden and Martina Havenith. Combining thz spectroscopy and md simulations to study protein-hydration coupling. *Methods*, 52(1):74–83, 2010.

[25] S Khodadadi, JE Curtis, and Alexei P Sokolov. Nanosecond relaxation dynamics of hydrated proteins: water versus protein contributions. *J. Phys. Chem. B*, 115(19):6222–6226, 2011.

[26] Jonathan D Nickels, Hugh O'Neill, Liang Hong, Madhusudan Tyagi, Georg Ehlers, Kevin L Weiss, Qiu Zhang, Zheng Yi, Eugene Mamontov, Jeremy C Smith, et al. Dynamics of protein and its hydration water: neutron scattering studies on fully deuterated gfp. *Biophys. J.*, 103(7):1566–1575, 2012.

[27] Naoki Yamamoto, Kaoru Ohta, Atsuo Tamura, and Keisuke Tominaga. Broadband dielectric spectroscopy on lysozyme in the sub-gigahertz to terahertz frequency regions: effects of hydration and thermal excitation. *J. Phys. Chem. B*, 120(21):4743–4755, 2016.

[28] Sheila Khodadadi and Alexei P Sokolov. Atomistic details of protein dynamics and the role of hydration water. *Biochim. Biophys. Acta, Gen. Subj.*, 1861(1):3546–3552, 2017.

[29] Fabio Novelli, Saeideh Ostovar Pour, Jonathan Tollerud, Ashkan Roozbeh, Dominique RT Appadoo, Ewan W Blanch, and Jeffrey A Davis. Time-domain thz spectroscopy reveals coupled protein–hydration dielectric response in solutions of native and fibrils of human lysozyme. *J. Phys. Chem. B*, 121(18):4810–4816, 2017.

[30] Ali Charkhesht, Chola K Regmi, Katie R Mitchell-Koch, Shengfeng Cheng, and Nguyen Q Vinh. High-precision megahertz-to-terahertz dielectric spectroscopy of protein collective motions and hydration dynamics. *J. Phys. Chem. B*, 122(24):6341–6350, 2018.

[31] Rafael de C Barbosa and Marcia C Barbosa. Hydration shell of the ts-kappa protein: Higher density than bulk water. *Phys. A*, 439:48–58, 2015.

[32] Simon Ebbinghaus, Konrad Meister, Benjamin Born, Arthur L DeVries, Martin Gruebele, and Martina Havenith. Antifreeze glycoprotein activity correlates

with long-range protein- water dynamics. *Journal of the American chemical society*, 132(35):12210–12211, 2010.

[33] Konrad Meister, Simon Ebbinghaus, Yao Xu, John G Duman, Arthur DeVries, Martin Gruebele, David M Leitner, and Martina Havenith. Long-range protein–water dynamics in hyperactive insect antifreeze proteins. *Proceedings of the National Academy of Sciences*, 110(5):1617–1622, 2013.

[34] PF Scholander, L Van Dam, JW Kanwisher, HT Hammel, and MS Gordon. Supercooling and osmoregulation in arctic fish. *Journal of Cellular and Comparative Physiology*, 49(1):5–24, 1957.

[35] Arthur L DeVries and Donald E Wohlschlag. Freezing resistance in some antarctic fishes. *Science*, 163(3871):1073–1075, 1969.

[36] Dang-Quan Zhang, Bing Liu, Dong-Ru Feng, Yan-Ming He, and Jin-Fa Wang. Expression, purification, and antifreeze activity of carrot antifreeze protein and its mutants. *Protein expression and purification*, 35(2):257–263, 2004.

[37] Christopher P Garnham, Jack A Gilbert, Christopher P Hartman, Robert L Campbell, Johanna Laybourn-Parry, and Peter L Davies. A ca2+-dependent bacterial antifreeze protein domain has a novel β-helical ice-binding fold. *Biochemical Journal*, 411(1):171–180, 2008.

[38] Jack A Gilbert, Philip J Hill, Christine ER Dodd, and Johanna Laybourn-Parry. Demonstration of antifreeze protein activity in antarctic lake bacteria. *Microbiology*, 150(1):171–180, 2004.

[39] Steffen P Graether, Michael J Kuiper, Stéphane M Gagné, Virginia K Walker, Zongchao Jia, Brian D Sykes, and Peter L Davies. β-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature*, 406(6793):325–328, 2000.

[40] Thomas Leya. Snow algae: adaptation strategies to survive on snow and ice. In *Polyextremophiles*, pages 401–423. Springer, 2013.

[41] Tamotsu Hoshino, Michiko Kiriaki, Satoru Ohgiya, Mineko Fujiwara, Hidemasa Kondo, Yoshiyuki Nishimiya, Isao Yumoto, and Sakae Tsuda. Antifreeze proteins from snow mold fungi. *Canadian Journal of Botany*, 81(12):1175–1181, 2003.

[42] Michael G Janech, Andreas Krell, Thomas Mock, Jae-Shin Kang, and James A Raymond. Ice-binding proteins from sea ice diatoms (bacillariophyceae) 1. *Journal of phycology*, 42(2):410–416, 2006.

[43] Eric Bender. Cell-based therapy: cells on trial. *Nature*, 540(7634):S106–S108, 2016.

[44] Jedediah K Lewis, John C Bischof, Ido Braslavsky, Kelvin GM Brockbank, Gregory M Fahy, Barry J Fuller, Yoed Rabin, Alessandro Tocchio, Erik J Woods, Brian G Wowk, et al. The grand challenges of organ banking: Proceedings from the first global summit on complex tissue cryopreservation. *Cryobiology*, 72(2):169–182, 2016.

[45] JE Lovelock and MWH Bishop. Prevention of freezing damage to living cells by dimethyl sulphoxide. *Nature*, 183(4672):1394–1395, 1959.

[46] Hak Jun Kim, Jun Hyuck Lee, Young Baek Hur, Chang Woo Lee, Sun-Ha Park, and Bon-Won Koo. Marine antifreeze proteins: structure, function, and application to cryopreservation as a potential cryoprotectant. *Marine drugs*, 15(2):27, 2017.

[47] Vanesa Robles, David G Valcarce, and Marta F Riesco. The use of antifreeze proteins in the cryopreservation of gametes and embryos. *Biomolecules*, 9(5):181, 2019.

[48] Peter L Davies. Ice-binding proteins: a remarkable diversity of structures for stopping and starting ice growth. *Trends in biochemical sciences*, 39(11):548–555, 2014.

[49] JA Raymond and AL DeVries. Freezing behavior of fish blood glycoproteins with antifreeze properties. *Cryobiology*, 9(6):541–547, 1972.

[50] James Anthony Raymond and Arthur L DeVries. Adsorption inhibition as a mechanism of freezing resistance in polar fishes. *Proc. Natl. Acad. Sci.*, 74(6):2589–2593, 1977.

[51] PW Wilson. Explaining thermal hysteresis by the kelvin effect. *CryoLetters*, 14:31–36, 1993.

[52] Denver G Hall and Alexander Lips. Phenomenology and mechanism of antifreeze peptide activity. *Langmuir*, 15(6):1905–1912, 1999.

[53] Qianzhong Li and Liaofu Luo. The kinetic theory of thermal hysteresis of a macromolecule solution. *Chemical physics letters*, 216(3-6):453–457, 1993.

[54] Qianzhong Li and Liaofu Luo. Further discussion on the thermal hysteresis of the ice growth inhibitor. *Chemical physics letters*, 223(3):181–184, 1994.

[55] CA Knight, CC Cheng, and AL DeVries. Adsorption of alpha-helical antifreeze peptides on specific ice crystal surface planes. *Biophysical journal*, 59(2):409–418, 1991.

[56] K Vanya Ewart, Daniel SC Yang, Vettai S Ananthanarayanan, Garth L Fletcher, and Choy L Hew. Ca2+-dependent antifreeze proteins: modulation of conformation and activity by divalent metal ions. *Journal of Biological Chemistry*, 271(28):16627–16632, 1996.

[57] Frank D Sönnichsen, Carl I DeLuca, Peter L Davies, and Brian D Sykes. Refined solution structure of type iii antifreeze protein: hydrophobic groups may be involved in the energetics of the protein–ice interaction. *Structure*, 4(11):1325–1337, 1996.

[58] JG Duman, N Li, D Verleye, FW Goetz, DW Wu, CA Andorfer, T Benjamin, and DC Parmelee. Molecular characterization and sequencing of antifreeze proteins from larvae of the beetle dendroides canadensis. *Journal of Comparative Physiology B*, 168(3):225–232, 1998.

[59] Sherry Y Gauthier, Cyril M Kay, Brian D Sykes, Virginia K Walker, and Peter L Davies. Disulfide bond mapping and structural characterization of spruce budworm antifreeze protein. *European journal of biochemistry*, 258(2):445–453, 1998.

[60] Hidemasa Kondo, Yuichi Hanada, Hiroshi Sugimoto, Tamotsu Hoshino, Christopher P Garnham, Peter L Davies, and Sakae Tsuda. Ice-binding site of snow mold fungus antifreeze protein deviates from structural regularity and high conservation. *Proc. Natl. Acad. Sci.*, 109(24):9360–9365, 2012.

[61] Bernard R Brooks, Robert E Bruccoleri, Barry D Olafson, David J States, S a Swaminathan, and Martin Karplus. Charmm: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2):187–217, 1983.

[62] Scott J Weiner, Peter A Kollman, Dzung T Nguyen, and David A Case. An all atom force field for simulations of proteins and nucleic acids. *Journal of computational chemistry*, 7(2):230–252, 1986.

[63] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

[64] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[65] Wilfred F van Gunsteren, SR Billeter, AA Eising, PH Hünenberger, PKHC Krüger, AE Mark, WRP Scott, and IG Tironi. Biomolecular simulation: the gromos96 manual and user guide. *Vdf Hochschulverlag AG an der ETH Zürich, Zürich*, 86:1–1044, 1996.

[66] Chris Oostenbrink, Alessandra Villa, Alan E Mark, and Wilfred F Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *Journal of computational chemistry*, 25(13):1656–1676, 2004.

[67] Loup Verlet. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical review*, 159(1):98, 1967.

[68] Loup Verlet. Computer" experiments" on classical fluids. ii. equilibrium correlation functions. *Physical Review*, 165(1):201, 1968.

[69] David Beeman. Some multistep methods for use in molecular dynamics calculations. *Journal of computational physics*, 20(2):130–139, 1976.

[70] C William Gear. Numerical initial value problems in ordinary differential equations. *Prentice-Hall series in automatic computation*, 1971.

[71] Michael P Allen and Dominic J Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.

[72] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[73] Herman JC Berendsen, JPM van Postma, Wilfred F van Gunsteren, ARHJ DiNola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.

[74] D Brown and JHR Clarke. Molecular dynamics computer simulation of polymer fiber microstructure. *The Journal of chemical physics*, 84(5):2858–2865, 1986.

[75] A Amadei, G Chillemi, MA Ceruso, A Grottesi, and A Di Nola. Molecular dynamics simulations with constrained roto-translational motions: theoretical basis and statistical mechanical consistency. *The Journal of Chemical Physics*, 112(1):9–23, 2000.

[76] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[77] Tamar Schlick, Eric Barth, and Margaret Mandziuk. Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation. *Annual review of biophysics and biomolecular structure*, 26(1):181–222, 1997.

[78] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341, 1977.

[79] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. Lincs: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[80] K Anton Feenstra, Berk Hess, and Herman JC Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry*, 20(8):786–798, 1999.

[81] Yaakov Levy and José N Onuchic. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.*, 35:389–415, 2006.

[82] Alexander M Klibanov. Improving enzymes by using them in organic solvents. *Nature*, 409(6817):241, 2001.

[83] David M Huang and David Chandler. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proc. Natl. Acad. Sci.*, 97(15):8324–8327, 2000.

[84] Henry S Ashbaugh and Michael E Paulaitis. Effect of solute size and solute-water attractive interactions on hydration water structure around hydrophobic solutes. *J. Am. Chem. Soc.*, 123(43):10721–10728, 2001.

[85] Nicolas Giovambattista, Peter J Rossky, and Pablo G Debenedetti. Effect of pressure on the phase behavior and structure of water confined between nanoscale hydrophobic and hydrophilic plates. *Phys. Rev. E*, 73(4):041604, 2006.

[86] Sapna Sarupria and Shekhar Garde. Quantifying water density fluctuations and compressibility of hydration shells of hydrophobic solutes and proteins. *Phys. Rev. Lett.*, 103(3):037803, 2009.

[87] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.

[88] George A Kaminski, Richard A Friesner, Julian Tirado-Rives, and William L Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.

[89] HJC Berendsen, JR Grigera, and TP Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[90] Isabella Daidone, Massimiliano Aschi, Laura Zanetti-Polzi, Alfredo Di Nola, and Andrea Amadei. On the origin of ir spectral changes upon protein folding. *Chem. Phys. Lett.*, 488:213–218, 2010.

[91] Andrea Amadei, Isabella Daidone, and Massimiliano Aschi. A general theoretical model for electron transfer reactions in complex systems. *Phys. Chem. Chem. Phys.*, 14:1360–1370, 2012.

[92] Jose LF Abascal and Carlos Vega. A general purpose model for the condensed phases of water: Tip4p/2005. *The Journal of chemical physics*, 123(23):234505, 2005.

[93] Frank Eisenhaber, Philip Lijnzaad, Patrick Argos, Chris Sander, and Michael Scharf. The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, 16(3):273–284, 1995.

[94] Guan-Rong Huang, Yangyang Wang, Changwoo Do, Lionel Porcar, Yuya Shinohara, Takeshi Egami, and Wei-Ren Chen. Determining gyration tensor of orienting macromolecules through their scattering signature. *J. Phys. Chem. Lett.*, 10(14):3978–3984, 2019.

[95] Mark Gerstein and Cyrus Chothia. Packing at the protein-water interface. *Proc. Natl. Acad. Sci.*, 93(19):10167–10172, 1996.

[96] Maya Bar Dolev, Ido Braslavsky, and Peter L Davies. Ice-binding proteins and their function. *Annual review of biochemistry*, 85, 2016.

[97] Daniel J Kozuch, Frank H Stillinger, and Pablo G Debenedetti. Combined molecular dynamics and neural network method for predicting protein antifreeze activity. *Proceedings of the National Academy of Sciences*, 115(52):13252–13257, 2018.

[98] CA Knight, E Driggers, and AL DeVries. Adsorption to ice of fish antifreeze glycopeptides 7 and 8. *Biophysical journal*, 64(1):252–259, 1993.

[99] Yih-Cherng Liou, Ante Tocilj, Peter L Davies, and Zongchao Jia. Mimicry of ice structure by surface hydroxyls and water of a β-helix antifreeze protein. *Nature*, 406(6793):322–324, 2000.

[100] Eeva K Leinala, Peter L Davies, and Zongchao Jia. Crystal structure of β-helical antifreeze protein points to a general ice binding model. *Structure*, 10(5):619–627, 2002.

[101] Natalya Pertaya, Christopher B Marshall, Yeliz Celik, Peter L Davies, and Ido Braslavsky. Direct visualization of spruce budworm antifreeze protein interacting with ice crystals: basal plane affinity confers hyperactivity. *Biophysical journal*, 95(1):333–341, 2008.

[102] Heman Chao, Michael E Houston, Robert S Hodges, Cyril M Kay, Brian D Sykes, Michele C Loewen, Peter L Davies, and Frank D Sönnichsen. A diminished role for hydrogen bonds in antifreeze protein binding to ice. *Biochemistry*, 36(48):14652–14660, 1997.

[103] ADJ Haymet, Leanne G Ward, Margaret M Harding, and Charles A Knight. Valine substituted winter flounderantifreeze': preservation of ice growth hysteresis. *FEBS letters*, 430(3):301–306, 1998.

[104] ADJ Haymet, Leanne G Ward, and Margaret M Harding. Winter flounder "antifreeze" proteins: synthesis and ice growth inhibition of analogues that probe the relative importance of hydrophobic and hydrogen-bonding interactions. *Journal of the American Chemical Society*, 121(5):941–948, 1999.

[105] Jason Baardsnes and Peter L Davies. Contribution of hydrophobic residues to ice binding by fish type iii antifreeze protein. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1601(1):49–54, 2002.

[106] Sean M Marks and Amish J Patel. Antifreeze protein hydration waters: Unstructured unless bound to ice. *Proceedings of the National Academy of Sciences*, 115(33):8244–8246, 2018.

[107] Zongchao Jia and Peter L Davies. Antifreeze proteins: an unusual receptor–ligand interaction. *Trends in biochemical sciences*, 27(2):101–106, 2002.

[108] Zuoyin Yang, Yanxia Zhou, Kai Liu, Yuhua Cheng, Ruozhuang Liu, Guangju Chen, and Zongchao Jia. Computational study on the function of water within a β-helix antifreeze protein dimer and in the process of ice-protein binding. *Biophysical journal*, 85(4):2599–2605, 2003.

[109] Cheng Yang and Kim A Sharp. The mechanism of the type iii antifreeze protein action: a computational study. *Biophysical chemistry*, 109(1):137–148, 2004.

[110] Cheng Yang and Kim A Sharp. Hydrophobic tendency of polar group hydration as a major force in type i antifreeze protein recognition. *PROTEINS: Structure, Function, and Bioinformatics*, 59(2):266–274, 2005.

[111] David R Nutt and Jeremy C Smith. Dual function of the hydration layer around an antifreeze protein revealed by atomistic molecular dynamics simulations. *Journal of the American Chemical Society*, 130(39):13066–13073, 2008.

[112] Nikolai Smolin and Valerie Daggett. Formation of ice-like water structure on the surface of an antifreeze protein. *The Journal of Physical Chemistry B*, 112(19):6193–6202, 2008.

[113] Jun Cui, Keith Battle, Andrzej Wierzbicki, and Jeffry D Madura. Investigations of structure and dynamics of water solvation of the type i antifreeze protein. *International Journal of Quantum Chemistry*, 109(1):73–80, 2009.

[114] Uday Sankar Midya and Sanjoy Bandyopadhyay. Hydration behavior at the ice-binding surface of the tenebrio molitor antifreeze protein. *The Journal of Physical Chemistry B*, 118(18):4743–4752, 2014.

[115] Pavithra M Naullage, Laura Lupi, and Valeria Molinero. Molecular recognition of ice by fully flexible molecules. *The Journal of Physical Chemistry C*, 121(48):26949–26957, 2017.

[116] Arpa Hudait, Daniel R Moberg, Yuqing Qiu, Nathan Odendahl, Francesco Paesani, and Valeria Molinero. Preordering of water is not needed for ice recognition by hyperactive antifreeze proteins. *Proceedings of the National Academy of Sciences*, 115(33):8266–8271, 2018.

[117] Kristofer Modig, Johan Qvist, Christopher B Marshall, Peter L Davies, and Bertil Halle. High water mobility on the ice-binding surface of a hyperactive antifreeze protein. *Physical Chemistry Chemical Physics*, 12(35):10189–10197, 2010.

[118] Joanna Grabowska, Anna Kuffel, and Jan Zielkiewicz. Structure of solvation water around the active and inactive regions of a type iii antifreeze protein and its mutants of lowered activity. *The Journal of Chemical Physics*, 145(7):075101, 2016.

[119] Yuchen Sun and Poul B Petersen. Solvation shell structure of small molecules and proteins by ir-mcr spectroscopy. *The journal of physical chemistry letters*, 8(3):611–614, 2017.

[120] Volkhard Helms. Protein dynamics tightly connected to the dynamics of surrounding and internal water molecules. *ChemPhysChem*, 8(1):23–33, 2007.

[121] Luyuan Zhang, Yi Yang, Ya-Ting Kao, Lijuan Wang, and Dongping Zhong. Protein hydration dynamics and molecular mechanism of coupled water- protein fluctuations. *Journal of the American Chemical Society*, 131(30):10677–10691, 2009.

[122] Michael Schauperl, Maren Podewitz, Teresa S Ortner, Franz Waibl, Alexander Thoeny, Thomas Loerting, and Klaus R Liedl. Balance between hydration enthalpy and entropy is important for ice binding surfaces in antifreeze proteins. *Scientific Reports*, 7(1):1–13, 2017.

[123] Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of computational chemistry*, 24(16):1999–2012, 2003.

[124] MM Conde, M Rovere, and P Gallo. High precision determination of the melting points of water tip4p/2005 and water tip4p/ice models by the direct coexistence technique. *The Journal of Chemical Physics*, 147(24):244506, 2017.

[125] C Vega, E Sanz, and JLF Abascal. The melting temperature of the most common models of water. *The Journal of chemical physics*, 122(11):114507, 2005.

[126] David Brown and JHR Clarke. A comparison of constant energy, constant temperature and constant pressure ensembles in molecular dynamics simulations of atomic liquids. *Molecular Physics*, 51(5):1243–1252, 1984.

[127] Alfred A Antson, Derek J Smith, David I Roper, Sally Lewis, Leo SD Caves, Chandra S Verma, Sarah L Buckley, Peter J Lillford, and Roderick E Hubbard. Understanding the mechanism of ice binding by type iii antifreeze proteins. *Journal of molecular biology*, 305(4):875–889, 2001.

[128] Eeva K Leinala, Peter L Davies, Daniel Doucet, Michael G Tyshenko, Virginia K Walker, and Zongchao Jia. A β-helical antifreeze protein isoform with increased activity structural and functional insights. *Journal of biological chemistry*, 277(36):33349–33352, 2002.

[129] Lada Biedermannova and Bohdan Schneider. Hydration of proteins and nucleic acids: Advances in experiment and theory. a review. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1860(9):1821–1835, 2016.

[130] Isabella Daidone, Martin B Ulmschneider, Alfredo Di Nola, Andrea Amadei, and Jeremy C Smith. Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding. *Proceedings of the National Academy of Sciences*, 104(39):15230–15235, 2007.

[131] Anneloes S Oude Vrielink, Antonio Aloi, Luuk LC Olijve, and Ilja K Voets. Interaction of ice binding proteins with ice, water and ions. *Biointerphases*, 11(1):018906, 2016.

[132] Z Faidon Brotzakis, Ilja K Voets, Huib J Bakker, and Peter G Bolhuis. Water structure and dynamics in the hydration layer of a type iii anti-freeze protein. *Physical Chemistry Chemical Physics*, 20(10):6996–7006, 2018.

[133] Erlend Kristiansen and Karl Erik Zachariassen. The mechanism by which fish antifreeze proteins cause thermal hysteresis. *Cryobiology*, 51(3):262–280, 2005.

[134] Yeliz Celik, Laurie A Graham, Yee-Foong Mok, Maya Bar, Peter L Davies, and Ido Braslavsky. Superheating of ice crystals in antifreeze protein solutions. *Proc. Natl. Acad. Sci.*, 107(12):5423–5428, 2010.

[135] Maya Bar Dolev, Ido Braslavsky, and Peter L Davies. Ice-binding proteins and their function. *Annu. Rev. Biochem.*, 85:515–542, 2016.

[136] Akash Deep Biswas, Vincenzo Barone, Andrea Amadei, and Isabella Daidone. Length-scale dependence of protein hydration-shell density. *Physical Chemistry Chemical Physics*, 22(14):7340–7347, 2020.

[137] Elise Duboué-Dijon and Damien Laage. Comparative study of hydration shell dynamics around a hyperactive antifreeze protein and around ubiquitin. *J. Phys. Chem.*, 141(22):12B633_1, 2014.

[138] Prasun Pal, Sandipan Chakraborty, and Biman Jana. Deciphering the role of non-ice-binding surface on the antifreeze activity of hyperactive antifreeze proteins. *J. Phys. Chem. B*, 2020.

[139] Joanna Grabowska, Anna Kuffel, and Jan Zielkiewicz. Role of the solvation water in remote interactions of hyperactive antifreeze proteins with the surface of ice. *J. Phys. Chem. B*, 123(38):8010–8018, 2019.

[140] Uday Sankar Midya and Sanjoy Bandyopadhyay. Role of polar and nonpolar groups in the activity of antifreeze proteins: a molecular dynamics simulation study. *J. Phys. Chem. B*, 122(40):9389–9398, 2018.

[141] Sara Del Galdo and Andrea Amadei. The unfolding effects on the protein hydration shell and partial molar volume: a computational study. *Physical Chemistry Chemical Physics*, 18(40):28175–28182, 2016.

[142] Daniela Russo, José Teixeira, Larry Kneller, John RD Copley, Jacques Ollivier, Stefania Perticaroli, Eric Pellegrini, and Miguel Angel Gonzalez. Vibrational density of states of hydration water at biomolecular sites: Hydrophobicity promotes low density amorphous ice behavior. *J. Am. Chem. Soc.*, 133(13):4882–4888, 2011.

[143] Arpa Hudait, Yuqing Qiu, Nathan Odendahl, and Valeria Molinero. Hydrogen-bonding and hydrophobic groups contribute equally to the binding of hyperactive antifreeze and ice-nucleating proteins to ice. *J. Am. Chem. Soc.*, 141(19):7887–7898, 2019.

List of Figures

1.1 Schematic illustration of an antifreeze protein adsorbed to an ice nucleus through the ice-binding-surface (colored in blue). The rest of the protein surface, colored in green, represents the non-ice-binding surface. 3

1.2 Antifreeze proteins structures. Representative configurations are drawn for the different types of AFPs from fish, insects and fungi. α helices are shown in purple, β strands in yellow, 3_{10} helices in blue and coil in cyan. The different AFPs are: fish AFP type I *Pseudopleuronectes americanus* (PaAFP), fish AFP type II *Hemitripterus americanus* (HaAFP), fish AFP type III *Zoarces americanus* (ZaAFP), insect AFP *Choristoneura fumiferana* isoform 337 (CfAFP337), insect AFP *Choristoneura fumiferana* isoform 501 (CfAFP501), insect AFP *Rhagium inquisitor* (RiAFP), snow mold fungus AFP *Typhula ishikariensis* isoform 6 (TisAFP6) and snow mold fungus *Typhula ishikariensis*isoform 8 (TisAFP8). 5

3.1 Schematic diagram showing the main parameters used in the model. The inner ellipsoid corresponds to the effective protein ellipsoid volume, $V_{ell,eff}$, while the outer ellipsoid corresponds to the volume of the ellipsoid defined by the boundary layer of the hydration shell, V_{sh}. The excluded protein volume, V_{ex}, is represented in red (dark red corresponds to V_{ex}^{in} while light red to V_{ex}^{out}); the solvent-accessible volume, V_{acc}, is represented in blue (dark blue corresponds to V_{acc}^{in} while light blue to V_{acc}^{out}). 14

3.2 Layer density profile of the hydration SPC molecules as a function of the distance from the protein ellipsoid surface calculated for the eighteen proteins: Trp-cage (black), Yeast-AFP (red), AFPI (green), Heliomicin (blue), GB1 (yellow), AFPIII (brown), BPTI (grey), TmAFP (violet), Ubiquitin (cyan), CfAFP (magenta), Barnase (orange), Lysozyme (indigo), RiAFP (maroon), AFPII (turquoise), Myoglobin (dark green), LEIBP (light red), HCAII (light blue), and COVID-19 (light green). . 21

3.3 Density variations with respect to the bulk density, $\rho/\rho_{b,fict}$, as a function of the distance from the protein ellipsoid surface for the eighteen proteins: Trp-cage (black), Yeast-AFP (red), AFPI (green), Heliomicin (blue), GB1 (yellow), AFPIII (brown), BPTI (grey), TmAFP (violet), Ubiquitin (cyan), CfAFP (magenta), Barnase (orange), Lysozyme (indigo), RiAFP (maroon), AFPII (turquoise), Myoglobin (dark green), LEIBP (light red), HCAII (light blue), and COVID-19 (light green). . 22

3.4 Hydrophobic (green) and hydrophilic (blue) atoms at the protein surface. 24

3.5 Hydrophobic (green) and hydrophilic (blue) atoms at the protein surface. 25

3.6 η as a function of the ellipsoid eccentricity, defined as $\epsilon^2 = 1 - a^2/c^2$ with a and c the smallest and largest ellipsoid axes, respectively (A), and of the fraction of hydrophobic atoms at the protein surface (B). . 26

3.7 A) V_{ex} as a function of the effective protein-ellipsoid volume, $V_{ell,eff}$, for all the studied proteins. From the linear fit $V_{ex} = mV_{ell,eff}$ we obtain $m = 0.95$. B) Mean number of solvent molecules inside the effective protein ellipsoid, n_{in}, as a function of the effective protein-ellipsoid volume, $V_{ell,eff}$, for all the studied proteins. From the linear fit $n_{in} = \beta V_{ell,eff}$ we obtain $\beta = 11.61$. The correlation coefficient is higher than 0.99 for both regressions. 27

3.8 (A) ρ_{in}/ρ_b as a function of $\alpha = V_{sh}/V_{ell,eff}$ for all the studied proteins. A linear regression of the ρ_{in}/ρ_b data yields a slope coefficient of -0.031 which is within one standard error of the coefficient (0.032), showing that ρ_{in}/ρ_b can be considered basically constant over the protein sample. (B) ρ_{out}/ρ_b as a function of $\alpha = V_{sh}/V_{ell,eff}$ for all the studied proteins. The black solid line is the analytical function $\rho_{out}/\rho_b = (a_0+x)/(x+a_1)$ with $a_0 = -0.63$ and $a_1 = -0.75$ as obtained from the fit of η with eq. 3.2.12 (see Results section). The analytical expression provides a good reproduction of the ρ_{out}/ρ_b values obtained from the MD simulations. 28

3.9 (A) η as a function of $\alpha = V_{sh}/V_{ell,eff}$ calculated for the different proteins. The data are fitted with eq. 3.2.12 (dashed line) providing $r_0 = \rho_{out}/\rho_b = 1.04$ and eq. 3.2.12 (full line) providing $a_0 = -0.63$ and $a_1 = -0.75$; the correlation coefficient is 0.97 for both models. The eight AFPs are highlighted in red. 29

3.10 Representative configuration of HCAII showing the water molecules within the protein ellipsoid in ball and stick (oxygen and hydrogen atoms are reported in red and white, respectively). Most of the water molecules are localized within grooves and pockets. The behaviour is qualitatively similar for all proteins. 30

4.1 Representative sketch of the eight surfaces defined. CfAFP IBS and NIBS; ZaAFP IBS and NIBS; Barnase Surface 1 and Surface 2; Ubiquitin Surface 1 and Surface 2. The hydrophobic and hydrophilic portions of the surfaces are highlighted in yellow and purple, respectively. The rest of the protein is represented as white surface. 36

4.2 Representative structure of the four proteins investigated inside their 1 nm-thick hydration shell at 300 K: A, CfAFP; B, ZaAFP, C, Barnase; D, Ubiquitin. The protein+solvent atoms inside the ellipsoid that approximates the protein are highlighted in licorice with blue cartoon backbone. The hydration shell solvent molecules outside such an ellipsoid are represented in transparent licorice and the protein atoms that protrude from the protein ellipsoid surface are represented in gray licorice. 37

4.3 Ratio ρ/ρ_{fict} as a function of the distance from the protein ellipsoid surface for the four proteins at 300 K (a,b) and at 245 K (c,d) with TIP4P/2005 water model. AFPs are reported in panels (a,c) (CfAFP, black; ZaAFP, red) and non-AFPs are reported in panels b and d (Bar, green; Ubi, blue). 38

4.4 Ratio ρ/ρ_{fict} as a function of the distance from the protein ellipsoid surface for the four proteins at 300 K (a,b) and at 200 K (c,d) with SPC water model. AFPs are reported in panels (a,c) (CfAFP, black; ZaAFP, red) and non-AFPs are reported in panels b and d (Bar, green; Ubi, blue). 39

4.5 Ratio ρ/ρ_{fict} as a function of the distance from the protein ellipsoid surface for the four proteins at 300 K (a,b) and at 200 K (c,d). AFPs are reported in panels (a,c) (CfAFP, black; ZaAFP, red) and non-AFPs are reported in panels b and d (Bar, green; Ubi, blue). 40

4.6 Layer density profile of the solvation TIP4P/2005 molecules for the four proteins at 300 K (a and b) and at 245 K (c and d). AFPs are reported in panels a and c (CfAFP, black; ZaAFP, red) and non-AFPs are reported in panels b and d (Bar, green; Ubi, blue). The insets show the convergence of the density profile, with its standard error, within the bulk density noise interval (displayed with two orange lines) corresponding to the 95% confidence. 42

4.7 Relative increment η^{surf} of the hydration surface density with respect to the bulk density *vs* the fraction of hydrophobic solvent-exposed area (S_{pho}/S) of the 8 surfaces selected for the four proteins. (A) data at 300 K, (B) data at 245 K. (C) Representative sketch of the IBS and NIBS defined for CfAFP and ZaAFP. The hydrophobic and hydrophilic portions of the surfaces are highlighted in yellow and purple, respectively. The rest of the protein is represented as white surface. 45

4.8 H–O–O angle distribution, P(Θ), of nearby water molecules (i.e., with a oxygen–oxygen distance below 0.35 nm) for waters in the first hydration layer (within 0.55 nm from the heavy atoms of the protein) of the IBS (solid lines) and NIBS (dashed lines) surfaces of CfAFP (black) and ZaAFP (red) at 245 K. The low-angle population is increased around the IBSs with respect to bulk water (blue dashed line) and with respect to the NIBSs. 47

5.1 SPC solvation molecules layer density profile at 300K for the seven AFPs. 54

5.2 The ellipsoidal model generated a representative solvation layer for the seven proteins. The color of the proteins is specified by the density profile shown in Figure S1. A) *Tis*AFP6, B) *Za*AFP, and C) *Tis*AFP8 are moderately-active AFPs with antifreeze activity ¡ 1 K. D) *Cf*AFP337, E) *Tm*AFP, F) *Ri*AFP, and G) *Cf*AFP501 are hyperactive AFPs with antifreeze activity ¿ 1 and ¡ 6 K. 55

5.3 Panel A: Ratio $\rho/\rho_{b,fict}$ as a function of the distance from the protein ellipsoid surface calculated for the seven proteins: *Tis*AFP6 (black), *Za*AFP (red), *Tis*AFP8 (green), *Cf*AFP337 (Blue), *Tm*AFP (violet), *Ri*AFP (cyan), and *Cf*AFP501 (magenta). Panel B: Relative surface density increment η as a function of partial molar volume v calculated for each protein molecule using eq 5.2.4. Panel C: η as a function of the antifreeze activity, $\Delta T(K)$, measured at a protein solution concentration of 0.3 g/L. [97] The moderatetly-active (Mod.) AFPs are represented with green squares whereas the hyperactive (Hyper.) AFPs are represented with purple triangles. 56

5.4 The ice-binding surfaces of the proteins are colored in orange and the rest of the protein considered as NIBS are colored in silver. A. *Tis*AFP6, B. *Za*AFP, C. *Tis*AFP8, D. *Cf*AFP337, E. *Tm*AFP, F. *Ri*AFP, and G. *Cf*AFP501. The respective residues of the IBS are also mentioned in the dark column. 57

5.5 Change in the relative water-density increment with respect to the bulk (η_{surf}) as a function of the hydrophobic fraction of the solvent-exposed-area ($\langle S_{pho}/S \rangle$) of the IBSs (represented in red) and NIBSs (represented in blue) of the seven antifreeze proteins. 59

5.6 (A) S_{pho}/S and (B) η_{surf} of IBS (red) and NIBS (blue) reported as a function of the antifreeze activity of the corresponding protein. The horizontal dashed lines in panels A and B indicate the average S_{pho}/S and average η_{surf}, respectively, calculated over the NIBSs of the four hyperactive AFPs 61

6.1 Schematic diagram showing the density of water molecules around the ice-binding surfaces and the non-ice-binding surfaces. 65

List of Tables

3.1 Residue number, effective protein-ellipsoid volume, $V_{ell,eff}$, ratio between the volume of the shell, V_{sh}, and the effective ellipsoid volume, $V_{ell,eff}$, and relative density increment, η, for the eighteen studied proteins. Volumes are given in nm^3 23

4.1 Mean values of the protein excluded volume $\langle V_{ex} \rangle$, hydration shell volume $\langle V_{shell} \rangle$ and number of SPC molecules inside the hydration shell $\langle N_{shell} \rangle$ as obtained from the MD trajectories of the four proteins at 300 K and 200 K[a]. 41

4.2 Mean values of the Protein-Excluded Volume $\langle V_{ex} \rangle$, Hydration Shell Volume $\langle V_{shell} \rangle$ and Number of TIP4P/2005 Molecules Inside the Hydration Shell $\langle N_{shell} \rangle$ as Obtained from the MD Trajectories of the Four Proteins at 300 K and 245 K[a]. 43

4.3 Mean Value of the Solvent Accessible Surface Area $\langle S \rangle$ (in nm^2), Surface Density ρ^{surf} (in Molecules per nm^2) and Relative Increment with Respect to the Bulk Surface Density η^{surf} Calculated for the Four Proteins at 300 and 245 K[a]. 46

5.1 Total number of residues (RN), experimental thermal hysteresis (ΔT), mean excluded volume ($\langle V_{ex} \rangle$), mean hydration shell volume ($\langle V_{shell} \rangle$), hydration-shell density increment relative to bulk density (η) and partial molar volume (v) as obtained from the MD trajectories of the seven proteins.[a] 58

5.2 Average solvent accessible surface area, $\langle S \rangle$, of the whole protein (WP) and of the ice binding surface (IBS) and average hydrophobic-surface fraction, $\langle S_{pho}/S \rangle$, for the WP, IBS, and the NIBS. 59

5.3 The relative surface density increment η_{surf} of the IBS and NIBS[a]. . . 60